ROZPRAWA DOKTORSKA

mgr inż. Igor Stępień

Metody oceny jakości obrazów cyfrowych z wykorzystaniem sieci neuronowych

w formie cyklu publikacji naukowych

Promotor
dr hab. inż. Mariusz Oszust, prof. PRz

Rzeszów, lipiec 2025

Podziękowania

Pragnę wyrazić serdeczne podziękowania Panu dr. hab. inż. Mariuszowi Oszustowi, prof. PRz, za nieocenione wsparcie, życzliwość oraz inspirujące wskazówki, które towarzyszyły mi na każdym etapie realizacji niniejszej rozprawy doktorskiej. Jego merytoryczne uwagi, cierpliwość i zaangażowanie w proces mojej pracy badawczej miały kluczowe znaczenie dla jej kształtu i jakości.

Dziękuję za okazaną wyrozumiałość i motywację, które pozwalały mi przezwyciężać trudności i konsekwentnie dążyć do naukowej rzetelności. Możliwość pracy pod Pańskim kierunkiem była dla mnie zaszczytem oraz cennym doświadczeniem.

Na szczególne podziękowania zasługuje moja żona, Joanna, która przez cały okres mojej pracy nad doktoratem była dla mnie niewyczerpanym źródłem wsparcia, cierpliwości i motywacji. Jej wyrozumiałość wobec mojej nieobecności – zarówno fizycznej, jak i myślowej – w najważniejszych chwilach tej naukowej drogi była bezcenna. Dziękuję za codzienną obecność, ciepło, wiarę w moje możliwości oraz za to, że nigdy nie pozwoliła mi się poddać – nawet wtedy, gdy sam w siebie wątpiłem.

Spis treści

Wykaz symboli, oznaczeń i akronimów

AI – Sztuczna inteligencja (*ang. Artificial Intelligence*)
DL – Głębokie uczenie (*ang. Deep Learning*)
DMOS – Subiektywny współczynnik jakości oparty o średnią ocen (*ang. Differential Mean Opinion Score*)
FR – Ocena z pełnym odniesieniem (*ang. Full-Reference*)
HVS – Ludzki układ wzrokowy (*ang. Human Visual System*)
IQA – Ocena jakości obrazu (*ang. Image Quality Assessment*)
KPCA – Jądrowa analiza głównych składowych (*ang. Kernel Principal Component Analysis*)
KRCC – Współczynnik Korelacji Rang Kendalla (*ang. Kendall Rank Correlation Coefficient*)
LLM – Duży model językowy (*ang. Large Language Model*)
LSTM – Długoterminowa pamięć krótkotrwała (*ang. Long Short-Term Memory*)
MEDQEMRI – *ang. Multi lEvel and multi-model Deep Quality Evaluator of MR Images*
MLPSIE – *ang. Multi-Level Pan-Sharpening Images Evaluator*
MOS – Średnia ocena subiektywna (*ang. Mean Opinion Score*)
MR – Rezonans magnetyczny (*ang. Magnetic Resonance*)
MRI – Obrazowanie za pomocą rezonansu magnetycznego (*ang. Magnetic Resonance Imaging*)
MSCN – Normalizacja kontrastu z odjętą średnią (*ang. Mean Subtracted Contrast Normalization*)
MSE – Błąd średniokwadratowy (*ang. Mean Square Error*)
NR – Ocena bez odniesienia, brak referencji (*ang. No-Reference*)
PC – technika nadawania kolorów obrazowi w skali szarości (*ang. Pseudo Color*)
PCA – Analiza głównych składowych (*ang. Principal Component Analysis*)
PLCC – Współczynnik Korelacji Liniowej Pearsona (*ang. Pearson Linear Correlation Coefficient*)
PS – Wyostrzanie obrazów satelitarnych (*ang. Pan-Sharpening*)
RMSE – Pierwiastek błędu średniokwadratowego (*ang. Root Mean Square Error*)
ROI – Region zainteresowania (*ang. Region of Interest*)
SGDM – Stochastyczny spadek gradientu z momentem (*Stochastic Gradient Descent with Momentum*)
RR – Ocena z częściowym odniesieniem, częściowe odniesienie (*ang. Reduced-Reference*)
SRCC – Współczynnik Korelacji Rang Spearmana (*ang. Spearman Rank Correlation Coefficient*)
SVM – Maszyna wektorów nośnych (*ang. Support Vector Machine*)
SVR – Regresja wektorów nośnych (*ang. Support Vector Regression*)

TBN-PSI – *ang. The Three-Branch Neural Network for No-Reference Quality Assessment of Pan-Sharpened Images*
TC – Technika nadawania kolorów obrazowi w skali RGB (*ang. True Color*)
TE – Czas echa (*ang. Echo Time*)
TIQA-MRI – Pakiet narzędzi do percepcyjnej oceny jakości obrazów rezonansu magnetycznego (*ang. Toolbox for Perceptual Image Quality Assessment of Magnetic Resonance Images*)
TIQA-PSI – Pakiet narzędzi do percepcyjnej oceny jakości wyostrzonych obrazów satelitarnych (*ang. Toolbox for Perceptual Image Quality Assessment of Pan-Sharpened Images*)
TR – Czas powtórzenia (*ang. Repetition Time*)
ViT – Architektura sieci neuronowej do przetwarzania obrazów oparta o model transformatora (*ang. Vision Transformer*)

1. Wprowadzenie

Niniejsza rozprawa doktorska stanowi jednotematyczny cykl publikacji naukowych poświęconych ocenie jakości obrazów cyfrowych z wykorzystaniem sieci neuronowych. Praca ma charakter interdyscyplinarny i bazuje na dyscyplinie Informatyka Techniczna i Telekomunikacja, wykorzystując bazy obrazów oraz metody z obszaru medycyny i teledetekcji. Podjęcie tematyki oceny jakości obrazów (IQA) w kontekście rezonansu magnetycznego (MRI) oraz przetwarzania wyostrzonych obrazów satelitarnych (PS) przyczyniło się do rozwoju nowych metod wykorzystujących architektury głębokich sieci neuronowych. Zaproponowane, dedykowane rozwiązania wykazują się dużą wrażliwością na charakter analizowanych obrazów i pozwalają przewidywać ich jakość. Ponadto, w obszarze oceny jakości obrazów rezonansu magnetycznego, została przygotowana praca przeglądowa, której celem było uporządkowanie istniejącej wiedzy odnośnie metod oceny jakości obrazów bez referencji (NR IQA) z uwzględnieniem specyfiki przetwarzanych obrazów. Opracowano również zestawy narzędzi programowych działających w środowisku MATLAB, służących jako wsparcie eksperymentów.

W skład prezentowanej rozprawy doktorskiej wchodzi siedem publikacji:

[A-1] **Stępień, I.**, Oszust, M. *A Brief Survey on No-Reference Image Quality Assessment Methods for Magnetic Resonance Images.* J. Imaging 2022, 8(6), 160. https://doi.org/10.3390/jimaging8060160, Impact Factor $(IF)_{2022} = 3{,}2$; Cite Score (CS): 5,9; wkład 50%; liczba punktów MNiSW: 20.

[A-2] **Stępień, I.**, Obuchowicz, R., Piórkowski, A., Oszust, M. *Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment.* Sensors 2021, 21(4), 1043. https://doi.org/10.3390/s21041043, $IF_{2021} = 3{,}847$; CS: 7,3; wkład 25%; liczba punktów MNiSW: 100.

[A-3] **Stępień, I.**, Oszust, M. *No-Reference Image Quality Assessment of Magnetic Resonance Images with Multi-Level and Multi-Model Representations Based on Fusion of Deep Architectures.* Engineering Applications of Artificial Intelligence 2023, 123, 106283. https://doi.org/10.1016/j.engappai.2023.106283, $IF_{2023} = 7{,}5$; CS: 9,6; wkład 50%; liczba punktów MNiSW: 140.

[A-4] **Stępień, I.**, *TIQA-MRI: Toolbox for Perceptual Image Quality Assessment of Magnetic Resonance Images.* SoftwareX 2025, 29, 102073, https://doi.org/10.1016/j.softx.2025.102073, $IF_{2023} = 2{,}4$; CS: 4,1; wkład 100%; liczba punktów MNiSW: 200.

[A-5] **Stępień, I.**, Oszust, M. *No-Reference Quality Assessment of Pan-Sharpening Images with Multi-Level Deep Image Representations.* Remote Sens. 2022, 14(5), 1119. https://doi.org/10.3390/rs14051119, $IF_{2022} = 5$; CS: 8,3; wkład 50%; liczba punktów MNiSW: 100.

[A-6] **Stępień, I.**, Oszust, M. *Three-Branch Neural Network for No-Reference Quality Assessment of Pan-Sharpened Images.* Engineering Applications of Artificial Intelligence 2025, 139(B), 109594. https://doi.org/10.1016/j.engappai.2024.109594, IF_{2023} = 7,5; CS: 9,6; wkład 50%; liczba punktów MNiSW: 140.

[A-7] **Stępień, I.**, Oszust, M. *TIQA-PSI: Toolbox for Perceptual Image Quality Assessment of Pan-Sharpened Images.* SoftwareX 2023, 23, 101494, https://doi.org/10.1016/j.softx.2023.101494, IF_{2023} = 2,4; CS: 4,1; wkład 50%; liczba punktów MNiSW: 200.

Wszystkie artykuły naukowe wchodzące w skład cyklu znajdują się w czasopismach wyszczególnionych na liście czasopism punktowanych Ministerstwa Nauki i Szkolnictwa Wyższego (MNiSW) i zostały opublikowane w latach 2021-2025. Sumaryczny Impact Factor zgodnie z rokiem ukazania się publikacji wynosi 31,847, zaś liczba punktów zgodnie z wykazem MNiSW wynosi 900.

1.1. Motywacja

Sztuczna inteligencja (AI) jest dziedziną, która odgrywa kluczową rolę w automatyzacji wielu procesów, takich jak analiza mowy, rozpoznawanie obrazów, diagnostyka medyczna czy wspieranie badań naukowych. Jej rozwój opiera się na zaawansowanych algorytmach i modelach matematycznych, które umożliwiają maszynom uczenie się na podstawie danych i adaptację do różnorodnych zastosowań. Historia głębokiego uczenia się (DL), jednej z najważniejszych metod w AI, ma swoje początki w 1943 roku, kiedy Walter Pitts i Warren McCulloch zaprezentowali pierwszy model sieci neuronowej inspirowany strukturą ludzkiego mózgu [55]. Koncepcja ta, oparta na „logice progowej", miała na celu odwzorowanie procesów myślowych za pomocą algorytmów matematycznych. Rozwój AI był utrudniony przez problemy z precyzyjnym ujęciem zadań wykonywanych przez ludzi w sposób intuicyjny, na przykład rozpoznawanie twarzy czy mowy. Rozwiązanie tego problemu zaproponowano poprzez umożliwienie maszynom uczenia się na podstawie doświadczeń i budowania hierarchii pojęć, gdzie bardziej skomplikowane struktury są tworzone z prostszych elementów. To podejście, nazwane głębokim uczeniem, stało się podstawą rozwoju nowoczesnych systemów AI [77].

W 1958 roku Frank Rosenblatt opracował perceptron – model zdolny do uczenia się wag na podstawie danych, co pozwalało na klasyfikację obiektów do odpowiednich kategorii [76]. W tym samym okresie Bernard Widrow i Marcian Hoff stworzyli adaptacyjny liniowy element (ADALINE), który umożliwiał przewidywanie wartości liczbowych na podstawie danych wejściowych [96]. Kolejny istotny krok miał miejsce w 1979 roku, kiedy Kunihiko Fukushima zaprojektował Neocognitron – pierwszą splotową sieć neuronową. Wykorzystując hierarchiczną strukturę warstw model ten pozwalał komputerom na rozpoznawanie wzorców wizualnych [22]. Lata 80. i 90. przyniosły przełomowe osiągnięcia w rozwoju sztucznej inteligencji. W 1989 roku Yann LeCun

zademonstrował skuteczność algorytmu wstecznej propagacji błędu, co umożliwiło trenowanie sieci neuronowych do rozpoznawania ręcznie pisanych cyfr – rozwiązanie to znalazło praktyczne zastosowanie m.in. przy automatycznym odczytywaniu numerów na czekach [45]. W 1995 roku Corinna Cortes i Vladimir Vapnik opracowali maszynę wektorów nośnych (SVM), która szybko stała się jednym z podstawowych narzędzi stosowanych w analizie danych i klasyfikacji [90]. Dwa lata później, w 1997 roku, Sepp Hochreiter i Jürgen Schmidhuber przedstawili architekturę długiej pamięci krótkotrwałej (LSTM), przeznaczoną dla rekurencyjnych sieci neuronowych, pozwalającą na skuteczne modelowanie sekwencji danych [33]. Rok 1997 okazał się szczególnie ważny dla rozwoju sztucznej inteligencji. Komputer Deep Blue firmy IBM pokonał mistrza świata w szachach Garriego Kasparowa, co ukazało potencjał AI w analizie i podejmowaniu decyzji w zamkniętych, dobrze zdefiniowanych środowiskach [35]. Jednak codzienne zadania, wymagające intuicyjnej wiedzy o świecie, pozostawały wyzwaniem. W odpowiedzi zaczęto rozwijać systemy uczące się, które mogły przyswajać wiedzę z danych i podejmować subiektywne decyzje. Przełom technologiczny w przetwarzaniu danych nastąpił pod koniec lat 90. wraz z wprowadzeniem jednostek przetwarzania grafiki (GPU)[1]. W 1999 roku szybkość obliczeniowa wzrosła tysiąckrotnie w porównaniu z dekadą wcześniejszą, co umożliwiło bardziej efektywne trenowanie sieci neuronowych. Zaczęły skutecznie przewyższać możliwości maszyn wektorów nośnych, oferując lepsze wyniki na tych samych danych. W 2001 roku raport META Group[2] (obecnie Gartner) zwrócił uwagę na wyzwania i możliwości związane z rosnącą ilością danych. Wskazano na konieczność przygotowania się na erę Big Data, która wymagała nowych narzędzi do analizy i przetwarzania informacji. Kluczową rolę w tej dziedzinie odegrały sieci neuronowe i algorytmy głębokiego uczenia. Kolejnym istotnym krokiem był rok 2009, w którym Fei-Fei Li uruchomiła projekt ImageNet[3] – bazę danych zawierającą ponad 14 milionów oznaczonych obrazów. Ten zasób umożliwił trenowanie algorytmów w skali wcześniej niedostępnej, przyspieszając rozwój metod do rozpoznawania obrazów. Z biegiem czasu standardem stało się osiąganie wysokiej wydajności algorytmów przy użyciu milionów oznaczonych przykładów. Przełom w tej dziedzinie nastąpił w 2015 roku, wraz z wprowadzniem architektury ResNet, która zwyciężyła w konkursie ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Model ten osiągnął 3,57% błędu w klasyfikacji obrazów, przewyższając szacowany poziom błędu człowieka wynoszący około 5% [30].

Głębokie uczenie znalazło zastosowanie w wielu dziedzinach nauki. W 2014 roku techniki te zostały wykorzystane do przewidywania interakcji między molekułami, co pomogło w opracowywaniu nowych leków [15]. Z powodzeniem zastosowano je także w badaniach nad cząstkami elementarnymi oraz w analizie obrazów mikroskopowych, umożliwiając tworzenie trójwymiarowych map ludzkiego mózgu [50].

[1] http://titancity.com/articles/gfxcards.html

[2] https://studiadecultura.uken.krakow.pl/article/download/4353/4090/14473

[3] https://www.historyofdatascience.com/imagenet-a-pioneering-vision-for-computers/

Kolejnym krokiem milowym w DL było wprowadzenie architektury transformera w 2017 roku. Model ten, opisany w pracy „Attention Is All You Need" przez Ashish Vaswani wraz z zespołem [91], całkowicie zmienił podejście do przetwarzania sekwencji danych, eliminując konieczność stosowania rekurencyjnych struktur. Rozwój transformerów przyczynił się do powstania dużych modeli językowych (LLM), takich jak BERT, GPT-4 czy T5, które dzięki analizie ogromnych zbiorów tekstów nauczyły się generować spójne i kontekstowo trafne wypowiedzi. Obecnie LLM wspierają wiele obszarów, od automatycznego tłumaczenia, przez chatboty, po systemy wspomagania pisania i analizy dokumentów [9].

Współczesne zastosowania AI obejmują nie tylko medycynę, gdzie algorytmy wspierają decyzje diagnostyczne, lecz także teledetekcję, w której sztuczna inteligencja wspomaga monitorowanie zagrożeń naturalnych oraz ocenę zmian środowiskowych, zwiększając tym samym skuteczność działań prewencyjnych i analitycznych. Reasumując, rozwój sztucznej inteligencji przeszedł długą drogę – od pierwszych modeli sieci neuronowych po zaawansowane algorytmy głębokiego uczenia. Charakteryzuje się dynamicznym postępem technologicznym i coraz szerszym spektrum zastosowań. Dzięki zaawansowanemu przetwarzaniu danych AI umożliwia nie tylko automatyzację procesów, ale także eksplorację nowych granic wiedzy i technologii, zmieniając sposób, w jaki ludzie postrzegają i rozwiązują problemy. Jednym z kluczowych obszarów napędzających ten rozwój jest rozpoznawanie obrazów – technologia, która leży u podstaw wielu nowoczesnych systemów, od diagnostyki medycznej po autonomiczne pojazdy. Co więcej, analiza jakości obrazu odgrywa coraz większą rolę zarówno w ocenie poprawności przetwarzania danych wizualnych, jak i w optymalizacji algorytmów wykorzystywanych do ich prezentacji i interpretacji. Z tego powodu atrakcyjne jest tworzenie metod oceniających obrazy, uwzględniając tzw. jakość percepcyjną, w oparciu o cechy wrażliwe na jakość lub mając do dyspozycji oceny subiektywne dla obrazów. Dotychczasowe osiągnięcia to jedynie wstęp do potencjału sztucznej inteligencji, której rozwój kształtuje przyszłość i wymaga dalszych badań, udoskonaleń oraz nowych zastosowań.

1.2. Hipoteza badawcza

Głównym celem pracy jest opracowanie i przedstawienie automatycznych metod percepcyjnej oceny jakości obrazów cyfrowych bez referencji z wykorzystaniem sieci neuronowych. Na podstawie tego założenia sformułowano następującą hipotezę badawczą:

> *Zastosowanie metod sztucznej inteligencji, w szczególności różnorodnych architektur sieci neuronowych, umożliwia ocenę jakości obrazów satelitarnych oraz obrazów rezonansu magnetycznego bez referencji, zapewniając wysoką zgodność z subiektywnymi ocenami ekspertów. Opracowane metody nie ustępują pod względem skuteczności predykcji jakości rozwiązaniom znanym z literatury.*

W celu potwierdzenia postawionej hipotezy sformułowano następujące zadania szczegółowe:

1) Analiza literatury dotyczącej najnowszych metod oceny jakości obrazów rezonansu magnetycznego, z uwzględnieniem typowych zniekształceń, protokołów zastosowań oraz wykorzystywanych baz danych.

2) Opracowanie narzędzi do percepcyjnej oceny jakości wyostrzonych obrazów satelitarnych oraz platformy do oceny jakości obrazów rezonansu magnetycznego.

3) Zastosowanie głębokich sieci neuronowych w ocenie jakości obrazów satelitarnych i medycznych bez referencji, w celu opracowania metod zgodnych z oceną subiektywną i konkurencyjnych względem rozwiązań literaturowych, z wykorzystaniem:

   - zewnętrznego połączenia architektur głębokich splotowych sieci neuronowych,
   - wewnętrznego połączenia dwóch architektur sieci,
   - struktury sieci złożonej z trzech podsieci,
   - informacji wyodrębnionych z warstw sieci.

4) Przeprowadzenie eksperymentów z wykorzystaniem opracowanych metod, porównanie wyników z metodami z literatury oraz analiza i interpretacja rezultatów.

Osiągnięcie celu dysertacji wymaga realizacji powyższych zadań, wiążąc się z pozyskaniem, przetworzeniem i analizą danych, a także sformułowaniem szczegółowych problemów badawczych. Proces ten obejmuje wybór odpowiednich narzędzi informatycznych do ich rozwiązania, opracowanie dedykowanego oprogramowania oraz przeprowadzenie testów porównawczych z analogicznymi metodami z literatury. Interpretacja uzyskanych wyników umożliwia sformułowanie końcowych wniosków.

2. Metody sztucznej inteligencji w opracowaniu metod percepcyjnej oceny jakości obrazów

W niniejszym rozdziale omówiono nowoczesne podejścia do oceny jakości obrazów rezonansu magnetycznego oraz wyostrzonych obrazów satelitarnych. Kolejno w podrozdziałach przedstawiono wyniki przeprowadzonej analizy literaturowej, w tym omówienie typowych zniekształceń, protokołów oceny oraz baz danych stosowanych w pracy nad oceną obrazów MR. Następnie opisano zastosowanie głębokich sieci neuronowych do oceny jakości obrazów, uwzględniając ich różnorodne połączenia, modyfikacje, czy wykorzystanie informacji wyodrębnionych z warstw sieci. Ostatnia część rozdziału prezentuje utworzone narzędzia do kompleksowej oceny jakości obrazów rezonansu magnetycznego i wyostrzonych obrazów satelitarnych.

2.1. Metody oceny jakości obrazu bez odniesienia dla obrazów rezonansu magnetycznego

Rezonans magnetyczny wyróżnia się na tle innych metod obrazowania medycznego, takich jak tomografia komputerowa, rentgen czy ultrasonografia, dzięki wysokiemu poziomowi bezpieczeństwa oraz zdolności do generowania szczegółowych obrazów w różnych płaszczyznach. Istotą jego działania jest oddziaływanie silnego pola magnetycznego na protony w tkankach, co pozwala na uzyskanie precyzyjnych obrazów opartych na zjawiskach relaksacji T1 i T2 [79]. Obrazy T1-zależne charakteryzują się krótkim czasem echa (TE) i krótkim czasem powtórzenia (TR). W takich sekwencjach sygnał pochodzący od tkanek bogatych w tłuszcz, jest zazwyczaj jasny, natomiast woda i płyny – ciemne. Dzięki temu obrazy T1-ważone dobrze oddają szczegóły anatomiczne oraz pozwalają na ocenę struktur zawierających tłuszcz. Z kolei obrazy T2-zależne uzyskiwane są przy użyciu dłuższego czasu echa i powtórzenia. W tych sekwencjach sygnał płynów – w tym wody – jest jasny, co ułatwia identyfikację patologicznych zmian takich jak obrzęki, stany zapalne czy zmiany nowotworowe. Kluczowe parametry techniczne obrazowania MR, takie jak czas echa oraz czas powtórzenia, mają istotny wpływ na końcowy wygląd obrazów i ich zastosowanie diagnostyczne. Czas echa określa okres między środkami impulsu wzbudzającego, a rejestracją sygnału echa, a jego wartość wpływa na kontrast obrazu – krótsze wartości stosuje się w obrazach T1-ważonych, natomiast dłuższe w obrazach T2-ważonych. Czas powtórzenia określa odstęp między kolejnymi impulsami wzbudzającymi, a jego długość decyduje o tym, czy obrazowanie będzie bardziej czułe na relaksację podłużną czy poprzeczną. Oprócz podstawowych obrazów T1 i T2-ważonych stosowane są również inne techniki obrazowania MR, pozwalające na uzyskanie bardziej szczegółowych informacji o strukturach anatomicznych i funkcjonalnych. Obrazy FLAIR, dzięki tłumieniu sygnału płynu mózgowo-rdzeniowego, umożliwiają lepsze uwidocznienie zmian patologicznych w mózgu, zaś obrazy DWI pozwalają na ocenę ruchu cząsteczek wody w tkankach, co jest niezwykle przydatne w diagnostyce udarów i guzów. Jednakże występowanie artefaktów i zniekształceń wynikających z operacji sygnałowych, cech

sprzętu, jak i potencjalnych błędów operatora mogą wpływać na jakość uzyskiwanych obrazów, a tym samym na poprawność diagnostyki. W związku z tym kluczowe staje się rozwijanie automatycznych metod oceny jakości obrazu, które mogą ograniczyć lub nawet zastąpić subiektywną ocenę człowieka.

Algorytmy oceny jakości obrazu (IQA) dzielą się na metody z pełną referencją (FR), zredukowaną referencją (RR) oraz bez referencji (NR) [95]. W kontekście obrazowania medycznego szczególne znaczenie mają metody NR pozwalające na ocenę jakości obrazu bez odniesienia do obrazu wzorcowego, który najczęściej nie istnieje. Rozwój tych technik, uwzględniających zarówno subiektywne opinie ekspertów, jak i w pełni obiektywne podejścia oparte na analizie statystycznej i uczeniu maszynowym, ma kluczowe znaczenie dla przyszłości diagnostyki obrazowej [12].

Metody oceny jakości obrazu porównuje się za pomocą miar statystycznych, które pozwalają ocenić zgodność wyników algorytmów z ocenami subiektywnymi. Do najczęściej stosowanych wskaźników należą: współczynnik korelacji rang Spearmana (SRCC), współczynnik korelacji rang Kendalla (KRCC), współczynnik korelacji liniowej Pearsona (PLCC) oraz pierwiastek średniego błędu kwadratowego. Każda z tych miar charakteryzuje się innymi właściwościami i pozwala na analizę różnych aspektów zależności między ocenami metody a ocenami referencyjnymi. Współczynnik korelacji rang Spearmana jest miarą monotonicznej zależności między dwiema zmiennymi, co oznacza, że pozwala ocenić, w jakim stopniu wartości jednej zmiennej rosną lub maleją wraz ze wzrostem drugiej. Jest to miara nieparametryczna, czyniąc ją odporną na wpływ wartości odstających i umożliwia jej zastosowanie w przypadku danych, które nie spełniają założeń normalności rozkładu. W kontekście oceny jakości obrazu wskaźnik SRCC wykorzystywany jest do określenia, na ile porządek ocen przyznanych przez model odpowiada kolejności ocen subiektywnych użytkowników. Podobną funkcję pełni współczynnik korelacji rang Kendalla, który również ocenia monotoniczną zależność między zmiennymi, ale bazuje na liczbie par obserwacji o zgodnym i niezgodnym uporządkowaniu. Jego zastosowanie jest szczególnie korzystne w przypadku małych zbiorów danych oraz gdy wiele obserwacji ma identyczne wartości, ponieważ jest mniej podatny na zakłócenia wynikające z takich powtórzeń. Innym istotnym wskaźnikiem jest współczynnik korelacji liniowej Pearsona, który mierzy siłę i kierunek liniowej zależności między dwiema zmiennymi. W przeciwieństwie do SRCC i KRCC, wskaźnik ten uwzględnia nie tylko kolejność wartości, ale także wielkość różnic między nimi — stopień, w jakim wartości jednej zmiennej zmieniają się liniowo względem drugiej. W przypadku metod IQA współczynnik Pearsona pozwala określić, jak dokładnie model odwzorowuje wartości ocen jakości obrazu. Oprócz współczynników korelacji stosuje się również pierwiastek średniego błędu kwadratowego (RMSE), który ocenia średnią wielkość błędów popełnianych przez model względem wartości referencyjnych. Wskaźnik ten mierzy różnicę między wartościami przewidywanymi a rzeczywistymi, wyrażoną w tych samych jednostkach, co oceniana cecha. Niższa wartość tego wskaźnika świadczy o lepszej dokładności modelu, gdyż oznacza mniejsze błędy predykcji.

Wszystkie te miary stosowane są w celu oceny skuteczności metod IQA i ich zdolności do przewidywania jakości obrazu w sposób zgodny z subiektywnymi ocenami użytkowników. Wysokie wartości SRCC, KRCC i PLCC oraz niskie RMSE wskazują na lepszą jakość algorytmu i jego większą zgodność z ludzką percepcją jakości obrazu [28]. Przykłady wybranych zniekształceń przedstawiono na Rys. 2.1 zaczerpniętym z [A-1].

Obrazy MRI mogą być obarczone różnymi zniekształceniami, które wpływają na ich jakość i przydatność diagnostyczną. Artefakty te mogą wynikać m.in. z zakłóceń w przestrzeni k, ruchu pacjenta, ograniczeń sprzętowych, różnic w czułości cewek, obecności metalu czy błędów rekonstrukcji [8, 10, 19, 29, 40, 42, 63, 71, 98, 103].

Ocena jakości obrazów medycznych, w tym MRI, odgrywa kluczową rolę w diagnostyce, ponieważ może bezpośrednio wpływać na dokładność interpretacji oraz podejmowanych decyzji klinicznych. Dlatego też rozwój metod automatycznej analizy jakości staje się nieodzowny w celu usprawnienia procesów diagnostycznych i eliminacji skanów o niewystarczającej jakości. Metody NR-IQA są stosowane w przypadkach, gdy nie ma dostępu do obrazu referencyjnego. Bhateja i in. [7, 23] zaproponowali metodę oceny opartej na ludzkim systemie wzrokowym (HVS), wykorzystującą fuzję dwóch obrazów w celu poprawy detekcji zniekształceń. Xu i in. [100] opracowali technikę częściowo nadzorowaną do oceny jakości obrazów MRI mózgu płodów, używając metody „mean teacher" oraz spójności w ROI. Liu i in. [47] zastosowali częściowo nadzorowane uczenie do rozwiązania problemu szumów w adnotacjach zadań segmentacji obrazów. Ich metoda hierarchicznej oceny jakości uwzględnia analizę na poziomie warstw, objętości oraz całego obiektu. Inne podejście wykorzystuje niesparowane generatywne sieci przeciwstawne (GAN) oraz słabo nadzorowany klasyfikator [37]. Z kolei HyS-net, wykorzystujący adaptacyjną architekturę przestrzenną zaprojektowano tak, by efektywnie wykorzystywać trójwymiarową informację przestrzenną [74]. Natomiast Chow i Rajagopal zmodyfikowali metodę BRISQUE, dokonującą oceny obrazu poprzez analizę statystyk sceny i normalizację kontrastu z odjętą średnią (MSCN, ang. Mean Subtracted Contrast Normalization) [13]. W Tabeli 2.1. zaczerpniętej z [A-1], przedstawiono porównanie różnych metod NR-IQA dla obrazów MRI. W kontekście prowadzonych badań kluczowe okazały się metody oceny jakości obrazu bazujące na cechach takich jak entropia oraz reprezentacje pochodzące z sieci neuronowych, ze względu na ich wysoką skuteczność w ocenie zniekształceń istotnych z diagnostycznego punktu widzenia. Ich walidacja, choć zazwyczaj oparta na ograniczonych bazach danych tworzonych przy udziale specjalistów, dostarczyła cennych wskazówek do opracowania nowych podejść.

Analiza dostępnych metod oceny jakości MRI wykazała, że metody automatycznej oceny jakości obrazów są szczególnie istotne w zastosowaniach medycznych.

[4] radiopaedia.org

Tabela 2.1. Porównanie metod NR-IQA pod względem stosowanych technik, cech oraz wykorzystywanych zbiorów danych.

Metoda	**Podejście i cechy**	**Liczba cech**	**Baza obrazów**
Dwustopniowa automatyczna ocena jakości obrazów rezonansem magnetycznym wątroby oparta na splotowej sieci neuronowej [94]	- Strategia oparta na wycinkach obrazu - Splotowa sieć neuronowa w segmentacji obszarów na obrazie - Obszar zainteresowania (ROI)	-	Nie określono w artykule
Uczenie półnadzorowane do oceny jakości obrazów rezonansem magnetycznym płodu z zachowaniem spójności obszaru zainteresowania [100]	- Uczenie półnadzorowane - Spójność obszaru zainteresowania - Mean teacher model	-	Skanowanie wykonano w Boston Children's Hospital
Ocena jakości obrazów T2-zależnych rezonansu magnetycznego bez odniesienia u pacjentów z rakiem prostaty [54]	- Generatywna sieć przeciwstawna - Uczenie słabo nadzorowane - Wytrenowany głęboki klasyfikator	-	- National Cancer Institute (NCI) - PIE-AAPM-NCI Prostate MR Gleason Grade Group Challenge - NIH Clinical Center
Metoda fuzji obrazów MR w dwóch etapach oparta na modelu logarytmiczno-parametrycznym przetwarzania obrazów (PLIP) [89]	- Fuzja obrazów MRI w dwóch etapach - PCA i PLIP - Stationary Wavelet Transform	-	- Whole Brain Atlas [17]
Hierarchiczna nielokalna sieć rezydualna do oceny jakości obrazów pediatrycznych MRI z ograniczonymi zniekształconymi adnotacjami [48]	- Ocena jakości obrazów pacjentów dla poszczególnych przekrojów i objętości - Nielokalna sieć rezydualna - Uczenie półnadzorowane	-	- Baza danych z Centrum Badań Rezonansu Magnetycznego (CMRR) na Uniwersytecie Minnesoty

Kontynuacja na następnej stronie

Tabela 2.1 – kontynuacja

Metoda	**Podejście i cechy**	**Liczba cech**	**Baza obrazów**
HyS-net [75]	- Sieć nadzorująca dostosowana do treści - Przestrzenna ekstrakcja cech - Predykcja jakości oparty na sieci	-	- Otwarty zbiór danych, MRIQC [21]
QEMDIM [38]	- Różnice cech statystycznych między obrazami testowymi - Współczynniki MSCN - MDFC – współczynniki filtrowania w wielu kierunkach	20	- ADNI [62] - ABIDE [16]
AQASB [60]	- Zniekształcenia tła - Odpowiedni poziom wokseli tła	-	- ADNI [62]
Multi-class cardiovascular magnetic resonance image quality assessment using unsupervised domain adaptation [6, 61]	- Adaptacja domeny bez nadzoru - Analiza w dziedzinie przestrzennej i częstotliwościowej - Manipulacja przestrzenią k	512	- UK Biobank - Zbiór danych MRI serca, York University [4]
MRIQC [20]	- Miary jakości - Klasyfikator binarny	64	- ABIDE [16] - OpenfMRI [73]
Obrazy MRI mózgu i serca w wieloośrodkowych badaniach klinicznych [66]	- Zachowanie momentów statystycznych - Pomiar różnic kontrastu tekstury	W zależności od obrazu	- NeuroRx research Inc. - BrainCare Oy - ADNI [62] - Department of Diagnostic Imaging of the Hospital for Sick Children in Toronto
Modified-BRISQUE [14]	- Luminancja, cechy obrazu - NSS - SVR	36	- Baza obrazów DICOM Sirix Viewer - Obrazy rezonansem magnetycznym z Uniwersytetu w Sheffield
R50GR18 [83]	- Fuzja architektur głębokich sieci - SVR	3584	- DB1 [70] - DB2 [65]

Kontynuacja na następnej stronie

Tabela 2.1 – kontynuacja

Metoda	**Podejście i cechy**	**Liczba cech**	**Baza obrazów**
ENMIQA [65]	- Progowane lokalne różnice intensywności uzyskane przez NMS - Entropia sekwencji ekstremów	1	- DB1 [70]
NOMRIQA [70]	- Cechy FAST - Histogramy binarnych deskryptorów - SVR	3840	- Symulowana Baza Danych Mózgu (SBD) [43] - DB1 [70]

Metody te są wrażliwe na obecność zniekształceń takich jak szumy, kompresja czy rozmycie, co znacząco wspiera proces diagnostyczny. Nowoczesne podejścia w tej dziedzinie coraz częściej wykorzystują zaawansowane techniki głębokiego uczenia, w tym hybrydowe sieci neuronowe i algorytmy częściowo nadzorowane, które dostosowują rozwiązania znane z oceny obrazów naturalnych do specyficznych potrzeb obrazowania medycznego. Opracowanie nowych metod wymaga jednak dostępu do dużych baz danych obrazów MRI, które zawierają subiektywne oceny jakości wykonane przez specjalistów. Brak takich baz danych stanowi jedno z największych wyzwań, którego rozwiązanie mogłoby znacząco przyspieszyć rozwój tej dziedziny. Artykuł zwraca uwagę na różnorodność metod – od oceny jakości na poziomie pojedynczych przekrojów po analizę pełnych obrazów 3D.

Udział własny autora w przygotowaniu artykułu dotyczącego oceny jakości obrazów rezonansu magnetycznego obejmował analizę istniejących metod oraz wyzwań związanych z oceną jakości obrazów NR, a także identyfikację potencjalnych kierunków dalszego rozwoju, jak na przykład wykorzystanie technik głębokiego uczenia.

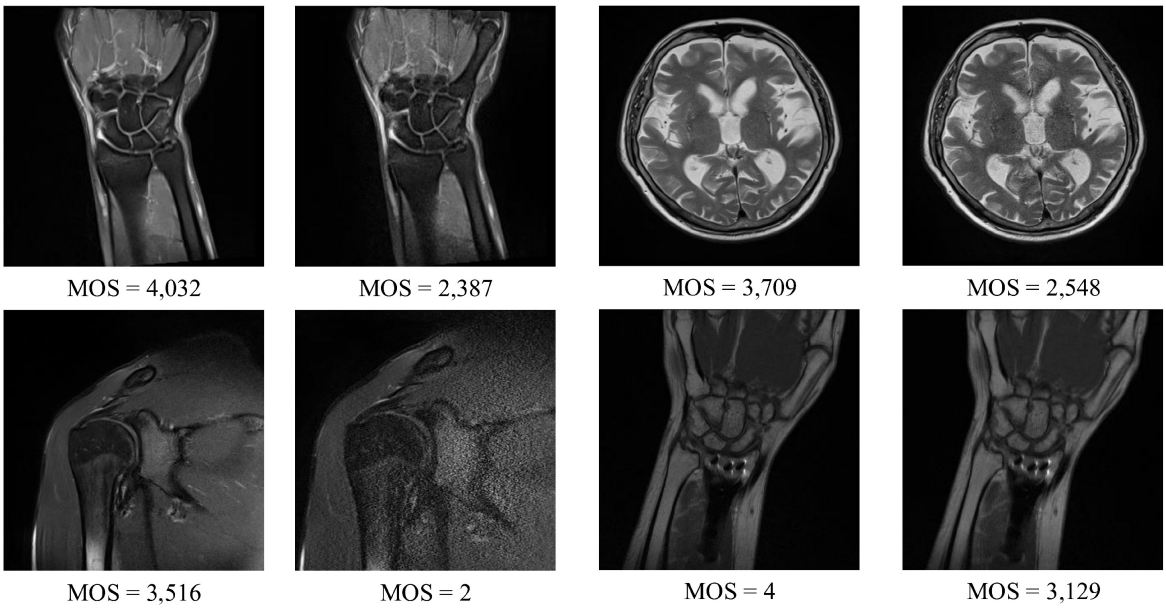
2.2. Zastosowanie fuzji głębokich splotowych sieci neuronowych do oceny jakości obrazów rezonansu magnetycznego

Rozwój technik IQA zależy od dostępności ocenionych baz obrazów, co jest szczególnie istotne w przypadku metod MR IQA wymagających opinii reprezentatywnej grupy radiologów [102]. Jednym z kluczowych problemów badawczych w tej dziedzinie jest opracowanie skutecznej metody oceny jakości obrazów MR, która będzie precyzyjna, powtarzalna i niezależna od subiektywnych ocen ekspertów. Dotychczasowe podejścia często adaptowały rozwiązania stosowane w analizie obrazów naturalnych, wzbogacając je o dodatkowe cechy lub techniki przetwarzania takie jak analiza entropii lokalnych ekstremów czy specyficzne filtrowanie [64, 69, 102]. Jednakże metody te nie zawsze uwzględniają unikalne cechy obrazów rezonansu magnetycznego. Z tego względu celem

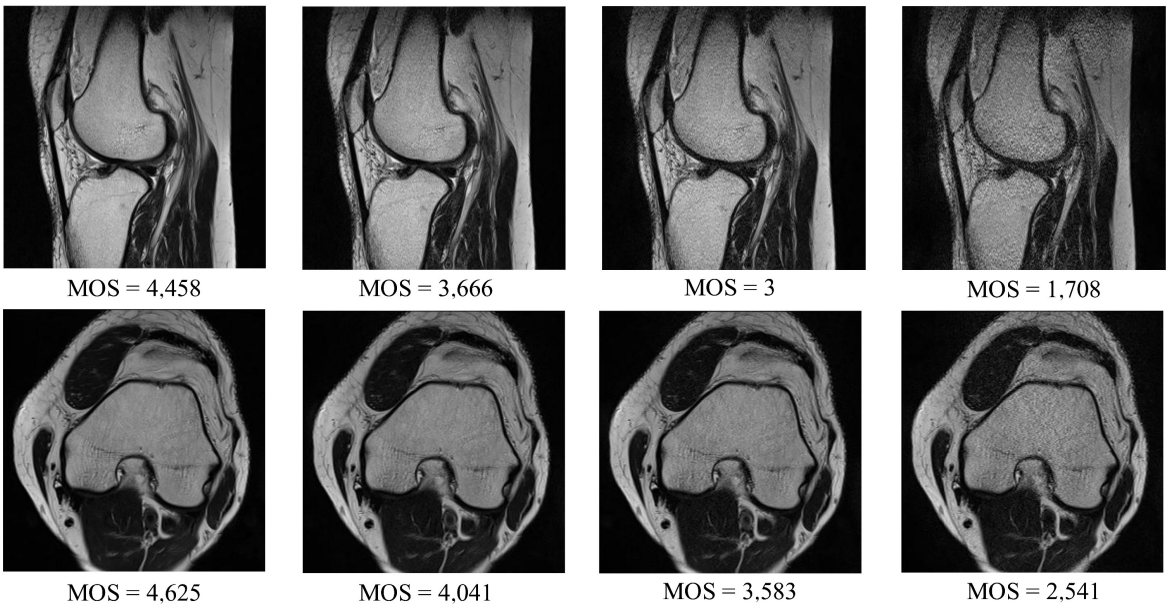
autora niniejszej pracy [A-2], było opracowanie metody dedykowanej obrazom MR.

W proponowanej metodzie zastosowano fuzję różnych architektur sieci neuronowych, szczególnie tych dedykowanych do rozpoznawania obrazów, które są szeroko wykorzystywane w zadaniach wizji komputerowej. Początkowo zastosowano pojedyncze modele, jednak ich wydajność w zadaniu MR IQA okazała się niewystarczająca. Z tego powodu wprowadzono fuzję sieci, próbując uchwycić zróżnicowane charakterystyki obrazów MR. Wśród sieci wybranych w celach porównawczych znalazły się: VGG [80], ResNet [31], GoogLeNet [87], MobileNet [78] oraz DenseNet [36, 88]. Sieci wstępnie przetrenowane na zbiorze ImageNet zostały zmodyfikowane, ich końcowe warstwy przeznaczone do klasyfikacji zastąpiono warstwami do regresji. Wyniki uzyskiwane przez różne sieci połączono za pomocą warstwy łączenia cech, co pozwalało na równoczesne uwzględnienie specyfiki każdej architektury sieci. Do predykcji jakości wykorzystano SVR, modelując relacje między cechami uzyskanymi z sieci a ocenami MOS. Obrazy wejściowe były w razie potrzeby dopasowywane do wymagań sieci, a następnie przechodziły proces ekstrakcji cech, fuzji wyników i uczenia. W treningu zastosowano metodę SGDM z pięcioma epokami. Ze względu na ograniczoną liczbę obrazów w benchmarkach MR IQA zastosowano uczenie transferowe oraz augmentację danych obejmującą rotacje obrazów w zakresie 360° z krokiem o 3° [53]. Fuzja została przetestowana w trzech wariantach: ResNet-50_GoogLeNet_ResNet-18 (R50GR18), ResNet-50_GoogLeNet_MobileNet-V2 (R18GR50M) oraz MobileNetV2_ResNet-50 (MR50). Modele te porównano z 17 metodami: NFERM [27], SEER [68], DEEPIQ [52], MEON [51], SNRTOI [105], NOREQI [67], BPRI [57], HOSA [99], NOMRIQA [69], IL-NIQE [104], GM-LOG [101], GWH-GLBP [46], BRISQUE [59], SISBLIM [26], metricQ [106], SINDEX [44] oraz ENMIQA [64]. NOMRIQA, ENMIQA i SNRTOI dedykowane są obrazom MR, zaś DEEPIQ i MEON to metody oparte na głębokim uczeniu, stworzone do oceny obrazów naturalnych. Ewaluacje przeprowadzono na dwóch zbiorach danych przy użyciu średnich wartości typowych w dyscyplinie. Eksperymenty realizowano w 10 powtórzeniach z losowym podziałem obrazów na 80% do trenowania i 20% do testów, uniemożliwiając powtarzanie się tych samych obrazów w obu zestawach. Wyniki wskazały na wyższą skuteczność fuzji w porównaniu do innych metod. Główne kroki obliczeniowe zaproponowanej metody przedstawiono na Rys. 2.4 ilustrującym całościowy przebieg procesu – od wejścia danych po końcową predykcję jakości. W pierwszym etapie (uczenie A) połączone sieci neuronowe są trenowane wspólnie w celu uchwycenia charakterystycznych cech obrazów MRI. Następnie, w etapie drugim (uczenie B), trenowany jest moduł regresji na podstawie sklejonych map cech, co prowadzi do utworzenia modelu predykcyjnego oceniającego jakość obrazów.

Przedstawiane podejście zostało ocenione na dwóch zbiorach danych MR IQA. Pierwszy, oznaczony jako DB1, zawiera 70 obrazów MR, a drugi, DB2, opracowany w ramach tego badania, obejmuje 240 obrazów MR [65, 70]. Przykładowe obrazy obu zbiorów zaprezentowano na Rys. 2.2. i 2.3 [A-2].



Rys. 2.2. Przykładowe obrazy rezonansu magnetycznego (MR) i ich subiektywne oceny w bazie obrazów DB1.

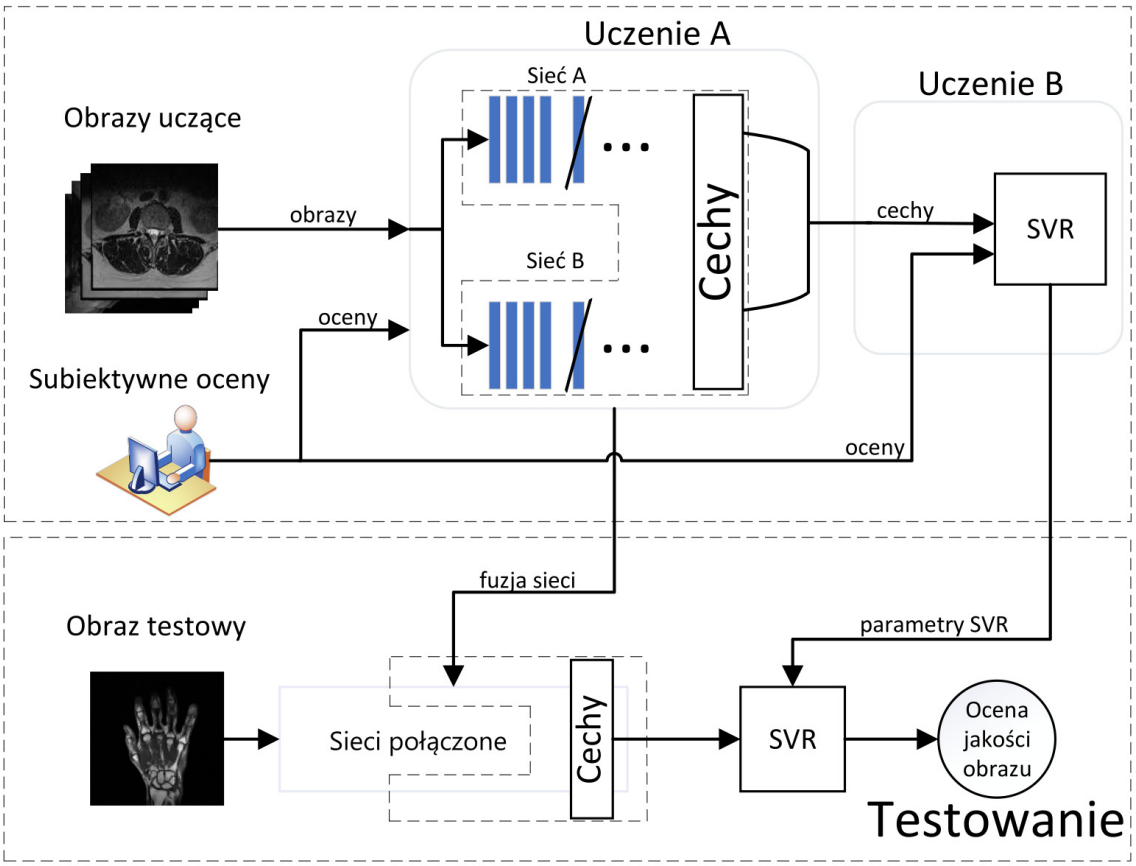


Rys. 2.3. Przykładowe obrazy dla bazy DB2 i ich subiektywne oceny.

Zbiór DB1 składa się z obrazów T2 z sekwencji wagowych uzyskanych za pomocą skanera 1.5T, obejmujących różne części ciała, takie jak kręgosłup, kolano, ramię, czy mózg. Rozdzielczość obrazów wynosi od 192 × 320 do 512 × 512 pikseli. Z uwagi na ograniczoną wielkość DB1 oraz brak dostępnych baz danych MR IQA odpowiednio ocenionych przez radiologów, w niniejszym badaniu stworzono nowy zbiór DB2. Obejmuje on obrazy MR ważone metodą T2 pozyskane w rutynowych badaniach diagnostycznych różnych partii ciała, wykonane skanerem Siemens Essensa 1.5T. Wykorzystano technikę obrazowania równoległego, GRAPPA, aby uzyskać obrazy o kontrolowanej degradacji

jakości. Oceny przeprowadzono wśród aż 24 doświadczonych radiologów, którzy przyznawali oceny od 1 do 5 odzwierciedlające jakość obrazów. Ostatecznie uzyskano oceny MOS.

Wyniki dla obu baz danych (Tabela 2.2, zaczerpnięta z [A-2]) pokazują, że zaproponowane techniki fuzji sieci przewyższają nowoczesne podejścia MR IQA. Modele R50GR18 i MR50 uzyskały najlepsze wyniki SRCC i KRCC dla pierwszej bazy, przewyższając NOMRIQA. Metody ENMIQA i SNRTOI wykazały słabszą korelację ze względu na brak uczenia i zaawansowanych algorytmów. Wizualizacja cech przedstawiona na Rys. 2.5 pokazuje, że warstwy w połączonych sieciach wyraźnie reagują na zniekształcenia. Obecność drugiej sieci w procesie treningu wpływa na sposób, w jaki sieci przetwarzają informacje, co odróżnia je od pojedynczych architektur. Przykładem tego jest fuzja GoogLeNet+ResNet-18, w której cechy obu sieci stają się bardziej podobne do siebie niż do swoich pojedynczych wersji. Fuzja sieci wykazuje również dużą wrażliwość na stopień zniekształcenia, co może ułatwiać predykcję jakości obrazów tego typu.

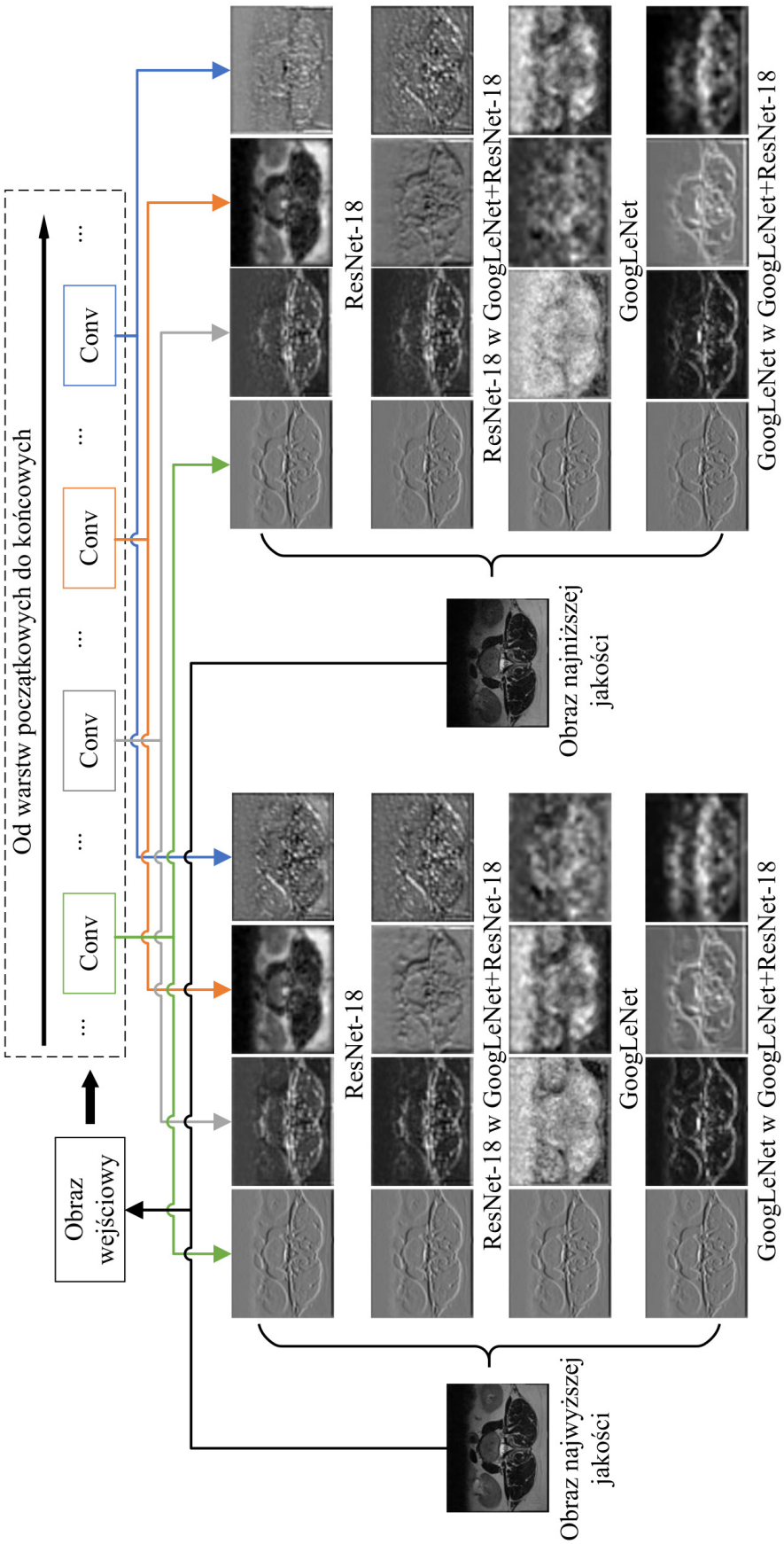


Rys. 2.4. Schemat blokowy proponowanego podejścia.

Tabela 2.2. Porównanie wydajności siedemnastu ocenianych metod na zestawach danych rezonansu magnetycznego.

Metody	DB1				DB2				Overall			
	SRCC	**KRCC**	**PLCC**	**RMSE**	**SRCC**	**KRCC**	**PLCC**	**RMSE**	**SRCC**	**KRCC**	**PLCC**	**RMSE**
NOMRIQA*	**0.7030**	**0.5527**	0.7978	0.4322	0.8040	0.6087	0.8737	0.4605	0.7535	0.5807	0.8358	0,4464
HOSA	0,4804	0,3909	0,6997	0,5318	0,8756	0,7052	**0,9276**	**0,3388**	0,6780	0,5481	0,8137	0,4353
NOREQI	0,4359	0,2922	**0,8045**	**0,4182**	0,8675	0,6984	0,9072	0,3833	0,6517	0,4953	0,8559	0,4008
IL-NIQE	0,1695	0,1275	0,3619	0,6674	0,1197	0,0836	0,3090	0,8821	0,1446	0,1056	0,3355	0,7748
GM-LOG	0,4673	0,3424	0,6515	0,4779	0,8854	0,7123	0,9010	0,4091	0,6764	0,5274	0,7763	0,4435
GWHGLBP	0,5075	0,3935	0,6886	0,5257	0,8726	0,6927	0,8947	0,4080	0,6901	0,5431	0,7917	0,4669
BRISQUE	0,4610	0,3648	0,6100	0,5311	0,8544	0,6738	0,8951	0,4076	0,6577	0,5193	0,7526	0,4694
SISBLIM	0,3976	0,2776	0,6240	0,5449	0,7216	0,5419	0,7592	0,6047	0,5596	0,4098	0,6916	0,5748
metricQ	0,2596	0,1657	0,2792	0,6709	0,5066	0,3701	0,5227	0,7791	0,3831	0,2679	0,4010	0,7250
BPRI	0,2412	0,1890	0,4785	0,5756	0,1317	0,0973	0,4928	0,7883	0,1865	0,1432	0,4857	0,6820
SINDEX	0,2939	0,2112	0,3243	0,7034	0,2673	0,1933	0,3185	0,8874	0,2806	0,2023	0,3214	0,7954
NFERM	0,5073	0,4091	0,7662	0,4491	0,8833	0,7087	0,9157	0,3872	0,6953	0,5589	0,8410	0,4182
SEER	0,4776	0,3574	0,7108	0,5267	**0,8938**	**0,7335**	0,9196	0,3594	0,6857	0,5455	0,8152	0,4431
MEON	0,2518	0,1879	0,3439	0,6428	0,5851	0,4001	0,6194	0,7426	0,4185	0,2940	0,4817	0,6927
DEEPIQ	0,1133	0,0827	0,5902	0,5707	0,2837	0,2078	0,5393	0,7822	0,1985	0,1453	0,5648	0,6765
SNRTOI*	0,1321	0,0728	0,4094	0,6784	0,1016	0,0720	0,3169	0,8930	0,1169	0,0724	0,3632	0,7857
ENMIQA*	0,3630	0,2479	0,5093	0,5873	0,7941	0,6119	0,8313	0,5130	0,5786	0,4299	0,6703	0,5502
*R18GR50M**	0,6299	0,5012	0,7999	0,4381	**0,8998**	**0,7398**	**0,9270**	**0,3465**	**0,7649**	**0,6205**	**0,8635**	**0,3923**
*R50GR18**	**0,7423**	**0,6039**	**0,8206**	**0,4238**	**0,9083**	**0,7490**	**0,9294**	**0,3479**	**0,8253**	**0,6765**	**0,8750**	**0,3859**
*MR50**	**0,7036**	**0,5740**	**0,8576**	**0,3865**	0,8919	0,7176	0,9241	0,3560	**0,7978**	**0,6458**	**0,8909**	**0,3712**

* oznacza podejście zaprojektowane do oceny jakości obrazów MR. Trzy najlepsze wyniki dla każdego kryterium są wyróżnione pogrubioną czcionką, nazwy metod powstałych w wyniku łączenia sieci, wprowadzonych w tym artykule, są oznaczone kursywą.

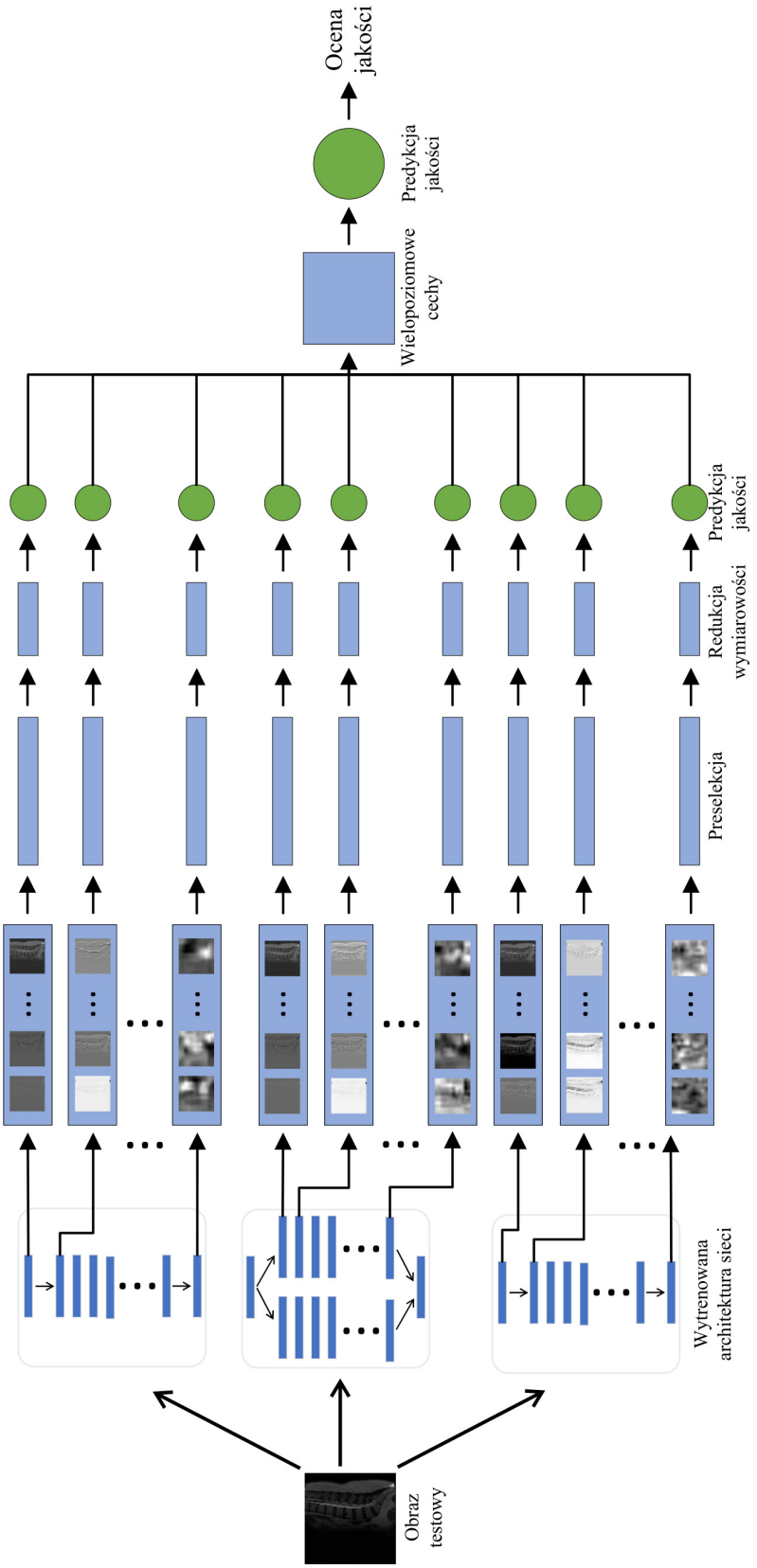


Rys. 2.5. Przedstawienie cech na różnych warstwach dla przykładowych obrazów MR najlepszej i najgorszej jakości. GoogLeNet, ResNet-18 i ich fuzja.

Udział własny autora niniejszego artykułu polegał na współtworzeniu koncepcji metody, doborze i implementacji odpowiednich algorytmów sztucznej inteligencji, adaptacji oraz łączeniu wybranych architektur sieci neuronowych, ich trenowaniu i dalszym douczaniu na potrzeby zadania MRI IQA. Autor brał również udział w realizacji eksperymentów obliczeniowych, analizie uzyskanych wyników, przygotowaniu grafik oraz współredakcji tekstu. Opracowana metoda stanowiła podstawę do dalszych badań nad integracją modeli neuronowych w celu zwiększenia zgodności uzyskiwanych wyników z oceną ekspertów.

2.3. Metoda z fuzją architektur sieci głębokich z wykorzystaniem wielopoziomowej i wielomodelowej reprezentacji obrazu

Jednym z istotnych wyzwań w ocenie jakości obrazów MR jest ich podatność na różnorodne zniekształcenia [34], co motywuje poszukiwanie skutecznych metod różnicowania obrazów pod względem jakości [34]. Prezentowana metoda, w przeciwieństwie do wcześniej omawianego rozwiązania, charakteryzuje się odmiennym sposobem łączenia sieci, w którym sieci wzajemnie na siebie oddziałują. To zjawisko znacząco zwiększa efektywność modelu oceny jakości. Jak wykazano w pracy [A-2], architektury sieci różnią się pod względem czułości na zniekształcenia obrazów. Zatem można założyć, że cechy wyodrębniane zarówno w poszczególnych warstwach, jak i w zintegrowanych modelach, umożliwiają skuteczne różnicowanie obrazów o różnej jakości. Proponowana metoda, MEDQEMRI [A-3], różni się od innych podejść dzięki wewnętrznemu łączeniu sieci z tej samej rodziny, co umożliwia wzbogacenie ekstrakcji cech związanych z jakością obrazów [86]. Obie sieci, ResNet-18 (R18) oraz ResNet-50 (R50) [32], standardowo używane w zadaniach klasyfikacji obrazów, zostały dostosowane do problemu regresji poprzez wymianę ostatnich warstw na warstwę w pełni połączoną oraz warstwę regresji. Następnie, po uczeniu transferowym, przeprowadzono dodatkowy etap ponownego trenowania sieci w celu uwzględnienia specyfiki obrazów MR, jak przedstawiono na Rys. 2.6. W kontekście analizy obrazów MRI, wielopoziomowość odnosi się do wykorzystania informacji pochodzących z różnych warstw sieci neuronowych, które reprezentują cechy o zróżnicowanym poziomie abstrakcji — od prostych struktur po złożone wzorce. Z kolei wielomodelowość polega na integracji wielu architektur sieciowych, co pozwala na uchwycenie różnorodnych charakterystyk obrazu wynikających z odmiennej czułości poszczególnych modeli na zniekształcenia. Istotnym elementem omawianego podejścia jest ponowny trening z uwzględnieniem fuzji sieci, polegający na wprowadzeniu warstwy łączącej cechy z ostatnich warstw obu sieci. Dzięki temu sieci mogą wzajemnie na siebie oddziaływać podczas procesu uczenia, prowadząc do wzbogacenia informacji możliwych do wydobycia. Aby dostosować sieć ResNet do analizy obrazów MR, konieczne było przekształcenie skanów z 16-bitowych macierzy dwuwymiarowych do obrazów z trzema kanałami o wymiarach 224×224, co umożliwiło ich przetwarzanie przez sieć. Proces ten obejmował zmniejszenie głębi bitowej oraz połączenie kanałów, stanowiących podstawę skutecznej integracji modeli.

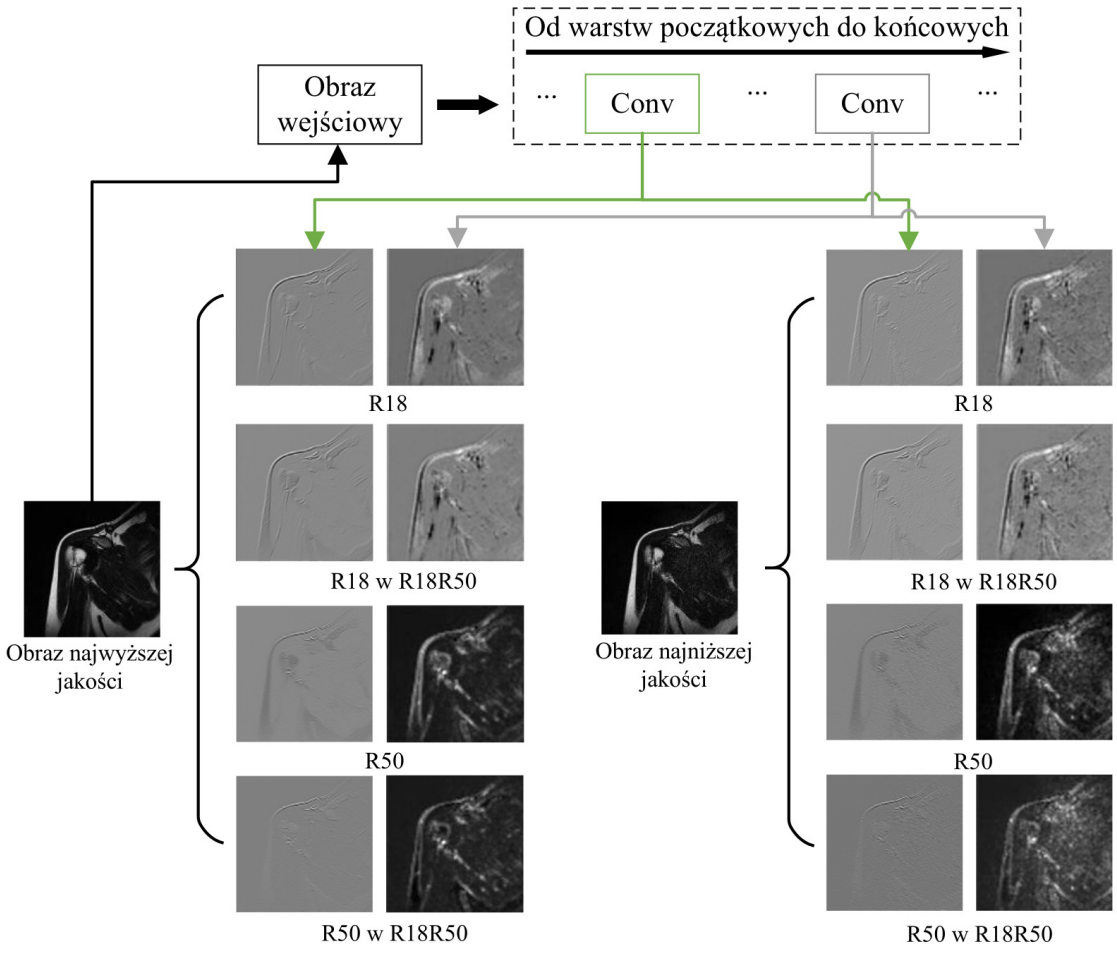


Rys. 2.6. Schemat proponowanego podejścia. Wielopoziomowe cechy pochodzące z warstw ponownie wytrenowanych architektur (R18, R18R50 i R50) są wstępnie wybierane i redukowane w celu utworzenia modeli jakości. Następnie ich wyniki są używane jako cechy wysokiego poziomu do trenowania modelu, który zapewnia ostateczną predykcję jakości.

Z wybranych warstw tych sieci wyodrębniono cechy, które następnie zostały wykorzystane w dalszym etapie przetwarzania. Ponieważ niektóre warstwy generują wektory o bardzo wysokim wymiarze, konieczne było zastosowanie wstępnej selekcji — cechy posortowano rosnąco według odchylenia standardowego, aby usunąć te o najmniejszym zróżnicowaniu. Tak przefiltrowane dane poddano redukcji wymiarowości za pomocą PCA, co pozwoliło uzyskać zwartą reprezentację przy zachowaniu istotnych informacji. Liczba zachowanych cech została dodatkowo dostosowana do ograniczeń pamięciowych środowiska obliczeniowego. Na tej bazie skonstruowano modele predykcyjne jakości, trenowane z wykorzystaniem hierarchicznej struktury sieci — podejście to umożliwiło uwzględnienie reprezentacji zarówno z niższych, jak i wyższych warstw, co pozwoliło lepiej uchwycić cechy istotne percepcyjnie. W efekcie powstał wysokopoziomowy model predykcyjny, zdolny do dokładnej oceny jakości obrazów. W przeciwieństwie do metod bazujących na pojedynczych warstwach takich jak te zaproponowane przez Kanga [39] czy Gao [24], MEDQEMRI wykorzystuje reprezentacje z wielu poziomów głębokości sieci, co przekłada się na wyższą skuteczność predykcji. Model został wytrenowany przy użyciu algorytmu SGDM oraz funkcji straty. Zastosowano współczynnik uczenia 10^{-4}, mini-batch size równy 32 oraz 5 epok. Ze względu na ograniczoną liczbę przykładów w zbiorze DB1 wprowadzono procedurę augmentacji danych.

Proces oceny jakości rozpoczyna się od przetworzenia obrazów MR przy użyciu trzech architektur sieciowych: R18, R50 oraz ich fuzji (R18R50). W badaniu wykorzystano dwie bazy obrazów MR: DB1 (70 obrazów ocenionych przez 31 radiologów) oraz DB2 (240 obrazów ocenionych przez 24 ekspertów). Obrazy podzielono zgodnie ze standardowym protokołem badawczym (80% do treningu, 20% do testowania, podział bazy był dziesięciokrotny, losowy). Skuteczność metody oceniano za pomocą korelacji SRCC, KRCC, PLCC oraz RMSE, porównując wyniki z ocenami MOS. Jak wskazują dane porównawcze z Tabeli 2.3 [A-3] metoda MEDQEMRI osiąga lepsze wyniki w większości metryk, takich jak SRCC, KRCC i PLCC, zarówno na danych DB1, jak i DB2, a także w zestawieniu ogólnym. Warto podkreślić, że dla bazy DB2 uzyskano wyjątkowo wysoką zgodność z medianą 10 wyników trenowania sieci (SRCC 0,9133, PLCC 0,9312). W celu statystycznej weryfikacji różnic zastosowano test istotności statystycznej Wilcoxona dla sparowanych prób. Uzyskany poziom istotności potwierdza, że różnice między wynikami dla modeli opartych na opartymi na pojedynczych sieciach a ich fuzją są statystycznie istotne. Rysunek 2.7 przedstawia wizualizację cech wyodrębnianych z kilku wybranych warstw sieci neuronowych — zarówno trenowanych osobno (R18, R50), jak i ich połączonej fuzji (R18R50). Zawarte obrazy cech ukazują jak różne architektury reagują na zniekształcenia i różnice w jakości obrazów MR. Celem tego rysunku jest zobrazowanie, że cechy generowane przez poszczególne sieci oraz ich fuzję różnicują obrazy o odmiennej jakości percepcyjnej. Warto zauważyć, że wartości i wzorce cech w warstwach sieci połączonej różnią się wyraźnie od tych uzyskanych w modelach trenowanych oddzielnie. Taka zmiana wskazuje na wzajemne oddziaływania sieci podczas procesu uczenia, które prowadzą do wzbogacenia reprezentacji obrazów. Analiza rysunku potwierdza,

że połączenie architektur R18 i R50 daje synergiczny efekt — modele uzupełniają się nawzajem, dostarczając bardziej zróżnicowanego, dyskryminatywnego opisu cech. Dzięki temu fuzja sieci przyczynia się do poprawy predykcji w zadaniu oceny jakości obrazów MRI, co znajduje odzwierciedlenie w lepszych wynikach modelu opartego na połączonych sieciach.



Rys. 2.7. Wizualizacja różnic w cechach trzech architektur sieciowych (R18, R50 i ich fuzji R18R50) na dwóch przykładowych warstwach dla obrazów MR o różnej jakości percepcyjnej

Wkład własny autora w przygotowanie niniejszego artykułu polegał na współautorstwie koncepcji nowatorskiej metody bezreferencyjnej oceny jakości obrazów MR opartej na fuzji dwóch komplementarnych architektur sieci neuronowych, wyborze odpowiednich algorytmów sztucznej inteligencji oraz adaptacji tych architektur do specyfiki obrazów MR. Ponadto autor uczestniczył w przeprowadzeniu eksperymentów obliczeniowych, analizie wyników opartych na danych z publicznych benchmarków oraz przygotowaniu grafik ilustrujących efekty metody.

Tabela 2.3. Porównanie wydajności metod NR na wzorcowych bazach danych.

Metoda	DB1				DB2				Overall			
	SRCC	**KRCC**	**PLCC**	**RMSE**	**SRCC**	**KRCC**	**PLCC**	**RMSE**	**SRCC**	**KRCC**	**PLCC**	**RMSE**
NOMRIQA	0,7030	0,5527	0,7978	0,4322	0,8040	0,6087	0,8737	0,4605	0,7812	0,5961	0,8566	0,4541
HOSA	0,4804	0,3909	0,6997	0,5318	0,8756	0,7052	**0,9276**	**0,3388**	0,7864	0,6342	0,8761	0,3824
IL-NIQE	0,1695	0,1275	0,3619	0,6674	0,1197	0,0836	0,3090	0,8821	0,1309	0,0935	0,3209	0,8336
GM-LOG	0,4673	0,3424	0,6515	0,4779	0,8854	0,7123	0,9010	0,4091	0,7910	0,6288	0,8447	0,4246
GWHGLBP	0,5075	0,3935	0,6886	0,5257	0,8726	0,6927	0,8947	0,4080	0,7902	0,6251	0,8482	0,4346
BRISQUE	0,4610	0,3648	0,6100	0,5311	0,8544	0,6738	0,8951	0,4076	0,7656	0,6040	0,8307	0,4355
SISBLIM	0,3976	0,2776	0,6240	0,5449	0,7216	0,5419	0,7592	0,6047	0,6484	0,4822	0,7287	0,5912
metricQ	0,2596	0,1657	0,2792	0,6709	0,5066	0,3701	0,5227	0,7791	0,4508	0,3239	0,4677	0,7547
BPRI	0,2412	0,1890	0,4785	0,5756	0,1317	0,0973	0,4928	0,7883	0,1564	0,1180	0,4896	0,7403
BLINDER	0,2756	0,1985	0,5122	0,5930	0,7037	0,5138	0,7379	0,6262	0,6070	0,4426	0,6869	0,6187
MEON	0,2518	0,1879	0,3439	0,6428	0,5851	0,4001	0,6194	0,7426	0,5098	0,3522	0,5572	0,7201
DEEPIQ	0,1133	0,0827	0,5902	0,5707	0,2837	0,2078	0,5393	0,7822	0,2452	0,1796	0,5508	0,7344
SNRTOI	0,1321	0,0728	0,4094	0,6784	0,1016	0,0720	0,3169	0,8930	0,1085	0,0722	0,3378	0,8445
ENMIQA	0,3630	0,2479	0,5093	0,5873	0,7941	0,6119	0,8313	0,5130	0,6968	0,5297	0,7586	0,5298
M1	0,6299	0,5012	0,7999	0,4381	**0,8998**	**0,7398**	0,9270	0,3465	0,8389	**0,6859**	0,8983	0,3672
M2	**0,7036**	**0,5740**	**0,8576**	**0,3865**	0,8919	0,7176	0,9241	0,3560	**0,8494**	0,6852	**0,9091**	**0,3629**
M3	0,6634	0,5194	**0,8299**	**0,4094**	0,8769	0,6999	0,9100	0,3902	0,8287	0,6591	0,8919	0,3945
MEDQEMRI	**0,7455**	**0,5857**	0,8204	0,4300	**0,9133**	**0,7573**	**0,9312**	**0,3386**	**0,8754**	**0,7185**	**0,9062**	**0,3592**

Dwa najlepsze wyniki dla każdego kryterium zostały wyróżnione pogrubioną czcionką.

2.4. Narzędzie do oceny jakości obrazów rezonansu magnetycznego oraz porównania metod

TIQA-MRI to wszechstronne narzędzie stworzone w środowisku Matlab z myślą o badaczach rozwijających nowe techniki MRI, które wspomaga ocenę jakości obrazów MR, umożliwiając analizę wyników i porównywanie skuteczności własnych algorytmów z istniejącymi metodami [84].

Interfejs aplikacji składa się z czterech zakładek, z których każda oferuje odrębną funkcjonalność odpowiadającą różnym etapom analizy danych. Pierwsza zakładka, *Image Assessment* (Ocena obrazów), umożliwia wybór metod oceny jakości obrazu oraz przeprowadzenie treningu modeli na podstawie cech wyodrębnionych z wybranych metod. Do tego celu wykorzystywana jest baza danych DB1, zawierająca 70 obrazów MRI wraz z ocenami subiektywnymi. W wyniku przeprowadzonego treningu użytkownik otrzymuje obiektywne oceny testowanych obrazów, bazujące na wyuczonych modelach. Druga zakładka, *Evaluation* (Ewaluacja), pozwala na ocenę skuteczności wytrenowanych modeli jakości. Obrazy z bazy danych mogą być dzielone na zbiory treningowe i testowe, na przykład w popularnej proporcji 80:20%. Wyniki prezentowane są w formie tabelarycznej, zawierającej zarówno oceny subiektywne, jak i obiektywne, a także współczynniki korelacji SRCC, KRCC i PLCC, umożliwiające ilościowe porównanie efektywności metod. Zakładka *Method Comparison* (Porównanie metod) oferuje możliwość zestawienia różnych technik IQA pod kątem skuteczności w ocenie jakości obrazów MRI. W tym przypadku wyniki prezentowane są w postaci wykresów oraz testu istotności statystycznej Wilcoxona, a także uzupełnione o zestawienie czasu obliczeń dla każdej z metod. Ostatnia zakładka, *Settings* (Ustawienia), służy do zarządzania listą dostępnych metod oraz umożliwia dodawanie nowych algorytmów do środowiska analizy.

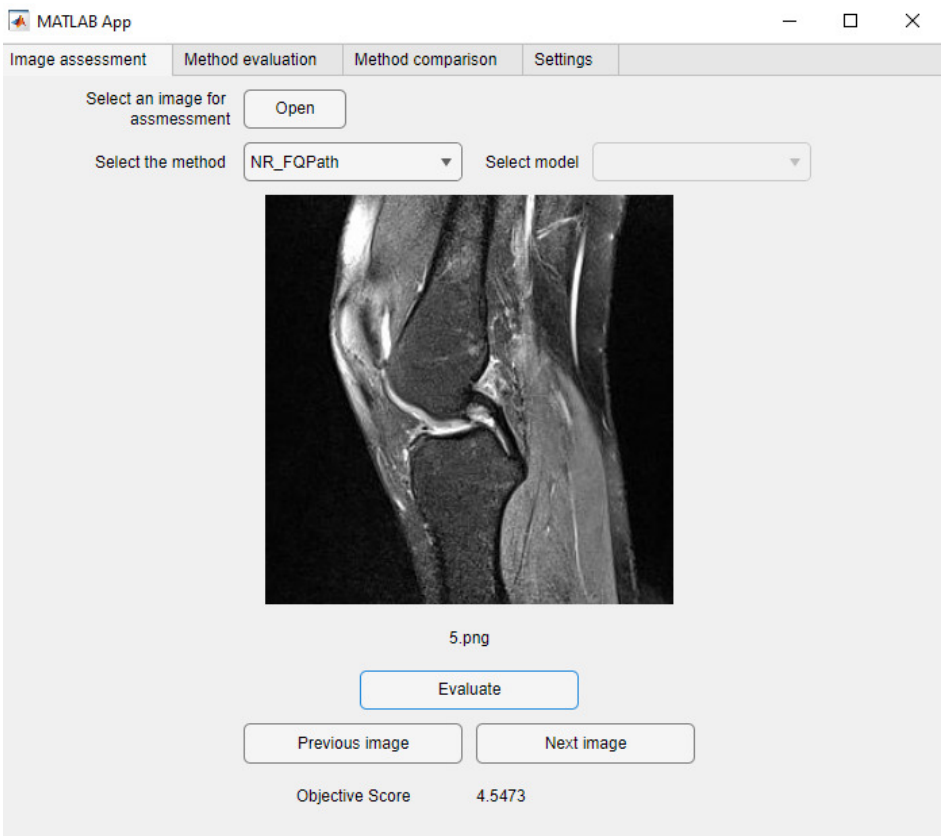
Procesy oceny i treningu modeli w narzędziu TIQA-MRI zostały zaprojektowane umożliwiając użytkownikom dopasowanie jego funkcji do konkretnych wymagań badawczych. Aplikacja opiera swoje działanie na bazie danych DB1, która zawiera obrazy MRI wraz z odpowiadającymi im ocenami subiektywnymi. Jak przedstawiono na Rys. 2.8, zaczerpniętym z [A-4], użytkownik może wybrać obraz do oceny oraz wskazać metodę IQA. Jeśli dostępny jest model treningowy, aplikacja umożliwia jego użycie w celu przeprowadzenia oceny pojedynczego obrazu. Wybrany obraz jest wyświetlany na ekranie, a proces oceny rozpoczyna się po kliknięciu przycisku *Evaluate* (Oceń). Po uruchamianiu algorytmu oceny aplikacja wyświetla przewidywaną jakość obrazu - ocenę obiektywną.

Dalsze możliwości, zaprezentowane na Rys. 2.9 zaczerpniętym z [A-4], obejmują proces treningu modeli oparty na podziale danych na zbiory treningowe i testowe. Użytkownik ma pełną kontrolę nad konfiguracją tego podziału, z opcjami określenia procentowego rozkładu danych, liczby losowych podziałów oraz liczby obrazów tego samego typu. Dodatkowo, dostępne są funkcje pozwalające na ustawienie wartości

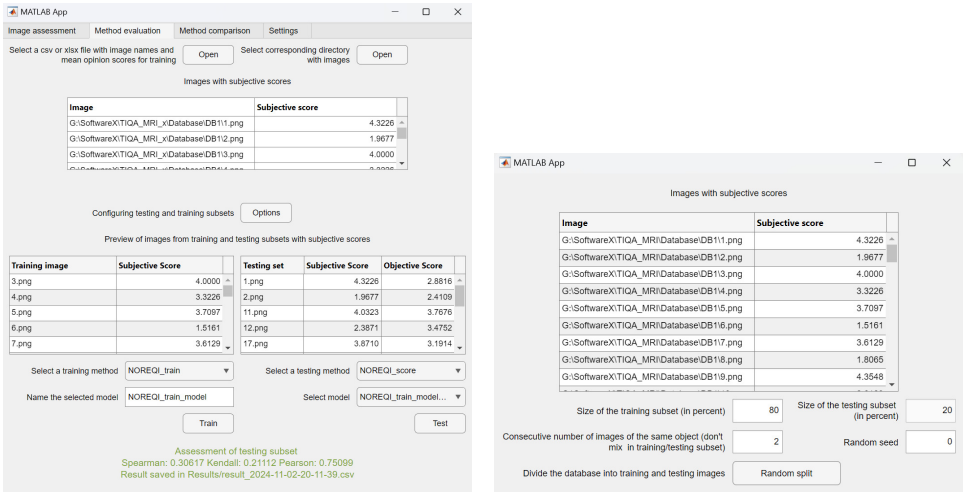
początkowych, które wpływają na sposób dzielenia danych. Po skonfigurowaniu podziału danych, proces jest inicjowany przyciskiem *Random split* (Losowy podział), a wyniki podziału są wyświetlane na ekranie.

Jeśli zachodzi potrzeba przeprowadzenia treningu modelu, aplikacja oferuje funkcję umożliwiającą użytkownikowi nadanie nazwynowo-powstałemu modelowi oraz zapisanie go w odpowiednim katalogu. Po kliknięciu przycisku *Train* (Trenuj) proces zostaje uruchomiony, a jego rezultatem jest uzyskanie wartości współczynników korelacji (SRCC, KRCC, PLCC) oraz obiektywnej oceny dla zbioru testowego. Wyniki te są prezentowane w przejrzystym układzie tabel.

Porównywanie skuteczności różnych metod oceny jakości obrazów MRI to kolejna kluczowa funkcjonalność TIQA-MRI, widoczna na Rys. 2.10 [A-4]. Użytkownik może wybrać rodzaj metody oraz bazę danych, na podstawie której zostaną wykonane obliczenia. Dla metod niewymagających treningu możliwość wyboru modelu jest niedostępna, co zwiększa przejrzystość procesu. Po ustaleniu parametrów i kliknięciu przycisku Train and test all aplikacja przeprowadza ocenę, prezentując postęp oraz finalne wyniki. W przypadku pominięcia któregokolwiek kroku konfiguracji użytkownik otrzymuje komunikat o błędzie, minimalizując ryzyko pomyłek. Wszystkie uzyskane wartości mogą zostać zapisane w dedykowanym katalogu Results. Użytkownik ma również możliwość ponownego wykorzystania zapisanych ustawień w przyszłych eksperymentach, co znacząco usprawnia pracę badawczą i umożliwia efektywne porównywanie wyników.

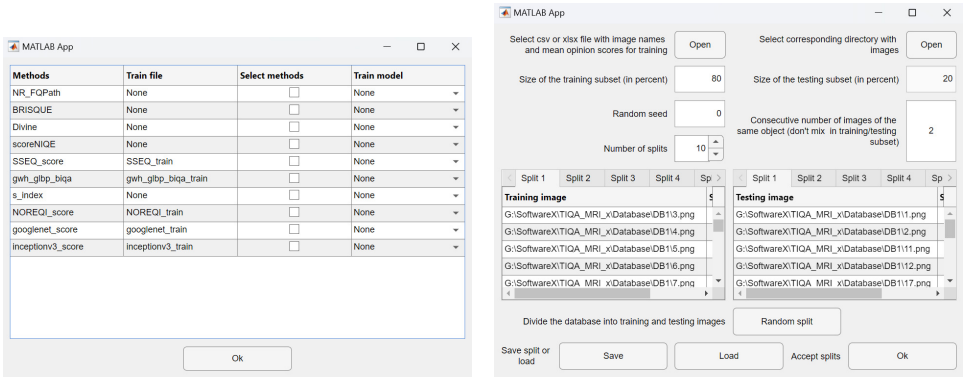


Rys. 2.8. Funkcjonalność narzędzia TIQA-MRI służąca do oceny obrazu.

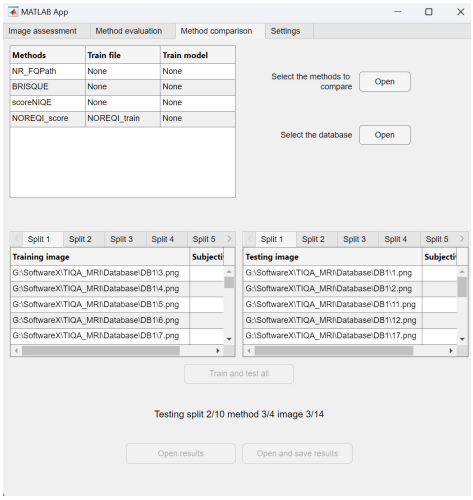


(a) **(b)**

Rys. 2.9. Funkcjonalność oceny metody obejmująca zarówno proces uczenia, jak i testowania, a) pomyślne ustawienia przebiegu procesu, b) konfiguracja podzbiorów bazy danych.



(a) **(b)**



(c)

Rys. 2.10. Funkcjonalność służąca do porównania metod: a) Pierwszy etap ustawień funkcjonalności z wyborem metod, b) Ustawienia dostępne dla wyboru bazy danych, c) ogólny widok zakładki w trakcie procesu trenowania i testowania.

TIQA-MRI to narzędzie, przeznaczone do obiektywnej oceny jakości obrazów MR. Umożliwia porównywanie różnych metod IQA, integrację nowych algorytmów oraz ich walidację względem ocen subiektywnych. Narzędzie wspiera również proces trenowania modeli oraz analizę wyników pod kątem czasu obliczeń i współczynników korelacji. Intuicyjny interfejs i rozbudowana funkcjonalność czynią TIQA-MRI kompleksowym wsparciem dla badaczy rozwijających techniki oceny jakości obrazów MRI.

Autor samodzielnie zaprojektował architekturę aplikacji, zaimplementował funkcjonalności umożliwiające ocenę i porównanie metod MR-IQA oraz przeprowadził testy walidacyjne w oparciu o wybrane miary jakości. Wkład obejmował również analizę uzyskanych wyników, przygotowanie materiałów graficznych oraz pełne opracowanie dokumentacji narzędzia.

2.5. Użycie wielopoziomowych reprezentacji obrazu w ocenie jakości wyostrzonych obrazów wielospektralnych

W dziedzinie teledetekcji obrazy satelitarne dostarczają cennych informacji o powierzchni Ziemi, jednak często występuje kompromis między rozdzielczością przestrzenną a spektralną. Obrazy panchromatyczne (PAN) charakteryzują się wysoką rozdzielczością przestrzenną, jednocześnie brakiem informacji o kolorze, podczas gdy obrazy multispektralne (MS) oferują bogate dane spektralne przy niższej rozdzielczości przestrzennej. Chcąc połączyć zalety obu typów obrazów stosuje się technikę pansharpeningu, która polega na fuzji obrazu PAN z obrazem MS w celu uzyskania obrazu o wysokiej rozdzielczości zarówno przestrzennej, jak i spektralnej. Jakość uzyskanych, wyostrzonych obrazów zależy jednak od zastosowanych algorytmów. Ocena jakości takich obrazów wciąż stanowi wyzwanie i jest przedmiotem otwartej dyskusji w środowisku naukowym. W przeciwieństwie do obrazów z innych dziedzin, dla których istnieje wiele różnych metod oceny jakości, w przypadku obrazów PS trudno odnaleźć dedykowane podejście bez referencji, które byłoby powszechnie akceptowane, zaś większość istniejących metod oceny jakości wymaga użycia do niedostępnego obrazu referencyjnego. Główną motywacją do przeprowadzenia tych badań była potrzeba opracowania skutecznej metody oceny jakości obrazów PS bez odniesienia (PS NR IQA), pozwalającej na ich analizę bez konieczności posiadania wzorcowego obrazu referencyjnego. W prezentowanym podejściu zaproponowano wykorzystanie głębokiego uczenia, co stanowi nowatorskie rozwiązanie w tej dziedzinie. Istotnym aspektem pracy była adaptacja metod oceny jakości obrazu stosowanych w innych dziedzinach. Przeprowadzone badania miały na celu dostosowanie znanych technik takich jak: BPRI [58], GWH-GLBP [46], MEON [51], ERGAS [92], czy BRISQUE [59] do specyfiki obrazów PS i stworzenie modelu jakości uwzględniającego ich charakterystyczne cechy. W ramach pracy wykonano eksperymenty na dużym zbiorze obrazów PS, co pozwoliło na dokładne zweryfikowanie skuteczności opracowanej metody i jej porównanie z istniejącymi rozwiązaniami. W omawianym artykule zaprezentowano nową metodę PS NR IQA – Multi-Level Pan-Sharpening Images Evaluator (MLPSIE) [81]. W odróżnieniu od innych podejść, technika ta wykorzystuje architektury głębokiego

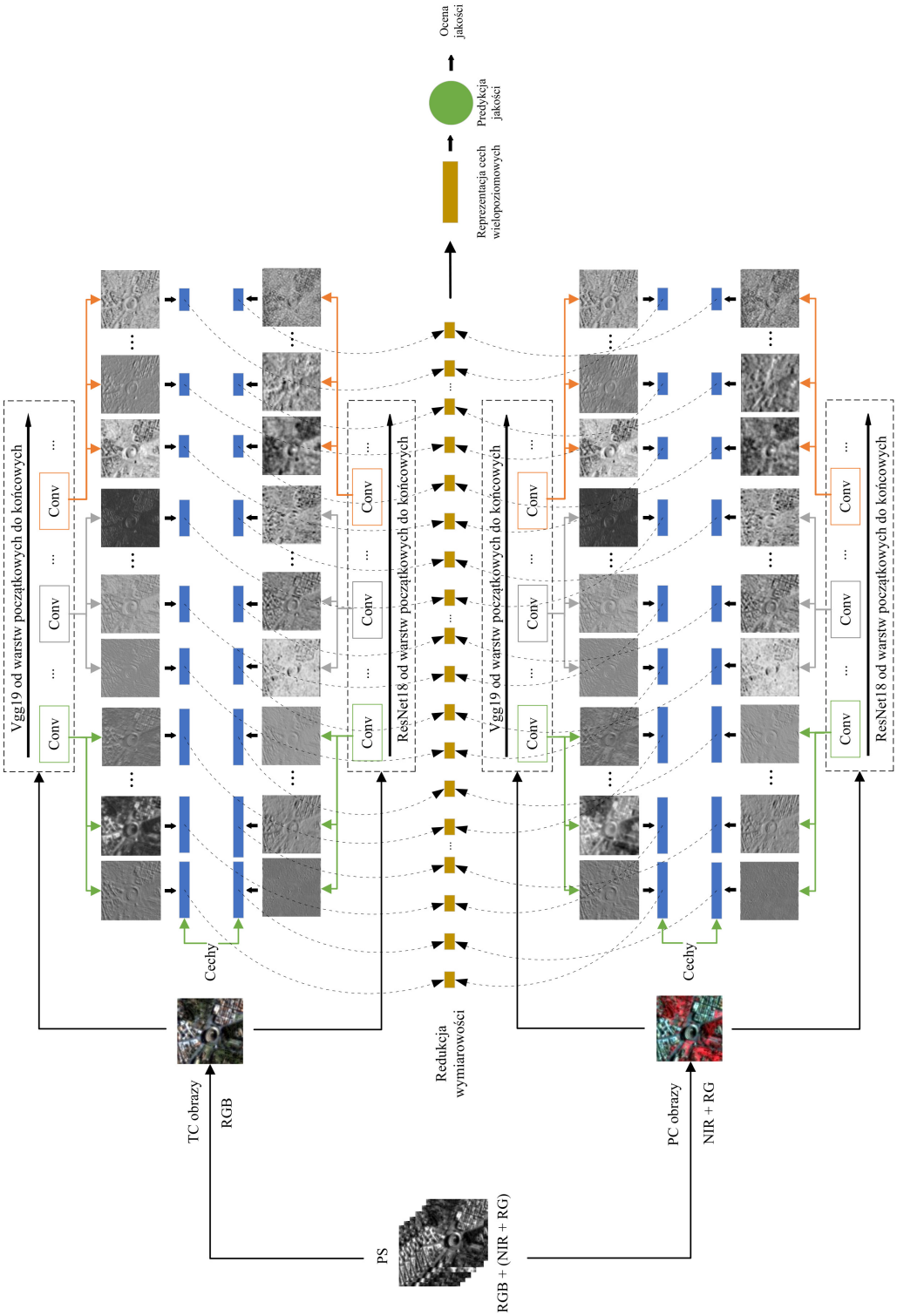
uczenia do oceny jakości obrazów PS w sposób skorelowany z ludzkimi ocenami. MLP-SIE jest pierwszą metodą, która łączy dwie sieci neuronowe w celu ekstrakcji istotnych cech obrazu, ich selekcji oraz organizacji w sposób umożliwiający ocenę jakości. Proces ten obejmuje trzy kluczowe etapy: wydobywanie cech o dużej wymiarowości opisujących strukturę i teksturę obrazu, redukcję ich liczby w celu eliminacji zbędnych informacji oraz budowę reprezentacji uwzględniającej zarówno ogólne, jak i szczegółowe właściwości obrazu. Takie wielopoziomowe podejście pozwala na stworzenie dokładniejszego modelu jakości, co wyróżnia tę metodę na tle istniejących rozwiązań i zwiększa jej skuteczność w ocenie obrazów PS.

Proponowana metoda wykorzystuje dwie sieci głębokiego uczenia, ResNet-18 [32] i VGG-19 [72], które umożliwiają przetwarzanie wielopoziomowych danych w celu uzyskania cech wrażliwych na zniekształcenia. Ze względu na ograniczoną wielkość bazy obrazów oraz obiecującą skuteczność metody, sieci nie są dodatkowo trenowane. Jak przedstawiono na Rys. 2.11 [A-5], metoda ta korzysta z obrazów PS składających się z kanałów RGB i bliskiej podczerwieni (NIR), które tworzą dwa rodzaje danych wejściowych: technika nadawania kolorów obrazowi w skali RGB (TC), czyli obrazy odwzorowujące kolory w sposób zbliżony do percepcji ludzkiego oka, składające się z trzech kanałów: czerwonego (R), zielonego (G) i niebieskiego (B), oraz technika nadawania kolorów obrazowi w skali szarości (PC) NIR+RG, w którym kanał bliskiej podczerwieni (NIR) jest łączony z kanałami czerwonym (R) i zielonym (G) w celu uzyskania obrazów, które uwydatniają cechy niewidoczne w świetle widzialnym, takie jak zdrowie roślinności, czy struktury terenu. Odpowiedzi sieci na różnych poziomach są wyodrębniane (niebieskie prostokąty) i redukowane za pomocą metody Kernel PCA (KPCA, brązowe prostokąty) [93]. Redukcja dotyczy połączonych informacji TC i PC (linie przerywane). Zredukowane cechy są następnie łączone (dłuższy brązowy blok) i wykorzystane przez model jakości bazujący na SVR do predykcji jakości obrazu PS (zielone koło).

Wykorzystana sieć VGG-19 składa się z 16 warstw konwolucyjnych i trzech w pełni połączonych, używających filtrów 3×3 i operacji maksymalnego próbkowania nad obszarami 2×2 piksele [41]. Wejście sieci ma rozmiar 224×224, a w pełni połączone warstwy zawierają kolejno 4096, 4096 i 1000 kanałów, z ostatnią funkcją SoftMax. ResNet-18 natomiast korzysta z połączeń rezydualnych, które eliminują problem zanikających gradientów, umożliwiając trenowanie głębszych sieci. Architektura ta rozpoczyna się warstwą konwolucyjną 7×7, maksymalnym próbkowaniem 3×3 i kończy na uśredniającym próbkowaniu oraz warstwach w pełni połączonych. Obie sieci, przystosowane do obrazów RGB, w tej metodzie umożliwiają ocenę jakości na podstawie dwóch obrazów trzykanałowych [97]. Pierwsze warstwy ResNet-18 i VGG-19 zawierają odpowiednio 802816 i 3211264 cech, co pokazuje skalę przetwarzanych danych.

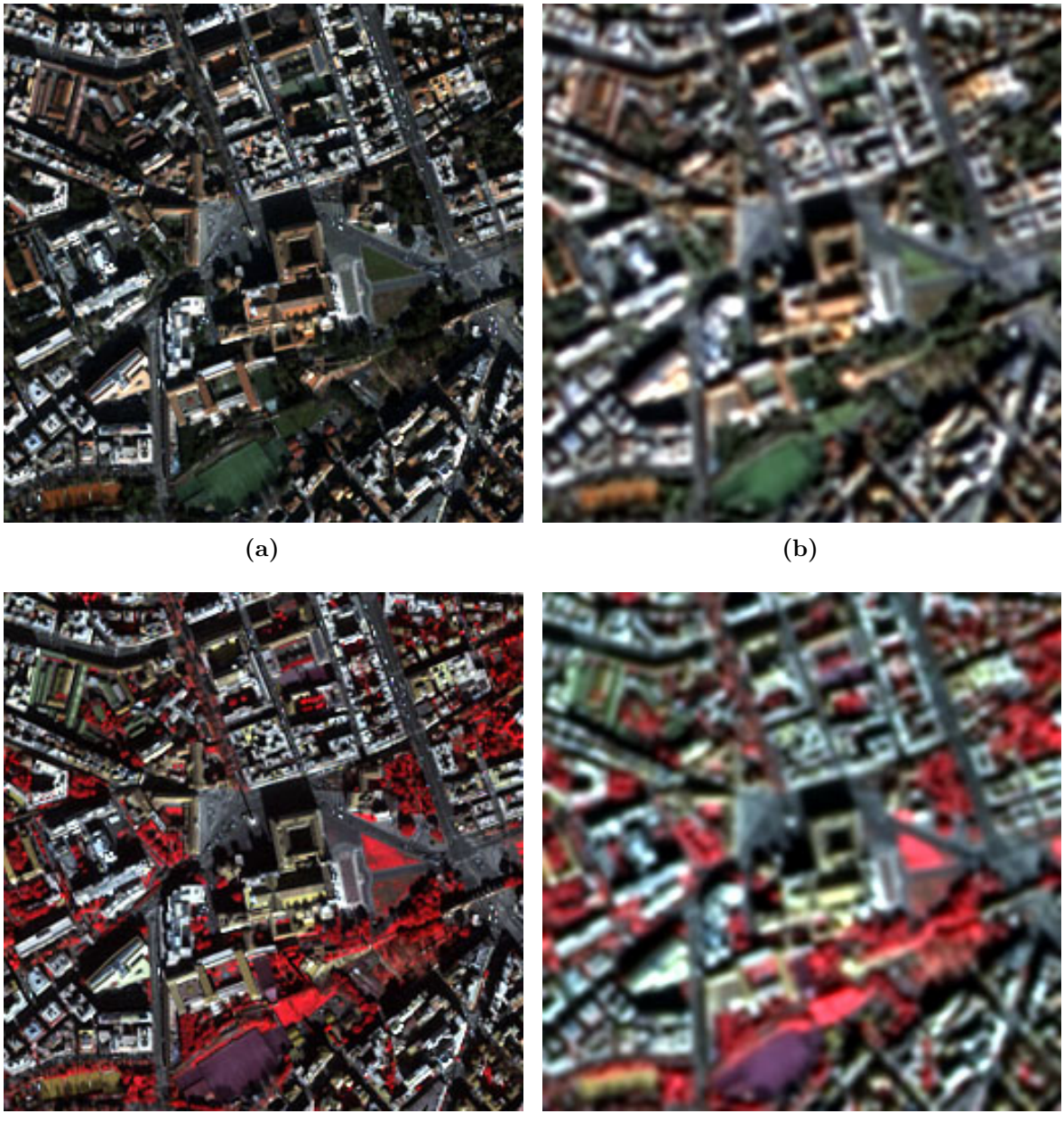
Dla każdego poziomu cechy są przetwarzane niezależnie za pomocą KPCA, co umożliwia uzyskanie zwartego wektora cech. KPCA działa jak klasyczna PCA, ale jest również skuteczna w przypadku problemów nieliniowych oraz sytuacji wymagają-

cych automatycznego określenia liczby komponentów [24]. KPCA zastosowano zamiast statycznej PCA, ponieważ zapewnia większą elastyczność w przetwarzaniu danych nieliniowych i bardziej efektywne odwzorowanie ich struktur.



Rys. 2.11. Schemat blokowy metody. Cechy wyodrębnione z obu sieci dla obrazów PC i TC wejściowego obrazu PS (przedstawione za pomocą niebieskich prostokątów) są redukowane za pomocą KPCA (brązowe bloki) i łączone w celu predykcji jakości przez SVR (zielone kółko).

Do eksperymentów użyto obrazów z bazy IKONOS ocenionych przez obserwatorów w badaniu Agudelo-Medina [2]. Baza, uzyskana po kontakcie z autorami, obejmuje 171 obrazów PS (par TC i PC) uzyskanych z czterech obrazów referencyjnych za pomocą sześciu metod PS (IHS, BDSD, PCA, MTF-GLP-CBD, HPF, ATWT-M2) oraz jednej metody interpolacyjnej (EXT). Dodatkowo obrazy zostały zniekształcone rozmyciem i szumem gaussowskim. Rozważane regiony to Koloseum, Droga, Teren miejski, Rzeka i Willa, jednak w badaniu wykorzystano tylko cztery z nich, ponieważ jedynie one zawierają pełne zestawy obrazów z subiektywnymi ocenami. Przykładowe obrazy zaczerpnięte z bazy IKONOS znajdują się na Rys. 2.12 [A-5]. Jest to największy obecnie zbiór PS oceniany przez ludzi, który pozwala na porównanie metod PS.



(a) **(b)**

(c) **(d)**

Rys. 2.12. Obrazy TC (a, c) i PC (b, d) bez zniekształceń (a, b) i rozmyte (c, d) z bazy IKONOS, scena miejska: Koloseum.

W procesie ewaluacji metody MLPSIE obrazy są podzielone na rozłączne zestawy treningowe i testowe, a kryteria są raportowane jako mediany dla zestawów testowych. Dla metod niewymagających treningu używa się wyłącznie obrazów testowych. Wyniki uzupełniają wykonane testy istotności statystycznej Wilcoxona oraz analiza zdolności najlepszych metod do sortowania wyników metod PS w porównaniu z ocenami ludzkimi (oznaczonych jako DMOS-GM).

Wprowadzona metoda została porównana z 17 nowoczesnymi metodami NR-IQA, w tym BRISQUE [59], CurveletQA [3], FRIQUEE [25], GMLOG [1], GWH-GLBP [46], NFERM [27], NOREQI [67], Oracle [5], SCORER [49], SISBLIM [26], BPRI [57], SINDEX [44], dipIQ [52], Blinder [24], Qoa [2], ERGAS [92] oraz R50GR18 [85]. Spośród nich Qoa, podobnie jak MLPSIE, jest metodą wymagającą uczenia, co pozwala na porównanie używanych w obu rozwiązaniach cech. R50GR18, dipIQ i Blinder bazują na głębokich sieciach neuronowych, jednak jedynie R50GR18 i Blinder można trenować. Metody CurveletQA, BPRI, SINDEX, czy ERGAS osobno oceniają obrazy PC i TC, a ich wyniki są uśredniane.

W pierwszym scenariuszu testowym obrazy z jednej z czterech scen były przeznaczone do testów, a pozostałe do uczenia się. Wyniki median z czterech testów Tabela 2.4 [A-5] pokazują, że MLPSIE znacząco przewyższa inne podejścia osiągając najlepsze wyniki w metrykach SRCC (0,9548), KRCC (0,8295) i PLCC (0,9803), a także uzyskując najniższą wartość błędu RMSE (1,5797). Warto również podkreślić, że metody takie jak CurveletQA i NFERM osiągnęły wysokie wyniki, jednak MLPSIE wykazała się największą wydajnością w tym scenariuszu testowym.

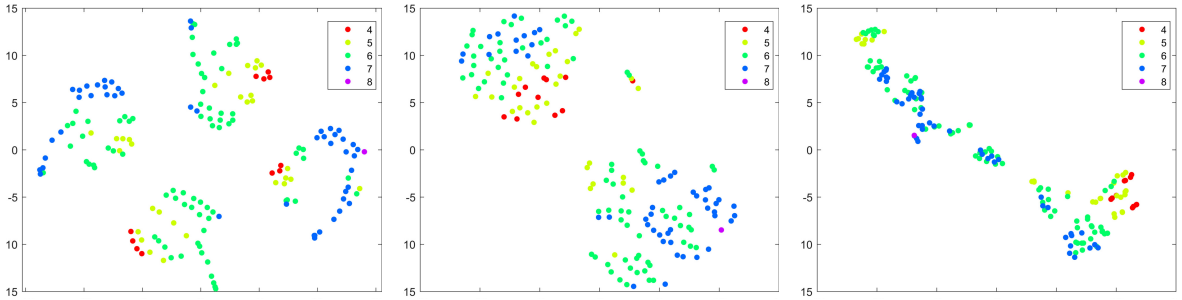
Tabela 2.4. Porównanie wydajności podejść dla różnych scen.

Method	**SRCC**	**KRCC**	**PLCC**	**RMSE**
BRISQUE	0,9007	0,7473	0,9655	1,9523
CurveletQA	**0,9443**	**0,8098**	**0,9720**	**1,6473**
FRIQUEE	0,8183	0,6336	0,9124	3,2830
GMLOG	0,9200	0,7698	0,9653	2,1162
GWH-GLBP	0,8766	0,7121	0,9377	2,7140
NFERM	**0,9373**	**0,7981**	**0,9698**	**1,8099**
NOREQI	0,8779	0,7110	0,9511	2,5011
Oracle	0,7882	0,5879	0,4137	7,1081
SCORER	0,7955	0,5857	0,7728	4,6842
SISBLIM	0,8026	0,6135	0,8808	3,8421
BPRI	0,4599	0,2957	0,6107	5,8090
SINDEX	0,8749	0,6848	0,9118	3,1256
dipIQ	0,6204	0,4762	0,6821	5,6095
Blinder	0,9115	0,7647	0,9650	1,8550
R50GR18	0,8003	0,6257	0,8136	4,0225
Q_{oa}	0,8971	0,7307	0,9381	2,7947
ERGAS	0,9265	0,7796	0,9658	1,9560
MLPSIE	**0,9548**	**0,8295**	**0,9803**	**1,5797**

Trzy najlepsze wyniki dla każdego kryterium zostały wyróżnione pogrubioną czcionką.

Zaproponowana metoda PS IQA wyróżnia się brakiem konieczności odniesienia do obrazu referencyjnego oraz wykorzystaniem wielopoziomowych reprezentacji obrazu generowanych przez gębokie sieci neuronowe. Istotnym aspektem tej metody jest zastosowanie KPCA w celu uzyskania skondensowanego, informacyjnie bogatego wektora cech. Dzięki połączeniu abstrakcyjnych i niskopoziomowych cech obrazu możliwe było uzyskanie wyników dobrze korelujących z subiektywnymi ocenami jakości, co sprawia, że metoda ta stanowi odporniejszą na zmienne warunki i efektywną obliczeniowo alternatywę dla tradycyjnych podejść wymagających obrazu referencyjnego. Kluczowym czynnikiem skuteczności tej metody jest zdolność sieci neuronowych, takich jak ResNet-18, do ekstrakcji cech na różnych poziomach.

Niższe warstwy przechowują szczegółowe informacje o krawędziach i teksturach, natomiast wyższe reprezentują bardziej abstrakcyjne informacje o strukturze obrazu. W celu zilustrowania skuteczności uzyskanych wielopoziomowych reprezentacji obrazów w rozróżnianiu obrazów o zróżnicowanej jakości zobrazowaano dwuwymiarowe osadzenia t-SNE cech wielopoziomowych Rys. 2.13, pozyskanych z różnych modeli głębokiego uczenia, w celu oceny ich zdolności do rozróżniania obrazów o różnej jakości. Zaprezentowano osadzenia dla: połączonych cech MLPSIE (a), oraz osobno dla metryk Q_{oa} (b) i BRISQUE (c). Kolory odpowiadają przeskalowanym wartościom DMOS-GM przypisanym do obrazów.



(a) **(b)** **(c)**

Rys. 2.13. Wykresy przedstawiają dwuwymiarowe osadzenia t-SNE cech uzyskanych przez MLPSIE (a) dla obrazów PS z czterech kategorii w zbiorze danych. Dodatkowo zaprezentowano osadzenia dla Q_{oa} (b) i BRISQUE (c). Kolory punktów odpowiadają zaokrąglonym wartościom DMOS-GM przypisanym do obrazów.

Widoczne w wykresach klastry sugerują, że cechy wyodrębnione przez sieci neuronowe pozwalają na skuteczne grupowanie obrazów według lokalizacji (Koloseum, Droga, Miasto, Rzeka) oraz wykazują wrażliwość na różnice jakości, mimo że jakość nie była jeszcze bezpośrednio oceniana (predykcja jakości następuje później na podstawie tych cech). Porównawczo, metryki BRISQUE i Q_{oa} gorzej radzą sobie z rozróżnianiem obrazów według jakości, choć Q_{oa} wykazuje lepsze rezultaty niż BRISQUE.

Wkład własny autora prezentowanego artykułu polegał na współautorstwie koncepcji pracy, opracowaniu metodologii oraz implementacji oprogramowania. Autor

brał udział w weryfikacji wyników, realizacji badań i analizie uzyskanych rezultatów, a także w przygotowaniu oraz redakcji tekstu publikacji.

2.6. Ocena jakości wyostrzonych obrazów satelitarnych za pomocą sieci o zmodyfikowanej architekturze

Kontynuując dotychczasowe prace nad oceną jakości wyostrzonych obrazów satelitarnych, opracowano nową metodę Three-Branch Neural Network for No-Reference Quality Assessment of Pan-Sharpened Images (TBN-PSI). Opiera się ona na trójgałęziowej architekturze sieci neuronowej, gdzie każda z gałęzi przetwarza obrazy PS, umożliwiając jednoczesne wydobywanie niskopoziomowych cech oraz informacji semantycznych wysokiego poziomu. Współdzielenie informacji między podsieciami realizowane jest poprzez dodatkową gałąź środkową, co wspomaga proces uczenia. Architektura została dostosowana specjalnie do charakterystyki obrazów PS, ponownie z uwzględnieniem zarówno aspektów spektralnych, jak i przestrzennych, co ma istotne znaczenie w kontekście złożoności tych danych. Dzięki temu możliwa jest trafna ocena jakości bez konieczności stosowania dodatkowych kroków przetwarzania takich jak SVR, czy KPCA. Istotnym elementem rozwiązania jest modyfikacja klasycznej struktury Inception-v3 poprzez wprowadzenie trzech równoległych gałęzi, co umożliwia skuteczniejsze przetwarzanie informacji obrazowych na różnych poziomach abstrakcji. Ocena jakości obrazów PS odgrywa kluczową rolę w zastosowaniach środowiskowych, takich jak monitorowanie wylesiania, urbanizacji, czy jakości wód, a także w analizie skutków katastrof naturalnych. Wiarygodna ocena jakości przekłada się bezpośrednio na jakość danych wykorzystywanych w systemach wspomagania decyzji, zwłaszcza tych opartych na sztucznej inteligencji, wpływając na ich skuteczność i niezawodność. W ramach badań przeprowadzono kompleksową walidację proponowanej metody na sześciu zestawach danych obrazów PS, porównując jej skuteczność z wynikami uzyskiwanymi przez uznane metody oceny jakości [18, 82].

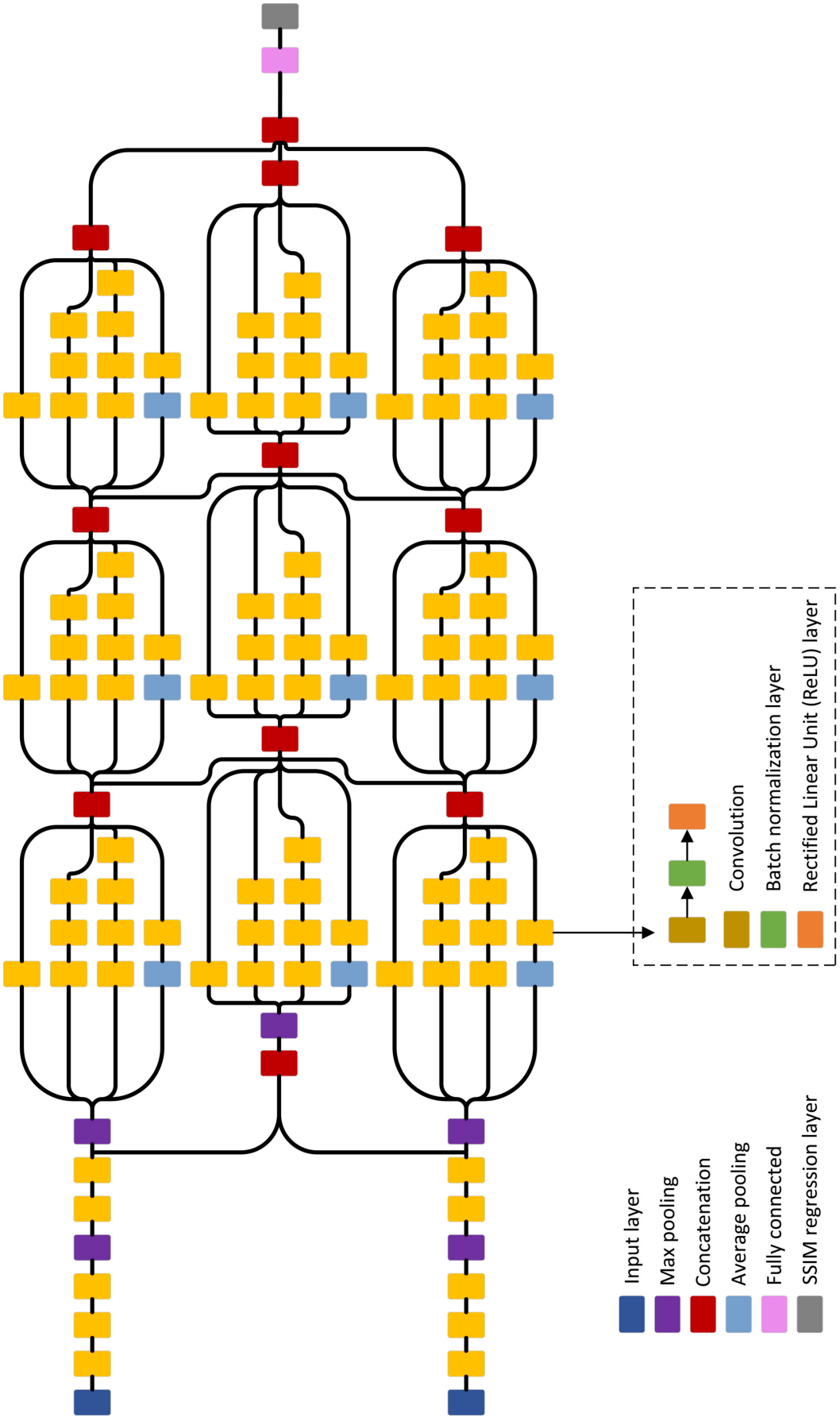
W artykule zastosowano bazę danych NBU PansharpRSData opracowaną przez Meng i in. [56], która obejmuje 13620 obrazów uzyskanych z 270 par obrazów PAN i MS, pochodzących z różnych satelitów: IKONOS, QuickBird, Gaofen, WorldView-4 (WV4), WorldView-3 (WV3) oraz WorldView-2 (WV2) utworzonych z wykorzystaniem sześciu metod PS: ATWT_M2, BDSD, GS, IHS, MTF-GLP-CBD i PCA. Każdy zestaw został przygotowany, w sposób umożliwiający ocenę jakości obrazów typu PS. Baza danych obejmuje obrazy różniące się zarówno charakterystyką spektralną, jak i przestrzenną, co zapewnia wszechstronny materiał do eksperymentów i porównań metod oceny jakości. Przykładowe obrazy przedstawiono na Rys. 2.14 [A-6]. Eksperymentalnie dobrane bloki zostały skopiowane, aby obsługiwać dwa wejścia: kanały RGB oraz NIR + GB, zgodne z wymaganym rozmiarem wejściowym sieci (299×299 pikseli). Wprowadzona metoda TBN-PSI wykorzystuje warstwy z sieci Inception-v3 jako bazę po selektywnym wyborze pierwszych 87 spośród 315 warstw, które zorganizowano w bloki.

Rys. 2.14. Przykładowe obrazy PS z bazy obrazów NBU PansharpRSData.

Jak przedstawiono na Rys. 2.15, zaczerpniętym z [A-6], kluczowym elementem jest dodatkowa gałąź integrująca informacje z obu wejść po przetworzeniu danych w blokach dolnej i górnej gałęzi. Takie podejście umożliwia łączenie cech o różnej granularności, od informacji niskopoziomowych po cechy semantyczne na wyższym poziomie.

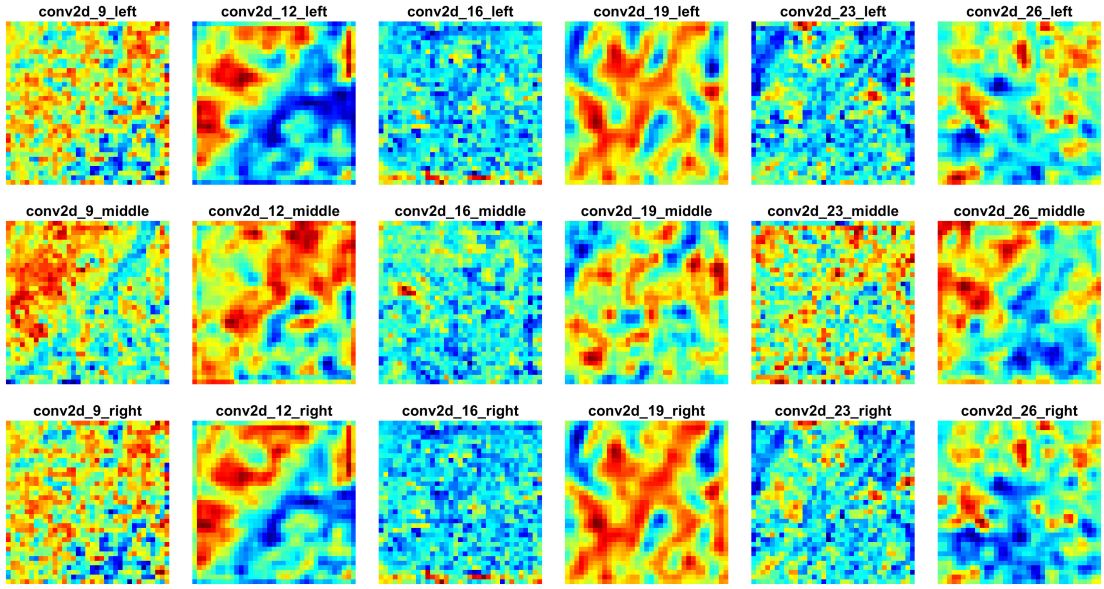
Jak w przypadku innych badań nad NR IQA, również w przypadku przedstawianego podejścia zastosowano standardowy protokół, w którym w eksperymentach 80% obrazów referencyjnych wraz z odpowiadającymi im zniekształconymi wersjami zostało wykorzystanych jako zbiór treningowy, natomiast pozostałe 20% przydzielono do zbioru testowego. Oba zbiory są rozłączne. Wyniki zostały przedstawione jako mediany uzyskane z dziesięciu losowych podziałów danych. Ze względu na brak subiektywnych ocen jakości tych danych, jako metryki walidacyjne w pracy zastosowano miary z referencją takie jak SAM, ERGAS oraz SCC, które pozwalają na obiektywną ocenę jakości przetworzonych obrazów. Analogiczne podejście stosowano przez Changa [11]. Ta różnorodność oraz wielkość bazy danych stanowi istotną zaletę przy ocenie skuteczności proponowanych metod oceny jakości obrazów. Fuzja informacji odbywa się za pomocą warstwy głębokiej konkatenacji, po której następują warstwy w pełni połączone, Global Average Pool oraz warstwy regresyjne. Zamiast tradycyjnej funkcji błędu średniokwadratowego (MSE), wykorzystano wskaźnik SSIM, co powoduje wzrost skuteczności modelu.

Rys. 2.16 przedstawia wizualizację aktywacji wybranych warstw w trzech gałęziach modelu TBN-PSI, uzyskanych na podstawie dwóch obrazów PS przedstawiających ten sam region – (a) obraz lepszej jakości oraz (b) obraz gorszej jakości.

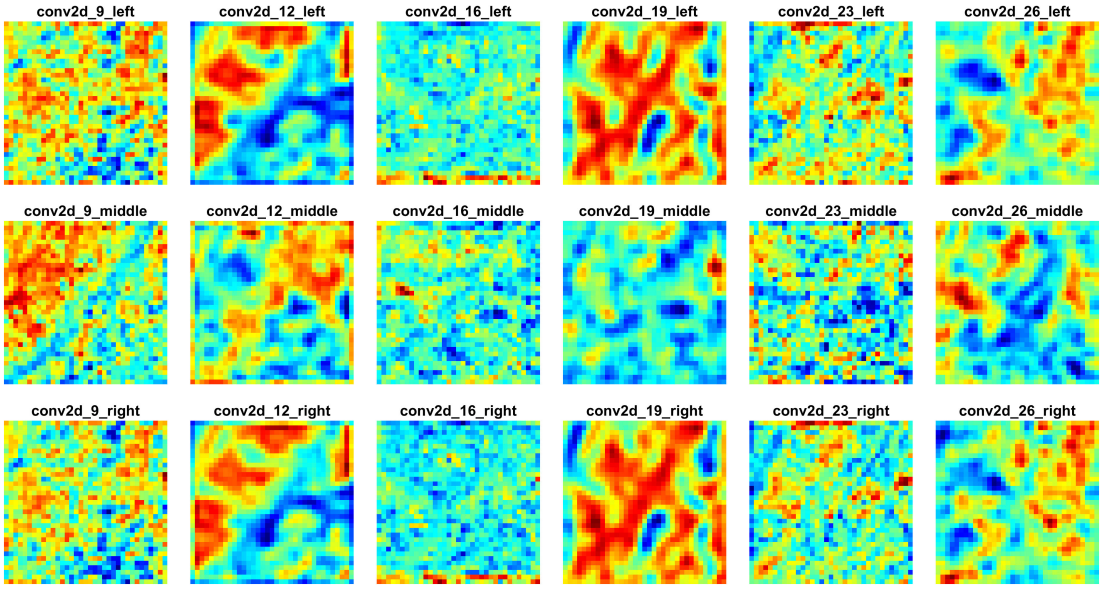


Rys. 2.15. Proponowana architektura sieci z trzema gałęziami podsieci.

Celem wizualizacji jest analiza i wskazanie różnic w odpowiedzi sieci w zależności od jakości obrazu wejściowego. Warstwy zaprezentowane na rysunku generują po 64 kanały aktywacji, przy czym do przedstawienia wyników wykorzystano pierwszy kanał każdej z nich. Zakres kolorów został znormalizowany – odcienie niebieskiego wskazują na niską aktywność, natomiast najbardziej czerwone obszary identyfikują punkty zainteresowania. Obserwacje wynikające z wizualizacji potwierdzają, że jakość obrazu istotnie wpływa na rozkład aktywacji – w przypadku obrazów o wyższej jakości (a), aktywacje są bardziej zorganizowane, wyraźne i wykazują silniejsze skupienie wokół struktur istotnych dla danego problemu.



(a)



(b)

Rys. 2.16. Wizualizacja aktywacji wybranych warstw konwolucyjnych w trzech gałęziach na podstawie dwóch obrazów PS tego samego regionu: (a) obraz lepszej i (b) gorszej jakości.

Natomiast w przypadku obrazów gorszej jakości (b), odpowiedzi sieci są mniej jednoznaczne i często bardziej rozproszone. Dodatkowo, zauważalna jest różnorodność aktywacji między trzema gałęziami sieci, co sugeruje, że zastosowanie architektury z wieloma ścieżkami przetwarzania umożliwia uzyskanie pełniejszej reprezentacji cech wejściowych. Potwierdza to założenie, że trójgałęziowa architektura modelu TBN-PSI sprzyja lepszej lokalizacji i interpretacji istotnych informacji zawartych w obrazie, niezależnie od jego jakości. Dodatkowo, wyniki testu istotności statystycznej Wilcoxona potwierdzają przewagę TBN-PSI nad pozostałymi metodami. W porównaniach parami metoda ta uzyskała największą liczbę istotnie lepszych wyników, wyprzedzając m.in. Inception-v3. Dla metryki ERGAS TBN-PSI została oceniona jako najlepsza spośród wszystkich metod, a dla metryki SAM uzyskano najlepsze rezultaty w 14 przypadkach oraz najwyższą średnią pozycję (1,56). Wyraźne różnice na korzyść TBN-PSI zaobserwowano m.in. w bazach IKONOS, WV3 i WV2. Łącznie, w analizie uwzględniającej wszystkie zbiory danych i metryki, metoda ta najczęściej osiągała wyniki statystycznie lepsze od pozostałych. Podsumowując TBN-PSI wykazuje dużą precyzję i stabilność w oszacowaniu jakości obrazów w różnych bazach danych i warunkach testowych.

TBN-PSI porównano z dziewięcioma innymi metodami IQA, w tym $Q2^n$, Q, Inception-v3, ResNet-50, NIQE, BRISQUE, GoogLeNet, ViT oraz Q_{oa}. Wyniki zaprezentowane w Tabela 2.5 [A-6] wskazują, że zaproponowana metoda osiąga znacząco lepsze wyniki w metrykach takich jak SRCC, KRCC oraz PLCC na wielu zestawach danych, w tym IKONOS, QuickBird, GaoFen-1, WorldView-2, WorldView-3 oraz WorldView-4.

Tabela 2.5. Porównanie wydajności dziewięciu algorytmów IQA i proponowanego TBN-PSI przy pomocy metryki ERGAS. Trzy najlepsze wyniki dla każdego zestawu danych są pogrubione.

Criterion	Dataset	$Q2^n$	Q	Inception-v3	ResNet-50	NIQE	BRISQUE	GoogleNet	ViT	Q_{oa}	TBN-PSI
SRCC	**IKONOS**	0.3400	0.3617	**0.8517**	0.7844	0.1052	0.7970	0.8159	0.3557	**0.8584**	**0.9198**
	QuickBird	0.3235	0.3036	**0.8848**	0.8407	0.2106	0.7741	**0.8528**	0.3012	0.8392	**0.8818**
	Gaofen	0.0435	0.2212	0.9620	**0.9716**	0.6933	**0.9724**	0,9554	0,7451	**0,9747**	0,9587
	WV4	0,0241	0,0190	**0,9024**	0,8815	0,0271	0,8138	**0,8898**	0,7206	0,8648	**0,9022**
	WV2	0,0847	0,0695	**0,9257**	0,9010	0,1428	0,7749	**0,9171**	0,7017	0,8795	**0,9215**
	WV3	0,0417	0,0307	**0,8657**	0,8510	0,1963	0,7376	**0,8727**	0,6253	0,8190	**0,9217**
KRCC	**IKONOS**	0,2574	0,2770	**0,6612**	0,5829	0,0797	0,5969	0,6235	0,2470	**0,6711**	**0,7542**
	QuickBird	0,2398	0,2244	**0,7031**	0,6405	0,1491	0,5742	**0,6604**	0,2048	0,6413	**0,6931**
	Gaofen	0,0562	0,2099	**0,9224**	**0,9273**	0,5647	0,9164	0,8747	0,5688	**0,9380**	0,8708
	WV4	0,0057	0,0011	**0,7446**	0,7118	0,0168	0,6228	**0,7257**	0,5308	0,6791	**0,7358**
	WV2	0,0315	0,0259	**0,7711**	0,7353	0,0921	0,5839	**0,7564**	0,4890	0,7058	**0,7667**
	WV3	0,0608	0,0552	**0,6961**	0,6859	0,1396	0,5659	**0,7072**	0,4512	0,6616	**0,7756**
PLCC	**IKONOS**	0,4144	0,4444	**0,8417**	0,7788	0,3102	0,7808	0,8173	0,3822	**0,8526**	**0,9076**
	QuickBird	0,4121	0,3990	**0,8685**	0,8223	0,3198	0,7838	0,8415	0,3376	**0,8500**	**0,8521**
	Gaofen	0,4165	0,5087	**0,9516**	0,9316	0,8486	0,9498	**0,9505**	0,8040	**0,9522**	0,8655
	WV4	0,2671	0,1164	**0,9498**	**0,9294**	0,1315	0,8475	**0,9378**	0,8129	0,9130	0,9039
	WV2	0,3653	0,1965	**0,9388**	**0,9135**	0,2595	0,7276	**0,9196**	0,6391	0,9025	0,9071
	WV3	0,3755	0,3812	**0,8414**	0,8374	0,4337	0,7739	**0,8575**	0,6025	0,7743	**0,9225**
Średnia ranga	**Wszystkie**	8,67	8,94	2,00	4,22	8,72	5,39	3,22	7,67	3,61	2,56

Wyniki przedstawione w tabeli pokazują, że TBN-PSI osiągnęło czołowe miejsca w większości scenariuszy testowych i kryteriów oceny jakości obrazów, co potwierdza jego skuteczność. Dla kryterium SRCC TBN-PSI uzyskało najlepsze wyniki na bazach danych IKONOS (0,9198), QuickBird (0,8818), WV4 (0,9022), WV2 (0,9215) oraz WV3 (0,9217), a jedynie na bazie GaoFen-1 zostało wyprzedzone przez metody Q_{oa} i NIQE, które osiągnęły marginalnie wyższe wartości. Warto zauważyć, że przedstawiana metoda wykazała przewagę w analizie scen charakteryzujących się bardziej zróżnicowanymi warunkami, co świadczy o jej wszechstronności.W przypadku KRCC, TBN-PSI również dominowało, osiągając najwyższe wartości dla baz IKONOS (0,7542), QuickBird (0,6931), WV4 (0,7358), WV2 (0,7667) i WV3 (0,7756). Podobnie jak w przypadku SRCC, jedynie dla GaoFen-1 metoda Q_{oa} przewyższyła TBN-PSI, ale różnica była minimalna. Dla PLCC TBN-PSI ponownie znalazło się w czołówce w większości baz danych, w tym w IKONOS (0,9076), QuickBird (0,8521), WV2 (0,9071) i WV3 (0,9225). Szczególnie znaczące wyniki odnotowano w bazach takich jak WV3, gdzie przewyższono inne metody w zauważalny sposób. Ogólnie wyniki TBN-PSI dowodzą jego zdolności do precyzyjnego i stabilnego oszacowania jakości obrazów w różnych scenariuszach testowych i bazach danych. Proponowane podejście pokazuje przewagę w metrykach opartych na korelacjach: SRCC, KRCC i PLCC, przewyższając metody takie jak BRISQUE, NIQE czy Transformer, co wskazuje na jego skuteczność w zadaniach IQA.

Wkład własny autora w przygotowaniu pracy polegał na współautorstwie koncepcji artykułu, zaprojektowaniu architektury metody TBN-PSI oraz adaptacji odpowiednich algorytmów sztucznej inteligencji do specyfiki oceny jakości obrazów PS. Autor implementował oraz trenował zaproponowany model, a także przeprowadzał eksperymenty obliczeniowe i analizę uzyskanych wyników. Dodatkowo autor opracował grafiki ilustrujące architekturę sieci oraz wyniki porównań, a także współredagował tekst artykułu.

2.7. Zestaw algorytmów służących do porównania metod oceny jakości wyostrzonych obrazów wielospektralnych

Toolbox for Image Quality Assessment of Pan-Sharpened Images (TIQA-PSI) to nowe narzędzie [86], obejmujące 30 metod oceny jakości obrazu. Zostało ono stworzone z myślą o wsparciu rozwoju nowych technik PS, które można dodawać i oceniać za pomocą najnowszych metod. Wzrost zapotrzebowania na metody dedykowane ocenie jakości w obrazach satelitarnych wymagał stworzenia narzędzia pozwalającego na porównanie nowych technik PS z najbardziej zaawansowanymi podejściami. TIQA-PSI umożliwia badaczom rozwijanie własnych technik PS-IQA i porównywanie ich z istniejącymi rozwiązaniami. Dodatkowo narzędzie to oferuje protokoły porównawcze takie jak indeksy jakości, testy istotności statystycznej Wilcoxona, wykresy typu boxplot oraz wizualizacje obrazów PS z powiększonymi obszarami.

Narzędzie TIQA-PSI zostało stworzone do testowania algorytmów oceny jakości

obrazów generowanych z wykorzystaniem technik PS. Jego kluczową funkcjonalnością jest możliwość łatwego dodawania nowych metod IQA i PS, co umożliwia porównanie własnych rozwiązań z zaawansowanymi algorytmami. TIQA-PSI wspiera również trenowanie modeli IQA na bazach obrazów zawierających subiektywne oceny jakości oraz ich ocenę na dodatkowych danych. Umożliwia ponadto kompleksowe porównanie różnych metod PS-IQA poprzez generowanie wyników w formie tabel, statystyk oraz wizualizację rezultatów.

TIQA-PSI zawiera trzy główne skrypty: *TIQA_PSI_FullDB.m*, *TIQA_PSI_Train.m* i *TIQA_PSI_PAN.m*. *TIQA_PSI_FullDB.m* odpowiada za trenowanie modeli IQA na bazie obrazów PS z ocenami subiektywnymi, generując obiektywne oceny jakości oraz rankingi obrazów testowych. Skrypt ten może wykorzystywać różne bazy danych, pod warunkiem że zawierają one subiektywne oceny jakości. Przykładowy kod przedstawiono na Listing 1 [A-7].

```
addpath(genpath('Tools'));
load Mos_imgs.mat
IQA = {%... Full-Reference methods 18
'SVD','ERGAS','Q2n','Q','SCC',...
%...'DISTS', 'HaarPSI', 'IFC', 'MDSI', 'MS-UNIQUE', 'ReSIFT', 'SAM', 'SUMMER','VIF', 'SSIM', 'MULTISSIM', 'PerSIM', 'UNIQUE',... No-Reference methods 13
'BRISQUE','MLPSIE','NOREQI','SSEQ'
%...'BLINDER', 'CEIQ', 'DIVINE', 'FRIQUEE', 'GWH-GLBP-BIQA', 'NIQE', 'SINQ','SINDEX', 'SR-metric'
};
PC_img = strcat(pwd,filesep,'Database',filesep,'SS',filesep,PC_img);
TC_img = strcat(pwd,filesep,'Database',filesep,'SS',filesep,TC_img);
%Create Svr models for IQA methods, where 0 is the svr model name
Create_Svr_Models(PC_img,TC_img,DMOS_PC_TC,0,IQA)
%% Inputs:
% images : true colour image (R+G+B), imagesRef : true colour reference image (R+G+B)
% images2 : pseudo colour image (NIR+R+G), imagesRef : pseudo colour reference image (NIR+R+G)
images ={strcat(pwd,filesep,'Database\SS\Test\TC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\TC_Test_ATWT-M2_AWGN_Level-1.tif'); strcat(pwd,filesep,'Database\SS\Test\TC_Test_ATWT-M2_AWGN_Level-2.tif'); strcat(pwd,filesep,'Database\SS\Test\TC_Test_ATWT-M2_AWGN_Level-3.tif')};
imagesRef ={strcat(pwd,filesep,'Database\SS\Test\TC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\TC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\TC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\TC_Test_REF_UD_UD.tif')};
images2 ={strcat(pwd,filesep,'Database\SS\Test\PC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\PC_Test_ATWT-M2_AWGN_Level-1.tif'); strcat(pwd,filesep,'Database\SS\Test\PC_Test_ATWT-M2_AWGN_Level-2.tif'); strcat(pwd,filesep,'Database\SS\Test\PC_Test_ATWT-M2_AWGN_Level-3.tif')};
imagesRef2 ={strcat(pwd,filesep,'Database\SS\Test\PC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\PC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\PC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\PC_Test_REF_UD_UD.tif')};
%% USAGE
% If you only have true colour image iqa_metrics(images,[],[],[],IQA,0), where 0 is the svr model name
% If you have true colour and reference images
iqa_metrics(images,imagesRef,[],[],IQA,0)
% If you have true colour and pseudo colour images
iqa_metrics(images,[],images2,[],IQA,0)
```

```
% If you have true colour and pseudo colour images and reference images
iqa_metrics(images,imagesRef,images2,imagesRef2,IQA,0)
score = iqa_metrics(images,imagesRef,images2,imagesRef2,IQA,0);
tableData = prepareScoreResult(score,[]);
```

Listing 1. Skrypt TIQA_PSI_FullDB.m

TIQA_PSI_Train.m odpowiada za podział obrazów na zbiory treningowe i testowe w proporcji 80:20. Na ich podstawie obliczane są mediany współczynników korelacji między wynikami algorytmów IQA a ocenami subiektywnymi, takich jak SRCC, KRCC i PLCC. Wyniki prezentowane są w formie tabel oraz wykresów pudełkowych (boxplot) i porównywanie przy użyciu testu istotności statystycznej Wilcoxona. Przykładowy kod przedstawiono na Listing 2.

```
rand('seed',0);
addpath(genpath('Tools'));
load Mos_imgs_2.mat
IQA = { % Full-Reference methods 18
'Q','Q2n','SSIM','ERGAS',...
...'DISTS','HaarPSI','IFC','MDSI','MS-UNIQUE',
...'ReSIFT','SAM','SCC','SUMMER','SVD','VIF','MULTISSIM','PerSIM','UNIQUE',...
... No-Reference methods 13
'MLPSIE','SINDEX','DIVINE','BRISQUE'
... 'BLINDER','CEIQ','FRIQUEE','GWH-GLBP-BIQA',
... 'NIQE','NOREQI','SINQ','SR-metric','SSEQ'
```

Listing 2. Skrypt TIQA_PSI_Train.m

TIQA_PSI_PAN.m uruchamia wybrane algorytmy PS i ocenia wyniki za pomocą metod IQA. Ostateczne rezultaty są prezentowane w tabelach, które zawierają zarówno oceny IQA, jak i rankingi algorytmów PS. Przykładowy kod przedstawiono na Listing 3.

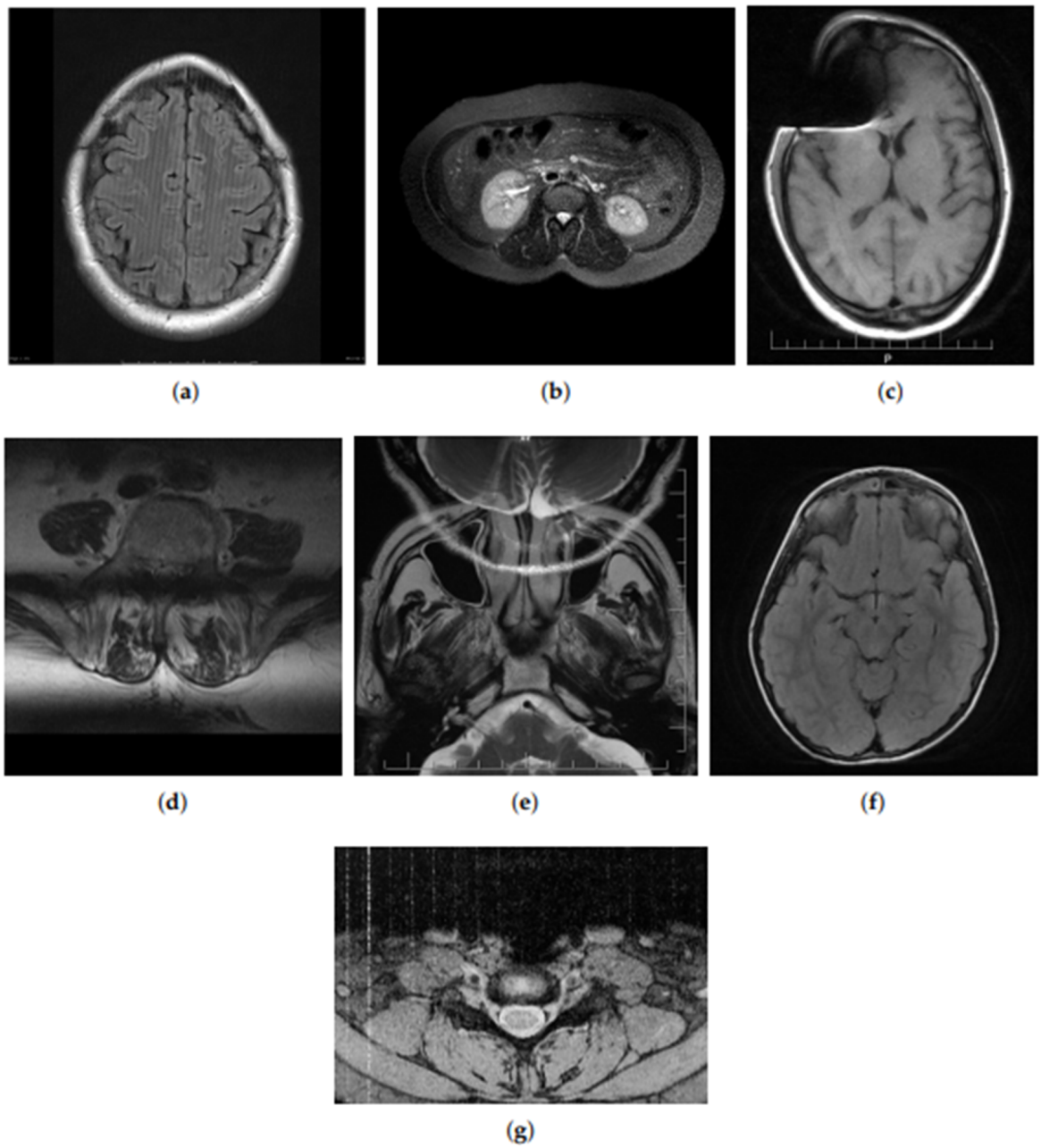
```
addpath(genpath('Tools'));
load Mos_imgs.mat
load(strcat(pwd,filesep,'Database',filesep,'Dataset',filesep,
'Test(HxWxC)_wv2_data1.mat'))
IQA = { % Full-Reference methods 18
'VIF','SSIM','SAM','HaarPSI','MDSI',...
...'DISTS', 'ERGAS', 'IFC', 'MS-UNIQUE', 'Q', 'Q2n', 'ReSIFT', 'SCC', 'SUMMER',
'SVD', 'MULTISSIM', 'PerSIM', 'UNIQUE',...
... No-Reference methods 13
'BLINDER','DIVINE','MLPSIE','SINQ'
... 'BRISQUE', 'CEIQ', 'FRIQUEE', 'GWH-GLBP-BIQA', 'NIQE', 'NOREQI', 'SINDEX',
'SR-metric', 'SSEQ',
};
PC_img = strcat(pwd,filesep,'Database',filesep,'SS',filesep,PC_img);
TC_img = strcat(pwd,filesep,'Database',filesep,'SS',filesep,TC_img);
% Create Svr models for IQA methods, where 0 is the svr model name
Create_Svr_Models(PC_img,TC_img,DMOS_PC_TC,0,IQA)
% Pan-Sharpening algorithms
algorithms = {... Require REF
'REF','EXP','AWLP','BDSD','MTF-GLP','MF'
... 'C-BDSD', 'BDSD_PC', 'BT-H', 'C-GSA', 'SR-D', 'MTF-GLP-FS', 'MTF-GLP-HPM',
'MTF-GLP-HPM-R', 'MTF-GLP-CBD', 'FE-HPM', 'PRACS','MTF-GLP-HPM-H'
};
ratio = 4; % Resize Factor
sensor = 'WV2'; % Sensor
[imageFused,imagesRef,VisibleImage,algorithms] = Pan_sharpening_algorithms(gt,lms,ms,
pan,ratio,algorithms,sensor);
```

```
%% Inputs:
% ImageFused : true colour fused image (R+G+B)
% imagesRef : true colour reference image (R+G+B)
score = iqa_metrics(imageFused,imagesRef,[],[],IQA,0);
tableData = prepareScoreResult(score,algorithms);
```

Listing 3. Skrypt TIQA_PSI_PAN.m

TIQA-PSI oferuje szeroki zakres funkcjonalności, czyniąc go wszechstronnym narzędziem do oceny jakości obrazów generowanych za pomocą technik PS. Dzięki elastycznej strukturze użytkownicy mogą integrować nowe algorytmy, wykorzystując zarówno dostępne w narzędziu metody, jak i własne rozwiązania. Pierwszą kluczową funkcjonalnością TIQA-PSI jest trenowanie modeli jakości dla metod IQA opartych na predykcji. Modele te mogą być uczone na pełnych zbiorach danych lub ich podzbiorach. Narzędzie wspiera także podejście end-to-end, umożliwiając bezpośrednie mapowanie obrazów wejściowych na oceny jakości. Drugą istotną funkcją jest porównywanie metod IQA. TIQA-PSI umożliwia ocenę skuteczności poszczególnych algorytmów na podstawie zgodności ocen obiektywnych z subiektywnymi, wykorzystując do tego współczynniki korelacji. Wyniki przedstawiane są w formie tabel, wykresów typu boxplot oraz analiz statystycznych, takich jak test istotności Wilcoxona. Trzecią funkcjonalnością jest ocena jakości metod PS. Narzędzie umożliwia przetwarzanie obrazów PAN i MS z wykorzystaniem wybranych algorytmów PS, generowanie obrazów PS oraz ocenę ich jakości przy użyciu dostępnych metod IQA.

Wkład własny autora w ramach pracy nad artykułem zawierał projekt koncepcji zestawu narzędzi, wybranie odpowiednich metody IQA oraz technik PS, a także opracowanie parametrów algorytmów w oparciu o zgromadzone dane. Wkład obejmował również implementację funkcjonalności umożliwiających generowanie wyników oceny metod IQA, a także statystyczne porównanie ich efektywności przy użyciu wskaźników korelacji i testów statystycznych. W ramach testów walidacyjnych przeprowadzono ocenę jakości obrazu na podstawie danych z bazy IKONOS oraz dodatkowych zbiorów, co pozwoliło na wyciągnięcie wniosków dotyczących efektywności poszczególnych metod. Autor współtworzył również analizę wyników testów, przygotował wizualizacje oraz opracował dokumentację prezentowanego narzędzia.



Rys. 2.1. Przykłady artefaktów na obrazach MR: (a) herringbone; (b) ghosting; (c) magnetic susceptibility; (d) slice overlap; (e) aliasing; (f) Gibbs effect; and (g) zipper. Przypadki zostały udostępnione dzięki uprzejmości: (a) Assoc. Prof. Frank Gaillard, rID: 19695; (b) Assis. Prof. Faeze Salahshour, rID: 81727; (c) Dr. Ayush Goel, rID: 22731; (d) Dr. Roberto Schubert, rID: 16705; (e) Dr. Prashant Mudgal, rID: 26927; (f) Dr. Prashant Mudgal, rID: 27302; i (g) Dr. Alan Nazerian, rID: 45665; (dostęp: 27 kwietnia 2022)[4]

3. Podsumowanie i wnioski

Niniejsza praca koncentruje się na opracowaniu i przedstawieniu automatycznych metod percepcyjnej oceny jakości obrazów cyfrowych bez refrencji z wykorzystaniem sieci neuronowych. Zaprezentowane badania potwierdzają skuteczność metod sztucznej inteligencji, w szczególności różnorodnych struktur sieci neuronowych w ocenie ich jakości. Postawiona hipoteza, że możliwe jest zastosowanie metod sztucznej inteligencji, w szczególności różnorodnych architektur sieci neuronowych, do oceny jakości obrazów satelitarnych oraz obrazów rezonansu magnetycznego bez referencji, zapewniając wysoką zgodność z subiektywnymi ocenami ekspertów została potwierdzona.

W świetle uzyskanych wyników można stwierdzić, że efektywność zaproponowanych rozwiązań jest porównywalna lub lepsza w porównaniu z istniejącymi metodami z literatury. Ich skuteczność została w znacznym stopniu potwierdzona w badanich eksperymentalnych na wielu zróżnicowanych zbiorach danych, poprzez realizację następujących zadań badawczych:

1. **Analiza literatury dotyczącej najnowszych metod oceny jakości obrazów rezonansu magnetycznego, z uwzględnieniem typowych zniekształceń, protokołów zastosowań oraz wykorzystywanych baz danych.**

Zadanie zostało zrealizowane poprzez analizę literatury dotyczącej najnowszych metod oceny jakości obrazów rezonansu magnetycznego, uwzględniając typy zniekształceń najczęściej występujących w tym rodzaju obrazowania, protokoły zastosowań tych metod oraz najczęściej używane bazy obrazów. W przypadku obrazów rezonansu magnetycznego przeanalizowano zniekształcenia związane z szumem termicznym, artefaktami ruchowymi, zniekształceniami geometrycznymi oraz efektami heterogeniczności pola magnetycznego, wskazując na ich wpływ na jakość i interpretację wyników. Omówiono specyfikę procedur oceny, podkreślając znaczenie odpowiedniego doboru wskaźników jakości.

Dodatkowo przedstawiono przegląd baz danych używanych w literaturze. Dzięki tak przeprowadzonej analizie literaturowej uzyskano pełny obraz aktualnych trendów i metod w ocenie jakości obrazowania MRI, umożliwiając wskazanie dalszych kierunków badań oraz potencjalnych wyzwań w tej dziedzinie.

2. **Opracowanie narzędzi do percepcyjnej oceny jakości wyostrzonych obrazów satelitarnych oraz platformy do oceny jakości obrazów rezonansu magnetycznego.**

Zadanie zostało zrealizowane dzięki opracowaniu dwóch narzędzi: TIQA-MRI oraz TIQA-PSI, które w nowoczesny sposób wspierają ocenę jakości obrazów w dwóch kluczowych obszarach: rezonansu magnetycznego oraz obrazów satelitarnych generowanych za pomocą technik wyostrzania. Oba rozwiązania umożliwiają przeprowadzanie złożonych analiz i porównań skuteczności metod oceny jakości obrazu, oferując badaczom narzędzia dostosowane do specyficznych potrzeb w ich dziedzinach. TIQA-MRI

jest dedykowaną platformą do oceny jakości obrazów rezonansu magnetycznego, integrująca dziewięć przykładowych metod IQA. Narzędzie zostało wyposażone w intuicyjny interfejs umożliwiający ocenę obrazów, trening modeli na podstawie danych z przykładowej bazy oraz analizę wyników w formie wykresów i tabel. Zaprojektowane z myślą o integracji nowych algorytmów, konfiguracji parametrów eksperymentów oraz analizie istotności statystycznej. Dzięki elastycznej architekturze aplikacji możliwe jest zapisywanie ustawień eksperymentów i ich ponowne wykorzystanie w przyszłych badaniach, co znacząco usprawnia proces analizy oraz porównania metod. TIQA-PSI koncentruje się na percepcyjnej ocenie jakości wyostrzonych obrazów satelitarnych. Narzędzie zawiera aż 30 metod oceny jakości, w tym 16 dedykowanych technikom PS, i umożliwia zarówno rozwój nowych algorytmów, jak i ich ocenę w odniesieniu do istniejących metod. TIQA-PSI wykorzystuje współczynniki korelacji (SRCC, KRCC, PLCC), testy istotności statystycznej Wilcoxona, a także umożliwia wizualizację wyników za pomocą wykresów boxplot i tabel. Narzędzie pozwala na trenowanie modeli na danych pochodzących z przykładowej bazy, przetwarzanie obrazów PAN i MS przy użyciu wybranych algorytmów PS oraz generowanie rankingów metod na podstawie obiektywnych wskaźników jakości. Dzięki integracji zaawansowanych technik uczenia maszynowego takich jak regresja wektorów nośnych, TIQA-PSI oferuje kompleksowe wsparcie w badaniach nad przetwarzaniem obrazów satelitarnych.

Zarówno TIQA-MRI, jak i TIQA-PSI zostały zaprojektowane z myślą o elastyczności i łatwości użytkowania. Oba narzędzia oferują wszechstronne funkcjonalności, pozwalając na wprowadzanie nowych algorytmów, dostosowywanie parametrów eksperymentów i generowanie raportów ułatwiających analizę wyników. Stanowią one kompleksowe rozwiązanie wspierające rozwój nowych technik.

3. **Zastosowanie architektur głębokich sieci neuronowych w ocenie jakości obrazów satelitarnych oraz rezonansu magnetycznego, w celu opracowania metod umożliwiających ocenę obrazów z wykorzystaniem:**

a) **Zewnętrznego połączenia architektur głębokich splotowych sieci neuronowych**

Zadania związane z oceną jakości obrazów MRI oraz PS zostały zrealizowane poprzez opracowanie i wdrożenie innowacyjnych metod opartych na głębokich sieciach neuronowych. W ramach badań zastosowano różnorodne podejścia takie jak wewnętrzne połączenie komplementarnych architektur oraz modyfikacje podsieci by dopasować je do wymagań analizy percepcyjnej. Prace te zaowocowały publikacjami naukowymi, które szczegółowo opisują opracowane metody i uzyskane wyniki, potwierdzając ich skuteczność w precyzyjnej ocenie jakości obrazów. Opublikowano dwa artykuły naukowe: [A-2] i [A-6]. W artykule [A-2] zaprezentowano metodę oceny jakości obrazów rezonansu magnetycznego, opartą na fuzji głębokich splotowych sieci neuronowych oraz uczeniu transferowym, co pozwoliło na skuteczną ocenę jakości obrazów MRI. Natomiast artykuł [A-6] przedstawia metodę MLPSIE, która łączy dwie sieci głębokiego uczenia

(ResNet-18 i VGG-19) w celu oceny jakości obrazów PS. W podejściu tym obie sieci trenowane są wspólnie, co umożliwia przetwarzanie wielopoziomowych danych wejściowych i uzyskiwanie cech wrażliwych na zniekształcenia, pozwalając na ocenę jakości obrazów satelitarnych bez potrzeby korzystania z oryginalnych obrazów referencyjnych. Wyniki badań zaprezentowane w obu artykułach wskazują na skuteczność zaproponowanych metod w obszarze oceny jakości obrazów w różnych dziedzinach.

b) **Informacji wyodrębnionych z warstw sieci**

Zadanie to zostało zrealizowane przez opracowanie metod opublikowanych w artykułach [A-3], [A-6] oraz [A-7], w których zaproponowano różne metody oceny jakości obrazów z wykorzystaniem głębokich sieci neuronowych i zaawansowanych technik analizy danych. W artykule [A-3] opracowano metodę MEDQEMRI, łączącą architektury ResNet-18 i ResNet-50 w celu fuzji wielowymiarowych danych z ich warstw. Kluczowym elementem tej metody jest ponowne trenowanie sieci z uwzględnieniem wzajemnych interakcji, co umożliwia skuteczne różnicowanie obrazów rezonansu magnetycznego o różnym poziomie jakości percepcyjnej. Natomiast w [A-6] zastosowano Kernel PCA do redukcji wymiarowości i integracji cech wyodrębnionych z warstw sieci ResNet-18 i VGG-19, co w połączeniu z modelem SVR pozwoliło na zdobycie wyników oceny jakości obrazów. Z kolei w [A-7] zaproponowano architekturę TBN-PSI, wykorzystującą specjalnie zaprojektowane warstwy sieci Inception-v3 do ekstrakcji cech z kanałów RGB oraz NIR+GB. Wyodrębnione cechy są łączone, a następnie przetwarzane przez warstwy regresyjne, co zapewnia wysoką dokładność w ocenie jakości obrazów PS. Wszystkie trzy metody wnoszą istotny wkład w rozwój technik oceny jakości wyostrzonych obrazów satelitarnych i rezonansu magnetycznego poprzez wykorzystanie głębokich sieci neuronowych do efektywnego wyodrębniania i przetwarzania wielowymiarowych danych.

c) **Wewnętrznego połączenia dwóch architektur sieci**

Zadanie zostało zrealizowane poprzez utworzenie nowej metody i publikację artykułów [A-3] i [A-6], w których zaprezentowano zaawansowane metody oceny jakości obrazów z wykorzystaniem głębokich sieci neuronowych i integracji uzupełniających się architektur. W artykule [A-3] opisano metodę MEDQEMRI, która łączy architektury ResNet-18 i ResNet-50. Zastosowano w niej fuzję dwóch sieci głębokiego uczenia, umożliwiającą wzajemne oddziaływanie między sieciami podczas procesu trenowania, co zwiększyło ich moc dyskryminacyjną. Dzięki wykorzystaniu uczenia transferowego oraz ponownego dostosowanego trenowania do obrazów MRI, możliwe było znaczne zwiększenie precyzji w zadaniu oceny jakości obrazów MRI (NR-IQA), co pozwoliło na skuteczniejsze rozróżnianie obrazów o różnym stopniu zniekształcenia. Z kolei artykuł [A-6] przedstawia metodę MLPSIE, która wykorzystuje dwie komplementarne architektury głębokich sieci neuronowych – ResNet-18 i VGG-19 – do ekstrakcji wielowymiarowych cech obrazów satelitarnych. Wyodrębnione cechy są następnie redukowane

za pomocą Kernel PCA i integrowane w jedną reprezentację, co pozwala na stworzenie modelu jakości opartego na SVR. Dzięki temu podejściu możliwe jest precyzyjne przewidywanie jakości obrazów satelitarnych w sposób skorelowany z ocenami ludzkimi. Obie metody w pełni realizują założenia zadania poprzez zastosowanie głębokich sieci neuronowych do opracowania zaawansowanych metod oceny jakości obrazów bazujących na integracji cech z uzupełniających się architektur głębokiego uczenia.

d) **Wykorzystanie struktury sieci złożonej z trzech zaproponowanych podsieci.**

Zadanie zostało zrealizowane poprzez utworzenie sieci i opis proponowanej metody oceny jakości wyostrzonych obrazów satelitarnych. W pracy zaproponowano trójgałęziową sieć neuronową TBN-PSI, umożliwiającą ocenę jakości obrazów PS z uwzględnieniem specyficznych cech przestrzennych i spektralnych. Zastosowano autorski sposób łączenia informacji o różnej granularności poprzez integrację niskopoziomowych cech i semantyki wysokiego poziomu w procesie fuzji. W pracy uwzględniono zarówno dobór architektury sieci opartej na modyfikacji Inception-v3, jak i zastosowanie specjalnych metryk takich jak SSIM. Model przetestowano na obszernej bazie danych NBU PansharpRSData uzyskując znaczącą przewagę nad istniejącymi metodami oceny jakości PS.

3.1. Wkład autora

Główny wkład autora rozprawy w działalność naukową w dyscyplinie Informatyka Techniczna i Telekomunikacja polega na:

1) Sformułowaniu problemów badawczych istotnych z perspektywy usprawnienia i poprawy jakości procesów oceny obrazów satelitarnych (PS IQA) oraz obrazów medycznych (MR IQA), w tym zastosowania sztucznej inteligencji do ich analizy.

2) Współautorstwie koncepcji oraz projektowaniu metodologii badawczej w publikacjach naukowych dotyczących zastosowania głębokich sieci neuronowych do oceny jakości obrazów.

3) Opracowaniu nowych metod integrujących architektury głębokiego uczenia takich jak:

   a) MLPSIE – metoda wykorzystująca architektury ResNet-18 i VGG-19 do oceny jakości obrazów satelitarnych (PS IQA), w tym implementacja Kernel PCA do redukcji wymiarowości i budowy wielopoziomowych reprezentacji obrazu,

   b) TBN-PSI – zaprojektowanie trójgałęziowej sieci neuronowej (Three-Branch Neural Network) do analizy obrazów wielospektralnych,

c) MEDQEMRI – zaprojektowanie podejścia łączącego architektury ResNet-18 i ResNet-50 w celu oceny jakości obrazów rezonansu magnetycznego,

d) R50GR18, R18GR50M, MR50 - metoda oparta na fuzji sieci neuronowych do MR IQA – obejmująca wykorzystanie różnych architektur sieci neuronowych, takich jak ResNet, GoogLeNet i MobileNet, przetestowana w trzech wariantach.

4) Adaptacji architektur sieci neuronowych oraz algorytmów sztucznej inteligencji do specyfiki zadań IQA, w tym zastosowanie transferowego uczenia i odpowiednich technik fuzji cech w procesie trenowania.

5) Znaczącym udziale w realizacji eksperymentów obliczeniowych, obejmujących:

a) Implementację algorytmów i funkcjonalności do testowania metod IQA,

b) Walidację wyników metod na bazach danych obrazów,

c) Analizę uzyskanych rezultatów oraz ich statystyczne porównanie za pomocą korelacji i testów istotności statystycznej.

6) Opracowaniu narzędzi umożliwiających ocenę i integrację nowych metod PS i IQA, w tym wizualizację wyników oraz ich dokumentację (TIQA-MRI, TIQA-PSI).

7) Przygotowaniu grafik ilustrujących architekturę zaproponowanych metod, ich działanie oraz wyniki badań porównawczych.

8) Znaczącym wkładzie w przygotowanie publikacji naukowych, w tym redagowanie tekstów, analizę literatury oraz interpretację wyników badań.

9) Wyciąganiu wniosków wynikających z przeprowadzonych badań, które otwierają perspektywy dla dalszego rozwoju zaawansowanych metod fuzji sieci neuronowych oraz ich zastosowań w ocenie obrazów.

3.2. Kierunki dalszych badań

Obiecujące wyniki przeprowadzonych eksperymentów potwierdzają zasadność kontynuacji prac, zaś interdyscyplinarny charakter przedstawionych rozwiązań otwiera szerokie możliwości ich dalszego rozwijania i doskonalenia. W oparciu o uzyskane wyniki,zidentyfikowano następujące kierunki dalszych badań:

1) Rozwój metod jakości obrazów MRI:

a) Eksploracja alternatywnych podejść do fuzji sieci głębokiego uczenia i optymalizacji ich architektury w kontekście predykcji jakości obrazów MRI,

b) Opracowanie nowej generacji metod bez referencji do oceny jakości obrazów medycznych, uwzględniających specyfikę innych modalności medycznych takich jak RTG czy CT,

c) Tworzenie dużych baz danych obrazów MRI zawierających subiektywne oceny radiologów, co umożliwi bardziej precyzyjne modele predykcji jakości.

2) Rozwój technologii oceny jakości obrazów PS:

a) Połączenie metod głębokiego uczenia z klasycznymi podejściami do oceny jakości obrazów w celu poprawy analizy obrazów PS,

b) Organizacja subiektywnych testów obejmujących szerszy zestaw obrazów satelitarnych, takich jak QuickBird czy WorldView, z uwzględnieniem wyników różnych metod PS,

c) Rozszerzenie narzędzi MATLAB w kierunku metod oceny jakości zaimplementowanych w Pythonie, aby zwiększyć ich uniwersalność.

d) Uwzględnienie nowych architektur sieci neuronowych takich jak Transformery lub podejść opartych o LLM.

3) Rozwój interdyscyplinarnych metod dla MRI i PS:

a) Przeprowadzenie szczegółowej analizy możliwości adaptacji metod IQA stosowanych w obrazach naturalnych do domeny obrazów medycznych i satelitarnych,

b) Optymalizacja technik redukcji wymiarowości dla wysokowymiarowych cech z sieci głębokiego uczenia, w celu zachowania kluczowych informacji jakościowych.

4) Zastosowania praktyczne i optymalizacja algorytmów:

a) Przeprowadzenie testów wydajnościowych pod kątem zużycia zasobów sprzętowych oraz czasu działania opracowanych metod na różnych zestawach danych,

b) Opracowanie algorytmów poprawy jakości obrazów opartych na regresji zdolnych do subtelnego rozróżniania różnic jakości.

Literatura

[1] Rafiq Abdul. *New Learning Frameworks for Blind Image Quality Assessment Model*. PhD thesis, University of Sheffield, Sheffield, UK, 2018.

[2] O. A. Agudelo-Medina, H. D. Benitez-Restrepo, G. Vivone, and A. Bovik. Perceptual quality assessment of pan-sharpened images. *Remote Sensing*, 11:877, 2019.

[3] Iqbal Tariq Ahmed and Chee Siong Der. Enhancement of no-reference image quality assessment for contrast-distorted images using natural scene statistics features in curvelet domain. In *Proceedings of the 2017 7th IEEE International Conference on System Engineering and Technology (ICSET)*, pages 128–133, Shah Alam, Malaysia, 2017.

[4] Alexander Andreopoulos and John K. Tsotsos. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac mri. *Medical Image Analysis*, 12(3):335–357, 2008.

[5] Frédéric Babonneau, Carlos Beltran, Alain Haurie, Christian Tadonki, and Jean-Philippe Vial. Proximal-accpm: A versatile oracle based optimisation method. In *Optimisation, Econometric and Financial Analysis*, pages 67–89. Springer, Berlin/Heidelberg, Germany, 2007.

[6] Wenjia Bai, Matthew Sinclair, Giacomo Tarroni, Ozan Oktay, Martin Rajchl, Ghislain Vaillant, Aaron M Lee, Nay Aung, Elena Lukaschuk, Mihir M Sanghvi, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*, 20(1):1–12, 2018.

[7] V. Bhateja, M. Nigam, A. S. Bhadauria, and A. Arya. Two-stage multi-modal mr images fusion method based on parametric logarithmic image processing (plip) model. *Pattern Recognition Letters*, 136:25–30, 2020.

[8] K. T. Block, M. Uecker, and J. Frahm. Suppression of mri truncation artifacts using total variation constrained data extrapolation. *International Journal of Biomedical Imaging*, 2008.

[9] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[10] Z. Caramanos, V. S. Fonov, S. J. Francis, S. Narayanan, G. B. Pike, D. L. Collins, and D. L. Arnold. Gradient distortions in mri: Characterizing and correcting for

their effects on siena-generated measures of brain volume change. *Neuroimage*, 49:1601–1611, 2010.

[11] Hong-Wei Chang and Ming-Hsuan Wang. Sparse correlation coefficient for objective image quality assessment. *Signal Processing: Image Communication*, 26:577–588, 2011.

[12] L. S. Chow and R. Paramesran. Review of medical image quality assessment. *Biomedical Signal Processing and Control*, 27:145–154, 2016.

[13] L. S. Chow and H. Rajagopal. Modified-brisque as no reference image quality assessment for structural mr images. *Magnetic Resonance Imaging*, 43:74–87, 2017.

[14] Li Sze Chow and Heshalini Rajagopal. Modified-brisque as no reference image quality assessment for structural mr images. *Magnetic resonance imaging*, 43:74–87, 2017.

[15] George E. Dahl, Navdeep Jaitly, and Ruslan Salakhutdinov. Multi-task neural networks for qsar predictions. *arXiv preprint*, arXiv:1406.1231, 2014.

[16] Adriana Di Martino, David O'connor, Bosi Chen, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Joshua H Balsters, Leslie Baxter, Anita Beggiato, Sylvie Bernaerts, et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii. *Scientific data*, 4(1):1–15, 2017.

[17] David Alexander Dickie, Susan D Shenkin, Devasuda Anblagan, Juyoung Lee, Manuel Blesa Cabez, David Rodriguez, James P Boardman, Adam Waldman, Dominic E Job, and Joanna M Wardlaw. Whole brain magnetic resonance image atlases: a systematic review of existing atlases and caveats for use in population imaging. *Frontiers in neuroinformatics*, 11:1, 2017.

[18] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words. *arXiv preprint arXiv:2010.11929*, 7, 2020.

[19] R. L. Ehman and J. P. Felmlee. Flow artifact reduction in mri: A review of the roles of gradient moment nulling and spatial presaturation. *Magnetic Resonance in Medicine*, 14:293–307, 1990.

[20] Oscar Esteban, Daniel Birman, Marie Schaer, Oluwasanmi O Koyejo, Russell A Poldrack, and Krzysztof J Gorgolewski. Mriqc: Advancing the automatic prediction of image quality in mri from unseen sites. *PloS one*, 12(9):e0184661, 2017.

[21] Oscar Esteban, Ross W Blair, Dylan M Nielson, Jan C Varada, Sean Marrett, Adam G Thomas, Russell A Poldrack, and Krzysztof J Gorgolewski. Crowdsourced mri quality metrics and expert quality annotations for training of humans and machines. *Scientific data*, 6(1):1–7, 2019.

[22] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, 1980.

[23] X. Gao, W. Lu, D. Tao, and X. Li. Image quality assessment and human visual system. In *SPIE Proceedings*, volume 7744, pages 316–325, 2010.

[24] Xin Gao, Jing Yu, Ying Zhu, Xia Huang, and Qi Tian. Blind image quality prediction by exploiting multi-level deep representations. *Pattern Recognition*, 81:432–442, 2018.

[25] Deepti Ghadiyaram and Alan C. Bovik. Perceptual quality prediction on authentically distorted images using a bag of features approach. *arXiv*, 2016.

[26] K. Gu, G. Zhai, X. Yang, and W. Zhang. Hybrid no-reference quality metric for singly and multiply distorted images. *IEEE Transactions on Broadcasting*, 60:555–567, 2014.

[27] K. Gu, G. Zhai, X. Yang, and W. Zhang. Using free energy principle for blind image quality assessment. *IEEE Transactions on Multimedia*, 17:50–63, 2015.

[28] R. Gupta, D. Bansal, and C. Singh. Image quality assessment using non-linear multi-metric fusion approach. *International Journal of Emerging Technology and Advanced Engineering*, 4:822–826, 2014.

[29] R. H. Hashemi, W. G. Bradley, and C. J. Lisanti. *MRI: The Basics: The Basics*. Lippincott Williams & Wilkins, Philadelphia, PA, USA, 2012.

[30] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[31] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[33] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[34] M. Hoff, J. Andre, and B. Stewart. Artifacts in magnetic resonance imaging. *Magnetic Resonance Imaging*, pages 165–190, 2016.

[35] Feng-Hsiung Hsu. *Behind Deep Blue: Building the Computer that Defeated the World Chess Champion*. Princeton University Press, 2002.

[36] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.

[37] T. Iqbal and H. Ali. Generative adversarial network for medical images (mi-gan). *Journal of Medical Systems*, 42:231, 2018.

[38] Jinseong Jang, Kihun Bang, Hanbyol Jang, Dosik Hwang, and Alzheimer's Disease Neuroimaging Initiative. Quality evaluation of no-reference mr images using multidirectional filters and image statistics. *Magnetic resonance in medicine*, 80(3):914–924, 2018.

[39] L. Kang, P. Ye, Y. Li, and D. Doermann. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1733–1740, 2014.

[40] P. Kaur, S. S. Kumaran, R. Tripathi, S. Khushu, and S. Kaushik. Protocol error artifacts in mri: Sources and remedies revisited. *Radiography*, 13:291–306, 2007.

[41] Ronald Kemker, Carl Salvaggio, and Christopher Kanan. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *arXiv*, 2018.

[42] W. Kucharczyk, A. P. Crawley, W. Kelly, and R. M. Henkelman. Effect of multislice interference on image contrast in t2- and t1-weighted mr images. *American Journal of Neuroradiology*, 9:443–451, 1988.

[43] RK-S Kwan, Alan C Evans, and G Bruce Pike. Mri simulation-based evaluation of image-processing and classification methods. *IEEE transactions on medical imaging*, 18(11):1085–1097, 1999.

[44] M. L. Leclaire, Arthur. No-reference image quality assessment and blind deblurring with sharpness metrics exploiting fourier phase information. *Journal of Mathematical Imaging and Vision*, 52:145–172, 2015.

[45] Yann LeCun, P. Haffner, L. Bottou, Y. Bengio, P. Howard, and P. Simard. Object recognition with gradient-based learning. *Feature Grouping*, 66:233–240, 1988.

[46] Qiang Li, Weisi Lin, and Yuming Fang. No-reference quality assessment for multiply-distorted images in gradient domain. *IEEE Signal Processing Letters*, 23:541–545, 2016.

[47] S. Liu, K. H. Thung, W. Lin, D. Shen, and P. T. Yap. Hierarchical nonlocal residual networks for image quality assessment of pediatric diffusion mri with limited and noisy annotations. *IEEE Transactions on Medical Imaging*, 39:3691–3702, 2020.

[48] Siyuan Liu, Kim-Han Thung, Weili Lin, Dinggang Shen, and Pew-Thian Yap. Hierarchical nonlocal residual networks for image quality assessment of pediatric diffusion mri with limited and noisy annotations. *IEEE Transactions on Medical Imaging*, 39(11):3691–3702, 2020.

[49] Tzu-Jui Liu and Kai-Hsiang Liu. No-reference image quality assessment by wide-perceptual-domain scorer ensemble method. *IEEE Transactions on Image Processing*, 27:1138–1151, 2018.

[50] Nick Lowe and et al. Analysis of the expression patterns, subcellular localisations and interaction partners of drosophila proteins using a pigp protein trap library. *Development*, 141(20):3994–4005, 2014.

[51] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo. End-to-end blind image quality assessment using deep neural networks. *IEEE Transactions on Image Processing*, 27:1202–1213, 2018.

[52] Kede Ma, Weisi Liu, Tong Liu, Zhou Wang, and Dacheng Tao. dipiq: Blind image quality assessment by learning-to-rank discriminable image pairs. *IEEE Transactions on Image Processing*, 26:3951–3964, 2017.

[53] M. Maqsood, F. Nazir, U. Khan, F. Aadil, H. Jamal, I. Mehmood, and O. Song. Transfer learning assisted classification and detection of alzheimer's disease stages using 3d mri scans. *Sensors*, 19:2645, 2019.

[54] Samira Masoudi, Stephanie Harmon, Sherif Mehralivand, Nathan Lay, Ulas Bagci, Bradford J Wood, Peter A Pinto, Peter Choyke, and Baris Turkbey. No-reference image quality assessment of t2-weighted magnetic resonance images in prostate cancer patients. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1201–1205. IEEE, 2021.

[55] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[56] Xin Meng, Yu Xiong, Feng Shao, Huanfeng Shen, Wei Sun, Guang Yang, Qiang Yuan, Rong Fu, and Hongyan Zhang. A large-scale benchmark data set for evaluating pansharpening performance: Overview and implementation. *IEEE Geoscience and Remote Sensing Magazine*, 9:18–52, 2021.

[57] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, and C. W. Chen. Blind quality assessment based on pseudo-reference image. *IEEE Transactions on Multimedia*, 20:2049–2062, 2018.

[58] Xiongkuo Min, Ke Gu, Guangtao Zhai, Jia Liu, Xiaokang Yang, and Chang Wen Chen. Blind quality assessment based on pseudo-reference image. *IEEE Transactions on Multimedia*, 20:2049–2062, 2017.

[59] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20:209–212, 2012.

[60] Bénédicte Mortamet, Matt A Bernstein, Clifford R Jack Jr, Jeffrey L Gunter, Chadwick Ward, Paula J Britson, Reto Meuli, Jean-Philippe Thiran, and Gunnar Krueger. Automatic quality assessment in structural brain magnetic resonance imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 62(2):365–372, 2009.

[61] Shahabedin Nabavi, Hossein Simchi, Mohsen Ebrahimi Moghaddam, Alejandro F Frangi, and Ahmad Ali Abin. Automatic multi-class cardiovascular magnetic resonance image quality assessment using unsupervised domain adaptation in spatial and frequency domains. *arXiv preprint arXiv:2112.06806*, 2021.

[62] Qiang Ning, Chao Ma, Fan Lam, and Zhi-Pei Liang. Spectral quantification for high-resolution mr spectroscopic imaging with spatiospectral constraints. *IEEE Transactions on Biomedical Engineering*, 64(5):1178–1186, 2016.

[63] P. Noël, R. Bammer, C. Reinhold, and M. A. Haider. Parallel imaging artifacts in body magnetic resonance imaging. *Canadian Association of Radiologists Journal*, 60:91–98, 2009.

[64] R. Obuchowicz, M. Oszust, M. Bielecka, A. Bielecki, and A. Piórkowski. Magnetic resonance image quality assessment by using non-maximum suppression and entropy analysis. *Entropy*, 22:220, 2020.

[65] Rafał Obuchowicz, Mariusz Oszust, Marzena Bielecka, Andrzej Bielecki, and Adam Piórkowski. Magnetic resonance image quality assessment by using non-maximum suppression and entropy analysis. *Entropy*, 22(2):220, 2020.

[66] Michael Osadebey, Marius Pedersen, Douglas Arnold, and Katrina Wendel-Mitoraj. Image quality evaluation in clinical research: A case study on brain and cardiac mri images in multi-center clinical trials. *IEEE journal of translational engineering in health and medicine*, 6:1–15, 2018.

[67] M. Oszust. No-reference image quality assessment using image statistics and robust feature descriptors. *IEEE Signal Processing Letters*, 24:1656–1660, 2017.

[68] M. Oszust. No-reference image quality assessment with local gradient orientations. *Symmetry*, 11:95, 2019.

[69] M. Oszust, A. Piórkowski, and R. Obuchowicz. No-reference image quality assessment of magnetic resonance images with high-boost filtering and local features. *Magnetic Resonance in Medicine*, 84:1648–1660, 2020.

[70] Mariusz Oszust, Adam Piórkowski, and Rafał Obuchowicz. No-reference image quality assessment of magnetic resonance images with high-boost filtering and local features. *Magnetic Resonance in Medicine*, 84(3):1648–1660, 2020.

[71] J. A. Patton, M. V. Kulkarni, J. K. Craig, O. H. Wolfe, R. Price, C. L. Partain, and A. E. Jr. James. Techniques, pitfalls and artifacts in magnetic resonance imaging. *Radiographics*, 7:505–519, 1987.

[72] Rodrigo Pires de Lima and Kurt Marfurt. Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. *Remote Sensing*, 12:86, 2020.

[73] Russell A Poldrack and Krzysztof J Gorgolewski. Openfmri: Open sharing of task fmri data. *Neuroimage*, 144:259–261, 2017.

[74] K. Qi, H. Li, C. Rong, Y. Gong, C. Li, H. Zheng, and S. Wang. Blind image quality assessment for mri with a deep three-dimensional content-adaptive hyper-network. *arXiv*, arXiv:2107.06888, 2021.

[75] Kehan Qi, Haoran Li, Chuyu Rong, Yu Gong, Cheng Li, Hairong Zheng, and Shanshan Wang. Blind image quality assessment for mri with a deep three-dimensional content-adaptive hyper-network. *arXiv preprint arXiv:2107.06888*, 2021.

[76] Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386, 1958.

[77] Stuart J Russell and Peter Norvig. *Artificial intelligence: a modern approach*. pearson, 2016.

[78] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. MobileNetV2: Inverted Residuals and Linear Bottlenecks . In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4510–4520, Los Alamitos, CA, USA, June 2018. IEEE Computer Society.

[79] Suraj D Serai. Basics of magnetic resonance imaging and quantitative parameters t1, t2, t2*, t1rho and diffusion-weighted imaging. *Pediatric radiology*, 52(2):217–227, 2022.

[80] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations (ICLR 2015)*, pages 1–14. Computational and Biological Learning Society, 2015.

[81] I. Stepien and M. Oszust. No-reference quality assessment of pan-sharpening images with multi-level deep image representations. *Remote Sensing*, 14, 2022.

[82] I. Stepien and M. Oszust. Three-branch neural network for no-reference quality assessment of pan-sharpened images. *Engineering Applications of Artificial Intelligence*, 139:109594, 2025.

[83] Igor Stępień, Rafał Obuchowicz, Adam Piórkowski, and Mariusz Oszust. Fusion of deep convolutional neural networks for no-reference magnetic resonance image quality assessment. *Sensors*, 21(4):1043, 2021.

[84] I. Stępień. Tiqa-mri: Toolbox for perceptual image quality assessment of magnetic resonance images. *SoftwareX*, 2025. (in press).

[85] I. Stępień, R. Obuchowicz, A. Piórkowski, and M. Oszust. Fusion of deep convolutional neural networks for no-reference magnetic resonance image quality assessment. *Sensors*, 21(4):1043, 2021.

[86] I. Stępień and M. Oszust. No-reference image quality assessment of magnetic resonance images with multi-level and multi-model representations based on fusion of deep architectures. *Engineering Applications of Artificial Intelligence*, 123:Part A, 106283, 2023.

[87] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, Boston, MA, USA, June 2015.

[88] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, Las Vegas, NV, USA, June 2016.

[89] Huixuan Tang, Neel Joshi, and Ashish Kapoor. Learning a blind measure of perceptual image quality. In *CVPR 2011*, pages 305–312, 2011.

[90] Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.

[91] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[92] Lucien Wald. Quality of high resolution synthesised images: Is there a simple criterion? page 1170, 2000.

[93] Wei Wang, Mingjie Zhang, Ding Wang, and Yuwu Jiang. Kernel pca feature extraction and the svm classification algorithm for multiple-status, through-wall, human being detection. *EURASIP Journal on Wireless Communications and Networking*, 2017:151, 2017.

[94] Yida Wang, Yang Song, Fang Wang, Jingjing Sun, Xinyi Gao, Zhe Han, Lei Shi, Guoliang Shao, Mingxia Fan, and Guang Yang. A two-step automated quality assessment for liver mr images based on convolutional neural network. *European Journal of Radiology*, 124:108822, 2020.

[95] Z. Wang and A. C. Bovik. *Modern Image Quality Assessment*, volume 2 of *Synthesis Lectures on Image, Video, and Multimedia Processing*. Morgan & Claypool Publishers, 2006.

[96] B. Widrow. Adaptive switching circuits. *IRE Wescon Convention Record*, 4:547–587, 1965.

[97] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1492–1500, 2017.

[98] J. Xu, C. Lin, W. Zhou, and Z. Chen. *Subjective Quality Assessment of Stereoscopic Omnidirectional Images*, pages 589–599. Springer International Publishing, Cham, Switzerland, 2018.

[99] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann. Blind image quality assessment based on high order statistics aggregation. *IEEE Transactions on Image Processing*, 25:4444–4457, 2016.

[100] Junshen Xu, Sayeri Lala, Borjan Gagoski, Esra Abaci Turk, P Ellen Grant, Polina Golland, and Elfar Adalsteinsson. Semi-supervised learning for fetal brain mri quality assessment with roi consistency. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 386–395. Springer, 2020.

[101] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng. Blind image quality assessment using joint statistics of gradient magnitude and laplacian features. *IEEE Transactions on Image Processing*, 23:4850–4862, 2014.

[102] S. Yu, G. Dai, Z. Wang, L. Li, X. Wei, and Y. Xie. A consistency evaluation of signal-to-noise ratio in the quality assessment of human brain magnetic resonance images. *BMC Medical Imaging*, 18:17, 2018.

[103] M. Zaitsev, J. Maclaren, and M. Herbst. Motion artifacts in mri: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, 42:887–901, 2015.

[104] L. Zhang, L. Zhang, and A. C. Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24:2579–2591, 2015.

[105] Z. Zhang, G. Dai, X. Liang, S. Yu, L. Li, and Y. Xie. Can signal-to-noise ratio perform as a baseline indicator for medical image quality assessment. *IEEE Access*, 6:11534–11543, 2018.

[106] X. Zhu and P. Milanfar. Automatic parameter selection for denoising algorithms using a no-reference measure of image content. *IEEE Transactions on Image Processing*, 19:3116–3132, 2010.

Dorobek naukowy autora

W niniejszej rozprawie omówiono wybrane z publikacji autora. Ten rozdział zawiera informacje o całym aktualnym dorobku naukowym autora dysertacji. Dodatkowo autor ten pełnił rolę recenzenta materiałów w czasopismach: *Engineering applications of artificial intelligence*, *Machine Vision and Applications*, *Measurement*, *Signal, Image and Video Processing*, *SoftwareX*.

Artykuły naukowe

[A-1] **Stępień, I.**, Oszust, M. *A Brief Survey on No-Reference Image Quality Assessment Methods for Magnetic Resonance Images.* J. Imaging 2022, 8(6), 160. https://doi.org/10.3390/jimaging8060160, IF_{2022} = 3,2; CS: 5,9; wkład 50%; liczba punktów: 20.

[A-2] **Stępień, I.**, Obuchowicz, R., Piórkowski, A., Oszust, M. *Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment.* Sensors 2021, 21(4), 1043. https://doi.org/10.3390/s21041043, IF_{2021} = 3,847; CS: 7,3; wkład 25%; liczba punktów: 100.

[A-3] **Stępień, I.**, Oszust, M. *No-Reference Image Quality Assessment of Magnetic Resonance Images with Multi-Level and Multi-Model Representations Based on Fusion of Deep Architectures.* Engineering Applications of Artificial Intelligence 2023, 123, 106283. https://doi.org/10.1016/j.engappai.2023.106283, IF_{2023} = 7,5; CS: 9,6; wkład 50%; liczba punktów: 140.

[A-4] **Stępień, I.**, *TIQA-MRI: Toolbox for Perceptual Image Quality Assessment of Magnetic Resonance Images.* SoftwareX 2025, 29, 102073, https://doi.org/10.1016/j.softx.2025.102073, IF_{2023} = 2,4; CS: 4,1; wkład 100%; liczba punktów MNiSW: 200.

[A-5] **Stępień, I.**, Oszust, M. *No-Reference Quality Assessment of Pan-Sharpening Images with Multi-Level Deep Image Representations.* Remote Sens. 2022, 14(5), 1119. https://doi.org/10.3390/rs14051119, IF_{2022} = 5; CS: 8,3; wkład 50%; liczba punktów: 100.

[A-6] **Stępień, I.**, Oszust, M. *Three-Branch Neural Network for No-Reference Quality Assessment of Pan-Sharpened Images.* Engineering Applications of Artificial Intelligence 2025, 139(B), 109594. https://doi.org/10.1016/j.engappai.2024.109594, IF_{2023} = 7,5; CS: 9,6; wkład 50%; liczba punktów: 140.

[A-7] **Stępień, I.**, Oszust, M. *TIQA-PSI: Toolbox for Perceptual Image Quality Assessment of Pan-Sharpened Images.* SoftwareX 2023, 23, 101494, https://doi.org/10.016/j.softx.2023.101494, IF_{2023} = 2,4; CS: 4,1; wkład 50%; liczba punktów: 200.

[A-8] Oszust, M., Bielecka, M., Bielecki, A., **Stępień, I.**, Obuchowicz, R., Piórkowski, A., *Blind image quality assessment of magnetic resonance images with statistics of local intensity extrema.* Information Sciences 2022, 606, 112-125. https://doi.org/10.1016/j.ins.2022.05.061, IF_{2022} = 8,1; CS: 13,4; wkład 16,67%; liczba punktów: 200.

Wystąpienia konferencyjne

[K-1] Kalandyk, D., **Stępień, I.**, *Reinforcement Learning with Hand Crafted and Deep Learning features for Video Quality Assessment Approach*, Selected Issues in Power Engineering, Electrical Engineering and Industry 4.0, grudzień 2023, Rzeszów, praca w komitecie organizacyjnym.

[K-2] Kalandyk, D., **Stępień, I.**, *Reinforcement Learning with Hand Crafted Deep Learning features and Transfer Learning for Image Quality Assessment Approach*, Selected Issues in Power Engineering, Electrical Engineering and Industry 4.0, grudzień 2023, Rzeszów, praca w komitecie organizacyjnym.

Artykuły naukowe wchodzące w skład cyklu (opublikowane w latach 2021-2025)

W niniejszym rozdziale zamieszczono pełną treść opublikowanych prac wchodzących w skład cyklu publikacji:

[A-1] **Stępień, I.**, Oszust, M. *A Brief Survey on No-Reference Image Quality Assessment Methods for Magnetic Resonance Images.* J. Imaging 2022, 8(6), 160. https://doi.org/10.3390/jimaging8060160, IF_{2022} = 3,2; CS: 5,9; wkład 50%; liczba punktów: 20.

[A-2] **Stępień, I.**, Obuchowicz, R., Piórkowski, A., Oszust, M. *Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment.* Sensors 2021, 21(4), 1043. https://doi.org/10.3390/s21041043, IF_{2021} = 3,847; CS: 7,3; wkład 25%; liczba punktów: 100.

[A-3] **Stępień, I.**, Oszust, M. *No-Reference Image Quality Assessment of Magnetic Resonance Images with Multi-Level and Multi-Model Representations Based on Fusion of Deep Architectures.* Engineering Applications of Artificial Intelligence 2023, 123, 106283. https://doi.org/10.1016/j.engappai.2023.106283, IF_{2023} = 7,5; CS: 9,6; wkład 50%; liczba punktów: 140.

[A-4] **Stępień, I.**, *TIQA-MRI: Toolbox for Perceptual Image Quality Assessment of Magnetic Resonance Images.* SoftwareX 2025, 29, 102073, https://doi.org/10.1016/j.softx.2025.102073 IF_{2023} = 2,4; CS: 4.1; wkład 100%; liczba punktów: 200.

[A-5] **Stępień, I.**, Oszust, M. *No-Reference Quality Assessment of Pan-Sharpening Images with Multi-Level Deep Image Representations.* Remote Sens. 2022, 14(5), 1119. https://doi.org/10.3390/rs14051119, IF_{2022} = 5; CS: 8.3; wkład 50%; liczba punktów: 100.

[A-6] **Stępień, I.**, Oszust, M. *Three-Branch Neural Network for No-Reference Quality Assessment of Pan-Sharpened Images.* Engineering Applications of Artificial Intelligence 2025, 139(B), 109594. https://doi.org/10.1016/j.engappai.2024.109594, IF_{2023} = 7,5; CS: 9,6; wkład 50%; liczba punktów: 140.

[A-7] **Stępień, I.**, Oszust, M. *TIQA-PSI: Toolbox for Perceptual Image Quality Assessment of Pan-Sharpened Images.* SoftwareX 2023, 23, 101494, https://doi.org/10.016/j.softx.2023.101494, IF_{2023} = 2,4; CS: 4.1; wkład 50%; liczba punktów: 200.

Journal of *Imaging*

MDPI

Article

A Brief Survey on No-Reference Image Quality Assessment Methods for Magnetic Resonance Images

Igor Stępień [1] and Mariusz Oszust [2,*]

[1] Doctoral School of Engineering and Technical Sciences, Rzeszow University of Technology, al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland; igorkrzysztofstepien@gmail.com
[2] Department of Computer and Control Engineering, Rzeszow University of Technology, Wincentego Pola 2, 35-959 Rzeszow, Poland
* Correspondence: marosz@kia.prz.edu.pl

Abstract: No-reference image quality assessment (NR-IQA) methods automatically and objectively predict the perceptual quality of images without access to a reference image. Therefore, due to the lack of pristine images in most medical image acquisition systems, they play a major role in supporting the examination of resulting images and may affect subsequent treatment. Their usage is particularly important in magnetic resonance imaging (MRI) characterized by long acquisition times and a variety of factors that influence the quality of images. In this work, a survey covering recently introduced NR-IQA methods for the assessment of MR images is presented. First, typical distortions are reviewed and then popular NR methods are characterized, taking into account the way in which they describe MR images and create quality models for prediction. The survey also includes protocols used to evaluate the methods and popular benchmark databases. Finally, emerging challenges are outlined along with an indication of the trends towards creating accurate image prediction models.

Keywords: survey; image quality assessment; no-reference image quality assessment; magnetic resonance images

Citation: Stępień, I.; Oszust, M. A Brief Survey on No-Reference Image Quality Assessment Methods for Magnetic Resonance Images. *J. Imaging* **2022**, *8*, 160. https://doi.org/10.3390/jimaging8060160

Academic Editor: Gabriel Cristobal

Received: 28 April 2022
Accepted: 1 June 2022
Published: 4 June 2022

1. Introduction

The advantages of magnetic resonance imaging (MRI) over other medical imaging methods, including computed tomography (CT), X-ray and ultrasound imaging, and positron emission tomography (PET) are based on its safety and ability to provide detailed images in axial, sagittal, and coronal planes [1]. Delving into the subject of improving the MRI quality, attention should be paid to the characteristics of the researched material. Protons that are normally and randomly oriented within the water nuclei of the examined tissue are arranged by a powerful, uniform, and external magnetic field. The most commonly used MRI sequences are T1-weighted and T2-weighted scans. T1 (longitudinal relaxation time) is a fixed time that determines the velocity at which excited protons return to equilibrium, while T2 (transverse relaxation time) is the constant that determines the rate at which excited protons leave the phase with each other or reach equilibrium. T1-weighted images are produced with the use of a short time to echo (TE) and repetition time (TR). Its properties determine the contrast and brightness of the image. On the other hand, the T2-weighted images are produced by using longer TE and TR. Another broadly used sequence is the fluid-attenuated inversion recovery (FLAIR). This sequence is similar to T2-weighted images aside from the fact that the TE and TR times are much longer. It is exceptionally sensitive to pathology and facilitates the differentiation between cerebrospinal fluid and an abnormality [2,3]. MR imaging is prone to distortions due to many reasons such as signal or image operations, equipment characteristics, or operator mistakes [4]. Examples of operations that cause distortions are acquisition, enhancement, compression, or reconstruction procedures. Furthermore, unexpected patient movement or an error made by the operator performing the test can introduce additional and unwanted

visual information to acquired images. These factors have an impact on the subjective assessment of the obtained image and the subsequent diagnosis. The scanners are used to provide a sequence of images valuable for further diagnostic purposes. They are created with varying magnetic field strength and measured in Tesla units (T). Nowadays, the most commonly utilized magnet strengths in MRI machines are 1.5T and 3T. The 1.5T scanner can be considered as commonly used in clinical settings, whereas 3T is then employed in research settings. Comparing two devices, some implants that are safe to go into a 1.5T scanner may not be safe for the patient in a 3T scanner. On the other hand, 3T scanner provides better image quality and faster scan speeds [5]. Therefore, automatic and objective image quality assessment (IQA) is particularly important. With the growing development of IQA research, several new algorithms have been created. This paper aimed to briefly elaborate on the MR-IQA methods suitable for the evaluation of MR images. To the best of the authors' knowledge, such methods have not been covered in a single survey since related surveys describe the general methods for IQA or methods designed for medical images with the focus on fetal brain MRI [6,7]. Furthermore, since their creation, new MR IQA methods have emerged.

The assessment performed by the method should be as close as possible to the subjective assessment, and with the development of reliable techniques, the human assessment in MRI may be minimized or, in the longer term, completely replaced. The objective approaches for IQA can be categorized into full reference (FR), reduced reference (RR), and no reference (NR)/blind image quality (BIQA) methods [8–11]. In the case of FR and RR, an undistorted reference image, or a part of an image to which the assessed image is compared, should be available. However, in the case of medical imaging, such an undistorted image does not exist [12]. Therefore, the development of BIQA methods is of particular importance in this field [13–16]. NR methods can be further classified into opinion-aware and opinion-unaware/completely blind approaches [17] depending on the access to subjective scores while creating a quality model.

The rest of the paper is organized as follows. In Section 2, selected MRI distortions are described. Section 3 presents NR-IQA features and briefly characterizes MRI methods. Section 4 introduces databases and MRI NR-IQA protocols. Finally, Section 5 summarizes the achievements in NR-IQA of MR images and indicates the main challenges and future directions.

2. MRI Distortions

here are several categories of distortions, sequence artifacts, or distortions related to image reconstruction that can be minimized by using an appropriate work protocol, system improvements, or enhancing image quality improvement and processing methods. Moreover, there are system artifacts due to the misused or faulty MRI equipment. However, sample artifacts or human-induced distortions are more complex as their elimination requires not only a thorough understanding of the patient's anatomy and psychology but also the use of specific pulse sequences [18]. The most common distortions are briefly described below and their examples are shown in Figure 1.

2.1. Spike (Herringbone) Artifact

The noise spike points in the k-space are generated by gradients that have been used in a very high duty cycle. The k-space is an extension of the Fourier concept, defined as an array of the numbers representing spatial frequencies in two or three dimensions of an object [19,20]. The distortions in the k-space are manifested by single or multiple points varying in intensity compared to their surroundings. The combination of information during the Fourier transform with the spike causes dark bands to be superimposed on the image. The direction, angle, and distance between the bands depend on the displacement of the noise spike from the center of the k-space. Spike noise usually occurs from loose electrical connections and is more noticeable when using high-duty-cycle sequences. In this

case, following the changes in the position of the pattern produced by the spike can be used in further diagnosis [21,22].

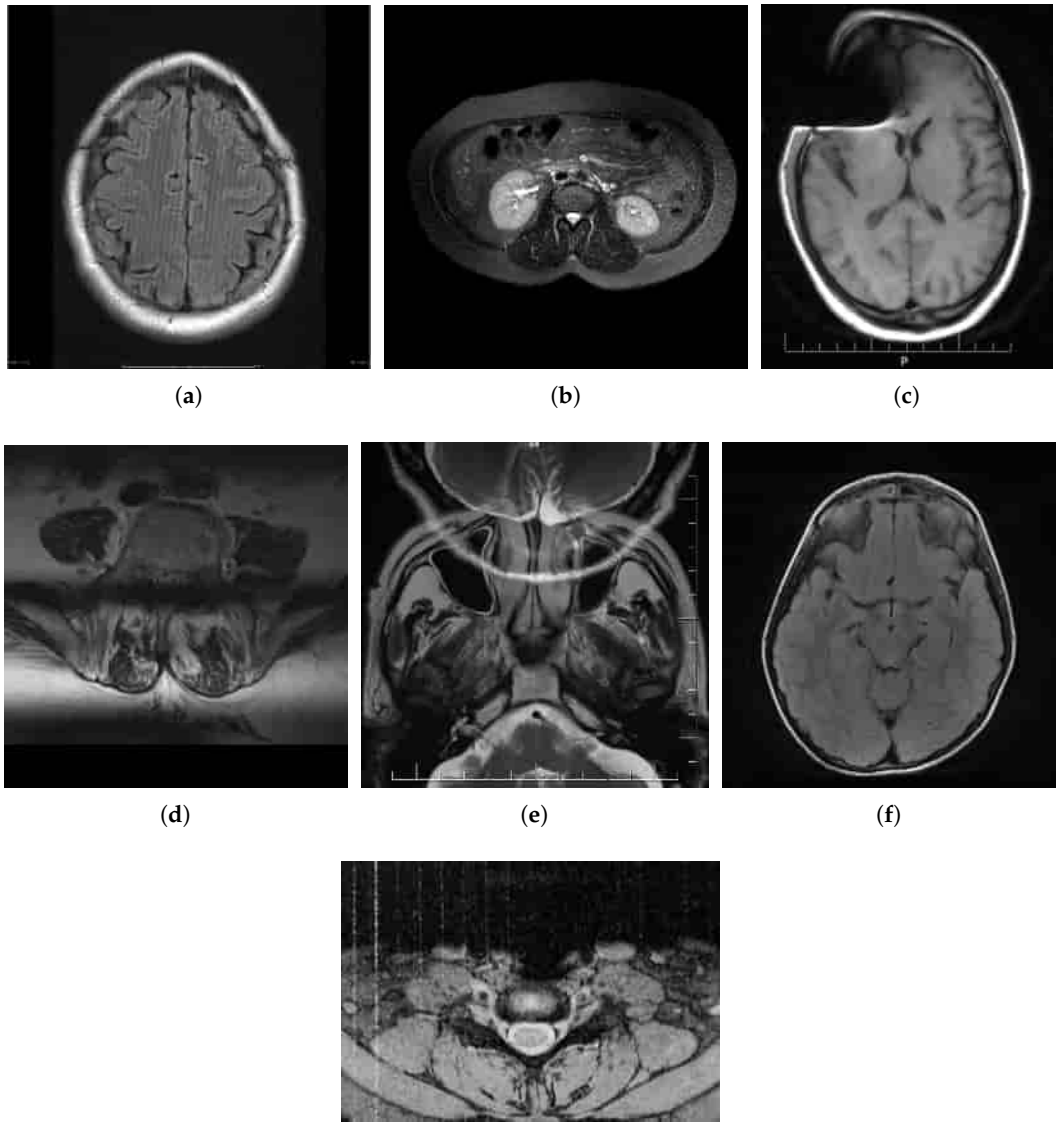


Figure 1. Examples of artifacts in MR images: (**a**) herringbone; (**b**) ghosting; (**c**) magnetic susceptibility; (**d**) slice overlap; (**e**) aliasing; (**f**) Gibbs effect; and (**g**) zipper. Cases were the courtesy of: (**a**) Assoc. Prof. Frank Gaillard, rID: 19695; (**b**) Assis. Prof. Faeze Salahshour, rID: 81727; (**c**) Dr. Ayush Goel, rID: 22731; (**d**) Dr. Roberto Schubert, rID: 16705; (**e**) Dr. Prashant Mudgal, rID: 26927; (**f**) Dr. Prashant Mudgal, rID: 27302; and (**g**) Dr. Alan Nazerian, rID: 45665; radiopaedia.org (accessed on 27 April 2022).

2.2. Zipper Artifact

A zipper artifact is an area of alternating light and dark pixels. This distortion is placed in the direction of the frequency coding and appears throughout the series of images. It is caused by the leakage of electromagnetic energy into the magnet room or by electronic equipment brought into the room and a breach of RF shielding in this equipment. Another source of the zipper artifact is an external light source reaching the receiver coil as a result of an open door in the room where the scanner is located [23,24].

2.3. Ghosting

Images are influenced not only by the physical reaction of the patient during the examination or environmental factors but also by the pulsating movement. For example, the ghosting artifact can be caused by cardiac beats, arterial or cerebrospinal fluid (CSF) pulsations, respiration, and implants. When the motion is strong enough, the distortions, resembling the shape of the imaged organ, overlap with the image itself. The arrangement may resemble parallel imaging artifacts whilst resembling clots or dissections. Parallel imaging is a method of improving MRI data acquisition that works by obtaining a reduced amount of k-space data with a series of receiver coils [25,26]. The distance between the distortions may depend on the frequency of movements and the repetition time. This artifact is also categorized as motion distortion [27,28].

2.4. Blurring

Blur is a motion-type distortion that occurs as unevenly distributed over the entire area of an image. It has a large impact on the strength of edge visibility, weakens the clarity of the image, and reduces the contrast between different anatomical structures [29,30].

2.5. Aliasing Artifacts

Aliasing artifacts appear when the field of view is smaller than the body part being imaged. The field of view is an area designated by the size of two- or three-dimensional spatial encoding space of the MR image [31]. Aliasing occurs in the direction of phase coding or cross-section in 3D acquisition. 'The moire' or 'fringe artifact' is a type of distortion that arises in the frontal lobe, where large fields of view produce these distortions. One way of reducing or aligning the distortion is to change the imaging axis so that the part of the body under study fits within the field of view. However, this can result in other types of artifacts. An alternative approach is to increase the field of view phase or apply spatial saturation pulses outside the field [32,33].

2.6. Gibbs Effect

Gibbs artifacts—also known as truncation artifacts or ringing artifacts—are a series of lines in the MR image that appear parallel to the area in which there has been a sudden and intense change in signal intensity. The Gibbs ring distortion is also mainly produced due to insufficient samples in the direction of phase coding or reading [34].

2.7. Slice-Overlap Artifact

The slice-overlap artifact is associated with the loss of the signal visible in the image from the multi-angle acquisition. With this distortion, the edge sections have a reduced signal intensity and do not create a section profile with a straight edge. To reduce the effect of the artifact, the angle of intersection between the ply groups should be decreased. If there are difficulties in the performed examination, e.g., damage occurs between the layers, the method of interleaving is used [35].

2.8. Gradient-Related Distortion

This distortion is characterized by image compression and the inadequate rebound of spins on the edges of images. It is created when the electric current receives the voltage from the gradient coil while tapering to the sides of the magnet [36,37].

2.9. Parallel Imaging Artifact

In general, parallel imaging is a widely used method for accelerating the acquisition of MRI data in which the distribution and sensitivity of the receiver coils play a major role. This has its positive results in reducing the imaging time but may cause distortions. As a consequence of the parallel imaging, in which each coil is at a different distance from the pixel, the signal recorded by such a coil changes and the closer coils have stronger signals. Professionals analyzing the MR images obtained from such an operation can

misdiagnose the approximate source of the MR signal. In parallel imaging, the signals from the individual coils are amplified and simultaneously processed along separate channels, keeping the identity to the end. Furthermore, it may suffer from inhomogeneous noise in reconstruction and unevenly distributed noise caused by the overlapping of different structures whilst lacking the core information [26,38].

2.10. Susceptibility Effect

By placement in a magnetic field, the tissues become temporarily magnetized. However, magnetization is not uniform and depends on the magnetic susceptibility of the tissue, as in the case of the air–tissue or bone–soft tissue interfaces. Bone and air are less magnetically susceptible which means that a low-intensity signal is generated at these sites. Such local variations in intensity give rise to geometric distortions in the images. Another problem, also causing the non-uniformity in the signal, is a magnetized implanted device. Metal implants have a much higher magnetic susceptibility than the rest of the tissues of the human body, leading to signal distortions related to higher signal intensity [2,39].

3. NR-IQA Approaches

NR-IQA approaches are used in cases in which a reference image is not available. Hence, such approaches support examination based on MR images by rejecting scans of unsuitable quality. There are NR-IQA methods for medical images taking into account distortions, e.g., noise, compression, or blur. For example, based upon the human visual system (HVS) [40,41], Bhateja et al. [42] used two-stage MRI fusion metrics for IQA, where two images are fused to improve the detection of distortion. With the objective of developing automatic deep learning methods, Xu et al. [7] introduced a semi-supervised technique devoted to the IQA of fetal brain MR images with the use of a mean teacher method and a region-of-interest (ROI) consistency. Furthermore, Liu et al. [43] used semi-supervised learning to solve the problem of creating noisy annotation in the image segmentation task. This three-staged quality assessment technique employs a hierarchical residual model, and as the consequence, provides an assessment of the slice, volume, and subject level. Another classification method uses unpaired generative adversarial network (GAN) and weakly supervised trained classifier to assess MR images [44]. To address the problem of wasting potentially important 3D spatial information, the HyS-net approach was created. It was based on a hyper-network and is capable of self-adaptation [45]. A more recent approach adapting the modified blind/referenceless image spatial quality evaluator (BRISQUE) was proposed by Chow and Rajagopal [46] in which BRISQUE [47] aims to quantify possible losses of 'naturalness' in the image using the scene statistics of mean subtracted contrast normalized (MSCN) coefficients.

Table 1 provides a brief comparison of NR-IQA approaches devoted to MR images. Several methods focus on specific artifacts, e.g., QEMDIM [48], or an approach introduced by Nabavi et al. [49]. However, most methods are designed to assess overall image quality considering the characteristics of MR images: modified-BRISQUE [46], R50GR18 [50], ENMIQA [51] or NOMRIQA [52]. The number of features depends on the way that the images are described from simple entropy to features extracted from the layers of a neural network. Most approaches were validated on one or two databases, among which large datasets that are assessed by more than several specialists are rare. A more detailed presentation of the methods is provided below. An experimental evaluation of recent approaches and related discussion on their performance can be found in [50] or [45].

3.1. A Two-Step Automated Quality Assessment for Liver MR Images Based on Convolutional Neural Network

In the method proposed by Wang et al. [53], a two-step method applied for image classification purposes was used. The method focuses on the regions of interest (ROI), i.e., each patch assessed by the radiologists as diagnostic or non-diagnostic is used to train a convolutional neural network (CNN) for segmentation purposes. Then, another CNN is

used to classify the quality of extracted patches. Overall image quality is assessed on the basis of the number of non-diagnostic patches in all the liver patches of the image.

Table 1. Comparison of NR-IQA methods in terms of employed techniques, features, and used datasets.

Method	Approach and Features	Number of Features	Datasets
A two-step automated quality assessment for liver MR images based on convolutional neural network [53]	• Patch-based strategy • CNN in image region segmentation • ROI	-	Not defined in the paper
Semi-supervised learning for fetal brain MRI quality assessment with ROI consistency [7]	• Semi-supervised learning • ROI consistency • Mean teacher model	-	Scans acquired at Boston Children's Hospital
No-reference image quality assessment of T2-weighted magnetic resonance images in prostate cancer patients [54]	• Generative adversarial network • Weakly supervised • Trained deep classifier	-	• National Cancer • Institute (NCI) • PIE-AAPM-NCI Prostate • MR Gleason • Grade Group Challenge • NIH Clinical Center
Two-stage multi-modal MR images fusion method based on parametric logarithmic image processing (PLIP) model [55]	• Two-stage MRI fusion • PCA and PLIP operators • Stationary wavelet transform	-	• Whole Brain Atlas [56]
Hierarchical non-local residual networks for image quality assessment of pediatric diffusion MRI with limited and noisy annotations [43]	• Slice-wise, volume-wise, and subject-wise IQA • Non-local residual networks • Semi-supervised learning	-	• Database from the Center for Magnetic Resonance Research (CMRR) at the University of Minnesota
HyS-net [45]	• Content-adaptive hypernetwork • A spatial feature extraction • Network-based quality predictor	-	• Open dataset, MRIQC [57]
QEMDIM [48]	• Difference of statistical features between test images • MSCN coefficients • Multi-directional filtered coefficients (MDFC)	20	• ADNI [58] • ABIDE [59]
AQASB [60]	• Background-connected distortions • Decent level of background voxels	-	• ADNI [58]

Table 1. *Cont.*

Method	Approach and Features	Number of Features	Datasets
Multi-class cardiovascular magnetic resonance image quality assessment using unsupervised domain adaptation [49,61]	• Unsupervised domain adaptation • Spatial and frequency domains • K-space manipulation	512	• UK Biobank • Cardiac MRI dataset, York University [62] • K-space manipulation
MRIQC [63]	• Quality measures • Binary classifier	64	• ABIDE [59] • OpenfMRI [64]
Brain and cardiac MRI images in multi-center clinical trials [65]	• The moments-preserving property application • Measures the differences in texture contrast	The number of features depends on the image	• NeuroRx research Inc. • BrainCare Oy • ADNI [58] • Department of Diagnostic Imaging of the Hospital for Sick Children in Toronto
Modified-BRISQUE [46]	• Luminosity, image characteristics • NSS	36	• Sirix DICOM Viewer MRI database • MR images from the Academy Unit of Radiology, University of Sheffield
R50GR18 [50]	• Fusion of deep network architectures • SVR	3584	• DB1 [52] • DB2 benchmarks [51]
ENMIQA [51]	• Thresholded local intensity differences obtained by using the non-maximum suppression (NMS) operation • Entropy of a sequence of extrema numbers	1	• DB1 [52]
NOMRIQA [52]	• FAST features • Histograms of binary descriptors	3840	• Simulated Brain Database (SBD) [66] • DB1 [52]

3.2. Semi-Supervised Learning for Fetal Brain MRI Quality Assessment with ROI Consistency

Semi-supervised learning for fetal brain MRI quality assessment with ROI consistency is a semi-supervised deep learning method that responds to the difficulties arising in fetal MRI [7]. The method is based on the average teacher model to control the consistency between the student–teacher approach, thanks to the aggregation of network parameters at different stages of training. It uses MR images classified according to the following criteria: diagnostic, non-diagnostic, and images without a brain region of interest. In this method, improved accuracy in detecting the non-diagnostic images of the brain's ROI during feature extraction is obtained. Importantly, the method is introduced and implemented on the MR scanner, which makes it possible to check the condition of the obtained image and if necessary, repeat the examination.

3.3. No-Reference Image Quality Assessment of T2-Weighted Magnetic Resonance Images in Prostate Cancer Patients

No-reference image quality assessment method of the T2-weighted magnetic resonance images in prostate cancer patients introduced by Masoudi et al. [54] is a classification method to determine non-diagnostic images with an artifact, diagnostic images with substantial noise or motion, and diagnostic images with trivial noise or motion. The model assumes the extension of NR-IQA scans to FR-IQA, after an improvement of image quality

by using CycleGAN. In the method, original images are compared with synthetic reference images. The poor quality images receives further correction from CycleGAN [54].

3.4. Two-Stage Multi-Modal MR Images Fusion Method Based on Parametric Logarithmic Image Processing Model

In the two-stage multi-modal MR images fusion method based on parametric logarithmic image processing model, image fusion is used to obtain a more accurate, final image [42]. It is a two-step process and uses decomposition based on stationary folk transformation (SWT) along with principal component analysis (PCA). The first and second fusion coefficients are combined using the HVS-based parameterized logarithmic image processing (PLIP) operator. To increase the accuracy, the obtained results are compared using different measures.

3.5. Hierarchical Non-Local Residual Networks for Image Quality Assessment of Pediatric Diffusion MRI with Limited and Noisy Annotations

The hierarchical non-local residual networks for image quality assessment of pediatric diffusion MRI with limited and noisy annotations is a deep-learning method based on local residual networks [43]. It consists of three stages, namely those involving the use of a slice-wise QA network (i.e., SQA-Net); the extracted slice features with the volume-wise QA network; and a compilation of the IQA results using the decision rule. Through these actions, this method makes the evaluation results available at different levels: namely those of slice, volume, and subject. SQA-Net is also constructed by implementing depthwise separable convolutions (DSConv) and non-local mean operation. To increase the effectiveness of the approach when working with a small amount of labeled data, semi-supervised learning and the subsequent slice with volume self-training are used.

3.6. HyS-Net

The spatial-related hyper-network-based MRI BIQA works on the MRIQC open dataset are based on the development of a hyper network that adapts to the content [45]. The structure of the 3D network was designed to explore spatial information of 3D images and improve the BIQA performance. In addition to relying on a hyper set that generates dynamic parameters, the method includes the extraction of spatial features and a combined network quality predictor.

3.7. QEMDIM

The quality evaluation using multi-directional filters for MRI (QEMDIM) is a method that enables the detection of distortions with different characteristics, e.g., Gaussian noise, motion artifacts, streaks, or aliasing [48]. It can be used not only in assessing the quality of medical images but also in assessing the performance of MR hardware and software. It is based on the feature difference between test and pre-scanned images. The performance of the method varies according to the slice position.

3.8. AQASB

The automatic quality assessment in structural brain magnetic resonance imaging (AQASB) is a method that focuses on the background area of the MRI image (i.e., the air) [60]. It specializes in analyzing images with distortions such as ghosting, motion, flow and wrap-around. The method consists of three steps: (1) background air region delineation; (2) computation of a model-free quality index; and (3) the computation of an additional quality index. It is developed to inform the operator about the poor quality of the measurement and to notify the need to perform an additional scan. The method has its limitations as it assumes that each scanned image has a sufficiently large percentage of background voxels to successfully perform the measurement. The method takes a limited number of artifacts into consideration. It aims to overcome the challenge of not having access to data labels and a reduction in computational time.

3.9. Multi-Class Cardiovascular Magnetic Resonance Image Quality Assessment Using Unsupervised Domain Adaptation

The multi-class cardiovascular magnetic resonance image quality assessment using unsupervised domain adaptation is a deep learning model for automatic cardiovascular magnetic resonance (CMR) IQA [49,61]. In the process of image quality assessment method evaluation, the distortions in the spatial domain of 2D sliced CMR images are identified. The domain adaptation is based on the trained model that is used to test the new dataset. Before the image reconstruction, the frequency domain was described to use the method on the data of k-space.

3.10. MRIQC

The MRI quality control tool (MRIQC) extracts quality measures and uses them as input to a binary classifier [63]. The classification is performed on the basis of binary labels from a set of MRI images. The model includes a selection of hyper-parameters in non-nested cross-validation, the training process on ABIDE dataset, evaluation on the held-out dataset, the normalization of features, and the elimination of features based on the site prediction. The release of an MRI quality control tool, MRIQC, leads to the extraction of a vector of 64 image quality measures (IQMs). The IQMs can be grouped into four categories—(1) measures based on noise measurements: the coefficient of joint variation of GM and WM (CJV), the contrast-to-noise ratio (CNR), the signal-to-noise ratio calculation (SNR) and the second quality index (QI2); (2) measures based on information theory: the entropy-focus criterion (EFC), the foreground-background energy ratio (FBER); (3) measures targeting specific artifacts: the bias field extracted estimated by the INU correction, the first quality index (QI1), the white-matter to maximum intensity ratio is the median intensity within the WM mask (WM2MAX); (4) other measures: the full-width half-maximum (FWHM), estimation of the ICVs, the residual partial volume effect feature (rPVE), several summary statistics such as the mean or standard deviation (SSTATs), and overlap of tissue probability maps (TPMs).

3.11. Brain and Cardiac MRI Images in Multi-Center Clinical Trials

The method assumes that MRI slices possess statistical properties describing different levels of contrast degradation [65]. Thus, to each level of contrast-distorted MRI slice, a set of pixel configurations is assigned. The IQA process is divided into four steps. Firstly, local contrast features are extracted from the test image, then the mean and standard deviation are computed. To obtain two separate z-scores, the mean and standard deviation are processed. As a result, the prediction of the contrast quality score, and the texture contrast quality score is performed. Focused on the labeling problem and the central limit theorem, the approach aims to describe each possible level of contrast degradation in an MRI slice. It perceives images with artifacts as darker than denoised ones, which also have lower contrast. The method predicts higher texture contrast quality score.

3.12. Modified Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE)

Modified blind/referenceless image spatial quality evaluator (BRISQUE) [47] is a model initially created with natural images in mind, but it was adapted to evaluate the quality of MR images [46]. It uses locally normalized luminosity coefficients for calculations, the MR image function and the difference mean opinion score (DMOS) for training. The goal of its development was to create a method that would be useful when working on images with all types of distortions.

3.13. R50GR18

This method presents a different approach to image quality assessment as it is based on the fusion of neural networks (ResNet50, GoogLeNet, and ResNet18) which then take part in the transfer-learning process [50]. The large diversity of the architectures selected for the study and fusion allows the assessment of a large spectrum of distortions. The method

uses support vector regression (SVR) [67] on features extracted from connected networks to improve the ability to evaluate the quality of images. Additionally, a network modification is applied, in which the last three layers of each network used are replaced with a fully connected layer and the regression layer to perform the regression task. Moreover, the size of the input image is not constant but is adapted to the size of the input network.

3.14. Entropy-Based Magnetic Resonance Image Quality Assessment Measure (ENMIQA)

In the method, the entropy of the extrema of local intensity differences representing filtered versions of an input image is used for quality prediction [51]. Specifically, the quality prediction is expressed by the entropy of a sequence of extrema numbers obtained with the thresholded non-maximum suppression (NMS) applied to filtered MR images.

3.15. No-Reference Image Quality Assessment of Magnetic Resonance Images with High-Boost Filtering and Local Features (NOMRIQA)

The method introduced by Oszust et al. [52], NOMRIQA, uses high-boost filtering to amplify high-frequency points allowing for the effective detection of distortions. Detected interest points in filtered images are described using the fast retina key-point (FREAK) descriptor and then represented by a histogram of such descriptors. The method builds a quality model with the SVR approach.

3.15.1. PSNR/SNR

The peak signal-to-noise ratio (PSNR) in numerous studies is used as an NR-IQA tool. For this reason, in this paper, PSNR and signal-to-noise ratio (SNR) [68–71] is also described. PSNR depends on the value of RMSE among the target image and the reference image. It is calculated as:

$$\mathrm{PSNR} = 10log\frac{m^2max}{RMSE^2}, \tag{1}$$

where m_{max} means the maximum pixel score. There is a poor correlation of PSNR with subjective quality assessment performed by human observers. However, this dissimilarity can be captured in each type of distortion by the perceptual complexity of the target image. There have been attempts to improve the PSNR performance by the use of a linear score mapping process with the use of factors such as image-free energy and type of distortion.

3.15.2. Maximum Difference

The maximum difference (MD) presents the maximum of the error signal [72]. Additionally, the quality of an image decreases with the growth of the MD value.

$$\mathrm{MD} = Max(|C_{ef} - D_{ef}|), \tag{2}$$

where $e = 1, 2, \ldots, n, f = 1, 2 \ldots, m$.

3.15.3. Normalized Cross-Correlation

The normalized cross-correlation (NCC or NK) measures the similarity of the image sets and detects patterns or the object of an image [72,73]. The metric is used in image registration.

$$\mathrm{NK} = \frac{\sum_{j=1}^{n}\sum_{i=1}^{m} z(j,i)Zs(j,i)}{\sum_{j=1}^{n}\sum_{i=1}^{m}(z(j,i))^2}, \tag{3}$$

where n denotes the number of pixels in the horizontal direction; m denotes the number of pixels in vertical direction; $s(j,i)$ denotes the filtered image at j and i coordinates; and $z(j,i)$ is the noisy image at j and i coordinates.

4. Evaluation of IQA Models

The development of image quality assessment methods is stimulated by the existence of suitable databases and widely accepted protocols, ensuring that the methods are fairly and thoroughly compared.

4.1. Databases

There are many medical databases created to develop and test methods related to MR. Most of them are assessed by authors of approaches to provide subjective scores for training the methods. Only rare examples contain images along with human scores. Below, the most commonly used databases are presented. Their summary and exemplary images are shown in Table 2 and Figure 2, respectively.

Table 2. Details of the MR image datasets.

Name	Year	No. of Images	Link (Accessed on 27 April 2022)
OpenfMRI	2010	Not specified/repository of datasets	openfmri.org
ADNI-1	2004–2010	200 elderly controls, 400 MCI, 200 AD	adni.loni.usc.edu
ADNI-GO	2009–2011	Existing ADNI-1 + 200 early MCI	adni.loni.usc.edu
ADNI-2	2011–2017	Existing ADNI-1 and ADNI-GO + additional images	adni.loni.usc.edu
ADNI-3	2017–2022	Existing ADNI-1, ADNI-GO, ADNI-2 + additional images	adni.loni.usc.edu
ABIDE I	2012	1112 datasets	fcon_1000.projects.nitrc.org
ABIDE II	2016	Existing ABIDE I and 1000 datasets	fcon_1000.projects.nitrc.org
DB1	2020	70	marosz.kia.prz.edu.pl/ENMIQA.html
DB2	2020	240	marosz.kia.prz.edu.pl/NOMRIQA.html

4.1.1. OpenfMRI

The OpenfMR collection contains the MR and EEG of human brain images [64]. It accepts all forms of data that include MR imaging, is perceived as simple in the organization, and has no data access limitations. The data usually consist of four-dimensional datasets. However, depending on the number and length of scanning runs, spatial resolution, and the number of slices acquired, an fMRI study can range from fifteen patients for most studies. The project was created to provide the open sharing of neuroimaging resources.

4.1.2. ADNI

The Alzheimer's Disease Neuroimaging Initiative (ADNI) was developed to contribute to the early detection and tracking of Alzheimer's disease [58]. It was founded in 2004 and is divided into four areas of study: ADNI-1, ADNI-GO, ADNI-2, and ADNI-3. ADNI-1 includes 800 patients and uses MRI and PET imaging measures. ADNI-GO consists of 1000 patients and its main purpose for evaluation is the examination of biomarkers in the earlier stages of Alzheimer's disease. The ADNI-2 and -3 studies expanded upon previously acquired image databases with hundreds of new examples. A new cohort was added, significant memory concern (SMC), in addition to brain scans that detect tau protein tangles (tau PET). In recent years, ADNI-3 has gathered data from scientists at 59 research centers in the United States and Canada.

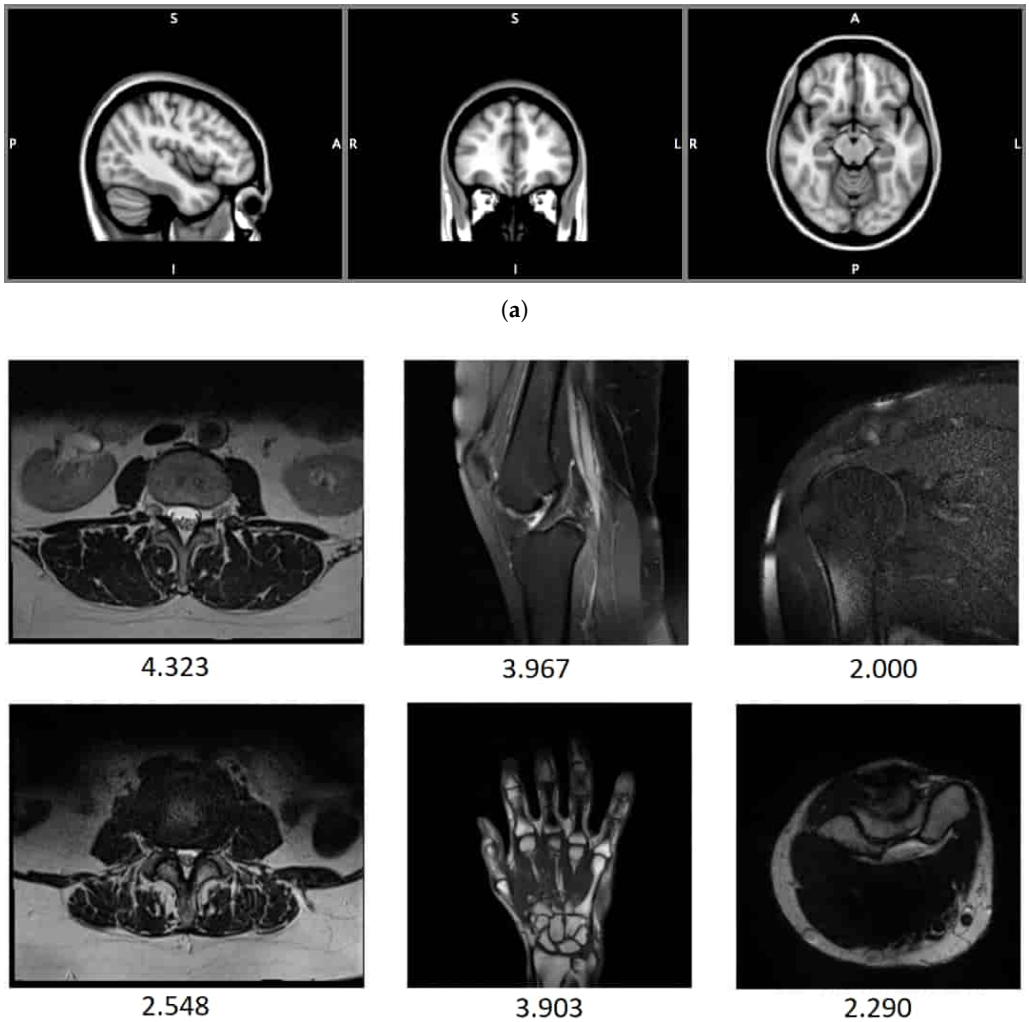


Figure 2. Exemplary MR images from OpenfMRI (**a**) and DB1 (**b**) databases. MOS values for images from DB1 database are also displayed.

4.1.3. National Resource for Quantitative Functional MRI

The National Resource for Quantitative Functional MRI project was created to design quantitative magnetic resonance acquisition and processing technology to track brain changes during neurodevelopment or neurodegeneration. This technology is developed in collaboration with a large community of specialists from several institutions in the USA. The overall scope is divided into four databases of MR images (D1–4). The D1 database concerns metabolic (S)I and operates on high-resolution MR spectroscopic imaging (MRSI) [58,74] in the research of brain and spine metabolism at the magnetic field strengths of 3 and 7 Tesla. The D2 database is related to psychological MRI that aims to discover tissue biomarkers to provide early information about physiological and metabolic changes in clinical imaging. The D3 database is devoted to functional MRI that uses information about the blood oxygenation level to assess changes in the brain that can be the cause of many diseases such as autism, ADHD, or Alzheimer's [75]. The objective of the last database, D4, is to develop image analysis technologies able to integrate different anatomical representations of the brain based on multi-modal MRI, including multi-contrast anatomical MRI, functional MRI (fMRI), and MR spectroscopy (MRS) into a common framework [76].

4.1.4. Autism Brain Imaging Data Exchange (ABIDE)

ABIDE is a collection of 16 worldwide imaging sites that are openly sharing 1112 datasets composed of structural and resting-state pre-processed MRI [59,77]. The age of the individuals under control varies from 7 to 64 years old. The pre-processing was performed using: the Connectome Computation System (CCS), the Configurable Pipeline for the Analysis of Connectomes (CPAC), the Data Processing Assistant for Resting-State fMRI (DPARSF) and the Neuroimaging Analysis Kit. Four pre-processing strategies were performed with each pipeline: all combinations of with and without filtering and with and without global signal correction as well as the statistical derivatives for each pipeline and strategy were calculated by the CPAC software. Three different pipelines were used: ANTS, CIVET and Free Surfer.

4.1.5. 1.5T T2-Weighted MR Image Databases: DB1, DB2

The DB1 [52] and DB2 [51] datasets contain 70 and 240 MR images, respectively. The DB1 benchmark consists of images selected from 1.5T MR T2-weighted sagittal sequences of several body locations, i.e., knee, shoulder, or spine. The resolution of images in the dataset is between 192 × 320 and 512 × 512. Images were taken under different conditions affecting the image quality (IPAT software to make generalized autocalibrating partially parallel acquisitions (GRAPPAs); GRAPPA3 [78]). The DB1 also contains the mean opinion score (MOS) ranging from 1 to 5 which was obtained in tests with a large group of radiologists. The greater MOS value denotes the better quality of the image. The DB2 collection contains T2-weighted MR images acquired during routine diagnostic exams of different body locations, including shoulders, knees, or cervical and lumbar spine. The dataset contains images of resolution from 192 × 320 to 512 × 512 and MOS (1–5). To acquire images of different quality in a controlled way, the parallel imaging technique was applied. Hence, a group of radiologists was also invited for the assessment of the image quality.

4.2. Evaluation Protocol

To compare the performance of regression-based IQA approaches, four performance criteria are typically used: Spearman rank-order correlation coefficient (SRCC); Kendall rank-order correlation coefficient (KRCC); and Pearson linear correlation coefficient (PLCC); and root mean square error (RMSE). The higher the SRCC, KRCC and PLCC, and the lower the RMSE, the better the output of the IQA method is [79–81]. The works on IQA methods that classify images often report the accuracy measured as the number of correctly classified images, receiver operating characteristic curves (ROC), or the area under the curve, showing the relation of the performance and a threshold. Furthermore, in the methods that classify images, the quality assessment is commonly formulated by the use of diagnostic and non-diagnostic labels from a trained observer [82–84]. The train–test split dataset protocol is used to calculate the performance of algorithms in cases they are used to make predictions on data not used to train the model. The results allow for comparing the performance of machine learning algorithms for chosen predictive modeling problems. Typically, IQA databases are randomly divided into a training set with 80% distorted images, and the remaining 20% of distorted images are used to test the model [85]. The division is then repeated and the metric of the quality measure is reported. For images with artificially created distortions, the images compatible with the same reference image are paired with the same set to provide the total separation of the training and testing content.

PLCC is used for prediction accuracy between the sets of values. It is calculated as follows:

$$\mathrm{PLCC} = \frac{\bar{\mathcal{E}^o}^T \bar{O^s}}{\sqrt{\bar{\mathcal{E}^o}^T \bar{\mathcal{E}^o} \bar{O^s}^T \bar{O^s}}}, \tag{4}$$

where $\bar{\mathcal{E}^o}$ and $\bar{O^s}$ are mean-removed vectors. RMSE is obtained as:

$$\mathrm{RMSE} = \sqrt{\frac{(\mathcal{E}^o - O^s)^T (\mathcal{E}^o - O^s)}{n}}, \tag{5}$$

where n means the total number of images. SRCC evaluates the prediction of monotonicity and is calculated as follows:

$$\mathrm{SRCC} = 1 - \frac{6\sum_{i=1}^{n} b_i^2}{n(n^2-1)}, \tag{6}$$

where b_i means the difference between the i-th image in E^o and O^s, $i = 1, 2, \ldots, n$. Finally, KRCC is obtained as:

$$\mathrm{KRCC} = \frac{n_c - n_d}{0.5n(n-1)}, \tag{7}$$

where n_c is the number of concordant pairs in the dataset, and n_d means the number of discordant pairs. Since the calculation of PLCC and RMSE involves the nonlinear mapping of objective scores e^o into subjective opinions O^s, the mapping model can be represented by:

$$e^0 = y_1\left(\frac{1}{2} - \frac{1}{1 + exp(y_2(e^0 - y_e))}\right) + y_4 e^0 + y_5, \tag{8}$$

where $y = [y_1, y_2, \ldots, y_5]$.

There are three error measurement metrics commonly used in model evaluation: RMSE [80], mean square error (MSE) [86] and mean absolute error (MAE) [87]. The metrics calculate the objective quality scores after regression, O, and the errors between the datasets, E:

$$\mathrm{MAE}(O, E) = \frac{1}{n} sum_{i=1}^{n} |O_i - E_i|, \tag{9}$$

$$\mathrm{MSE}(O, E) = \frac{1}{n} sum_{i=1}^{n} (O_i - E_i)^2, \tag{10}$$

where n denotes the size of the dataset.

5. Conclusions

In this paper, diverse approaches to the automatic NR quality assessment of MR images were presented, and detailed characteristics of recent methods, used features, and image benchmarks that stimulate the development in this field were provided. As presented, the number of works devoted to the assessment of MR images is relatively low, despite their usability in practice. First, such approaches adapt methods from the IQA of natural images or perform image processing steps that address the characteristics of MR images. More recent methods often use or introduce powerful deep learning architectures. Despite the present division of the methods into quality prediction and quality classification methods, both types aim to indicate which images should not be used for diagnostic purposes. However, the methods based on regression can be used for the development of image enhancement algorithms due to the capability of distinguishing small quality differences instead of binary classifiers that only reject unsuitable images.

Approaches are also divided into methods that assess slices and methods that can be employed for 3D quality evaluation. However, the development of both types of techniques requires access to large-scale MR image databases that also contain subjective scores obtained in tests with human participants. Nowadays, the creation of such databases can be considered the most challenging problem in MR-IQA. Its solution would lead to the emergence of more accurate quality prediction approaches.

Author Contributions: Conceptualization, I.S. and M.O.; investigation, I.S.; writing and editing, I.S. and M.O.; supervision, M.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable as this is a survey paper covering the descrption of the methods and appraoches found in the literature. We have not experimented with humans or animals.

Informed Consent Statement: Not applicable as this study did not involve humans.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Krupinski, E.A.; Jiang, Y. Anniversary paper: Evaluation of medical imaging systems. *Med. Phys.* **2008**, *35*, 645–659. [CrossRef] [PubMed]
2. Kolind, S.H.; MacKay, A.L.; Munk, P.L.; Xiang, Q.S. Quantitative evaluation of metal artifact reduction techniques. *J. Magn. Reson. Imaging: Off. J. Int. Soc. Magn. Reson. Med.* **2004**, *20*, 487–495. [CrossRef]
3. Roy, S.; Whitehead, T.D.; Quirk, J.D.; Salter, A.; Ademuyiwa, F.O.; Li, S.; An, H.; Shoghi, K.I. Optimal co-clinical radiomics: Sensitivity of radiomic features to tumour volume, image noise and resolution in co-clinical T1-weighted and T2-weighted magnetic resonance imaging. *EBioMedicine* **2020**, *59*, 102963. [CrossRef]
4. Westbrook, C.; Talbot, J. *MRI in Practice*; John Wiley and Sons: Hoboken, NJ, USA, 2018; pp. 56, 125, 141, 260–280. 331–332.
5. Soher, B.J.; Dale, B.M.; Merkle, E.M. A Review of MR Physics: 3T versus 1.5T. *Magn. Reson. Imaging Clin. N. Am.* **2007**, *15*, 277–290. [CrossRef]
6. Largent, A.; Kapse, K.; Barnett, S.D.; De Asis-Cruz, J.; Whitehead, M.; Murnick, J.; Zhao, L.; Andersen, N.; Quistorff, J.; Lopez, C.; et al. Image quality assessment of fetal brain MRI using multi-instance deep learning methods. *J. Magn. Reson. Imaging* **2021**, *54*, 818–829. [CrossRef]
7. Xu, J.; Lala, S.; Gagoski, B.; Abaci Turk, E.; Grant, P.E.; Golland, P.; Adalsteinsson, E. Semi-supervised learning for fetal brain MRI quality assessment with ROI consistency. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Virtual Conference, 4–8 October 2020; pp. 386–395.
8. Okarma, K. Current Trends and Advances in Image Quality Assessment. *Elektron. Ir Elektrotechnika* **2019**, *25*, 77–84. [CrossRef]
9. Wang, Z.; Bovik, A.C. Modern Image Quality Assessment. *Synth. Lect. Image Video, Multimed. Process.* **2006**, *2*, 1–156. [CrossRef]
10. Wang, Z.; Bovik, A.C. Reduced- and No-Reference Image Quality Assessment. *IEEE Signal Process. Mag.* **2011**, *28*, 29–40. [CrossRef]
11. Zhai, G.; Min, X. Perceptual image quality assessment: A survey. *Sci. China Inf. Sci.* **2020**, *63*, 1–52. [CrossRef]
12. Chow, L.S.; Paramesran, R. Review of medical image quality assessment. *Biomed. Signal Process. Control* **2016**, *27*, 145–154. [CrossRef]
13. Zhang, W.; Li, D.; Ma, C.; Zhai, G.; Yang, X.; Ma, K. Continual Learning for Blind Image Quality Assessment. *arXiv* **2021**. [CrossRef] [PubMed]
14. Varga, D. No-Reference Image Quality Assessment with Convolutional Neural Networks and Decision Fusion. *Appl. Sci.* **2022**, *12*, 101. [CrossRef]
15. Garcia Freitas, P.; Da Eira, L.P.; Santos, S.S.; Farias, M.C.Q.d. On the Application LBP Texture Descriptors and Its Variants for No-Reference Image Quality Assessment. *J. Imaging* **2018**, *4*. [CrossRef]
16. Varga, D. Analysis of Benford's Law for No-Reference Quality Assessment of Natural, Screen-Content, and Synthetic Images. *Electronics* **2021**, *10*, 2378. [CrossRef]
17. Leonardi, M.; Napoletano, P.; Schettini, R.; Rozza, A. No Reference, Opinion Unaware Image Quality Assessment by Anomaly Detection. *Sensors* **2021**, *21*, 994. [CrossRef]
18. Tsougos, I. *Image Principles, Neck, and the Brain*; CRC Press: Boca Raton, FL, USA, 2016; pp. 93–120. [CrossRef]
19. Patton, J.A.; Kulkarni, M.V.; Craig, J.K.; Wolfe, O.H.; Price, R.; Partain, C.L.; James, A.E., Jr. Techniques, pitfalls and artifacts in magnetic resonance imaging. *Radiographics* **1987**, *7*, 505–519. [CrossRef]
20. Mezrich, R. A perspective on K-space. *Radiology* **1995**, *195*, 297–315. [CrossRef]
21. Graves, M.J.; Mitchell, D.G. Body MRI artifacts in clinical practice: A physicist's and radiologist's perspective. *J. Magn. Reson. Imaging* **2013**, *38*, 269–287. [CrossRef]
22. Zhuo, J.; Gullapalli, R.P. MR Artifacts, Safety, and Quality Control. *RadioGraphics* **2006**, *26*, 275–297. [CrossRef]
23. Hashemi, R.H.; Bradley, W.G.; Lisanti, C.J. *MRI: The Basics: The Basics*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2012.
24. Allisy-Roberts, P.J.; Williams, J. *Farr's Physics for Medical Imaging*; Elsevier Health Sciences: Amsterdam, The Netherlands, 2007.
25. Deshmane, A.; Gulani, V.; Griswold, M.A.; Seiberlich, N. Parallel MR imaging. *J. Magn. Reson. Imaging* **2012**, *36*, 55–72. [CrossRef]
26. Hamilton, J.; Franson, D.; Seiberlich, N. Recent advances in parallel imaging for MRI. *Prog. Nucl. Magn. Reson. Spectrosc.* **2017**, *101*, 71–95. [CrossRef]
27. Ehman, R.L.; Felmlee, J.P. Flow artifact reduction in MRI: A review of the roles of gradient moment nulling and spatial presaturation. *Magn. Reson. Med.* **1990**, *14*, 293–307. [CrossRef]
28. Wood, M.L.; Henkelman, R.M. The magnetic field dependence of the breathing artifact. *Magn. Reson. Imaging* **1986**, *4*, 387–392. [CrossRef]
29. Zaitsev, M.; Maclaren, J.; Herbst, M. Motion artifacts in MRI: A complex problem with many partial solutions. *J. Magn. Reson. Imaging* **2015**, *42*, 887–901. [CrossRef]

30. Osadebey, M.E.; Pedersen, M.; Arnold, D.L.; Wendel-Mitoraj, K.E. Blind blur assessment of MRI images using parallel multiscale difference of Gaussian filters. *BioMedical Eng. OnLine* **2018**, *17*, 76. [CrossRef] [PubMed]
31. Xu, J.; Lin, C.; Zhou, W.; Chen, Z. Subjective Quality Assessment of Stereoscopic Omnidirectional Image. In Proceedings of the Advances in Multimedia Information Processing—PCM 2018, Hefei, China, 21–22 September 2018; Hong, R., Cheng, W.H., Yamasaki, T., Wang, M., Ngo, C.W., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 589–599.
32. Morelli, J.N.; Runge, V.M.; Ai, F.; Attenberger, U.; Vu, L.; Schmeets, S.H.; Nitz, W.R.; Kirsch, J.E. An image-based approach to understanding the physics of MR artifacts. *Radiographics* **2011**, *31*, 849–866. [CrossRef]
33. Pusey, E.; Lufkin, R.B.; Brown, R.; Solomon, M.A.; Stark, D.D.; Tarr, R.; Hanafee, W. Magnetic resonance imaging artifacts: Mechanism and clinical significance. *Radiographics* **1986**, *6*, 891–911. [CrossRef]
34. Block, K.T.; Uecker, M.; Frahm, J. Suppression of MRI truncation artifacts using total variation constrained data extrapolation. *Int. J. Biomed. Imaging* **2008**, *2008*, 184123. [CrossRef] [PubMed]
35. Kucharczyk, W.; Crawley, A.P.; Kelly, W.; Henkelman, R.M. Effect of multislice interference on image contrast in T2-and T1-weighted MR images. *Am. J. Neuroradiol.* **1988**, *9*, 443–451.
36. Doran, S.J.; Charles-Edwards, L.; Reinsberg, S.A.; Leach, M.O. A complete distortion correction for MR images: I. Gradient warp correction. *Phys. Med. Biol.* **2005**, *50*, 1343. [CrossRef]
37. Caramanos, Z.; Fonov, V.S.; Francis, S.J.; Narayanan, S.; Pike, G.B.; Collins, D.L.; Arnold, D.L. Gradient distortions in MRI: Characterizing and correcting for their effects on SIENA-generated measures of brain volume change. *Neuroimage* **2010**, *49*, 1601–1611. [CrossRef]
38. Noël, P.; Bammer, R.; Reinhold, C.; Haider, M.A. Parallel imaging artifacts in body magnetic resonance imaging. *Can. Assoc. Radiol. J.* **2009**, *60*, 91–98. [CrossRef]
39. Kaur, P.; Kumaran, S.S.; Tripathi, R.; Khushu, S.; Kaushik, S. Protocol error artifacts in MRI: Sources and remedies revisited. *Radiography* **2007**, *13*, 291–306. [CrossRef]
40. Gao, X.; Lu, W.; Tao, D.; Li, X. Image quality assessment and human visual system. *SPIE Proc.* **2010**, *7744*, 316–325. [CrossRef]
41. Suthaharan, S. No-reference visually significant blocking artifact metric for natural scene images. *Signal Process.* **2009**, *89*, 1647–1652. [CrossRef]
42. Bhateja, V.; Nigam, M.; Bhadauria, A.S.; Arya, A. Two-stage multi-modal MR images fusion method based on parametric logarithmic image processing (PLIP) model. *Pattern Recognit. Lett.* **2020**, *136*, 25–30. [CrossRef]
43. Liu, S.; Thung, K.H.; Lin, W.; Shen, D.; Yap, P.T. Hierarchical nonlocal residual networks for image quality assessment of pediatric diffusion MRI With Limited and Noisy Annotations. *IEEE Trans. Med. Imaging* **2020**, *39*, 3691–3702. [CrossRef]
44. Iqbal, T.; Ali, H. Generative Adversarial Network for Medical Images (MI-GAN). *J. Med. Syst.* **2018**, *42*, 231. [CrossRef] [PubMed]
45. Qi, K.; Li, H.; Rong, C.; Gong, Y.; Li, C.; Zheng, H.; Wang, S. Blind Image Quality Assessment for MRI with A Deep Three-dimensional content-adaptive Hyper-Network. *arXiv* **2021**, arXiv:2107.06888.
46. Chow, L.S.; Rajagopal, H. Modified-BRISQUE as no reference image quality assessment for structural MR images. *Magn. Reson. Imaging* **2017**, *43*, 74–87. [CrossRef]
47. Mittal, A.; Moorthy, A.K.; Bovik, A.C. Blind/Referenceless Image Spatial Quality Evaluator. In Proceedings of the 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, USA, 6–9 November 2011; pp. 723–727. [CrossRef]
48. Jang, J.; Bang, K.; Jang, H.; Hwang, D.; Initiative, A.D.N. Quality evaluation of no-reference MR images using multidirectional filters and image statistics. *Magn. Reson. Med.* **2018**, *80*, 914–924. [CrossRef]
49. Nabavi, S.; Simchi, H.; Moghaddam, M.E.; Frangi, A.F.; Abin, A.A. Automatic Multi-Class Cardiovascular Magnetic Resonance Image Quality Assessment using Unsupervised Domain Adaptation in Spatial and Frequency Domains. *arXiv* **2021**, arXiv:2112.06806.
50. Stępień, I.; Obuchowicz, R.; Piórkowski, A.; Oszust, M. Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment. *Sensors* **2021**, *21*, 1043. [CrossRef] [PubMed]
51. Obuchowicz, R.; Oszust, M.; Bielecka, M.; Bielecki, A.; Piórkowski, A. Magnetic resonance image quality assessment by using non-maximum suppression and entropy analysis. *Entropy* **2020**, *22*, 220. [CrossRef] [PubMed]
52. Oszust, M.; Piórkowski, A.; Obuchowicz, R. No-reference image quality assessment of magnetic resonance images with high-boost filtering and local features. *Magn. Reson. Med.* **2020**, *84*, 1648–1660. [CrossRef]
53. Wang, Y.; Song, Y.; Wang, F.; Sun, J.; Gao, X.; Han, Z.; Shi, L.; Shao, G.; Fan, M.; Yang, G. A two-step automated quality assessment for liver MR images based on convolutional neural network. *Eur. J. Radiol.* **2020**, *124*, 108822. [CrossRef] [PubMed]
54. Masoudi, S.; Harmon, S.; Mehralivand, S.; Lay, N.; Bagci, U.; Wood, B.J.; Pinto, P.A.; Choyke, P.; Turkbey, B. No-Reference Image Quality Assessment Of T2-Weighted Magnetic Resonance Images In Prostate Cancer Patients. In Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Virtual Conference, 13–16 April 2021; pp. 1201–1205.
55. Tang, H.; Joshi, N.; Kapoor, A. Learning a blind measure of perceptual image quality. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 305–312. [CrossRef]
56. Dickie, D.A.; Shenkin, S.D.; Anblagan, D.; Lee, J.; Blesa Cabez, M.; Rodriguez, D.; Boardman, J.P.; Waldman, A.; Job, D.E.; Wardlaw, J.M. Whole brain magnetic resonance image atlases: A systematic review of existing atlases and caveats for use in population imaging. *Front. Neuroinformatics* **2017**, *11*, 1. [CrossRef] [PubMed]

57. Esteban, O.; Blair, R.W.; Nielson, D.M.; Varada, J.C.; Marrett, S.; Thomas, A.G.; Poldrack, R.A.; Gorgolewski, K.J. Crowdsourced MRI quality metrics and expert quality annotations for training of humans and machines. *Sci. Data* **2019**, *6*, 1–7. [CrossRef]
58. Ning, Q.; Ma, C.; Lam, F.; Liang, Z.P. Spectral quantification for high-resolution MR spectroscopic imaging with spatiospectral constraints. *IEEE Trans. Biomed. Eng.* **2016**, *64*, 1178–1186. [CrossRef]
59. Di Martino, A.; O'connor, D.; Chen, B.; Alaerts, K.; Anderson, J.S.; Assaf, M.; Balsters, J.H.; Baxter, L.; Beggiato, A.; Bernaerts, S.; et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. *Sci. Data* **2017**, *4*, 1–15. [CrossRef]
60. Mortamet, B.; Bernstein, M.A.; Jack, C.R., Jr.; Gunter, J.L.; Ward, C.; Britson, P.J.; Meuli, R.; Thiran, J.P.; Krueger, G. Automatic quality assessment in structural brain magnetic resonance imaging. *Magn. Reson. Med. Off. J. Int. Soc. Magn. Reson. Med.* **2009**, *62*, 365–372. [CrossRef] [PubMed]
61. Bai, W.; Sinclair, M.; Tarroni, G.; Oktay, O.; Rajchl, M.; Vaillant, G.; Lee, A.M.; Aung, N.; Lukaschuk, E.; Sanghvi, M.M.; et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J. Cardiovasc. Magn. Reson.* **2018**, *20*, 1–12. [CrossRef]
62. Andreopoulos, A.; Tsotsos, J.K. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. *Med. Image Anal.* **2008**, *12*, 335–357. [CrossRef] [PubMed]
63. Esteban, O.; Birman, D.; Schaer, M.; Koyejo, O.O.; Poldrack, R.A.; Gorgolewski, K.J. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE* **2017**, *12*, e0184661. [CrossRef]
64. Poldrack, R.A.; Gorgolewski, K.J. OpenfMRI: Open sharing of task fMRI data. *Neuroimage* **2017**, *144*, 259–261. [CrossRef]
65. Osadebey, M.; Pedersen, M.; Arnold, D.; Wendel-Mitoraj, K. Image quality evaluation in clinical research: A case study on brain and cardiac mri images in multi-center clinical trials. *IEEE J. Transl. Eng. Health Med.* **2018**, *6*, 1–15. [CrossRef]
66. Kwan, R.S.; Evans, A.C.; Pike, G.B. MRI simulation-based evaluation of image-processing and classification methods. *IEEE Trans. Med. Imaging* **1999**, *18*, 1085–1097. [CrossRef]
67. Awad, M.; Khanna, R. Support vector regression. In *Efficient Learning Machines*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 67–80.
68. Kaufman, L.; Kramer, D.M.; Crooks, L.E.; Ortendahl, D.A. Measuring signal-to-noise ratios in MR imaging. *Radiology* **1989**, *173*, 265–267. [CrossRef]
69. Henkelman, R.M. Measurement of signal intensities in the presence of noise in MR images. *Med. Phys.* **1985**, *12*, 232–233. [CrossRef]
70. Dietrich, O.; Raya, J.G.; Reeder, S.B.; Reiser, M.F.; Schoenberg, S.O. Measurement of signal-to-noise ratios in MR images: Influence of multichannel coils, parallel imaging, and reconstruction filters. *J. Magn. Reson. Imaging Off. J. Int. Soc. Magn. Reson. Med.* **2007**, *26*, 375–385. [CrossRef]
71. Erdogmus, D.; Larsson, E.G.; Yan, R.; Principe, J.C.; Fitzsimmons, J.R. Measuring the Signal-to-Noise Ratio in Magnetic Resonance Imaging: A Caveat. *Signal Process.* **2004**, *84*, 1035–1040. [CrossRef]
72. Rajkumar, S.; Malathi, G. A comparative analysis on image quality assessment for real time satellite images. *Indian J. Sci. Technol* **2016**, *9*, 1–11. [CrossRef]
73. Zhao, F.; Huang, Q.; Gao, W. Image matching by normalized cross-correlation. In Proceedings of the 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, France, 14–19 May 2006; Volume 2, p. II.
74. Chen, A.P.; Cunningham, C.H.; Kurhanewicz, J.; Xu, D.; Hurd, R.E.; Pauly, J.M.; Carvajal, L.; Karpodinis, K.; Vigneron, D.B. High-resolution 3D MR spectroscopic imaging of the prostate at 3 T with the MLEV-PRESS sequence. *Magn. Reson. Imaging* **2006**, *24*, 825–832. [CrossRef]
75. Lu, W.; Dong, K.; Cui, D.; Jiao, Q.; Qiu, J. Quality assurance of human functional magnetic resonance imaging: A literature review. *Quant. Imaging Med. Surg.* **2019**, *9*, 1147. [CrossRef]
76. Rinck, P. Magnetic resonance: A critical peer-reviewed introduction. In *Magnetic Resonance in Medicine. The Basic Textbook of the European Magnetic Resonance Forum*; BoD: Norderstedt, Germany, 2014.
77. Di Martino, A.; Yan, C.G.; Li, Q.; Denio, E.; Castellanos, F.X.; Alaerts, K.; Anderson, J.S.; Assaf, M.; Bookheimer, S.Y.; Dapretto, M.; et al. The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* **2014**, *19*, 659–667. [CrossRef] [PubMed]
78. Griswold, M.A.; Heidemann, R.M.; Jakob, P.M. Direct parallel imaging reconstruction of radially sampled data using GRAPPA with relative shifts. In Proceedings of the 11th Annual Meeting of the ISMRM, Toronto, ON, Canada, 10–16 July 2003; Volume 2349.
79. Gupta, R.; Bansal, D.; Singh, C. Image Quality Assessment Using Non-Linear MultiMetric Fusion Approach. *Int. J. Emerg. Technol. Adv. Eng.* **2014**, *4*, 822–826.
80. Sheikh, H.R.; Bovik, A.C.; De Veciana, G. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* **2005**, *14*, 2117–2128. [CrossRef]
81. Mahajan, P.; Jakhetiya, V.; Abrol, P.; Lehana, P.K.; Subudhi, B.N.; Guntuku, S.C. Perceptual quality evaluation of hazy natural images. *IEEE Trans. Ind. Inform.* **2021**, *17*, 8046–8056. [CrossRef]
82. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
83. Esses, S.J.; Lu, X.; Zhao, T.; Shanbhogue, K.; Dane, B.; Bruno, M.; Chandarana, H. Automated image quality evaluation of T2-weighted liver MRI utilizing deep learning architecture. *J. Magn. Reson. Imaging* **2018**, *47*, 723–728. [CrossRef] [PubMed]

84. Ma, J.J.; Nakarmi, U.; Kin, C.Y.S.; Sandino, C.M.; Cheng, J.Y.; Syed, A.B.; Wei, P.; Pauly, J.M.; Vasanawala, S.S. Diagnostic image quality assessment and classification in medical imaging: Opportunities and challenges. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 337–340.
85. Yan, J.; Li, J.; Fu, X. No-reference quality assessment of contrast-distorted images using contrast enhancement. *arXiv* **2019**, arXiv:1904.08879.
86. Mathieu, M.; Couprie, C.; LeCun, Y. Deep multi-scale video prediction beyond mean square error. *arXiv* **2015**, arXiv:1511.05440.
87. Willmott, C.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]

sensors

MDPI

Article

Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment

Igor Stępień [1], Rafał Obuchowicz [2], Adam Piórkowski [3] and Mariusz Oszust [4,*]

1 Doctoral School of Engineering and Technical Sciences at the Rzeszow University of Technology, al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland; igorkrzysztofstepien@gmail.com
2 Department of Diagnostic Imaging, Jagiellonian University Medical College, 19 Kopernika Street, 31-501 Cracow, Poland; rafalobuchowicz@su.krakow.pl
3 Department of Biocybernetics and Biomedical Engineering, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Cracow, Poland; pioro@agh.edu.pl
4 Department of Computer and Control Engineering, Rzeszow University of Technology, W. Pola 2, 35-959 Rzeszow, Poland
* Correspondence: marosz@kia.prz.edu.pl

Abstract: The quality of magnetic resonance images may influence the diagnosis and subsequent treatment. Therefore, in this paper, a novel no-reference (NR) magnetic resonance image quality assessment (MRIQA) method is proposed. In the approach, deep convolutional neural network architectures are fused and jointly trained to better capture the characteristics of MR images. Then, to improve the quality prediction performance, the support vector machine regression (SVR) technique is employed on the features generated by fused networks. In the paper, several promising network architectures are introduced, investigated, and experimentally compared with state-of-the-art NR-IQA methods on two representative MRIQA benchmark datasets. One of the datasets is introduced in this work. As the experimental validation reveals, the proposed fusion of networks outperforms related approaches in terms of correlation with subjective opinions of a large number of experienced radiologists.

Keywords: image quality assessment; deep learning; network fusion; magnetic resonance images

Citation: Stępień, I.; Obuchowicz, R.; Piórkowski, A.; Oszust, M. Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment. *Sensors* **2021**, *21*, 1043. https://doi.org/10.3390/s21041043

Academic Editor: Rafal Scherer
Received: 23 December 2020
Accepted: 29 January 2021
Published: 3 February 2021

1. Introduction

Image quality assessment (IQA) of magnetic resonance images (MR) plays a vital part in the diagnosis and successful treatment [1–3]. The IQA methods aim to provide automatic, repeatable, and accurate evaluation of images that would replace tests with human subjects. Such tests are often time-consuming, difficult to organize, and their output may depend on the considered group of participants. Therefore, the progress in the development of IQA techniques depends on the availability of assessed image databases. This is particularly important for MRIQA methods that require MR image databases with opinions of a representative number of radiologists, i.e., the databases are used for their comparison and stimulate the emergence of new approaches in the field. The IQA approaches are divided into three groups, depending on whether distortion-free images are used: full-reference (FR), reduced-reference (RR), and no-reference (NR). The availability of unaltered, distortion-free, reference images is a basis for their differentiation. Nevertheless, such pristine images, or their partial characteristics, are unavailable for MR images, limiting the practical application of FR and RR methods. Therefore, the NR MRIQA measures are highly desired, while FR and RR approaches are mostly employed for artificially distorted, high-quality MR images.

There are several approaches to the FR medical IQA [4]. Among them, the Peak Signal-to-Noise Ratio (PSNR) is the most popular. However, it might not be accurate enough to give a proper measure between distorted and reference images, concerning

their characteristics and known inability of the PSNR to reliably reflect human subjective opinions. In the assessment of medical images, its derivative approaches, i.e., the signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) [1], are often used. Furthermore, some early methods adapt solutions from the IQA of natural images [5], train them on images assessed by the SNR [2], or add additional features to characterize MR images [6]. Some approaches employ the entropy of local intensity extrema [7] or specific image filtering to facilitate the usage of local features [8]. Other works are devoted to a binary classification of noisy MR images [9,10] or for classification of images with prior detection of selected distortion types [11,12].

In this paper, taking into account the lack of NR IQA measures devoted to MR images in the literature, a novel NR method is proposed in which deep learning architectures are fused and the transfer-learning process is jointly performed. The resulted fusion allows the network to better characterize distorted MR images due to the diverse backgrounds of employed architectures. Furthermore, to improve the quality prediction of the approach, the SVR is used on features extracted from fused networks. As most of the databases used in the literature are not publicly available, contain artificially distorted images, and/or were assessed by a few radiologists, in this paper, a novel IQA MRI database with images assessed by a large number of radiologists is introduced.

The main contributions of this work are as follows: (i) Fusion of deep convolutional network architectures for MRIQA. (ii) The usage of the SVR with features obtained in joint transfer learning of networks to improve the performance of the method. (iii) Novel large IQA database of MR images assessed by a large number of radiologists. (iv) Extensive evaluation of the numerous deep learning architectures and related techniques.

The remainder of the paper is organized as follows. In Section 2, previous work on NR-IQA is reviewed, while in Section 3, the proposed approach is introduced. Then, in Section 4, the experimental validation of the method and related techniques is presented. Finally, Section 5 concludes the paper and indicates future directions of the research.

2. Related Work

The introduced method belongs to the category of NR techniques that use a deep learning approach to predict the quality of assessed images. However, before such approaches were possible for the IQA of natural images, many handcrafted IQA measures were proposed. For example, the method of Moorthy and Bovik [13] employs a two-stage framework in which distortion type is predicted and used for the quality evaluation. A framework in which a probabilistic model with DTC-based natural scene statistics (NSS) is trained was proposed by Saad et al. [14]. Then, the popular BRISQUE technique [15] was introduced, which uses the training of the Generalized Gaussian Distribution (GGD) with Mean Substracted Contrast Normalization (MSCN) coefficients. The Gabor features and the soft-assignment coding with the max-pooling are employed in the work of Ye et al. [16]. In the High Order Statistics Aggregation (HOSA) [17] method, low and high order statistics for the description of normalized image patches obtained from codebook using the soft assignment is presented. In the method, the codebook was obtained with the k-means approach. As gradient-based features can effectively describe distorted images, many approaches use them for quality prediction. They employ global distributions of gradient magnitude maps [18], relative gradient orientations or magnitude [19], and local gradient orientations captured by Histogram of Oriented Gradient (HOG) technique for variously defined neighborhoods [20]. A histogram of local binary patterns (LBP) characterizing a gradient map of an image is used in the GWHGLBP approach [21]. In the NOREQI [22] measure, an image is filtered with gradient operators and described using speeded-up robust feature (SURF) descriptor. Then, the descriptors are characterized by typical statistics. The joint statistics of the gradient magnitude map and the Laplacian of Gaussian (LOG) response are used to characterize images in the GM-LOG technique [18].

Most learning-based NR-IQA techniques devoted to natural images employ the SVR method to create a quality model. However, some methods do not require training. For ex-

ample, in the Integrated Local Natural Image Quality Evaluator (IL-NIQE) [23], natural image statistics derived from multiple cues are modeled by the multivariate Gaussian model and used for the quality prediction without additional training step. In the BPRI [24], a pseudo-reference image is created and compared with the assessed image using quality metrics that measure blockiness, sharpness, and noisiness.

In recent years, more complex IQA approaches have been introduced that use deep neural network architectures. They do not contain a separate image description and prediction stages. However, their training requires a large amount of data or an adaptation of architectures developed for computer vision tasks not related to the IQA. Some of the early models address those challenges by using image patches [25,26], training with scores generated by FR-measures or [26,27], or fine-tuning of popular networks [28].

Considering the quality assessment of MR images, the number of approaches is much less diverse. Here, only several works have been published, revealing the lack of successful techniques and the scarcity of the IQA MRI benchmarks that could be used to stimulate their development. Furthermore, most clinical applications use the SNR and CNR [1] measures to assess images or calibrate scanners. However, they require an indication of disjoint image regions with noise and tissue, despite providing an inferior quality evaluation of images in comparison with modern methods. Some of NR IQA measures designed for the assessment of MR images adapt solutions devoted to natural images. For example, Chow and Rajagopal [5] trained the BRISQUE on MR images, Yu et al. [2] used SNR scores to train BRISQUE and three other IQA methods, while Jang et al. [6] used MSCN multidirectional-filtered coefficients. In the work of Esteban et al. [9], image quality was not predicted but binarily classified based on a set of simple measures. Taking into account the inclusion of neural network architectures for processing MR images, Kustner et al. [11] and Sujit et al. [12] detected motion artifacts and performed binary classification of structural brain images, respectively. Volumetric and artifact-specific features were used by Pizarro et al. [10] to train the SVM classifier. In previous authors' works on the MRIQA, the entropy of local intensity extrema was used for direct quality prediction [7] or high-boost filtering followed by the detection and description of local features [8] was used with an SVR-based quality model.

Taking into account the lack of deep learning architectures for the MRIQA, it can be stated that their effectiveness remains largely uninvestigated, and the introduction of their effective fusion can be seen as a promising area of research.

3. Proposed Method

In the proposed approach, a fusion of deep network architectures is considered. Such architectures are mostly devoted to image recognition tasks and were propagated to other areas of computer vision [29]. Among popular deep learning networks, the approach of Simonyan and Zisserman [30] (VGG) uses 3×3 convolutional filters and achieves outstanding performance at the ImageNet Large Scale Visual Recognition Competition (ILSVRC) 2014 competition. Another solution, Resnet [31], introduces a residual learning framework, with a shortcut connection between layers to address the overfitting experienced by the VGG. With the same purpose, Szegedy et al. [32,33] introduced the Inception module in the GoogLeNet model. In other works, Howard et al. [34] created Mobilenet aiming to reduce the computational costs, or Huang et al. [35], in the DenseNet, used network layers with inputs from all preceding layers.

Deep learning models were also used for the IQA of natural images [26,36–39]. Most of such adaptations employ transfer learning, making the networks aware of domain-specific characteristics. Therefore, in this study, a similar approach was applied at the beginning of the research. Thus, adapted single models can be seen as counterparts of the first approaches with deep learning models to the IQA of natural images. However, the performance of a single network turned out to be insufficient to provide superior performance in the MRIQA task (see Section 4.6). Therefore, the approach introduced in this paper considers an internal fusion of networks, assuming that the fusion of different

network types can capture characteristics of MR images, leading to outstanding IQA accuracy across the benchmark datasets.

Considering an image I_n that belongs to a set of N training images, $n = 1, \ldots, N$, each of which is associated with subjective score q_n^s obtained in tests with human subjects. Note that the subjective scores are denoted as Mean Opinion Scores (MOS) or Difference MOS (DMOS). The input image $I = \mathrm{R}^{h \times w \times c}$, where h, w, and c corresponds to the height, width, and the number of channels, respectively, is processed by the network. The network often consists of stacked modules of convolutional and pooling layers, and several fully connected layers. The convolutional layers extract features from earlier layers. This can be written as $O_k = \mathtt{f}(W_k \circledast I)$, where W_k denotes a k-th filter or weights, $'\circledast'$ is convolutional operator, and $\mathtt{f}$ is the nonlinear activation function, often represented by rectified linear units (ReLUs). The pooling layers reduce feature maps, introducing average or maximum values of inputs to the next layers. For example, the max pooling $O_k(a, b) = max(I_k(i, j))$, where the $(i, j) \in \mathtt{P}(a, b)$ denotes location of the element in the pooled region $\mathtt{P}(a, b)$. The fully connected layers are used to interpret features and provide high-level problem representations. Their outputs are further processed by the softmax or regression layer for the classification or regression problems, respectively.

Network Fusion

As the number of images in MRIQA benchmarks is not large enough for efficient training of considered network architectures or their fusions, in this study, transfer learning [40] is applied instead of learning from scratch. The considered networks are pretrained on ImageNet dataset and classify images into 1000 categories. However, the IQA task requires solving the regression problem, which forces the modification of the architecture of the network towards the quality prediction purpose. Therefore, in this study, the last three network layers of each used network, configured for 1000 classes, are replaced with a fully connected layer and the regression layer. Note that the replacement is performed regarding all network architectures, either single or fused. Here, the networks share the inputs and are connected to each other with a feature concatenation layer that adds outputs from multiple layers in an element-wise manner. If needed, the input image is resized to match the input size of a network. As layers responsible for average pooling remain in each network architecture after the removal of the part associated with image classification, they are used as inputs to the concatenation (addition) layer. An exemplary fusion of ResNet-18, ResNet-50, and GoogLeNet is presented in Figure 1. In the network graph, each network is represented by connected sets of convolution layers (yellow blocks). Among the last elements in each network are the global average pooling layer, addition layer that fuses their outputs, fully connected layer, and regression layer.

For the training of the resulted network architecture, N images are used. However, as MR images are often 2-dimensional 16-bit matrices, they are concatenated to form three channels (c = 3) to facilitate processing by the pretrained networks. To estimate network parameters, the half Mean Squared Error (MSE) loss function L is applied. Considering that $Q^o = (q_1^o, \ldots, q_N^o)$ is the vector of objective scores and $Q^s = (q_1^s, \ldots, q_N^s)$ represents subjective scores, L is calculated as

$$L(Q^s, Q^o) = \frac{1}{N} \sum_{n=1}^{N} (q_n^s - q_n^o)^2. \quad (1)$$

Typically, transfer learning of the network assumes freezing the original layers to prevent back-propagation of errors. However, as in this study MR images are processed and have different characteristics from natural images, the weights of fused networks were modified using a small learning rate.

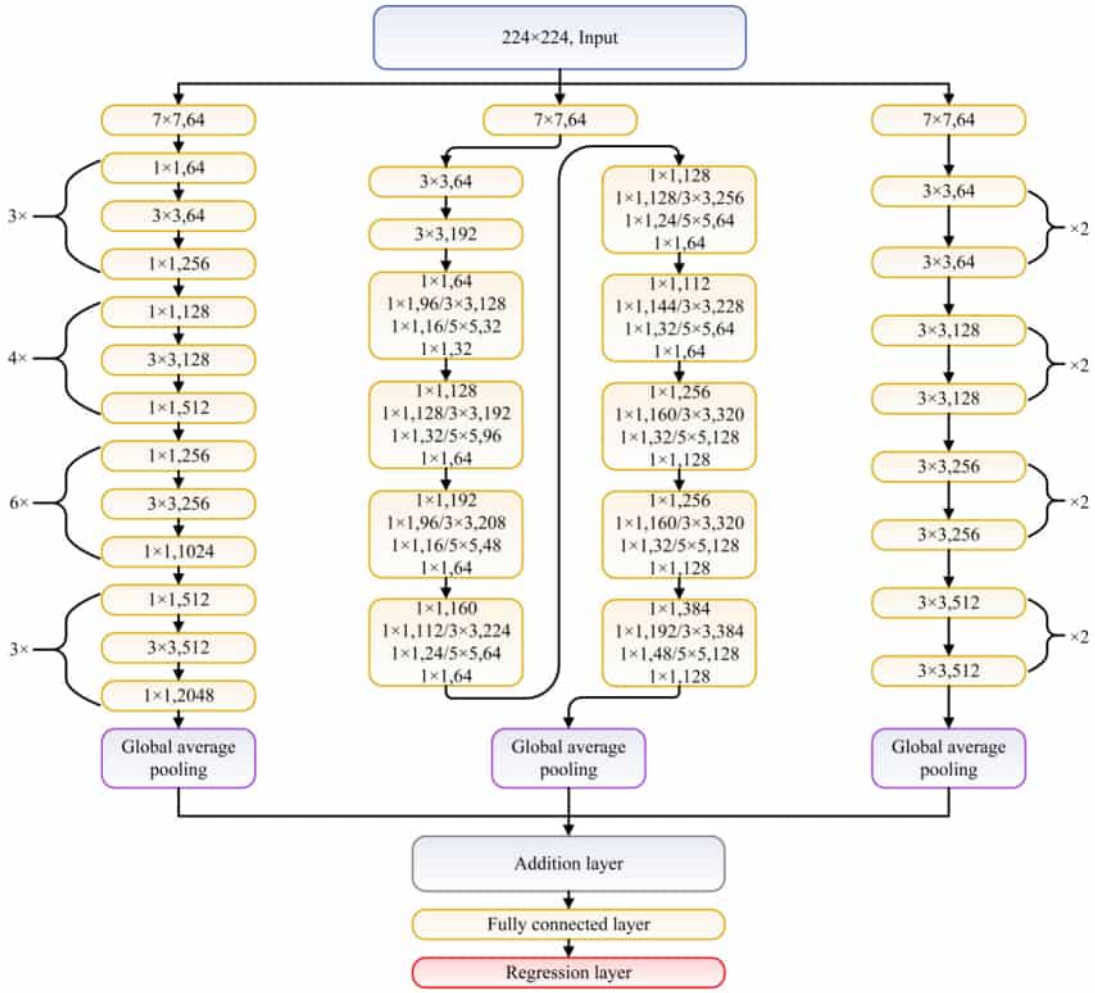


Figure 1. Example of fusion network composed of ResNet-50 (**left**), GoogleNet (**center**), and ResNet-18 (**right**).

Once the training of the network is finished, a second step of the approach is executed in which concatenated feature vectors (see the addition layer in Figure 1) are used as an input x to the SVR module to obtain a quality prediction model, $x = x_{net_1} \oplus x_{net_2} \oplus \cdots \oplus x_{net_M}$, where $\oplus$ is the concatenation operator and M is the number of fused deep learning architectures.

The SVR technique is commonly used to map perceptual features to MOS. In this paper, the ε-SVR is employed for training the regression model. Given training data $(X, Q_s) = \{(x_1, q_1^s),...,(x_N, q_N^s)\}$, where x_n is the feature vector and q_n^s is its MOS, a function $f(x) = \langle \omega, x \rangle + b$ is determined in which $\langle \cdot, \cdot \rangle$ denotes the inner product, ω is the weight vector, and b is a bias parameter. Introducing the slack variables ξ_n and ξ_n^*, ω and b can be computed by solving the following optimization problem,

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|\omega\|^2 + C\sum_{n=1}^{N}(\xi_n + \xi_n^*) \\ &subject\ to \begin{cases} \langle \omega, x_n \rangle - (q_n^s - b) \le \varepsilon + \xi_n \\ q_n^s - b - \langle \omega, x_n \rangle \le \varepsilon + \xi_n^* \\ \xi_n, \xi_n^* \ge 0, \end{cases} \end{aligned} \tag{2}$$

where C is a the constant parameter to balance ω and the slack variables. The $\omega = \sum_{n=1}^{N} t_n x_n$, where t_n is a combination coefficient. Usually, in the first step, the input feature vector is mapped into a high-dimensional feature space $\Phi(x)$, and then the regression function is obtained:

$$\mathfrak{f}(x) = \left\langle \sum_{n=1}^{N} t_n \Phi(x_n), \Phi(x) \right\rangle + b$$
$$= \sum_{n=1}^{N} t_n \langle \Phi(x_n), \Phi(x) \rangle + b. \quad (3)$$

The inner product $\langle \Phi(x_n), \Phi(x) \rangle$ can be written as a kernel function $\mathfrak{c}(x_n, x)$. Therefore,

$$\mathfrak{f}(x) = \sum_{n=1}^{N} t_n \mathfrak{c}(x_n, x) + b. \quad (4)$$

The radial base function (RBF) is often used as $\mathfrak{c}$, $\mathfrak{c}(x_n, x) = exp(-\gamma(|x_n - x|)^2)$, where the γ is the precision parameter [18].

The main computational steps of the approach are shown in Figure 2. As it can be seen, networks are fused and trained together (*training A*) to capture MR-specific characteristics. Then, the SVR module is trained with concatenated feature maps (*training B*) to obtain the quality model used in the prediction.

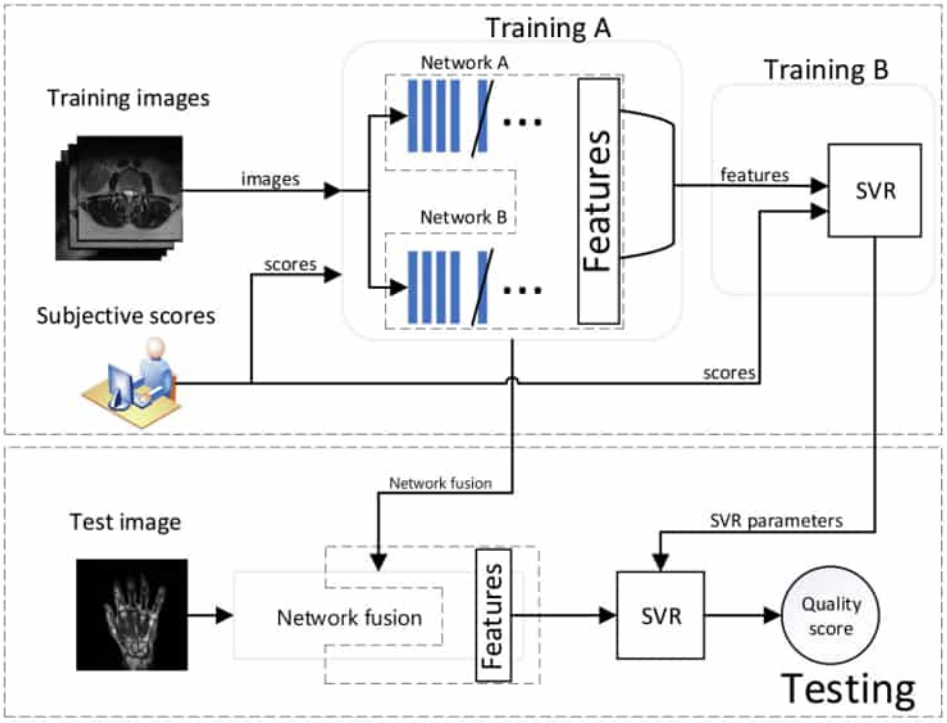


Figure 2. Block diagram of the proposed approach.

In this paper, the following networks are considered in the fusion: DenseNet-201 [35], GoogLeNet [32], Inception-v3 [33], MobileNet-V2 [34], ResNet-101 [31], ResNet-18 [31], and ResNet-50 [31]. The networks process 224 × 224 images (instead of Inception-v3 that works with 299 × 299 images). The ResNet employs 18, 50, or 101 layers, while GoogLeNet, Inception-v3, MobileNet-V2, and DenseNet-201, use 22, 48, 53, and 201 layers, respectively.

To further justify the need for network fusion proposed in this paper and show its sensitivity to distortions, a visualization of exemplary features processed by single and fused networks using DeepDream (https://www.tensorflow.org/tutorials/generative/deepdream (accessed on 23 December 2020)) technique is provided in Figure 3. The technique is often used to show what a given network has learned at a given stage. In the experiment, the best and worst quality images of the same body part were used. As presented, the features in

fused networks distinctively respond to distortions that are propagated through the architecture. Furthermore, their features seem affected by the existence of another network in the training which modifies their response comparing to single architectures. Interestingly, the features of GoogLeNet or ResNet in the GoogLeNet+ResNet-18 fusion are more similar to each other than to features from single network architectures. This can be attributed to joint transfer learning. The fusion exhibits a different response to different distortion severity. Consequently, it can be assumed that the quality prediction model based on the fusion would be able to reliably evaluate MR images of different quality.

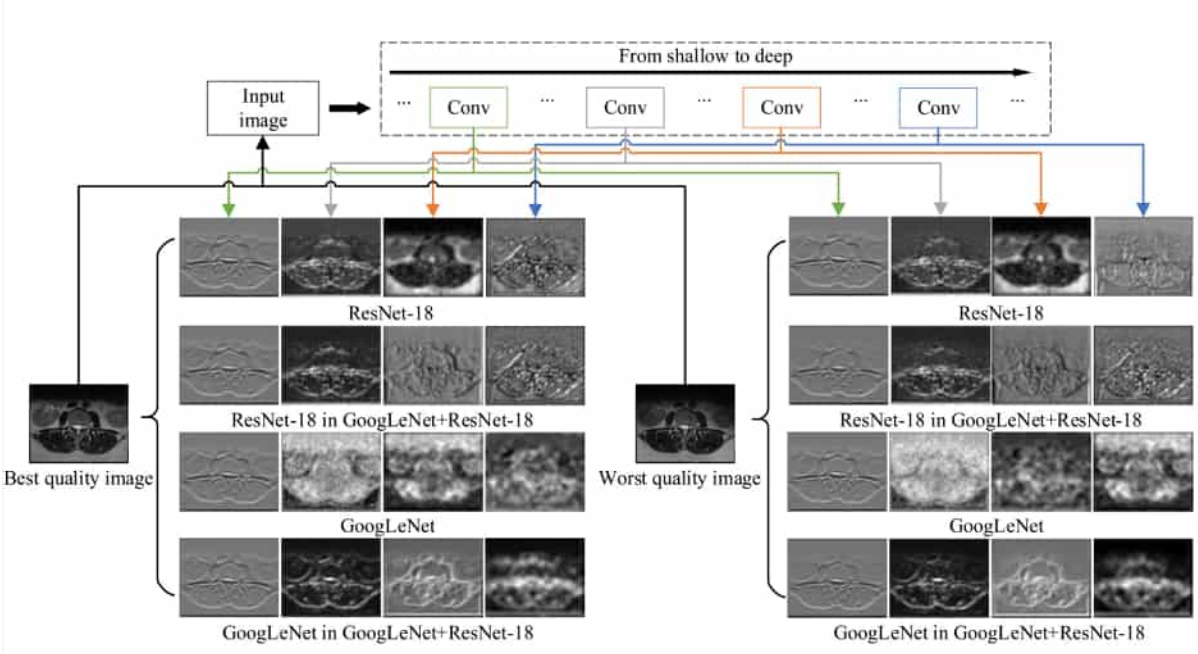


Figure 3. Visualization of features at different layers for the best and worst quality MR images. The GoogLeNet, ResNet-18, and their fusion are shown.

4. Results and Discussion

4.1. Experimental Data

The proposed approach is evaluated on two MRIQA benchmark datasets. The first dataset, denoted for convenience as *DB1*, contains 70 MR images [8], while the second one (*DB2*) has been created for the needs of this study and contains 240 MR images.

The DB1 benchmark contains images selected from 1.5T MR T2-weighted sagittal sequences: the spine (14 images), knee (14), shoulder (16), brain (8), wrist (6), hip (4), pelvis (4), elbow (2), and ankle (2). The collection consists of images captured under different conditions affecting the image quality (IPAT software to made GeneRalized Autocalibrating Partially Parallel Acquisitions (GRAPPA); GRAPPA3 [41]). Apart from the images, the benchmark contains the MOS, ranging from 1 to 5, which was obtained in tests with a group of radiologists [8]. The resolution of images in the dataset is between 192 × 320 and 512 × 512. Exemplary images that can be found in the DB1 are shown in Figure 4.

As the DB1 collection is relatively small and databases that can be found in the literature were created for different purposes, are not available, or were assessed by a small number of radiologists, a novel dataset has been introduced in this study. The DB2 collection contains T2 weighted MR images acquired during routine diagnostic exams of the shoulders, knees, and cervical and lumbar spine. Patients aged 29–51 yo participated in the study. Siemens Essensa 1.5 Tesla scanner equipped with table coils, 6-channel—knee and 4-channel—shoulder coils, were used to obtain two-dimensional images in axial, coronal, and sagittal planes. The gradient strength and a slew rate of the scanner were set to 30 mT/m, and 100 T/m/s, respectively. The following parameters were also used:

the echo time TE ranged from 3060 up to 6040 ms, repetition time TR (77 to 101 ms), phase oversampling of 20, distance factor of 30, and flip angle of 150°. The dataset was made on matrices from 192 × 320 to 512 × 512, using a voxel of nonisotropic resolution at 0.8 mm × 0.8 mm × 3 mm. To obtain images of different quality in a controlled way, the parallel imaging technique was applied (Siemens IPAT software) [41], reducing the number of echoes. The parallel imaging shortens the acquisition time [42] as it is commonly employed to increase patient comfort. However, in this study, it was applied to obtain degraded images during the routine imaging process. The T2 sequences were obtained using the GRAPPA approach, repeated in four modes with gradually increased severity of echo reduction. Here, GRAPPA1, GRAPPA2, GRAPPA3, and GRAPPA4 were consecutively applied, resulting in up to a 4-minute increase of total patient examination time. Finally, 30%, 40%, 60%, and 80% of the signals were lost with GRAPPA 1–4, respectively [43,44]. Obtained images were anonymized and saved ensuring the highest standards of personal data safety. Then, the DB2 was created with images of different patients. The subset of 30 exams was further investigated: knee (images of eight patients), shoulder (10 patients), cervical spine (three patients), and lumbar spine (nine patients). Then, from sequences of a better and worse quality associated to a given patient, two images per sequence were automatically selected. The selected scans were located at 1/3 and 2/3 of the length of each sequence. Once the 240 images were selected for subjective tests, a large group of 24 radiologists with more than 6 years of experience in MR study reading was invited for the assessment of their quality. The group was gathered in a room with limited access to daylight, reflecting typical conditions of such examination. Images of the different quality were displayed in pairs and presented to radiologists for 30 s (each pair) on Eizo Radi-Force high-resolution monitors (2600 × 1800) connected to dedicated graphic cards. Each radiologist was introduced to the assessment procedure and assigned scores from 1 to 5 to each evaluated image on paper evaluation cards. In the assessment, a higher grade reflects a better quality of an image. Then, subjective scores were processed and averaged to obtain MOS. Exemplary images from the DB2 are presented in Figure 5.

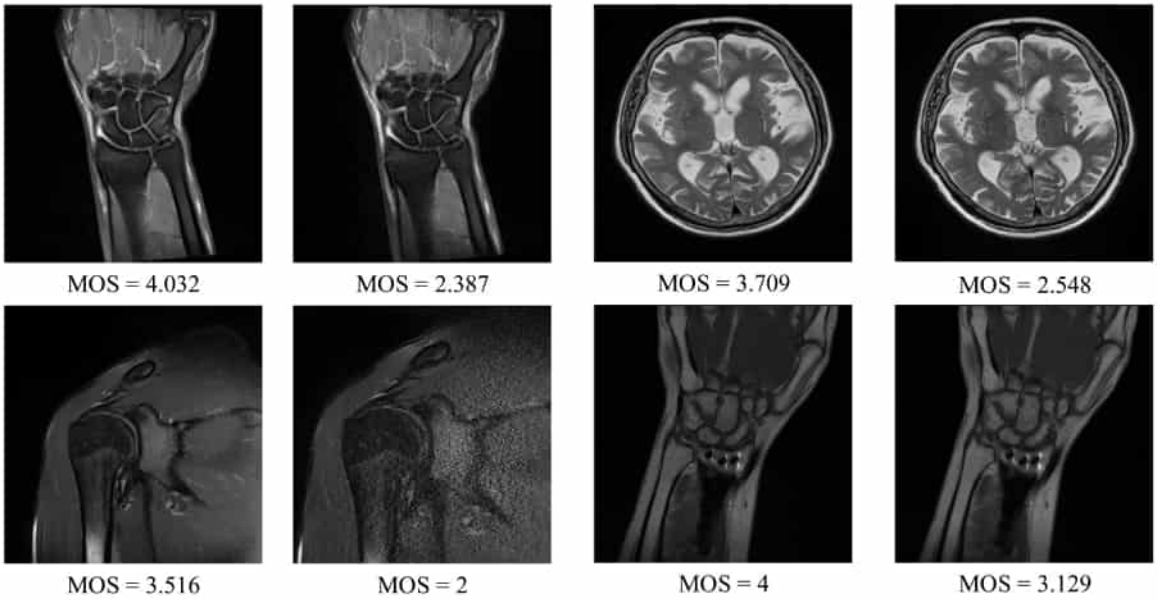


Figure 4. DB1 benchmark: Exemplary magnetic resonance (MR) images and their subjective scores.

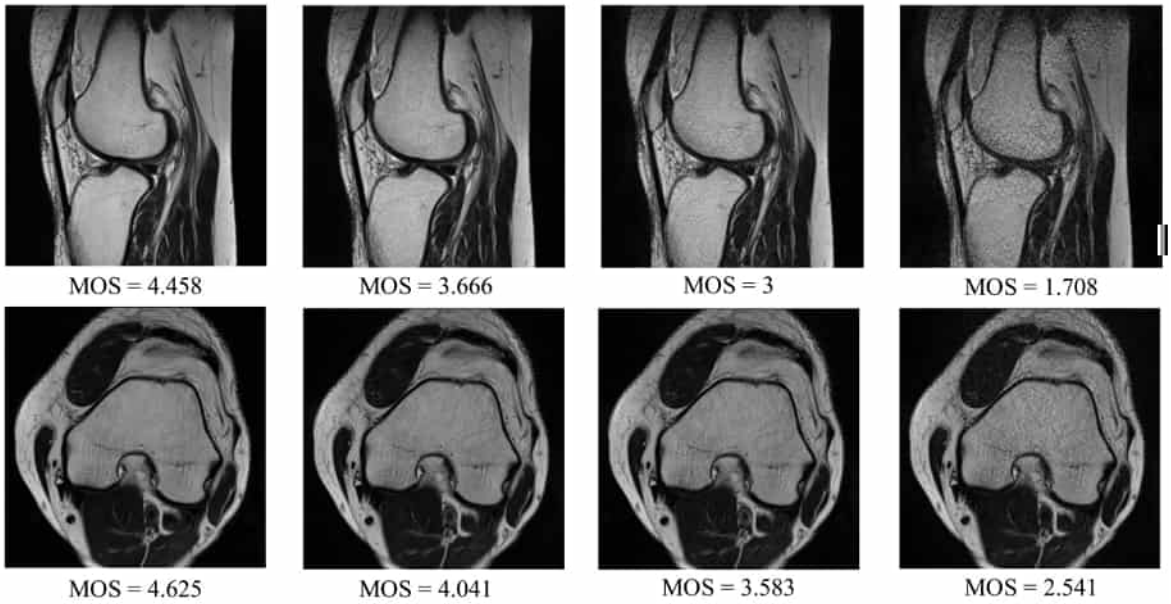


Figure 5. Introduced MRIQA benchmark (DB2): Exemplary MR images and their subjective scores.

4.2. Evaluation Methodology

Image quality assessment techniques are evaluated and compared on benchmark datasets using four performance criteria [45]: Spearman rank-order correlation coefficient (SRCC), Kendall rank-order correlation coefficient (KRCC), Pearson linear correlation coefficient (PLCC), and Root Mean Square Error (RMSE). The higher SRCC, KRCC, and PLCC, and lower RMSE, the better output of objective IQA approach. The calculation of the PLCC (Equation (5)) and RMSE (Equation (6)) require a nonlinear mapping of objective scores Q^o fitted with a regression model, $\mathcal{Q}^o$, and subjective opinions Q^s. The model employed for the mapping is expressed as $\mathcal{Q}^o = \beta_1\left(\frac{1}{2} - \frac{1}{1+\exp(\beta_2(Q^o-\beta_3))}\right) + \beta_4 Q^o + \beta_5$, where $\beta = [\beta_1, \beta_2, \ldots, \beta_5]$ [45]. The PLCC is calculated as

$$\text{PLCC} = \frac{\bar{\mathcal{Q}}^{o\,T}\bar{Q}^s}{\sqrt{\bar{\mathcal{Q}}^{o\,T}\bar{\mathcal{Q}}^o\bar{Q}^{s\,T}\bar{Q}^s}}, \tag{5}$$

where $\bar{\mathcal{Q}}^o$ and $\bar{Q}^s$ are mean-removed vectors. The RMSE is calculated as

$$\text{RMSE} = \sqrt{\frac{(\mathcal{Q}^o - Q^s)^T(\mathcal{Q}^o - Q^s)}{m}}, \tag{6}$$

where m is the total number of images. The SRCC is defined as

$$\text{SRCC} = 1 - \frac{6\sum_{i=1}^{m} d_i^2}{m(m^2-1)}, \tag{7}$$

where d_i is the difference between i-th image in Q^o and Q^s, $i = 1, 2, \ldots, m$. Consequently, the KRCC calculated as

$$\text{KRCC} = \frac{m_c - m_d}{0.5m(m-1)}, \tag{8}$$

where m_c is the number of concordant pairs in the dataset, and m_d denotes the number of discordant pairs.

As the proposed approach should be trained to obtain a quality model for the prediction, a widely accepted protocol for the evaluation of related methods is used in which 80% of randomly selected images of the dataset are selected for the training and the remaining

20% of images test the approach [16,24]. Both image subsets are disjoint based on the experiment that leads to the acquisition of images of a given body part. Then, to avoid bias, the performance of an NR method is reported in terms of the median values of SRCC, KRCC, PCC, and RMSE over 10 training–testing iterations [27,46].

4.3. Comparative Evaluation

The introduced approach is represented by three fusion models: Resnet-50_GoogLeNet _ResNet-18 (R50GR18), ResNet-50_GoogLeNet_MobileNet-V2 (R18GR50M), and MobileNet-V2_ResNet-50 (MR50). They are experimentally compared with 17 state-of-the-art techniques: NFERM [47], SEER [20], DEEPIQ [27], MEON [46], SNRTOI [48], NOREQI [22], BPRI [24], HOSA [17], NOMRIQA [8], IL-NIQE [23], GM-LOG [18], GWHGLBP [21], BRISQUE [15], SISBLIM [49], metricQ [50], SINDEX [51], and ENMIQA [7]. The NOMRIQA, ENMIQA, and SNRTOI are designed for MR images, whereas DEEPIQ and MEON are deep learning approaches devoted to natural images. Interestingly, as Chow and Rajagopal [5] trained the BRISQUE on MR images in their approach, the BRISQUE in this study, as well as other methods trained on considered benchmark databases, can be seen as adaptations to the MR domain.

For a fair comparison, all methods were run in Matlab with their default parameters, while the SVR parameters [52] were determined aiming at their best quality prediction performance. NR techniques that process color images were assessing MR images concatenated to form three channels. For the training of the proposed fusion architectures, stochastic gradient descent with momentum (SGDM) was used with a learning rate of 10^{-4}, mini-batch size of 32, and 5 epochs. Furthermore, as the number of images in the first dataset is relatively low (70), data augmentation was employed in which each distorted image was rotated up to 360° with the step of 3°. The approaches were run in Matlab R2020b, Windows 10, on PC with i7-7700k CPU, 32GB RAM, and GTX 1080 Ti graphic card.

The results for both databases are presented in Table 1. Their analysis indicates that the proposed network fusion techniques outperform the state-of-the-art approaches. Specifically, the R50GR18 (composed of three network architectures) and MR50 (fusing two networks) obtained the best SRCC and KRCC performances for the first database, outperforming the recently introduced NOMRIQA. This is also confirmed by the results for the remaining criteria, i.e., PLCC and RMSE. It is worth noticing that NR MRIQA methods that do not require training, ENMIQA and SNRTOI, produce poorly correlated quality opinions while compared with outputs of learning-based approaches as they are not equipped with various perceptual features and powerful machine learning algorithms that efficiently map them with subjective scores.

Table 1. Performance comparison of twenty evaluated methods on both datasets.

Method	DB1				DB2				Overall			
	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
NOMRIQA *	**0.7030**	**0.5527**	0.7978	0.4322	0.8040	0.6087	0.8737	0.4605	0.7535	0.5807	0.8358	0.4464
HOSA	0.4804	0.3909	0.6997	0.5318	0.8756	0.7052	**0.9276**	**0.3388**	0.6780	0.5481	0.8137	0.4353
NOREQI	0.4359	0.2922	**0.8045**	**0.4182**	0.8675	0.6984	0.9072	0.3833	0.6517	0.4953	0.8559	0.4008
IL-NIQE	0.1695	0.1275	0.3619	0.6674	0.1197	0.0836	0.3090	0.8821	0.1446	0.1056	0.3355	0.7748
GM-LOG	0.4673	0.3424	0.6515	0.4779	0.8854	0.7123	0.9010	0.4091	0.6764	0.5274	0.7763	0.4435
GWHGLBP	0.5075	0.3935	0.6886	0.5257	0.8726	0.6927	0.8947	0.4080	0.6901	0.5431	0.7917	0.4669
BRISQUE	0.4610	0.3648	0.6100	0.5311	0.8544	0.6738	0.8951	0.4076	0.6577	0.5193	0.7526	0.4694
SISBLIM	0.3976	0.2776	0.6240	0.5449	0.7216	0.5419	0.7592	0.6047	0.5596	0.4098	0.6916	0.5748
metricQ	0.2596	0.1657	0.2792	0.6709	0.5066	0.3701	0.5227	0.7791	0.3831	0.2679	0.4010	0.7250
BPRI	0.2412	0.1890	0.4785	0.5756	0.1317	0.0973	0.4928	0.7883	0.1865	0.1432	0.4857	0.6820
SINDEX	0.2939	0.2112	0.3243	0.7034	0.2673	0.1933	0.3185	0.8874	0.2806	0.2023	0.3214	0.7954
NFERM	0.5073	0.4091	0.7662	0.4491	0.8833	0.7087	0.9157	0.3872	0.6953	0.5589	0.8410	0.4182
SEER	0.4776	0.3574	0.7108	0.5267	**0.8938**	**0.7335**	0.9196	0.3594	0.6857	0.5455	0.8152	0.4431
MEON	0.2518	0.1879	0.3439	0.6428	0.5851	0.4001	0.6194	0.7426	0.4185	0.2940	0.4817	0.6927
DEEPIQ	0.1133	0.0827	0.5902	0.5707	0.2837	0.2078	0.5393	0.7822	0.1985	0.1453	0.5648	0.6765
SNRTOI *	0.1321	0.0728	0.4094	0.6784	0.1016	0.0720	0.3169	0.8930	0.1169	0.0724	0.3632	0.7857
ENMIQA *	0.3630	0.2479	0.5093	0.5873	0.7941	0.6119	0.8313	0.5130	0.5786	0.4299	0.6703	0.5502
R18GR50M *	0.6299	0.5012	0.7999	0.4381	**0.8998**	**0.7398**	**0.9270**	**0.3465**	**0.7649**	**0.6205**	**0.8635**	**0.3923**
R50GR18 *	**0.7423**	**0.6039**	**0.8206**	**0.4238**	**0.9083**	**0.7490**	**0.9294**	**0.3479**	**0.8253**	**0.6765**	**0.8750**	**0.3859**
MR50 *	**0.7036**	**0.5740**	**0.8576**	**0.3865**	0.8919	0.7176	0.9241	0.3560	**0.7978**	**0.6458**	**0.8909**	**0.3712**

Note: * denotes the approach designed for the evaluation of MR images. The best three results for each criterion are written in bold, the names of fusion architectures introduced in this paper are written in italics.

For the database introduced in this paper, DB2, the R18GR50M, and R50GR18 present superior performance, followed by the SEER and MR50. Here, more techniques obtain better results in comparison to those obtained for the DB1 due to the larger representation of distorted images in the dataset. The methods designed for MR images—ENMIQA and NOMRIQA—are outperformed by well-established learning methods devoted to natural images (SEER, HOSA, GM-LOG, or NFERM). Deep learning models—MEON and DEEPIQ—were pretrained by their authors and do not capture characteristics of MR images, leading to inferior prediction accuracy. The overall results, averaging criteria over both databases, indicate the superiority of introduced fusion approaches. In this case, the NOMRIQA is fourth in terms of the SRCC values, followed by NFERM and GWHGLBP after a large performance gap.

To compare relative differences between methods and determine whether they are statistically significant, the Wilcoxon rank-sum test is used. The test measures the equivalence of the median values of independent samples with a 5% significance level [20]. Here, the SRCC values are taken into consideration. In the experiment, the method with a significantly greater SRCC median obtained the score of "1". Consequently, the worse and indistinguishable method obtained "−1" and "0", respectively. Finally, scores were added and displayed in cells in Figure 6 to characterize methods in rows. The figure also contains sums of scores to indicate globally best approaches. As reported, all three introduced fusion architectures offer promising and stable performance across both datasets, outperforming the remaining approaches. Other learning-based methods designed for the MRIQA, i.e., ENMIQA or NOMRIQA, exhibit inferior relative performance in comparison to the proposed models and methods with rich image representations (GM-LOG, SEER, or NFERM).

	NOMRIQA*	HOSA	NOREQI	IL-NIQE	GM-LOG	GWHGLBP	BRISQUE	SISBLIM	metricQ	BPRI	SINDEX	NFERM	SEER	MEON	DEEPIQ	SNRTOI*	ENMIQA*	**R18GR50M***	**R50GR18***	**MR50***	Sum
NOMRIQA*	0	0	-1	2	-1	0	1	2	2	2	2	-1	0	2	2	2	1	0	-2	-1	12
HOSA	0	0	0	2	0	0	0	1	2	2	1	0	0	2	2	2	1	0	-1	-1	13
NOREQI	1	0	0	2	0	0	0	1	2	2	2	0	0	2	2	2	1	-1	-1	-1	14
IL-NIQE	-2	-2	-2	0	-2	-2	-2	-2	-1	0	-2	-2	-2	-1	-1	0	-2	-2	-2	-2	-31
GM-LOG	1	0	0	2	0	0	0	1	2	2	2	0	0	2	2	2	2	0	-1	-1	16
GWHGLBP	0	0	0	2	0	0	0	1	2	2	2	0	0	2	2	2	1	0	-1	-1	14
BRISQUE	-1	0	0	2	0	0	0	1	2	2	2	0	0	2	2	2	0	-2	-1	-1	10
SISBLIM	-2	-1	-1	2	-1	-1	-1	0	1	1	1	-1	-1	2	1	2	0	-2	-2	-2	-5
metricQ	-2	-2	-2	1	-2	-2	-2	-1	0	1	1	-2	-2	0	1	1	-1	-2	-2	-2	-19
BPRI	-2	-2	-2	0	-2	-2	-2	-1	-1	0	-1	-2	-2	-1	-1	0	-2	-2	-2	-2	-29
SINDEX	-2	-1	-2	2	-2	-2	-2	-1	-1	1	0	-1	-2	-1	0	2	-1	-2	-2	-2	-19
NFERM	1	0	0	2	0	0	0	1	2	2	1	0	0	2	2	2	1	0	-1	-1	14
SEER	0	0	0	2	0	0	0	1	2	2	2	0	0	2	2	2	1	-1	-1	-1	13
MEON	-2	-2	-2	1	-2	-2	-2	-2	0	1	1	-2	-2	0	1	1	-2	-2	-2	-2	-21
DEEPIQ	-2	-2	-2	1	-2	-2	-2	-1	-1	1	0	-2	-2	-1	0	1	-2	-2	-2	-2	-24
SNRTOI*	-2	-2	-2	0	-2	-2	-2	-2	-1	0	-2	-2	-2	-1	-1	0	-2	-2	-2	-2	-31
ENMIQA*	-1	-1	-1	2	-2	-1	0	0	1	2	1	-1	-1	2	2	2	0	-2	-2	-2	-2
R18GR50M*	0	0	1	2	0	0	2	2	2	2	2	0	1	2	2	2	2	0	-1	-1	**20**
R50GR18*	2	1	1	2	1	1	1	2	2	2	2	1	1	2	2	2	2	1	0	1	**29**
MR50*	1	1	1	2	1	1	1	2	2	2	2	1	1	2	2	2	2	1	-1	0	**26**

Figure 6. Summary of statistical significance tests on both databases. The approach designed for the evaluation of MR images is indicated with *. The names of the best three methods and sums of their scores are written in bold.

4.4. Computational Complexity

The computational complexity of methods, reflected by the average time taken to assess an image from DB2, was also investigated (Table 2). The time spent on extracting and predicting the quality by a method based on the fusion of networks depends on the number of networks. However, the proposed models are of moderate complexity, being on par with the fastest and more reliable approaches. Further reduction of the computation time can be achieved by a parallel feature extraction process or providing a native implementation (e.g., C++).

Table 2. Run-time comparison. The approach designed for the evaluation of MR images is denoted by *. The names of fusion architectures introduced in this paper are written in italics.

Method	Time (s)
NOMRIQA *	0.1564
HOSA	0.2992
NOREQI	0.1315
IL-NIQE	4.4956
GM-LOG	0.0138
GWHGLBP	0.0336
BRISQUE	0.0232
SISBLIM	0.7821
metricQ	0.1994
BPRI	0.1473
SINDEX	0.0141
NFERM	9.5449
SEER	0.3473
MEON	0.0775
DEEPIQ	1.2746
SNRTOI *	0.0018
ENMIQA *	0.0737
R18GR50M *	0.0293
R50GR18 *	0.0237
MR50 *	0.0226

4.5. Cross—Database Experiments

The performances of R18GR50M, R50GR18, and MR50 are compared with those of related IQA methods in the cross-database experiment. In the experiment, learning-based methods are trained on one database and tested on another. The methods that do not require training are only tested on the second database. The obtained results are shown in Table 3. As reported, the introduced fusion models outperform other techniques and exhibit stable prediction accuracy. Here, the method for IQA of MR images, NOMRIQA, is close to the proposed architectures. The values of performance indices of fusion models trained on the small DB1 and tested on much larger DB2 are only several percent lower than values reported for the DB2 in the first experiment (see Table 1). This confirms their capability of successful extraction of MR image characteristics needed for the quality prediction. Overall, all three introduced architectures provide superior performance, followed by NOMRIQA and SEER.

Table 3. Cross-database performance of twenty evaluated NR approaches.

Method	Training on Database 1 Testing on Database 2				Training on Database 2 Testing on Database 1				Overall			
	SRCC	**KRCC**	**PLCC**	**RMSE**	**SRCC**	**KRCC**	**PLCC**	**RMSE**	**SRCC**	**KRCC**	**PLCC**	**RMSE**
NOMRIQA *	0.7348	0.5280	0.7861	0.5979	**0.6116**	**0.4436**	0.7113	0.5116	0.6732	0.4858	0.7487	0.5548
HOSA	0.7625	0.5612	0.7968	0.5845	0.4550	0.3311	0.6428	0.5574	0.6088	0.4462	0.7198	0.5710
NOREQI	0.7259	0.5312	0.7436	0.6468	0.5082	0.3719	0.7019	0.5183	0.6171	0.4516	0.7228	0.5826
IL-NIQE	0.0050	0.0044	0.1773	0.9520	0.1796	0.1162	0.3465	0.6826	0.0923	0.0603	0.2619	0.8173
GM-LOG	0.7064	0.5134	0.7420	0.6486	0.2721	0.1774	0.1379	0.7207	0.4893	0.3454	0.4400	0.6847
GWHGLBP	0.6247	0.4315	0.6656	0.7220	0.5207	0.3694	0.6189	0.5716	0.5727	0.4005	0.6423	0.6468
BRISQUE	0.6528	0.4640	0.7294	0.6618	0.4895	0.3353	0.6172	0.5725	0.5712	0.3997	0.6733	0.6172
SISBLIM	0.6836	0.5037	0.6746	0.7140	0.2885	0.1820	0.5733	0.5962	0.4861	0.3429	0.6240	0.6551
metricQ	0.4642	0.3271	0.3931	0.8942	0.2300	0.1520	0.2243	0.7091	0.3471	0.2396	0.3087	0.8017
BPRI	0.0747	0.0558	0.4592	0.8593	0.1515	0.1120	0.3440	0.6832	0.1131	0.0839	0.4016	0.7713
SINDEX	0.2807	0.1935	0.3604	0.9024	0.2802	0.1962	0.3307	0.6869	0.2805	0.1949	0.3456	0.7947
NFERM	0.6718	0.4856	0.7240	0.6672	0.4660	0.3536	0.4637	0.6447	0.5689	0.4196	0.5939	0.6560
SEER	0.7855	0.5960	0.8356	0.5314	0.5397	0.4053	**0.7341**	**0.4941**	0.6626	0.5007	0.7849	0.5128
MEON	0.5314	0.3701	0.5148	0.8293	0.1247	0.0771	0.1401	0.7205	0.3281	0.2236	0.3275	0.7749
DEEPIQ	0.3620	0.2528	0.5778	0.7895	0.3030	0.2037	0.4041	0.6656	0.3325	0.2283	0.4910	0.7276
SNRTOI *	0.0681	0.0443	0.1033	0.9622	0.1828	0.1245	0.2262	0.7088	0.1255	0.0844	0.1648	0.8355
ENMIQA *	0.7631	0.5736	0.8040	0.5753	0.3540	0.2428	0.6741	0.5375	0.5586	0.4082	0.7391	0.5564
R18GR50M *	**0.8451**	**0.6574**	**0.8911**	**0.4390**	0.6098	0.4402	0.7231	0.5026	**0.7275**	**0.5488**	**0.8071**	**0.4708**
R50GR18 *	**0.8638**	**0.6684**	**0.8930**	**0.4354**	**0.6163**	**0.4502**	**0.7306**	**0.4968**	**0.7401**	**0.5593**	**0.8118**	**0.4661**
MR50 *	**0.8568**	**0.6709**	**0.8941**	**0.4332**	**0.6299**	**0.4686**	**0.7345**	**0.4938**	**0.7434**	**0.5697**	**0.8143**	**0.4635**

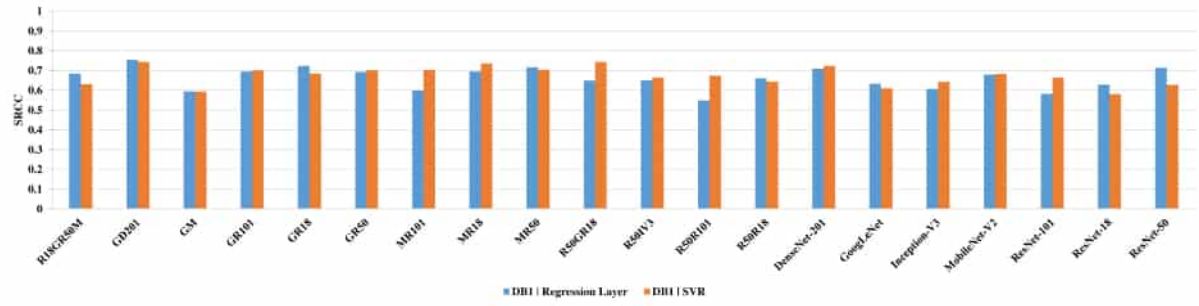
Note: * denotes the approach designed for the evaluation of MR images. The best three results for each criterion are written in bold, the names of fusion architectures introduced in this paper are written in italics.

4.6. Ablation Tests

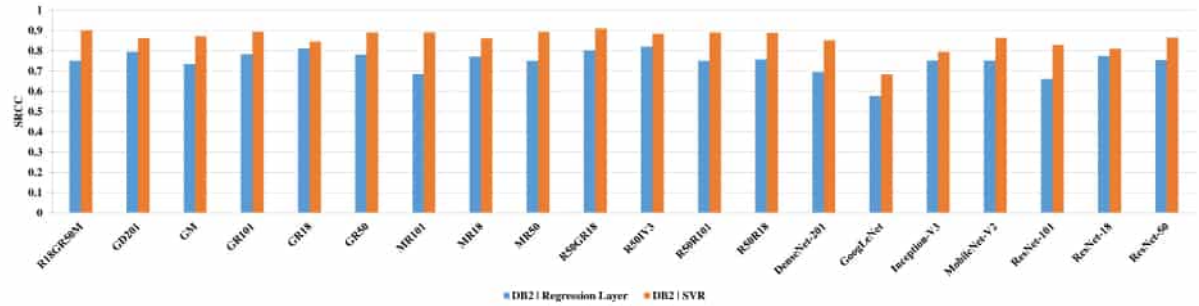
As in the literature many different network architectures have been introduced, in this section, several proposing fusion approaches are reported and discussed. Fur-

thermore, the inclusion of the SVR method is also supported experimentally to show that it improves the results of the networks. In experiments, single deep learning networks or their fusions were considered. The following fusions took part in the study: ResNet-50_GoogLeNet_MobileNet-V2 (R18GR50M), GoogLeNet_DenseNet-201 (GD201), GoogLeNet_MobileNet-V2 (GM), GoogLeNet_ResNet-101 (GR101), GoogLeNet_ResNet-18 (GR18), GoogLeNet_ResNet-50 (GR50), MobileNet-V2_ResNet-101 (MR101), MobileNet-V2_ResNet-18 (MR18), MobileNet-V2_ResNet-50 (MR50), ResNet-50_GoogLeNet_ResNet-18 (R50GR18), ResNet-50_Inception-V3 (R50Iv3), ResNet-50_ResNet-101 (R50R101), and ResNet-50_ResNet-18 (R50R18).

The results presented in Figure 7 reveal that the usage of the SVR module improves the results of the half networks for the DB1 and in all cases for the DB2. Interestingly, most network architectures outperform other state-of-the-art IQA methods on both databases (see Table 1), showing that they can be successfully used for the quality prediction of MR images. However, network architectures that are based on the proposed fusion of single models offer better performance than it can be seen for single networks. Among single architectures, ResNet-50, MobileNet-V2, and DenseNet-201 yield promising results. Therefore, two of them—ResNet-50 and MobileNet-V2—were fused together with ResNet-18 and GoogLeNet obtaining the best performing fusion architectures: (R18GR50M, R18GR50M, and MR50). Here, the fusion with the worst-performing GoogLeNet seems beneficial as its features turned out to be complementary with those of other networks.



(**a**)



(**b**)

Figure 7. Spearman rank-order correlation coefficient (SRCC) performance of compared single and fused networks for the DB1 (**a**) and DB2 (**b**) databases. The results involve quality prediction performed by networks or support vector machine regression (SVR) modules.

5. Conclusions

In this study, a novel no-reference image quality assessment approach for automatic quality prediction of MR images has been presented. In the approach, deep learning architectures are fused, suited to the regression problem, and, after joint transfer learning, their concatenated feature maps are used for quality prediction with the SVR technique. The usage of two or more network architectures, the way they are fused, and their application to the no-reference IQA of MR images are among contributions of this work. Furthermore,

several promising fusion models are proposed and investigated as well as a novel dataset for the development and evaluation of NR methods. The dataset contains 240 distorted images assessed by a large number of experienced radiologists. The comprehensive experimental evaluation of fusion models against 17 state-of-the-art NR techniques, including methods designed for NR IQA of MR images, reveals the superiority of the presented approach in terms of typical performance criteria.

Future work will focus on the investigation of alternative network fusion approaches or developing NR measures for IQA of medical images with different specificity, e.g., CT or RTG.

Author Contributions: Conceptualization, I.S. and M.O.; methodology, M.O.; software, I.S. and M.O.; validation, I.S., R.O., A.P., and M.O.; investigation, I.S., R.O., A.P., and M.O.; writing and editing, I.S., R.O., A.P., and M.O.; data curation: R.O. and A.P.; supervision: R.O., A.P., and M.O. All authors have read and agreed to the published version of the manuscript.

Funding: The research work disclosed in this publication was partially funded by Ultragen, Dabska 18, 30-572 Krakow, Poland.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and the Good Clinical Practice Declaration Statement, and approved by the Local Medical Ethics Committee no. 155/KBL/OIL/2017 dated 22 September 2017.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: http://marosz.kia.prz.edu.pl/fusionMRIQA.html.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Welvaert, M.; Rosseel, Y. On the Definition of Signal-To-Noise Ratio and Contrast-To-Noise Ratio for fMRI Data. *PLoS ONE* **2013**, *8*, e77089. [CrossRef] [PubMed]
2. Yu, S.; Dai, G.; Wang, Z.; Li, L.; Wei, X.; Xie, Y. A consistency evaluation of signal-to-noise ratio in the quality assessment of human brain magnetic resonance images. *BMC Med. Imaging* **2018**, *18*, 17. [CrossRef] [PubMed]
3. Baselice, F.; Ferraioli, G.; Grassia, A.; Pascazio, V. Optimal configuration for relaxation times estimation in complex spin echo imaging. *Sensors* **2014**, *14*, 2182–2198. [CrossRef]
4. Chow, L.S.; Paramesran, R. Review of medical image quality assessment. *Biomed. Signal Process. Control* **2016**, *27*, 148. [CrossRef]
5. Chow, L.S.; Rajagopal, H. Modified-BRISQUE as no reference image quality assessment for structural MR images. *Magn. Reson. Imaging* **2017**, *43*, 74–87. [CrossRef] [PubMed]
6. Jang, J.; Bang, K.; Jang, H.; Hwang, D.; Initiative, A.D.N. Quality evaluation of no-reference MR images using multidirectional filters and image statistics. *Magn. Reson. Med.* **2018**, *80*, 914–924. [CrossRef]
7. Obuchowicz, R.; Oszust, M.; Bielecka, M.; Bielecki, A.; Piórkowski, A. Magnetic Resonance Image Quality Assessment by Using Non-Maximum Suppression and Entropy Analysis. *Entropy* **2020**, *22*, 220. [CrossRef]
8. Oszust, M.; Piórkowski, A.; Obuchowicz, R. No-reference image quality assessment of magnetic resonance images with high-boost filtering and local features. *Magn. Reson. Med.* **2020**, *84*, 1648–1660. [CrossRef]
9. Esteban, O.; Birman, D.; Schaer, M.; Koyejo, O.O.; Poldrack, R.A.; Gorgolewski, K.J. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE* **2017**, *12*, e0184661. [CrossRef]
10. Pizarro, R.A.; Cheng, X.; Barnett, A.; Lemaitre, H.; Verchinski, B.A.; Goldman, A.L.; Xiao, E.; Luo, Q.; Berman, K.F.; Callicott, J.H.; et al. Automated Quality Assessment of Structural Magnetic Resonance Brain Images Based on a Supervised Machine Learning Algorithm. *Front. Neuroinform.* **2016**, *10*, 52. [CrossRef]
11. Kustner, T.; Liebgott, A.; Mauch, L.; Martirosian, P.; Bamberg, F.; Nikolaou, K.; Yang, B.; Schick, F.; Gatidis, S. Automated reference-free detection of motion artifacts in magnetic resonance images. *Magn. Reson. Mater. Phys. Biol. Med.* **2018**, *31*, 243–256. [CrossRef] [PubMed]
12. Sujit, S.J.; Gabr, R.E.; Coronado, I.; Robinson, M.; Datta, S.; Narayana, P.A. Automated Image Quality Evaluation of Structural Brain Magnetic Resonance Images using Deep Convolutional Neural Networks. In Proceedings of the 9th Cairo International Biomedical Engineering Conference (CIBEC), Cairo, Egypt, 20–22 December 2018; pp. 33–36. [CrossRef]
13. Moorthy, A.K.; Bovik, A.C. Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality. *IEEE Trans. Image Process.* **2011**, *20*, 3350–3364. [CrossRef] [PubMed]
14. Saad, M.A.; Bovik, A.C.; Charrier, C. Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain. *IEEE Trans. Image Process.* **2012**, *21*, 3339–3352. [CrossRef] [PubMed]

15. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [CrossRef]
16. Ye, P.; Kumar, J.; Kang, L.; Doermann, D. Unsupervised feature learning framework for no-reference image quality assessment. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1098–1105. [CrossRef]
17. Xu, J.; Ye, P.; Li, Q.; Du, H.; Liu, Y.; Doermann, D. Blind Image Quality Assessment Based on High Order Statistics Aggregation. *IEEE Trans. Image Process.* **2016**, *25*, 4444–4457. [CrossRef]
18. Xue, W.; Mou, X.; Zhang, L.; Bovik, A.C.; Feng, X. Blind Image Quality Assessment Using Joint Statistics of Gradient Magnitude and Laplacian Features. *IEEE Trans. Image Process.* **2014**, *23*, 4850–4862. [CrossRef]
19. Liu, L.; Hua, Y.; Zhao, Q.; Huang, H.; Bovik, A.C. Blind image quality assessment by relative gradient statistics and adaboosting neural network. *Signal Process. Image* **2016**, *40*, 1–15. [CrossRef]
20. Oszust, M. No-Reference Image Quality Assessment with Local Gradient Orientations. *Symmetry* **2019**, *11*, 95, doi:10.3390/sym11010095. [CrossRef]
21. Li, Q.; Lin, W.; Fang, Y. No-Reference Quality Assessment for Multiply-Distorted Images in Gradient Domain. *IEEE Signal Process. Lett.* **2016**, *23*, 541–545. [CrossRef]
22. Oszust, M. No-Reference Image Quality Assessment Using Image Statistics and Robust Feature Descriptors. *IEEE Signal Process. Lett.* **2017**, *24*, 1656–1660. [CrossRef]
23. Zhang, L.; Zhang, L.; Bovik, A.C. A Feature-Enriched Completely Blind Image Quality Evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. [CrossRef] [PubMed]
24. Min, X.; Gu, K.; Zhai, G.; Liu, J.; Yang, X.; Chen, C.W. Blind Quality Assessment Based on Pseudo-Reference Image. *IEEE Trans. Multimed.* **2018**, *20*, 2049–2062. [CrossRef]
25. Bosse, S.; Maniry, D.; Wiegand, T.; Samek, W. A deep neural network for image quality assessment. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3773–3777. [CrossRef]
26. Kim, J.; Lee, S. Fully Deep Blind Image Quality Predictor. *IEEE J. Sel. Top. Signal* **2017**, *11*, 206–220. [CrossRef]
27. Ma, K.; Liu, W.; Liu, T.; Wang, Z.; Tao, D. dipIQ: Blind Image Quality Assessment by Learning-to-Rank Discriminable Image Pairs. *IEEE Trans. Image Process.* **2017**, *26*, 3951–3964. [CrossRef] [PubMed]
28. Zeng, H.; Zhang, L.; Bovik, A.C. A Probabilistic Quality Representation Approach to Deep Blind Image Quality Prediction. *arxiv* **2017**, arxiv:1708.08190.
29. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [CrossRef]
30. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556.
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
32. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.E.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
33. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2818–2826.
34. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *arXiv* **2019**, arXiv:1801.04381.
35. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* **2018**, arXiv:1608.06993.
36. Li, Y.; Ye, X.; Li, Y. Image quality assessment using deep convolutional networks. *AIP Adv.* **2017**, *7*, 125324. [CrossRef]
37. He, Q.; Li, D.; Jiang, T.; Jiang, M. Quality Assessment for Tone-Mapped HDR Images Using Multi-Scale and Multi-Layer Information. In Proceedings of the 2018 IEEE International Conference on Multimedia Expo Workshops (ICMEW), San Diego, CA, USA, 23–27 July 2018; pp. 1–6. [CrossRef]
38. Lin, H.; Hosu, V.; Saupe, D. DeepFL-IQA: Weak Supervision for Deep IQA Feature Learning. *arXiv* **2020**, arXiv:2001.08113.
39. Ieremeiev, O.; Lukin, V.; Okarma, K.; Egiazarian, K. Full-Reference Quality Metric Based on Neural Network to Assess the Visual Quality of Remote Sensing Images. *Remote Sens.* **2020**, *12*, 2349. [CrossRef]
40. Maqsood, M.; Nazir, F.; Khan, U.; Aadil, F.; Jamal, H.; Mehmood, I.; Song, O. Transfer Learning Assisted Classification and Detection of Alzheimer's Disease Stages Using 3D MRI Scans. *Sensors* **2019**, *19*, 2645. [CrossRef] [PubMed]
41. Griswold, M.; Heidemann, R.; Jakob, P. Direct parallel imaging reconstruction of radially sampled data using GRAPPA with relative shifts. In Proceedings of the 11th Annual Meeting of the ISMRM, Toronto, ON, Canada, 10–16 July 2003.
42. Breuer, F.; Griswoldl, M.; Jakob, P.; Kellman, P. Dynamic autocalibrated parallel imaging using temporal GRAPPA (TGRAPPA). *Off. J. Int. Soc. Magn. Reson. Med.* **2005**, *53*, 981–985. [CrossRef] [PubMed]
43. Reykowski, A.; Blasche, M. Mode Matrix—A Generalized Signal Combiner For Parallel Imaging Arrays. In Proceedings of the 12th Annual Meeting of the International Society for Magnetic Resonance in Medicine, Kyoto, Japan, 15–21 May 2004.
44. Deshmane, A.; Gulani, V.; Griswold, M.; Seiberlich, N. Parallel MR imaging. *J. Magn. Reson. Imaging* **2012**, *36*, 55–72. [CrossRef]

45. Sheikh, H.R.; Sabir, M.F.; Bovik, A.C. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Trans. Image Process.* **2006**, *15*, 3440–3451. [CrossRef]
46. Ma, K.; Liu, W.; Zhang, K.; Duanmu, Z.; Wang, Z.; Zuo, W. End-to-End Blind Image Quality Assessment Using Deep Neural Networks. *IEEE Trans. Image Process.* **2018**, *27*, 1202–1213. [CrossRef]
47. Gu, K.; Zhai, G.; Yang, X.; Zhang, W. Using Free Energy Principle For Blind Image Quality Assessment. *IEEE Trans. Multimed.* **2015**, *17*, 50–63. [CrossRef]
48. Zhang, Z.; Dai, G.; Liang, X.; Yu, S.; Li, L.; Xie, Y. Can Signal-to-Noise Ratio Perform as a Baseline Indicator for Medical Image Quality Assessment. *IEEE Access* **2018**, *6*, 11534–11543. [CrossRef]
49. Gu, K.; Zhai, G.; Yang, X.; Zhang, W. Hybrid No-Reference Quality Metric for Singly and Multiply Distorted Images. *IEEE Trans. Broadcast.* **2014**, *60*, 555–567. [CrossRef]
50. Zhu, X.; Milanfar, P. Automatic Parameter Selection for Denoising Algorithms Using a No-Reference Measure of Image Content. *IEEE Trans. Image Process.* **2010**, *19*, 3116–3132. [PubMed]
51. Leclaire Arthur, M.L. No-Reference Image Quality Assessment and Blind Deblurring with Sharpness Metrics Exploiting Fourier Phase Information. *J. Math. Imaging Vis.* **2015**, *52*, 145–172. [CrossRef]
52. Chang, C.C.; Lin, C.J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [CrossRef]

Engineering Applications of Artificial Intelligence 123 (2023) 106283

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

No-Reference Image Quality Assessment of Magnetic Resonance images with multi-level and multi-model representations based on fusion of deep architectures

Igor Stępień [a], Mariusz Oszust [b,*]

[a] Doctoral School of Engineering and Technical Sciences at the Rzeszow University of Technology, al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland
[b] Department of Computer and Control Engineering, Rzeszow University of Technology, Wincentego Pola 2, 35-959 Rzeszow, Poland

ARTICLE INFO

Keywords:
Deep neural networks
Deep features
Image Quality Assessment
Magnetic Resonance images
Quality models
Network fusion

ABSTRACT

Accurate quality assessment of Magnetic Resonance (MR) images is essential for effective medical diagnostics, as it impacts the time spent on image acquisition and image interpretation by radiologists. Therefore, this study aims to provide a novel deep learning-based No-Reference (NR) MR Image Quality Assessment (IQA) method for the quality prediction of MR images. Therefore, in this work, an internal fusion of two complementary deep learning architectures is introduced that offers MR-specific quality-aware features. Apart from obtaining multi-level image representations of joint networks, to further enhance the quality prediction performance, features from re-trained single deep learning architectures are also used. Finally, as one of the main findings of this study, the quality assessment is performed by a high-level quality model trained on scores of quality models obtained for layers of the networks. The superiority of the method against the state-of-the-art techniques is verified by using two publicly available MRIQA benchmarks containing MR images and subjective scores provided by a large number of radiologists. As reported, the method offers superior IQA of MR images as the obtained scores are highly correlated with subjective opinions of medical specialists. It is characterized by the weighted average values of the Spearman Rank-order Correlation Coefficient, Kendall Rank-order Correlation Coefficient, and Pearson Linear Correlation Coefficient of 0.8754, 0.7185, and 0.9062, respectively.

1. Introduction

Magnetic Resonance (MR) imaging supports diagnostics of diseases by offering a visualization of anatomical structures or physiological functions of the human body (Zhu and Qiu, 2021). However, due to the long acquisition time and the way the recording of the energy are performed, it is susceptible to imaging artifacts (Hoff et al., 2016). In MR imaging, there are tissue-, motion-, and technique-related artifacts. For example, the spatial displacement of water and fat molecules, visible as an additional area along boundary edges, is related to differences in resonant frequencies. Also, the flow of body fluids or patient movement causes motion artifacts such as ghosting or blurring, which emerge on scans as additional body areas or blurred image locations. Furthermore, the techniques used for creating MR images may produce scans difficult to interpret due to e.g., overlapped slices in a single acquisition or aliasing. Another source of noise or artifacts is attributed to the magnetic field homogeneity or problems related to the intensity of the signal (Gonzalez-Jaime et al., 2016). In practice, the quality evaluation is performed manually by a technician or radiologist. This procedure leads to the rejection of unsuitable scans (Pontoriero et al., 2021) followed by the repetition of the image acquisition with different parameters of the scanner. Consequently, the time of the procedure is prolonged. Hence, MR image quality assessment (IQA) techniques have been developed to support the automatic evaluation of MR scans and image processing techniques devoted to their restoration and enhancement, or methods improving the MR-based diagnostics. They also aim to replace or support time-consuming manual tests which output can be influenced by a subjective manner of the quality evaluation.

Objective quality assessment methods are divided into Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR) approaches (Okarma, 2019; Yan et al., 2019; Hu et al., 2021; Ullah et al., 2020). The FR techniques are based on the availability of distortion-free pristine images, whereas the RR methods require a part of the data from a reference image for the comparison. However, due to the specificity of MR imaging, the FR and RR methods are used for the assessment of artificially distorted high-quality MR images (Masoudi et al., 2021), while the NR techniques are highly demanded in practice.

Among approaches to NR-IQA of MR images that can be found in the literature, a few adapt solutions devoted to natural images (Chow

* Corresponding author.
E-mail addresses: istepien@kia.prz.edu.pl (I. Stępień), marosz@kia.prz.edu.pl (M. Oszust).
URL: http://marosz.kia.prz.edu.pl (M. Oszust).

https://doi.org/10.1016/j.engappai.2023.106283
Received 26 January 2023; Received in revised form 12 March 2023; Accepted 3 April 2023
Available online 13 April 2023

and Rajagopal, 2017; Jang et al., 2018; Oszust et al., 2020), while other propose MR-specific architectures (Qi et al., 2021; Stepien et al., 2021). However, taking into account the scarcity of IQA methods designed for MR images and the lack of approaches that utilize rich features extracted at different levels of image processing offered by deep learning models that are aware of MR characteristics, in this paper, a novel NR-MRIQA technique is introduced. In the approach, two complementary convolutional networks are internally fused together with single such networks, fine-tuned, and suitably re-trained. It is shown that each network in the fusion is affected by the presence of the second model, while all architectures offer features representing different levels of quality-aware information, from low-level edges to high-level semantics. Contrary to other methods from literature, rich multi-level and multi-model (or multi-architecture) information is extracted and used to train quality models for features of involved networks. This can be regarded as a thorough approach to capturing information from entire networks instead of relying only on their last layers (Stepien et al., 2021). Finally, quality models for layers are treated as perceptual features that train a high-level quality model used to provide the final quality prediction. Since multi-level and multi-model feature vectors extracted from layers are of high-dimensionality, a preselection of features and their subsequent reduction are also introduced.

The main contributions of this work are as follows: (1) Efficient fine-tuning and re-training of internally fused network architectures aiming at effective quality prediction and making the deep learning architectures better transferable to the MR domain. (2) Automatic processing of high-dimensional features captured by networks from numerous layers, i.e., their preselection and extraction. (3) Creation of a high-level quality model using novel secondary training on multi-level multi-model representations build from quality models for preselected and extracted features from layers of the fused networks. (4) Extensive experimental comparison of the introduced approach with more than 15 related methods on two public MRIQA benchmark datasets, including methods devoted to the MRIQA purposes.

The rest of the paper is organized as follows. In Section 2, related works are described, while in Section 3, the proposed method is introduced. Section 4 presents results and analysis of conducted experiments, Section 5 discusses limitations of the study and future directions. Finally, Section 6 concludes the paper.

2. Related works

Despite the importance of the IQA of MR images, expressed in various studies on the interpretation of scans of insufficient quality (Backhausen et al., 2016), there are only a few works devoted to the development of new measures. First such measures adapt IQA popular methods for the IQA of natural images, hoping that they would capture characteristics of MR images and acceptably predict their quality. For example, in the method introduced by Chow and Rajagopal (2017), Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) is trained on MR images instead of natural grayscale images. The method uses the Asymmetric Generalized Gaussian Model (AGGM) distribution to fit Mean Subtracted Contrast Normalized (MSCN) coefficients to capture the severity of image distortions. The parameters of AGGM are used as perceptual features employed to train the Support Vector Regression (SVR) quality model and provide objective scores. Jang et al. (2018) used MSCN multidirectional-filtered coefficients, while Yu et al. (2018) employed a similar approach with the Signal-to-Noise Ratio (SNR) instead of subjective ratings to train four IQA techniques, including the BRISQUE. In the approach of Qi et al. (2021), a 3D spatial-related hyper-network is proposed that takes into account information of neighboring MR scans that create 3D structures. In previous authors' work, local features in MR images filtered using high-boost operators are described to train the SVR model image quality prediction (Oszust et al., 2020). Also, considering other approaches, objective scores are obtained using the entropy characterizing the local intensity extrema (Obuchowicz et al., 2020) or the SVR model created for features extracted from global average layer composed of last layers of up to four internally connected deep learning networks (Stepien et al., 2021).

In the literature, there are methods focused on the usage of various machine learning techniques to extract features and perform the classification of MRI images. They aim to reject low-quality scans, but their usability in the calibration of scanners or the assessment of algorithms that process MR images or signals is limited. For example, Esteban et al. (2017) used a set of simple quality indicators as features used to train a classifier responsible for images that can be further used for diagnostic purposes. In the approach of Kustner et al. (2018), a classification-based NR method is proposed to estimate the quality classes on a 5-point Likert scale. The method uses features from diverse feature pool indicated by an active learning technique. Then, after the dimensionality reduction with Principal Component Analysis (PCA), the extracted features are used to train the Support Vector Machine (SVM) and deep learning classifiers. An IQA of pediatric diffusion MR images is considered in an approach introduced by Liu et al. (2020) that classifies MR images into three classes (i.e., "pass", "fail", and "questionable") using introduced hierarchical non-local residual networks. In other work, Masoudi et al. (2021) transformed the NR problem to the FR domain by creating new images with an unpaired generative adversarial network and using deep learning-based FR technique to classify images into three categories. Three image categories are also used by an approach of Gupta et al. (2020) in which a dense neural network is trained with 64 features. It is worth noticing that those classifiers of MR images cannot be used to develop image enhancement or restoration techniques since the binary output, or even the assignment of labels out of the 5-point Likert scale, cannot compare methods with similar outputs in a way it can be achieved with a regression-based technique. Hence, as it is present in the mature field of the IQA of natural images and dependent research, regression-based MRIQA approaches are desirable.

The scarcity of MR IQA approaches, involving methods that efficiently extract information from the networks, lead to emergence of the method introduced in this paper. The approach extracts rich multi-level and multi-model information of fused and jointly trained networks and introducing high-level model allows for superior MR image quality prediction.

3. Methods

The proposed technique, **M**ulti l**E**vel and multi-model **D**eep **Q**uality **E**valuator of **MR I**mages (MEDQEMRI), efficiently fuses two deep learning networks of different complexities that belong to the same network family, i.e., ResNet18 and ResNet50 (He et al., 2015). Both models are often used to capture various image characteristics in state-of-the-art methods from computer vision field based on deep learning architectures and are released with pre-trained implementation ready to perform the image classification task (He et al., 2015). However, to obtain MR-related quality-aware features, they are fine-tuned in this work to address the considered regression problem. Then, the networks are internally connected and the resulted fusion architecture is re-trained. Since the employed networks are designed to perform object recognition tasks, their single performance in MRIQA may be limited due to the used architecture and the specificity of MR images. The proposed fusion increases the discriminative power of their features by making a model aware of the second network while training. Consequently, the fusion enriches the quality-aware information that can be extracted from networks. Furthermore, to improve the results, the effective re-training of the fusion architecture is employed. The pipeline of the method is presented in Fig. 1. As shown, an input MR image is fed into trained ResNet18 (R18), ResNet50 (R50), and their fusion (R18R50). Then, extracted rich features from a layer of each network are preselected, reduced, and used as input to a quality model. These layer-based quality prediction models reflect the hierarchical

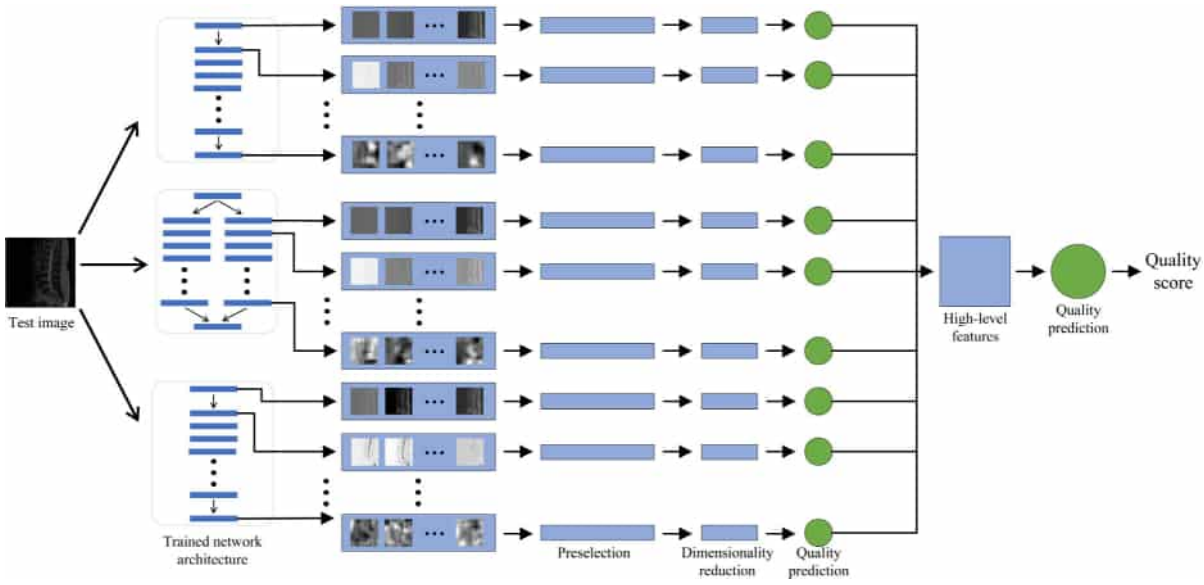


Fig. 1. Pipeline of the proposed approach. High-dimensional features of re-trained architectures (R18, R18R50, and R50) are preselected and reduced to create quality models. Then, their scores are used as high-level features to train a model that provides final quality prediction.

structure of the networks and are based on multi-level multi-model image representations. Finally, objective scores from the quality models are seen together as a perceptual feature vector or high-level features used to predict image quality with a high-level quality model.

3.1. Network fusion

In this paper, two ResNet networks are fused to address the MRIQA problem. They are employed due to their efficiency in image classification, confirmed by the outstanding results in the ILSVRC 2015 classification challenge (He et al., 2015), and complementarity of implementations. Since there are many variants of ResNet networks with a different number of layers, the ResNet18 and ResNet50 are employed taking into account their complexity and performance in classification tasks that require high discriminative power. The ResNet architecture uses shortcut connections to skip two or three layers containing ReLU and batch normalization, addressing the problem of vanishing gradients that causes an increase of the training error with a growing number of layers. The identity connections used for skipping the layers allow for training deeper networks while maintaining the same computation complexity as simple deep networks. Initially, the architecture performs the convolution (7×7) and max-pooling (3×3) with the stride of 2. Then, depending on the number of layers, four stacks of residual blocks that contain two (R18) or three layers (R50) are employed (He et al., 2015). Finally, the average pooling and fully connected layer are used as the output of the network.

In this work, to train R18, R50, and their fusion architecture on MR images, fine-tuning should consider the regression problem. Therefore, the last three layers of each network are replaced with a fully connected layer and the regression layer. In the case of the internally fused networks, a feature concatenation layer is added to combine outputs from the last layers of the networks in an element-wise manner. Then, to make the networks aware of MR characteristics, after the transfer-learning of the layers left in original networks by the fine-tuning, the weights are not frozen as it can be seen in a typical transfer-learning approach, but the architectures are trained for a small number of epochs to update them. Such re-training combines the fine-tuning of the pre-trained network with early stopping to make the models better transferable to the MR domain (Azizpour et al., 2015; Tajbakhsh et al., 2016).

It is worth noticing that the MR scans are two-dimensional 16-bit matrices. Hence, to use the ResNet architectures that process 224×224 RGB images, scans are resized, reduced to 8-bit matrices, and concatenated to form three channels.

The applicability of the fusion of the networks and their separately re-trained architectures for the MRIQA is justified by the sensitivity of the considered architectures to the image distortion severity. For example, as shown in Fig. 2 with the visualization of several layers (features), features of separate and fused networks indicate images of different perceptual quality. This observation supports their usage in the quality prediction as they provide discriminative description of assessed MR images of different quality. Interestingly, it can be seen that values in layers of both networks in the fusion are visibly different from their separately trained versions, highlighting the mutual influences of the networks. Consequently, the three considered network architectures (i.e., R18, R50, and R18R50) complement each other, resulting in outstanding MRIQA performance of the obtained quality model (see Section 4.3).

3.2. Deep features

The feature vector for a layer in one of the network architectures contains values resulting from the network training. It is subjected to a preselection performed to obtain a vector furtherly used for creating the quality prediction model for the layer. This step reduces the dimensionality of the vector by the factor of ten in the case of the first layers of the ResNet architectures. Contrary to the approaches of Kang et al. (2014) or Gao et al. (2018) to the NRIQA of natural images with simply extracted features of the pre-trained VGG16 network, feature maps that belong to a layer are not processed separately by using their minimum and maximum values but processed jointly, assuming that richer layer representation would provide more distinctive quality models. Note that the first convolutional layer of ResNet18, or ResNet50, contains 802 816 values (64 feature maps), and they store in each layer about 120k and 200k features on average, respectively. Such large dimensionality prohibits the creation of the quality model. Therefore, after the preselection and further extraction, a compact feature vector is obtained.

Let V_l denote a feature vector of the lth layer of a network, with $l = 1, 2, \ldots, L$, where L is the number of layers. Hence, $V_l = [V_{l,1}, V_{l,2}, \ldots, V_{l,N_l}]$, where the N_l denotes the length of the vector of lth layer. Taking into account high dimensionality of the V_l, expressed by the large differences in $\{N_1, N_2, \ldots, N_L\}$, the V_l is reordered, i.e., the K of its dimensions, $K \leq N_l$, are selected for further processing

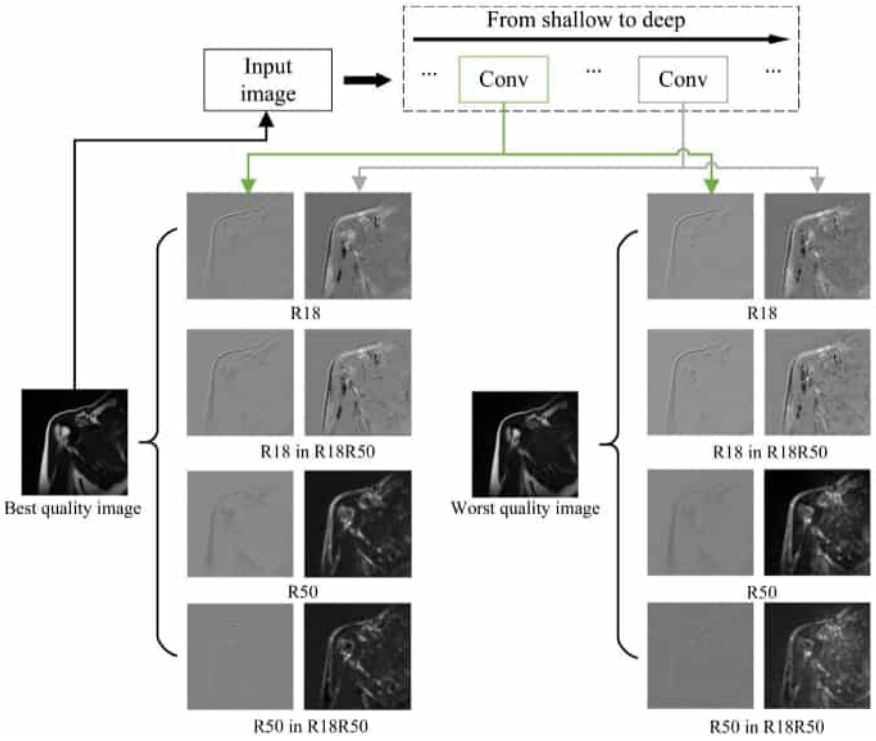


Fig. 2. Visualization of differences in features of three network architectures (R18, R50, and their fusion R18R50) at two exemplary layers for MR images of different perceptual quality.

based on their standard deviation calculated for a set of training MR scans. Consequently, a set of K indices ($\mathbf{k} = \{k_1, k_2, \ldots, k_K\}$, $\mathbf{k} \subset [1, 2, \ldots, L]$) is obtained. Then, the $V_l(\mathbf{k}) = [V_{l,k_1}, V_{l,k_2}, \ldots, V_{l,k_K}]$ is reduced by the Principal Component Analysis (PCA) (Jolliffe, 2011) due to the high number of preselected features ($V_l(\mathbf{k})$). Furthermore, they may contain redundant and correlated information. Consequently, for a given layer, the PCA method extracts R-dimensional feature vectors V^R from the $V_l(\mathbf{k})$, based on the projection of their set for training images onto R-dimensional subspace, spanned by the R dominant eigenvectors and calculated for the covariance matrix. Finally, the resultant V_l^R is used as a quality-aware vector to train the quality model for the lth layer. For simplicity, in further parts of this paper, the V_l^R will be denoted as V.

3.3. Quality models

For the lth layer, a quality model Q_l is obtained using the popular SVR technique with the radial basis function (Xue et al., 2014). It is used in this work due to its ability to create reliable quality prediction models even with deep learning features (Gao et al., 2018). The used ε-SVR maps feature vector for an image (V) into its subjective score (S). Given the training data ($\mathbf{V}, \mathbf{S}$), where $\mathbf{V}$ denotes reduced feature vectors of lth layer of M training images ($V_1, V_2, \ldots, V_M$) and S contains their subjective scores in a form of Mean Opinion Scores (MOS) ($S_1, S_2, \ldots, S_M$), a function $\mathfrak{f}(V) = \langle \omega, V \rangle + b$ is determined in which $\langle \cdot, \cdot \rangle$, ω, and b, are the inner product, weight vector, and a bias parameter, respectively. Once the slack variables ξ_m and ξ_m^* are introduced, the ω and b can be computed by solving the following optimization problem:

$$\text{minimize } \frac{1}{2}\|\omega\|^2 + C\sum_{m=1}^{M}(\xi_m + \xi_m^*)$$
$$subject\ to \begin{cases} \langle \omega, V_m \rangle - (S_m - b) \le \varepsilon + \xi_m \\ S_n - b - \langle \omega, V_m \rangle \le \varepsilon + \xi_m^* \\ \xi_m, \xi_m^* \ge 0, \end{cases} \tag{1}$$

where C balances ω and ξ_m and ξ_m^*. The $\omega = \sum_{m=1}^{M} t_m V_m$, where t_m is a combination coefficient. After mapping the V into high-dimensional space $\Phi(V)$,

$$\mathfrak{f}(V) = \left\langle \sum_{m=1}^{M} t_m \Phi(V_m), \Phi(V) \right\rangle + b = \sum_{m=1}^{M} t_m \langle \Phi(V_m), \Phi(V) \rangle + b. \tag{2}$$

Furthermore, since the RBF is employed,

$$\mathfrak{f}(V) = \sum_{m=1}^{M} t_m exp(-\gamma(|V_m - V|)^2) + b, \tag{3}$$

where, the γ is the precision parameter (Xue et al., 2014). The parameters of the SVR are typically determined using a grid search.

The multi-level representation of the assessed image is reflected by the obtained quality models for the L layers of deep learning architectures (multi-model). However, to provide a global quality score, contrary to popular average pooling (Gao et al., 2018), a high-level quality model is created using the scores for layers. Specifically, the perceptual vector of an image is built from an ensemble of quality models, i.e., the vector $V = [Q_1^{R18}, Q_2^{R18}, \ldots, Q_{L=len(R18)}^{R18}, Q_1^{R50}, Q_2^{R50}, \ldots, Q_{L=len(R50)}^{R50}, Q_1^{R18R50}, Q_2^{R18R50}, \ldots, Q_{L=len(R18R50)}^{R18R50}]$, where $len(*)$ indicates the number of layers in the network, is mapped into its subjective score (S). Finally, taking into account that multi-level quality scores can come from jointly and separately trained ResNet architectures, the objective quality of the image (Q) is calculated.

4. Results and discussion

4.1. Benchmark databases

In this work, experiments are conducted on two publicly available MR image benchmark databases. They contain scans from 1.5T MR T2-weighted sequences of different quality and subjective scores obtained

Table 1
Characteristics of MRIQA benchmark datasets used in experiments.

	Dataset	
	DB1 (Oszust et al., 2020)	DB2 (Stepien et al., 2021)
No. of images	70	240
Locations	Spine, Knee, Shoulder, Brain, Wrist, Hip, Pelvis, Elbow, Ankle	Spine, Knee, Shoulder
No. of radiologists	31	24
Image resolution	From 204 × 256 to 512 × 512	320 × 320, 384 × 384

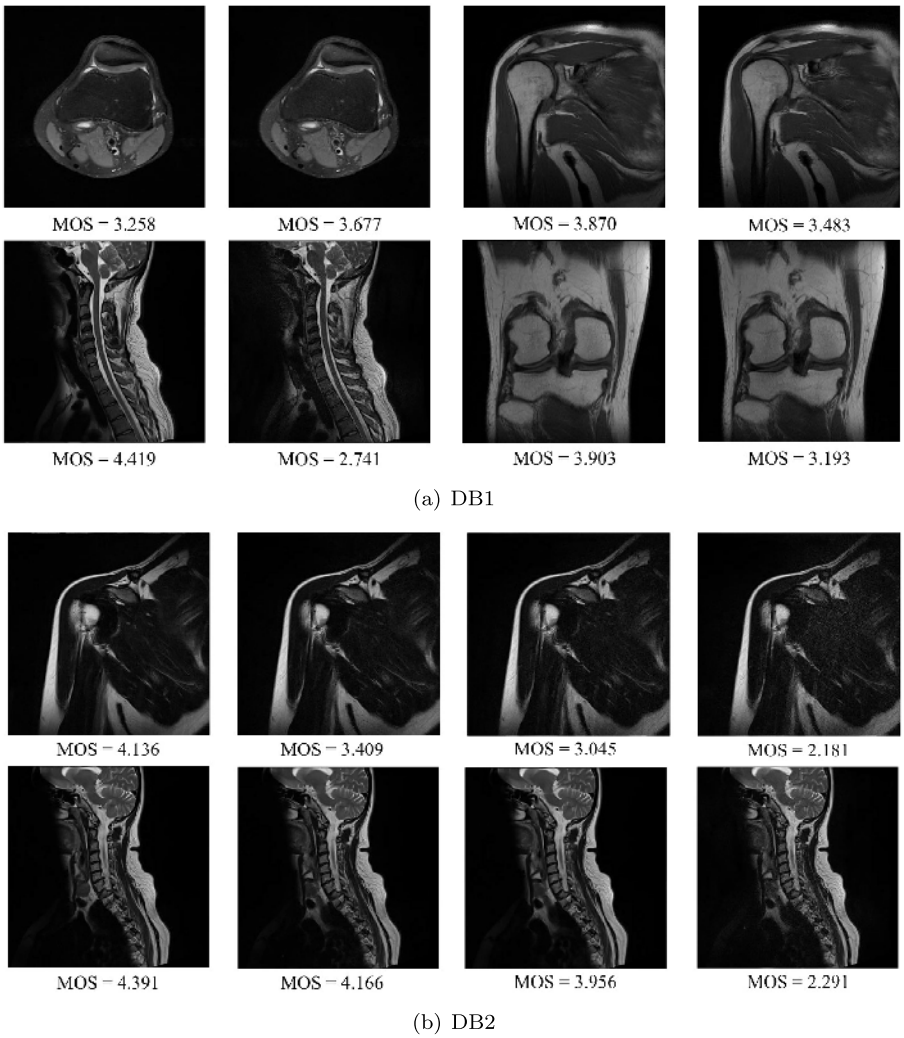


Fig. 3. Sample images and their subjective opinions from used MR image benchmarks.

in tests with radiologists (i.e., MOS). The lower quality images in those databases resulted from shortened sequences acquired using Process Analytical Technology (PAT) I software (Siemens) with the GeneRalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) (Breuer et al., 2005). The databases are the largest publicly available MR benchmarks that contain MOS values obtained in subjective tests with a large group of radiologists (Oszust et al., 2020; Stepien et al., 2021). The first benchmark (denoted in this work as DB1) contains 70 MR images assessed by 31 radiologists (Oszust et al., 2020), while the second image collection (DB2) contains 240 images assessed by 24 radiologists (Stepien et al., 2021). Both databases are composed of images of different parts of the body. It is worth noticing that, contrary to image benchmarks used for the development of the IQA methods for natural images, the DB1 and DB2 contain scans with various resolutions. They are characterized in Table 1, while exemplary images of different quality and their MOSs are shown in Fig. 3.

4.2. Experimental protocol

To evaluate the MEDQEMRI and related methods, four commonly used evaluation criteria are employed: Spearman Rank-order Correlation Coefficient (SRCC), Kendall Rank-order Correlation Coefficient (KRCC), Pearson Linear Correlation Coefficient (PLCC), and Root Mean Square Error (RMSE) (Sheikh et al., 2006). Higher values of SRCC, KRCC, and PLCC mean a better agreement between the output of the presented approach and the subjective opinions of specialists, expressed by the MOS. However, in the case of RMSE, the best performance is indicated by the lower values of this criterion. In this work, a widely accepted protocol for evaluating related methods is employed, in which 80% of randomly selected images from a dataset are used for training the network, and the remaining 20% of images test it (Min et al., 2018; Ye et al., 2012). Furthermore, it is ensured that testing and training

Table 2
Performance comparison of NR methods on benchmark databases.

Method	DB1				DB2				Overall weighted			
	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
NOMRIQA*°	0.7030	0.5527	0.7978	0.4322	0.8040	0.6087	0.8737	0.4605	0.7812	0.5961	0.8566	0.4541
HOSA°	0.4804	0.3909	0.6997	0.5318	0.8756	0.7052	**0.9276**	**0.3388**	0.7864	0.6342	0.8761	0.3824
IL-NIQE	0.1695	0.1275	0.3619	0.6674	0.1197	0.0836	0.3090	0.8821	0.1309	0.0935	0.3209	0.8336
GM-LOG°	0.4673	0.3424	0.6515	0.4779	0.8854	0.7123	0.9010	0.4091	0.7910	0.6288	0.8447	0.4246
GWHGLBP°	0.5075	0.3935	0.6886	0.5257	0.8726	0.6927	0.8947	0.4080	0.7902	0.6251	0.8482	0.4346
BRISQUE°	0.4610	0.3648	0.6100	0.5311	0.8544	0.6738	0.8951	0.4076	0.7656	0.6040	0.8307	0.4355
SISBLIM	0.3976	0.2776	0.6240	0.5449	0.7216	0.5419	0.7592	0.6047	0.6484	0.4822	0.7287	0.5912
metricQ	0.2596	0.1657	0.2792	0.6709	0.5066	0.3701	0.5227	0.7791	0.4508	0.3239	0.4677	0.7547
BPRI	0.2412	0.1890	0.4785	0.5756	0.1317	0.0973	0.4928	0.7883	0.1564	0.1180	0.4896	0.7403
BLINDER°	0.2756	0.1985	0.5122	0.5930	0.7037	0.5138	0.7379	0.6262	0.6070	0.4426	0.6869	0.6187
MEON	0.2518	0.1879	0.3439	0.6428	0.5851	0.4001	0.6194	0.7426	0.5098	0.3522	0.5572	0.7201
DEEPIQ	0.1133	0.0827	0.5902	0.5707	0.2837	0.2078	0.5393	0.7822	0.2452	0.1796	0.5508	0.7344
SNRTOI*	0.1321	0.0728	0.4094	0.6784	0.1016	0.0720	0.3169	0.8930	0.1085	0.0722	0.3378	0.8445
ENMIQA*	0.3630	0.2479	0.5093	0.5873	0.7941	0.6119	0.8313	0.5130	0.6968	0.5297	0.7586	0.5298
M1*°	0.6299	0.5012	0.7999	0.4381	**0.8998**	**0.7398**	0.9270	0.3465	0.8389	**0.6859**	0.8983	0.3672
M2*°	**0.7036**	**0.5740**	**0.8576**	**0.3865**	0.8919	0.7176	0.9241	0.3560	**0.8494**	0.6852	**0.9091**	**0.3629**
M3*°	0.6634	0.5194	**0.8299**	**0.4094**	0.8769	0.6999	0.9100	0.3902	0.8287	0.6591	0.8919	0.3945
MEDQEMRI*°	**0.7455**	**0.5857**	0.8204	0.4300	**0.9133**	**0.7573**	**0.9312**	**0.3386**	**0.8754**	**0.7185**	**0.9062**	**0.3592**

Note: Two best results for each criterion are written in bold. Methods designed for MRIQA and trained on MR images are indicated with * and °, respectively.

images do not share images from the same body location. The performance indices are reported as median values over ten training–testing iterations.

4.3. Comparison of NR methods

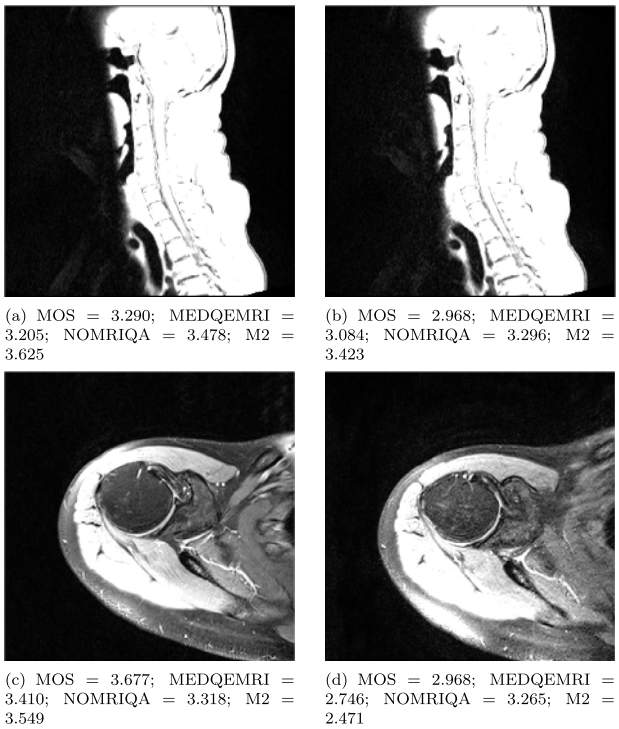
The introduced MEDQEMRI is experimentally compared with the following state-of-the-art techniques with publicly available source codes: NOMRIQA (Oszust et al., 2020), HOSA (Xu et al., 2016), IL-NIQE (Zhang et al., 2015), GM-LOG (Xue et al., 2014), GWHGLBP (Li et al., 2016), BRISQUE (Mittal et al., 2012), metricQ (Ilsche et al., 2019), SISBLIM (Gu et al., 2014), BPRI (Min et al., 2018), MEON (Ma et al., 2018), DEEPIQ (Ma et al., 2017), SNRTOI (Zhang et al., 2018), ENMIQA (Obuchowicz et al., 2020), BLINDER (Gao et al., 2018), M1 (ResNet18, Googlenet, Resnet50, and MobileNet) (Stepien et al., 2021), M2 (Resnet50 and MobileNet) (Stepien et al., 2021), and M3 (ResNet18 and Resnet50) (Stepien et al., 2021). The NOMRIQA, ENMIQA, M1-3, SNRTOI, and MEDQEMRI are methods dedicated to MR images. The MEON, DEEPIQ, BLINDER, and M1-3 are deep learning approaches. However, the MEON and DEEPIQ do not contain training source code and are pre-trained on natural images. Consequently, they assess images similarly to the methods that does not require training, such as IL-NIQE, metricQ, SISBLIM, BPRI, SNRTOI, or ENMIQA. In the case of BLINDER, its implementation created by the authors of this paper is used. In experiments, all methods are run in Matlab with their default parameters, while the SVR parameters, if needed, are determined using a grid search. Since some methods process color images, the MR images are merged to form three channels to facilitate their usage. The approach introduced in this work is trained with the Stochastic Gradient Falling (SGDM) (Ruder, 2017) using the L_2 loss, a learning rate of 10e−4, mini-batch size of 32, and five epochs. The feature vectors in layers of size larger than $K = 100,000$ values are subjected to the preselection leaving K features, while the PCA further extracts $R = 75$ of them. The value of the K is set taking into account the memory limits of the PC that run the computations. Since the DB1 contains a small number of MR images, to provide efficient training, a data augmentation procedure is applied in which each training image is rotated to 360° with the step of 3° (Stepien et al., 2021). The approaches are run in Matlab R2021a, Windows 10, on a PC with an i7-7700k CPU, 32 GB RAM, and a GTX 1080 Ti graphic card.

The results of compared methods are presented in Table 2. The proposed MEDQEMRI outperforms related techniques on the DB1, taking into account the SRCC and KRCC criteria and is after M2 with M3 for the PLCC and RMSE. For the larger DB2 database, the MEDQEMRI exhibits superior performance indices, followed by the M1 (SRCC, KRCC) and HOSA (PLCC, RMSE). The overall weighted results favor the MEDQEMRI and place the M2 as the second-best technique. It can be seen that in the case of methods M1-3 that use raw features from networks in the SVR model, the selection of the best model is difficult due to uneven performance of the M1, with inferior results on the DB1 and acceptable performance on the DB2. It seems that the junction of three models (M1) is more beneficial than the usage of two network architectures (M2, M3) for the MRIQA in larger databases with such an approach (Stepien et al., 2021). Interestingly, the BLINDER in which simplified information from all layers of the VGG16 architecture is employed, shows inferior performance on the DB1, on par with methods that do not require training. Its results on the DB2 are better, however, it also is not able to accurately capture the influence of MR image distortions and their severity on the subjective perception of radiologists. Similarly, popular NR methods designed for natural images and trained on MR scans (e.g., HOSA, BRISQUE, GWHGLBP, or GM-LOG) yield promising performance on the DB2 but cannot acceptably assess images from the smaller DB1 benchmark. It is worth noticing that the SNRTOI, related to the SNR, which is often used in practice to characterize MR scans, poorly correlates with opinions of human observers, justifying the need for the development of new MRIQA approaches.

To determine whether the obtained results are statistically significant, the Wilcoxon rank-sum test is used that measures the equivalence of the median values of independent samples with a 5% significance level (Stepien et al., 2021). In the test, the method obtained a score of "1" in the case the median of SRCC values is significantly greater than those of the compared method. The worse method obtains "−1" and indistinguishable approaches are assigned "0". Consequently, the scores for both databases are added and displayed as cells in Fig. 4. Furthermore, the figure contains an additional column with sums of scores in cells to indicate the best solution in overall comparison. Among M1-M3 approaches, only M1 variant is considered due to its better performance on the second dataset. It can be noticed from the summary that the introduced MEDQEMRI outperforms the remaining methods taking into account the number of overall comparisons with statistically better results, with the M1 approach as the second-best technique. M1 is unable to outperform approaches with hand-crafted features (GM-LOG, GWHGLBP, and HOSA), while MEDQEMRI obtains statistically better results in comparison with them on at least one database (GM-LOG, GWHGLBP) or both databases (HOSA). Considering other NR methods designed for the assessment of MR images,

	BLINDER	BPRI	BRISQUE	DEEPIQ	ENMIQA	GM-LOG	GWHGLBP	HOSA	IL-NIQE	M1	MEON	NOMRIQA	SISBLIM	SNRTOI	metricQ	**MEDQEMRI**	SUM
BLINDER	0	1	-1	1	0	-1	-1	-1	2	-2	2	-2	0	2	1	-2	-1
BPRI	-1	0	-2	-1	-2	-2	-2	-2	0	-2	-1	-2	-1	0	-1	-2	-21
BRISQUE	1	2	0	2	0	0	0	0	2	-2	2	-1	1	2	2	-2	9
DEEPIQ	-1	1	-2	0	-2	-2	-2	-2	1	-2	-1	-2	-1	1	-1	-2	-17
ENMIQA	0	2	0	2	0	-2	-1	-1	2	-2	2	-1	0	2	1	-2	2
GM-LOG	1	2	0	2	2	0	0	0	2	0	2	1	1	2	2	-1	16
GWHGLBP	1	2	0	2	1	0	0	0	2	0	2	0	1	2	2	-1	14
HOSA	1	2	0	2	1	0	0	0	2	0	2	0	1	2	2	-2	13
IL-NIQE	-2	0	-2	-1	-2	-2	-2	-2	0	-2	-1	-2	-2	0	-1	-2	-23
M1	2	2	2	2	2	0	0	0	2	0	2	1	2	2	2	0	21
MEON	-2	1	-2	1	-2	-2	-2	-2	1	-2	0	-2	-2	1	0	-2	-16
NOMRIQA	2	2	1	2	1	-1	0	0	2	-1	2	0	2	2	2	-1	15
SISBLIM	0	1	-1	1	0	-1	-1	-1	2	-2	2	-2	0	2	1	-2	-1
SNRTOI	-2	0	-2	-1	-2	-2	-2	-2	0	-2	-1	-2	-2	0	-1	-2	-23
metricQ	-1	1	-2	1	-1	-2	-2	-2	1	-2	0	-2	-1	1	0	-2	-13
MEDQEMRI	2	2	2	2	2	1	1	2	2	0	2	1	2	2	2	0	25

Fig. 4. Summary of statistical significance tests over two databases. A significantly better approach in the row that in the column obtains the score of '1', while worse and indistinguishable obtains '−1' and '0', respectively. Scores are summed for both databases and summarized in the last column.



(a) MOS = 3.290; MEDQEMRI = 3.205; NOMRIQA = 3.478; M2 = 3.625

(b) MOS = 2.968; MEDQEMRI = 3.084; NOMRIQA = 3.296; M2 = 3.423

(c) MOS = 3.677; MEDQEMRI = 3.410; NOMRIQA = 3.318; M2 = 3.549

(d) MOS = 2.968; MEDQEMRI = 2.746; NOMRIQA = 3.265; M2 = 2.471

Fig. 5. Exemplary images with opinion scores and corresponding scores predicted by three best methods on DB1.

NOMRIQA, being learning-based approach, much better predicts image quality than simpler ENMIQA or popular SNRTOI.

To further show the scores produced by the compared methods, the three best approaches designed for NR MRI (MEDQEMRI, M2, and NOMRIQA) are compared on exemplary images from the DB1 dataset (Fig. 5). Apart from obtained scores, the figure also contains the MOS values provided for the images in the dataset. As reported, all three methods provided similar ranges of values in comparison with MOS for (Fig. 5ab) with MEDQEMRI offering closer results. However, in the second case, MEDQEMRI lowers the quality preserving the range while the remaining methods narrow the range of predicted scores (NOMRIQA) or significantly lower the score of the second image from the pair (Fig. 5cd).

4.4. Cross-database evaluation

To determine the generalization capability of the introduced approach and compare it with those of the state-of-the-art methods, the results of cross-database experiments are reported. As reported (Table 3), MEDQEMRI maintains stable performance across databases, outperforming the remaining NR methods. It is followed by the M1-3 models and NOMRIQA which also are designed for the MRIQA. Among the methods for the IQA natural images, the BRISQUE and GWHGLBP exhibit relatively good performance. To summarize, MEDQEMRI is insensitive to the employed MR image databases and can be successfully used for quality prediction.

4.5. Ablation tests

4.5.1. Feature extraction

To examine the effect of the length of the feature vector extracted by the PCA (R) on the results, experiments that involve the ResNet50 network and DB2 are carried out. The length of the resulted feature vector ranges from 5 to 150. As reported in Fig. 6, the values greater than 75 do not improve the results obtained for the network. Therefore, the $R = 75$ is used by default as a perceptual vector (V) to train quality models.

4.5.2. Multi-level quality models

Once the PCA extracts R-dimensional vectors for the considered layers of trained networks, it is worth examining the influence of each layer on the results. Therefore, the PLCC performance of layers of the R18R50 network on DB2 is presented in Fig. 7. Despite promising values of some layers, it is difficult to indicate and justify a possible usage of only some of them, taking into account that the PLCC value for the high-level model with all layers is 0.9327. Hence, all objective scores produced by quality models for layers constituting multi-level representations are further used in the ensemble to provide a quality score of the approach. The multi-level layers provide different outlooks on the image, from low-level features to information associated with their patterns or even semantics, expressed by further layers. The reported performance of the high-level model of the R18R50 architecture (Fig. 7) is greater than those of separate layers, evidencing the complementary role of layers that took part in the final model.

4.5.3. Separate and fused networks

The introduced MEDQEMRI is based on two ResNet architectures, i.e., ResNet18 and ResNet50. Hence, to indicate their influence on the results and justify the employed usage of several architectures, the quality prediction performance of involved models on DB1 and DB2 is reported. Table 4 shows the advantage of combined networks over single network architectures. The fusion of networks (R18R50) performs better than separately trained architectures (R18+R50) or any of the state-of-the-art methods (see Table 2). Despite promising correlation values for a single benchmark database, only the conjunction of separately trained networks with their fusion offers the overall best performance, supporting the way the networks are managed in this work (R18+R50+R18R50). Interestingly, the addition of the separately

Table 3
Cross-database performance of evaluated NR approaches.

Method	Training on DB1 Testing on DB2				Training on DB2 Testing on DB1				Overall weighted			
	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
NOMRIQA*°	0.7348	0.5280	0.7861	0.5979	0.6116	0.4436	0.7113	0.5116	0.6394	0.4627	0.7282	0.5311
HOSA°	0.7625	0.5612	0.7968	0.5845	0.4550	0.3311	0.6428	0.5574	0.5244	0.3831	0.6776	0.5635
IL-NIQE	0.0050	0.0044	0.1773	0.9520	0.1796	0.1162	0.3465	0.6826	0.1402	0.0910	0.3083	0.7434
GM-LOG°	0.7064	0.5134	0.7420	0.6486	0.2721	0.1774	0.1379	0.7207	0.3702	0.2533	0.2743	0.7044
GWHGLBP°	0.6247	0.4315	0.6656	0.7220	0.5207	0.3694	0.6189	0.5716	0.5442	0.3834	0.6294	0.6056
BRISQUE°	0.6528	0.4640	0.7294	0.6618	0.4895	0.3353	0.6172	0.5725	0.5264	0.3644	0.6425	0.5927
SISBLIM	0.6836	0.5037	0.6746	0.7140	0.2885	0.1820	0.5733	0.5962	0.3777	0.2546	0.5962	0.6228
metricQ	0.4642	0.3271	0.3931	0.8942	0.2300	0.1520	0.2243	0.7091	0.2829	0.1915	0.2624	0.7509
BPRI	0.0747	0.0558	0.4592	0.8593	0.1515	0.1120	0.3440	0.6832	0.1342	0.0993	0.3700	0.7230
BLINDER°	0.2541	0.1705	0.3271	0.9142	0.4145	0.2936	0.542	0.6115	0.3783	0.2658	0.4935	0.6799
MEON	0.5314	0.3701	0.5148	0.8293	0.1247	0.0771	0.1401	0.7205	0.2165	0.1433	0.2247	0.7451
DEEPIQ	0.3620	0.2528	0.5778	0.7895	0.3030	0.2037	0.4041	0.6656	0.3163	0.2148	0.4433	0.6936
SNRTOI*	0.0681	0.0443	0.1033	0.9622	0.1828	0.1245	0.2262	0.7088	0.1569	0.1064	0.1984	0.7660
ENMIQA*	0.7631	0.5736	0.8040	0.5753	0.3540	0.2428	0.6741	0.5375	0.4464	0.3175	0.7034	0.5460
M1*°	0.8451	0.6574	0.8911	0.4390	0.6098	0.4402	0.7231	0.5026	0.6629	0.4892	0.7610	0.4882
M2*°	**0.8568**	**0.6709**	**0.8941**	**0.4332**	**0.6299**	**0.4686**	**0.7345**	0.4938	**0.6811**	**0.5143**	**0.7705**	0.4801
M3*°	0.8358	0.6490	0.8781	0.4550	0.5592	0.4133	0.7225	**0.4845**	0.6217	0.4665	0.7576	**0.4778**
MEDQEMRI*°	**0.8584**	**0.6715**	**0.8970**	**0.4277**	**0.6465**	**0.4827**	**0.7581**	**0.4746**	**0.6943**	**0.5254**	**0.7894**	**0.4640**

Note: Two best results for each criterion are written in bold. Methods designed for MRIQA and trained on MR images are indicated with * and °, respectively.

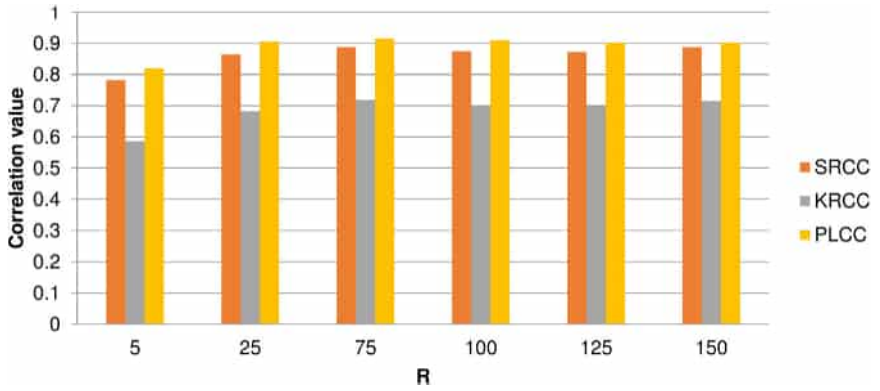


Fig. 6. Influence of the length of the feature vector on the results of the ResNet50 network.

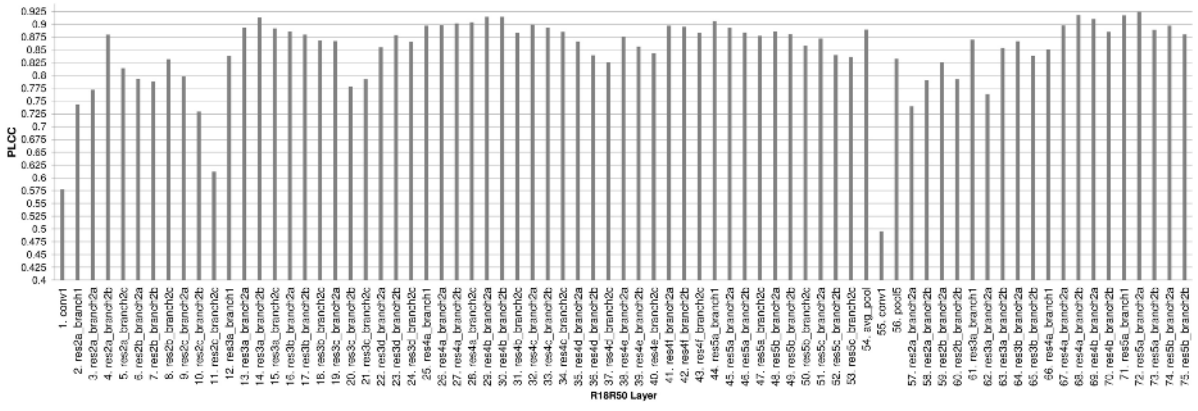


Fig. 7. PLCC performance of quality models of R18R50 architecture on DB2.

trained R18 to the R18R50 (R18+R18R50) achieves better SRCC for the DB1 than it can be observed in the case of the R50 (R50+R18R50), while the R50 alone is much better than the R18 on this benchmark (0.7408 vs. 0.6432). This indicates complementarity of the R18 with features of R18R50 and supports the multi-model approach. The table also contains the results for the R18 and R50 networks trained in the end-to-end fashion. They are worse than those obtained with the help of the presented approach by a large margin due to the design of both networks to perform object recognition tasks and the relatively small size of the available MRI databases for the training.

It is worth noticing that the R18R50 can be also compared with the M3 technique (Stepien et al., 2021) (see Table 2) that also uses these two networks. As presented, the rich information from all layers and creating quality models for them in the fusion R18R50, introduced in this work, greatly affects the quality prediction performance in comparison with the M3, with its raw features from two networks

Table 4
Performance of separately and jointly trained networks in the MEDQEMRI.

Architecture	DB1				DB2				Overall weighted			
	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
R18 (end-to-end)	0.5903	0.4333	0.6334	0.5713	0.6851	0.5243	0.7249	0.6340	0.6637	0.5038	0.7042	0.6198
R50 (end-to-end)	0.3421	0.2793	0.3479	0.6569	0.7015	0.5210	0.7244	0.6378	0.6203	0.4664	0.6394	0.6421
R18	0.6432	0.5013	0.7706	0.4559	0.8904	0.7323	0.9141	0.3775	0.8346	0.6802	0.8817	0.3952
R50	0.7408	0.5485	0.8187	0.4305	0.8872	0.7182	0.9150	0.3773	0.8542	0.6799	0.8933	0.3893
R18R50	0.7400	0.5675	**0.8328**	**0.4277**	0.9076	0.7481	**0.9327**	**0.3401**	0.8697	0.7073	**0.9101**	**0.3599**
R18+R50	0.7015	0.5525	0.8022	**0.4072**	0.9057	0.7482	0.9276	0.3551	0.8596	0.7040	0.8993	0.3668
R18+R18R50	**0.7584**	0.5818	0.8106	0.4310	**0.9095**	**0.7517**	0.9279	0.3494	**0.8754**	**0.7133**	0.9014	0.3678
R50+R18R50	0.7283	**0.5857**	0.8007	0.4365	0.9029	0.7449	0.9289	0.3518	0.8634	0.7089	0.8999	0.3709
R18+R50+R18R50	**0.7455**	**0.5857**	**0.8204**	0.4300	**0.9133**	**0.7573**	**0.9312**	**0.3386**	**0.8754**	**0.7185**	**0.9062**	**0.3592**

Note: Two best results for each criterion are written in bold.
The MEDQEMRI is represented by the R18+R50+R18R50.

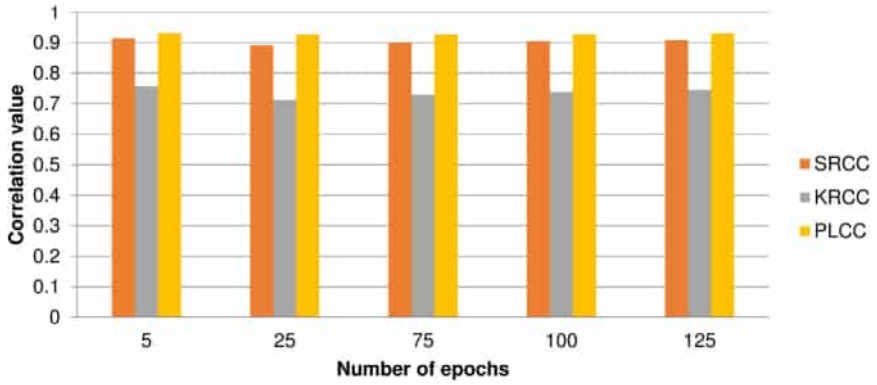


Fig. 8. Re-training of MEDQEMRI with a different number of epochs.

that fed the quality model. This additionally justifies the proposed processing of multi-level and multi-model quality information in the MEDQEMRI.

4.5.4. Re-training with early stopping

In the MEDQEMRI, after the fine-tuning and transfer-learning, the layers of networks are not frozen as it can be seen in a typical scenario (Tajbakhsh et al., 2016). Here, the architectures are trained for a small number of epochs to update them and make the networks aware of MR-specific characteristics. It can be seen as a fine-tuning of the pre-trained network with early stopping (Azizpour et al., 2015; Tajbakhsh et al., 2016). To show the influence of the number of epochs used to train an architecture on its results, Fig. 8 shows the performance of the MEDQEMRI on the DB2 with the number of epochs ranging from 5 to 100. It is observed that the longer training of the method (i.e., over the 100 epochs) does not provide the performance gain. As presented, the PLCC results of the method with early stopping after the several epochs indicate both the discriminative power of pre-trained features and superior quality prediction.

5. Limitations of the study and future directions

As reported in the ablation tests, the introduced MEDQEMRI addresses the limitations of the network structures from the object recognition field using their joint training and employing features from their layers with high-level quality models that gather quality-aware layer-based models. Despite better performance in comparison with the remaining approaches, the method does not employ MR-specific information (Obuchowicz et al., 2020), relying on the powerful network structures able to efficiently learn it. Consequently, it can be assumed that further improvement of such an approach should be based on a combination of prior knowledge (Zhang et al., 2023), designed MRIQA network architecture, or an application of an architecture designed for other MR-related tasks, e.g., compressed sensing magnetic resonance image reconstruction (Sun et al., 2018; Yan et al., 2020). Interestingly, the results obtained using deep learning approaches from the compressed sensing domain are often compared using image quality indices, which evidences the relationship between image quality and optimization task solved while reconstructing high-quality images from sub-Nyquist sampling signals (Sun et al., 2018). Another direction of research, extending the capabilities of related MRIQA methods is to combine deep features with features of promising hand-crafted approaches to describe MR scans for quality prediction purposes.

Furthermore, with the emergence of larger MRIQA datasets, the problem of training acceleration of deep learning models would require additional efforts or the usage of new promising solutions like MultiLayer Extreme Learning Machines (ML-ELM) (Zhang et al., 2020).

6. Conclusion

In this study, a new MRIQA approach is introduced in which a fusion of deep learning architectures is proposed and employed. It is shown that the fine-tuning and effective re-training of the fusion architecture and two separate networks results in a complementary solution that offers superior quality prediction performance. In the approach, the quality-aware deep features emerge from the obtained multi-level multi-model representations of MR images. The method is validated on two representative MR-IQA datasets containing MR images and subjective scores of a large number of radiologists. The method obtained the weighted average values of the Spearman Rank-order Correlation Coefficient, Kendall Rank-order Correlation Coefficient, and Pearson Linear Correlation Coefficient of 0.8754, 0.7185, and 0.9062, respectively. As the extensive experiments on two MRIQA benchmarks reveal, MEDQEMRI achieves better quality prediction than related methods, including popular techniques trained on MR scans, other deep learning representatives, or recently introduced MRIQA techniques. Overall, the presented results across the benchmark databases indicate the usability of MEDQEMRI in practice.

The Matlab code of MEDQEMRI is available at https://marosz.kia.prz.edu.pl/MEDQEMRI.html.

CRediT authorship contribution statement

Igor Stępień: Software, Validation, Writing, Visualization, Investigation. **Mariusz Oszust:** Conceptualization, Methodology, Writing, Investigation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Azizpour, H., Razavian, A., Sullivan, J., Maki, A., Carlsson, S., 2015. Factors of transferability for a generic convnet representation. IEEE Trans. Pattern Anal. Mach. Intell. 38, 36–45.

Backhausen, L.L., Herting, M.M., Buse, J., Roessner, V., Smolka, M.N., Vetter, N.C., 2016. Quality control of structural MRI images applied using FreeSurfer-a hands-on workflow to rate motion artifacts. Front. Neurosci. 10, 558.

Breuer, F.A., Kellman, P., Griswold, M.A., Jakob, P.M., 2005. Dynamic autocalibrated parallel imaging using temporal grappa (tgrappa). Magn. Reson. Med. 53 (4), 981–985.

Chow, L.S., Rajagopal, H., 2017. Modified-BRISQUE as no reference image quality assessment for structural MR images. Magn. Reson. Imaging 43, 74–87.

Esteban, O., Birman, D., Schaer, M., Koyejo, O.O., Poldrack, R.A., Gorgolewski, K.J., 2017. MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. PLOS ONE 12 (9), 1–21.

Gao, F., Yu, J., Zhu, S., Huang, Q., Tian, Q., 2018. Blind image quality prediction by exploiting multi-level deep representations. Pattern Recognit. 81, 432–442.

Gonzalez-Jaime, L., Vegas-Sanchez-Ferrero, G., Kerre, E.E., Aja-Fernandez, S., 2016. Spatially-variant noise filtering in magnetic resonance imaging: A consensus-based approach. Knowl.-Based Syst. 106, 264–273.

Gu, K., Zhai, G., Yang, X., Zhang, W., 2014. Hybrid no-reference quality metric for singly and multiply distorted images. IEEE Trans. Broadcast. 60 (3), 555–567.

Gupta, A., Sadri, A.R., Viswanath, S.E., Tiwari, P., 2020. Quality assessment of brain MRI scans using a dense neural network model and image metrics. In: Chen, G.-H., Bosmans, H. (Eds.), Medical Imaging 2020: Physics of Medical Imaging, Vol. 11312. International Society for Optics and Photonics, SPIE.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition.

Hoff, M., Andre, J., Stewart, B., 2016. Artifacts in magnetic resonance imaging. pp. 165–190.

Hu, R., Liu, Y., Gu, K., Min, X., Zhai, G., 2021. Toward a no-reference quality metric for camera-captured images. IEEE Trans. Cybern. 1–14.

Ilsche, T., Hackenberg, D., Schone, R., Bielert, M., Hopfner, F., Nagel, W.E., 2019. Metricq: A scalable infrastructure for processing high-resolution time series data. pp. 7–12.

Jang, J., Bang, K., Jang, H., Hwang, D., 2018. Quality evaluation of no-reference mr images using multidirectional filters and image statistics. Magn. Reson. Med. 80 (3), 914–924.

Jolliffe, I., 2011. Principal component analysis. pp. 1094–1096.

Kang, L., Ye, P., Li, Y., Doermann, D., 2014. Convolutional neural networks for no-reference image quality assessment. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition.

Kustner, T., Gatidis, S., Liebgott, A., Schwartz, M., Mauch, L., Martirosian, P., Schmidt, H., Schwenzer, N., Nikolaou, K., Yang, B., Schick, F., 2018. A machine-learning framework for automatic reference-free quality assessment in MRI. Magn. Reson. Imaging 53, 134–147.

Li, Q., Lin, W., Fang, Y., 2016. No-reference quality assessment for multiply-distorted images in gradient domain. IEEE Signal Process. Lett. 23 (4), 541–545.

Liu, S., Thung, K.-H., Lin, W., Shen, D., Yap, P.-T., 2020. Hierarchical nonlocal residual networks for image quality assessment of pediatric diffusion MRI with limited and noisy annotations. IEEE Trans. Med. Imaging 39 (11), 3691–3702.

Ma, K., Liu, W., Liu, T., Wang, Z., Tao, D., 2017. Dipiq: Blind image quality assessment by learning-to-rank discriminable image pairs. IEEE Trans. Image Process. 26 (8), 3951–3964.

Ma, K., Liu, W., Zhang, K., Duanmu, Z., Wang, Z., Zuo, W., 2018. End-to-end blind image quality assessment using deep neural networks. IEEE Trans. Image Process. 27 (3), 1202–1213.

Masoudi, S., Harmon, S., Mehralivand, S., Lay, N., Bagci, U., Wood, B.J., Pinto, P.A., Choyke, P., Turkbey, B., 2021. No-reference image quality assessment of t2-weighted magnetic resonance images in prostate cancer patients. In: 2021 IEEE 18th International Symposium on Biomedical Imaging. ISBI.

Min, X., Gu, K., Zhai, G., Liu, J., Yang, X., Chen, C.W., 2018. Blind quality assessment based on pseudo-reference image. IEEE Trans. Multimed. 20 (8), 2049–2062.

Mittal, A., Moorthy, A.K., Bovik, A.C., 2012. No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. 21 (12), 4695–4708.

Obuchowicz, R., Oszust, M., Bielecka, M., Bielecki, A., Piorkowski, A., 2020. Magnetic resonance image quality assessment by using non-maximum suppression and entropy analysis. Entropy 22 (2), 220.

Okarma, K., 2019. Current trends and advances in image quality assessment. Elektron. Elektrotech. 25 (3), 77–84.

Oszust, M., Piorkowski, A., Obuchowicz, R., 2020. No-reference image quality assessment of magnetic resonance images with high-boost filtering and local features. Magn. Reson. Med. 84 (3), 1648–1660.

Pontoriero, A.D., Nordio, G., Easmin, R., Giacomel, A., Santangelo, B., Jahuar, S., Bonoldi, I., Rogdaki, M., Turkheimer, F., Howes, O., Veronese, M., 2021. Automated data quality control in fdopa brain pet imaging using deep learning. Comput. Methods Programs Biomed. 208, 106239.

Qi, K., Li, H., Rong, C., Gong, Y., Li, C., Zheng, H., Wang, S., 2021. Blind image quality assessment for MRI with a deep three-dimensional content- adaptive hyper-network. arXiv:2107.06888 https://arxiv.org/pdf/2107.06888.pdf.

Ruder, S., 2017. An overview of gradient descent optimization algorithms.

Sheikh, H., Sabir, M., Bovik, A., 2006. A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Process. 15 (11), 3440–3451.

Stepien, I., Obuchowicz, R., Piorkowski, A., Oszust, M., 2021. Fusion of deep convolutional neural networks for no-reference magnetic resonance image quality assessment. Sensors 21 (4), 1043.

Sun, L., Fan, Z., Huang, Y., Ding, X., Paisley, J., 2018. Compressed sensing MRI using a recursive dilated network. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans. Med. Imaging 35 (5), 1299–1312.

Ullah, H., Irfan, M., Han, K., Lee, J.W., 2020. DLNR-SIQA: deep learning-based no-reference stitched image quality assessment. Sensors 20 (22), 6457.

Xu, J., Ye, P., Li, Q., Du, H., Liu, Y., Doermann, D., 2016. Blind image quality assessment based on high order statistics aggregation. IEEE Trans. Image Process. 25 (9), 4444–4457.

Xue, W., Mou, X., Zhang, L., Bovik, A.C., Feng, X., 2014. Blind image quality assessment using joint statistics of gradient magnitude and laplacian features. IEEE Trans. Image Process. 23 (11), 4850–4862.

Yan, J., Chen, S., Zhang, Y., Li, X., 2020. Neural architecture search for compressed sensing magnetic resonance image reconstruction. Comput. Med. Imaging Graph. 85, 101784.

Yan, Q., Gong, D., Zhang, Y., 2019. Two-stream convolutional networks for blind image quality assessment. IEEE Trans. Image Process. 28 (5), 2200–2211.

Ye, P., Kumar, J., Kang, L., Doermann, D., 2012. Unsupervised feature learning framework for no-reference image quality assessment. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition.

Yu, S., Dai, G., Wang, Z., Li, L., Wei, X., Xie, Y., 2018. A consistency evaluation of signal-to-noise ratio in the quality assessment of human brain magnetic resonance images. BMC Med. Imaging 18 (1), 17.

Zhang, Z., Dai, G., Liang, X., Yu, S., Li, L., Xie, Y., 2018. Can signal-to-noise ratio perform as a baseline indicator for medical image quality assessment. IEEE Access 6, 11534–11543.

Zhang, J., Li, Y., Xiao, W., Zhang, Z., 2020. Non-iterative and fast deep learning: Multilayer extreme learning machines. J. Franklin Inst. B 357 (13), 8925–8955.

Zhang, L., Zhang, L., Bovik, A.C., 2015. A feature-enriched completely blind image quality evaluator. IEEE Trans. Image Process. 24 (8), 2579–2591.

Zhang, J., Zhao, Y., Shone, F., Li, Z., Frangi, A.F., Xie, S.Q., Zhang, Z.-Q., 2023. Physics-informed deep learning for musculoskeletal modeling: Predicting muscle forces and joint kinematics from surface EMG. IEEE Trans. Neural Syst. Rehabil. Eng. 31, 484–493.

Zhu, D., Qiu, D., 2021. Residual dense network for medical magnetic resonance images super-resolution. Comput. Methods Programs Biomed. 209, 106330.

SoftwareX 29 (2025) 102073

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original software publication

TIQA-MRI: Toolbox for Perceptual Image Quality Assessment of Magnetic Resonance Images

Igor Stępień

Doctoral School of Engineering and Technical Sciences at the Rzeszow University of Technology, al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland

ARTICLE INFO

Keywords:
Image quality assessment
Magnetic resonance images
Perceptual image quality
Toolbox
Benchmark

ABSTRACT

Magnetic Resonance Imaging (MRI) plays a pivotal role in medical diagnostics and research as a non-invasive imaging tool. The accuracy and reliability of clinical evaluations depend heavily on the quality of MRI images, making high-quality imaging indispensable. Unfortunately, many MRI scans are compromised by noise, artifacts, or distortions, highlighting the need for no-reference Image Quality Assessment (IQA) methods to evaluate MR image quality. The proposed Toolbox for Perceptual Image Quality Assessment of Magnetic Resonance Images (TIQA-MRI) addresses this issue by offering a dedicated platform for assessing MR image quality. TIQA-MRI supports the development, testing, and evaluation of MR-IQA techniques through a collection of objective IQA methods and a curated dataset of MR images with subjective scores, facilitating comprehensive quality assessment. The toolbox allows users to evaluate the quality of their images and compare their MR-IQA techniques with existing methods. Implemented in MATLAB and available on GitHub, TIQA-MRI offers a standardized and accessible framework to advance research in MR image quality assessment.

Required Metadata

Current code version	v1.0
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-24-00605
Code Ocean compute capsule	none
Legal Code License	MIT
Code versioning system used	git
Software code languages, tools, and services used	Matlab 2024a, libsvm, MatlabPyrTools
Compilation requirements, operating environments & dependencies	Matlab toolbox: Image Processing, Statistics and Machine Learning, Deep Learning, curve fitting, Computer Vision, Parallel Computing, Signal Processing, Deep Learning Toolbox Model: GoogLeNet, Inception-v3
If available Link to developer documentation/manual	https://github.com/Igorles/TIQA-MRI/blob/main/README.md
Support email for questions	istepien@kia.prz.edu.pl

1. Motivation and significance

Magnetic Resonance Imaging (MRI) has established itself as a reliable diagnostic tool, particularly for soft tissue evaluation, while avoiding harmful radiation exposure. MRI allows for the inspection of ligaments and muscles, and due to its ability to differentiate between white and grey matter in the brain, it can also be used to diagnose conditions such as aneurysms and tumors [1]. MRI data are typically acquired in a matrix known as k-space, which is subsequently transformed into an MR image using Fourier techniques. Each pixel in the image corresponds to a weighted sum of points in k-space, meaning that any mislocalization can introduce artifacts [2]. Understanding these artifacts is essential to prevent misdiagnoses and to develop effective minimization strategies. Common sources of artifacts include physical distortions, sequence-related artifacts, and image reconstruction errors. For instance, motion artifacts like ghosting and blurring are caused by patient movement or physiological pulsations, while instrumental errors such as zipper artifacts result from electromagnetic interference or equipment malfunctions [3]. Sampling inaccuracies can lead to aliasing artifacts, while insufficient data resolution produces

E-mail address: istepien@kia.prz.edu.pl.

https://doi.org/10.1016/j.softx.2025.102073
Received 16 November 2024; Received in revised form 7 January 2025; Accepted 26 January 2025
Available online 3 February 2025

truncation artifacts like the Gibbs effect. Magnetic susceptibility variations at tissue interfaces or near implants create additional distortions. Thus, achieving optimal image quality requires careful design of MR sequences, thoughtful parameter selection, and favorable acquisition conditions. To date, no single method has been identified that can address all quality objectives; therefore, the goal remains to optimize image quality for specific applications [4,5].

The quality of MRI scans is critically important for accurate clinical evaluations and research. However, many scans suffer from issues such as noise, artifacts, and distortions, making the development of effective Image Quality Assessment (IQA) methods essential [6,7]. IQA has primarily relied on manual processes involving human observers and their ratings. Subjective quality metrics are typically categorized into two groups: (1) score-based ratings, which utilize visual quality score metrics such as the Mean Opinion Score (MOS) [8], and (2) class-based ratings, where the quality spectrum is classified into categories such as Excellent/Very Good/Good/Fair/Poor/Unusable [9] or Pass (Excellent to Very Good)/Questionable (Good to Fair)/Fail (Poor to Unusable) [10]. However, subjective assessment has several limitations, including being time-consuming, labor-intensive, costly, and error-prone [11]. Moreover, clinicians prioritize the clinical utility of an MR image's features over its perceptual quality, underscoring the need for objective quality assessment methods that provide reliable and repeatable evaluations [12]. Objective IQA methods can be divided into three categories based on the availability of reference images: full-reference (FR), reduced-reference (RR), and no-reference (NR). FR methods compare the image against a distortion-free reference, RR methods utilize only a portion of the reference image, and NR methods assess quality without any reference image. NR methods are particularly critical for MRI applications because high-quality reference images for specific regions of the examined patient are often unavailable. Additionally, clinical settings frequently require real-time assessments of image quality, making NR methods essential for timely and effective decision-making.

Progress in developing advanced IQA measures for MRI images has evolved from early adaptations of natural image techniques to more domain-specific approaches. Methods such as the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) and its extensions by Chow et al. [13] and Jang et al. [14] adapted quality models to MRI by analyzing image statistics and multidirectional-filtered coefficients. Yu et al. [15] used Signal-to-Noise Ratio for objective scoring, while Qi et al. [16] introduced 3D spatial networks incorporating neighboring scan information for more robust assessments.

Feature extraction techniques have also advanced, with methods using high-boost filters, entropy-based scores, and deep learning frameworks to improve quality prediction [17–19]. Machine learning models have been applied to classify image quality, as in Kustner et al. [20], who used active learning and PCA for no-reference classification, and Liu et al. [21], who applied hierarchical networks for pediatric diffusion MR images. Hybrid approaches integrating perceptual modeling and statistical rigor demonstrate improvements in correlation with human judgments [22]. Recent innovations also include subjective image processing methods such as FQPath and HVS_MaxPol, introduced by Hosseini et al. [23–25], which complement objective metrics by sharpening images based on the Human Visual System's sensitivity [26,27]. Metrics like the Objective Measure of Quality of Denoised Images (OMQDI), which balances edge preservation and noise suppression, and GAN-based quality assessments incorporating perceptual nuances have further enhanced NR IQA [28,29].

Additionally, a novel NR method proposed by [30] combines two complementary deep learning architectures to extract MR-specific quality-aware features. This approach integrates multi-level image representations from joint networks with additional features derived from retrained single architectures, improving quality prediction accuracy. A high-level quality model, trained on quality scores from intermediate network layers, enables robust and reliable assessments. This method exemplifies the growing importance of deep learning in NR IQA, enhancing real-time assessment capabilities and the reliability of distinguishing low-quality diagnostic scans [31–34].

In this paper, a new toolbox is introduced which incorporates nine MRI IQA methods. The toolbox is designed to facilitate the development of novel MRI techniques, enabling them to be benchmarked and evaluated alongside the state-of-the-art methods. Additionally, the inclusion of a database of MR images with subjective scores supports the training of IQA methods, which often yield superior results for MR images. The main contributions of this MATLAB-based software toolbox are: (1) simplifying the training and comparison of new IQA approaches for MR images using nine state-of-the-art methods, (2) evaluating selected MRI methods with IQA approaches that incorporate advanced perceptual techniques trained on human-assessed images, and (3) automating the generation of statistical significance test results or values of evaluation criteria.

2. Software description

The *TIQA-MRI* toolbox is designed to evaluate algorithms that assess the quality of MR images or generate objective scores for MR image quality. Users can easily compare their methods with existing approaches or obtain the objective quality scores required for their research. The application is divided into four tabs, each serving distinct functionalities.

The *Image Assessment* tab allows users to select IQA methods for evaluating MRI images. The script trains models based on features generated by the chosen IQA methods. The dataset used consists of 70 exemplary MRI images with corresponding human assessments [22]. The results are displayed as objective scores for the test images. The TIQA-MRI tool is designed to support specialized MRI formats, such as DICOM.

In the *Method Evaluation* tab, images are divided into training and testing subsets using a typical 80:20% split. The IQA methods are then applied to these subsets. The tab allows users to independently select parameters and methods, such as random seed values or split sizes for training and testing data. Results are presented in a table that includes subjective scores for both the training and testing sets. Furthermore, Spearman's Rank Correlation Coefficient (SRCC), Kendall's Order Correlation Coefficient (KRCC), and Pearson's Linear Correlation Coefficient (PLCC) values are displayed at the bottom of the window.

The *Method Comparison* tab provides functionality of selecting IQA methods for the comparison. The selected MRI images are evaluated using specified IQA methods, and the results are displayed as boxplots. Additionally, a summary of Wilcoxon's statistical significance test is provided, along with SRCC, KRCC, and PLCC correlation values, as well as a table presenting the computational time for each method.

The *Settings* tab enables users to manage the list of training methods. Once the list of test methods is established, users can select a training method from a drop-down menu within the table.

2.1. Database

In the experiments, a representative database of 70 MR images is used [22]. The benchmark includes images from various anatomical regions: the spine (14 images), knee (14), shoulder (16), brain (8), wrist (6), hip (4), pelvis (4), elbow (2), and ankle (2). The images were captured under diverse conditions that influence image quality. The database also includes Mean Opinion Score (MOS) values obtained from assessments conducted by a group of radiologists. The resolutions of the images in the dataset range from 192×320 to 512×512 [22].

Fig. 1. Image Assessment functionality.

2.2. Software architecture

The TIQA-MRI toolbox is developed and run using MATLAB, a widely used programming language for data analysis. To launch the application, users execute the *main.mlapp* file located in the main directory. The toolbox allows users to create new IQA, MR-IQA, or MRI methods by implementing advanced techniques available within TIQA-MRI.

2.3. Software functionalities

The TIQA-MRI toolbox includes scripts organized into application tabs that facilitate three main functionalities. As illustrated in Fig. 1, the first functionality utilizes the database to assess images and present objective scores. Users can select an image for the evaluation. If a method that require training is selected, an option to choose its previously trained model becomes available, enabling evaluation of a single image. The selected image is displayed in the center of the screen. To initiate the assessment process, the user clicks the *Evaluate* button, which triggers the image evaluation and displays the resulting objective scores.

In the second functionality, presented in Fig. 2a, quality models are trained on a portion of the database, while its remaining part is reserved for testing the methods. Users can choose a CSV or XLSX file containing images for training or testing methods and specify a directory for the database. This directory selection feature allows users to add paths to images that may not be located in a common folder, providing a clear overview of image paths for subsequent experiments. Selecting file paths is optional. After this step, a list of selected images and their corresponding subjective scores is displayed at the top of the window. The script offers a further range of setting options under *Options* button, as shown in Fig. 2b. The user can specify the desired database percentage division into training/testing data, choose the number of images of the same type, or enter a random seed value, which determines the randomness of the dataset split. It should be noted that if any number is entered in the *Size of the training subset* field, the corresponding field with testing subset will be filled in automatically. The entire process is completed by dividing the database into training and testing images by pressing the *Random split* button. The numerical split results are displayed in the main window; however, after making modifications, the tab with additional options does not close automatically and can be closed by clicking the *x* in the upper right corner of the window.

At the bottom of the window, the results of the image split are displayed, with training images on the left and test images on the right. If the objective is solely to test a method — either because it is not a training method or a pre-trained model is already available — the user should select the appropriate method and corresponding model. If model training is necessary, the application offers this option. After selecting a model, specify its name in the *Name the selected model* field on the left, which will determine how it is saved in the methods directory. Then, press the *Train* button to initiate the training process. To test a trained method, select the method and the newly created model. Upon pressing the *Test* button, the SRCC, KRCC, and PLCC values will be displayed at the bottom of the page, along with information about the file name where the evaluation results are saved. The objective scores will appear in the *Testing set* table.

The third functionality enables comparison of the performance of the image database trained on selected IQA methods. The method type and database must be selected to perform the calculations as presented in Fig. 3a. It should be noted that a model cannot be added for a non-training method. Therefore, if *None* appears in the *Train file* column, the option to select a model will not be available in the *Train model* column. Selecting a model means that in the main tab *Method comparison*, the chosen method will only be tested based on the existing model when the *Train and test all* button is pressed.

The database is chosen as a CSV or XLSX file containing image names and MOS values, followed by specifying the directory with the images. Selecting the file path is required to proceed with image processing. At this stage, the application allows setting the size of the training and testing subsets, the number of random seed, the number of splits, and the number of consecutive images of the same object. Upon specifying the number of splits, selecting the *Random split* button loads the directory into the split tabs. Once all settings are selected, they can be saved for future use. The application also supports loading previously saved settings, with all options visualized in Fig. 3b. After confirming the parameters with the *OK* button, the main *Method comparison* tab is displayed, where the selected methods, the image database (divided into training and testing sets), and a preview of subjective scores are shown (Fig. 3c).

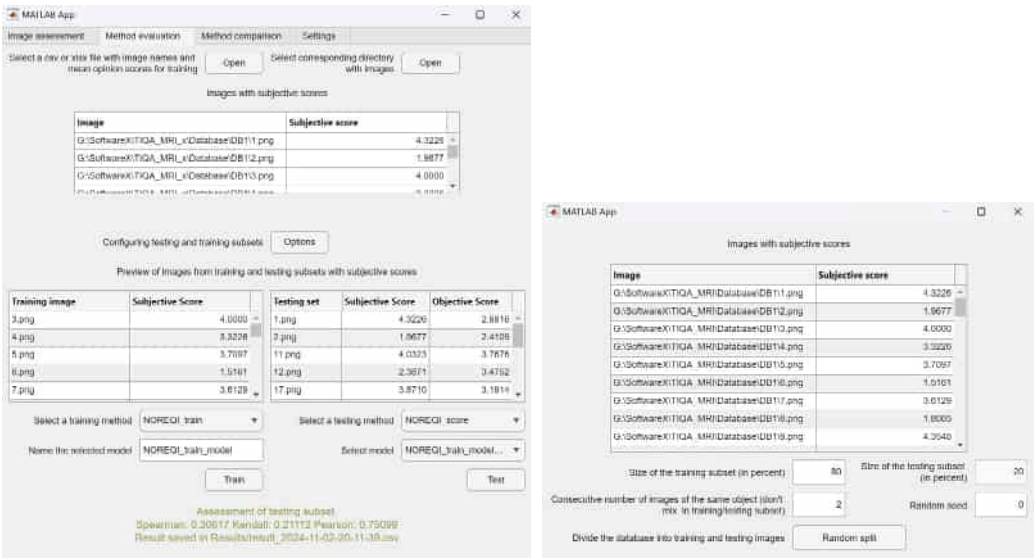


Fig. 2. *Method evaluation* functionality, (a) running the experiment, (b) database division into training and testing subsets.

To initiate the assessment process, pressing the *Train and test all* button displays the evaluation's progress. Depending on the selected method type, information are displayed regarding the ongoing training or testing process. If any required setup step is missed, an error message will be displayed. Once completed, results can be viewed and saved using the *Open results/Open and save results* button. The saved results can be found in a directory named *Results*.

2.3.1. Methods

TIQA-MRI contains nine IQA approaches. To facilitate the application of more IQA methods based on natural image processing, the NR (BLINDER [35], BRISQUE [36], DIIVINE [37], NIQE [38], SSEQ [39], SINDEX [40], SNRTOI [41], FQPath [42], and NOREQI [43]) methods are included. Among them, NR perceptual approaches are represented by BRISQUE, BLINDER, GWH-GLBP-BIQA [44], SSEQ, and NOREQI. Additionally, deep learning networks such as Inception-v3 [45] and GoogLeNet [46], adapted to regression tasks, are incorporated. Furthermore, when developing a new measure, it can be compared with other algorithms using SRCC, KRCC, and PLCC. In the proposed toolbox, these evaluation criteria can be found in the *Training methods* tab, while new methods can be added in the *Settings* tab.

3. Illustrative examples

3.1. Image assessment

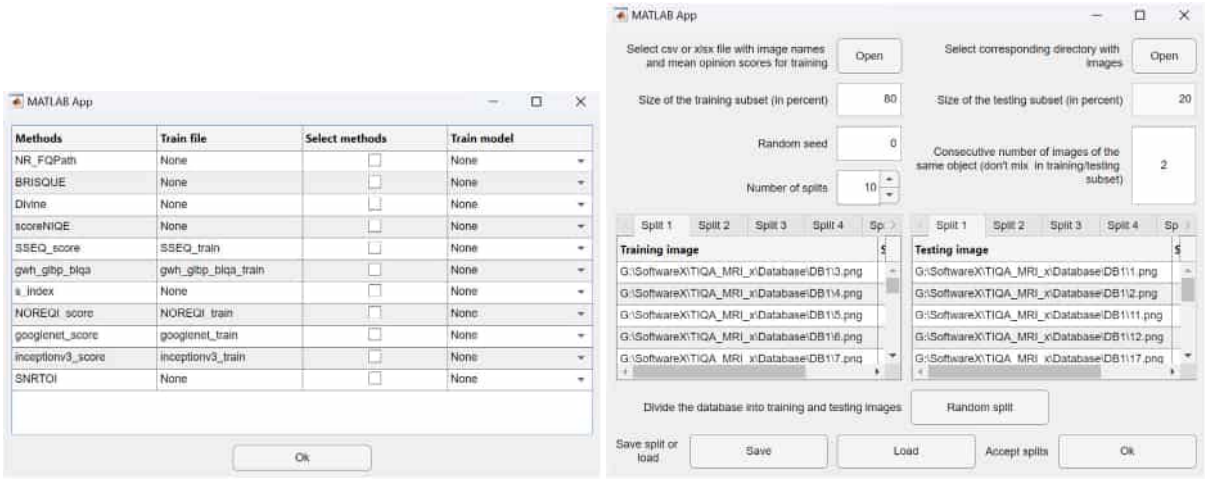
In the *Image assessment* tab, after the user uploads an image and selects a training method, the image is evaluated for quality (see Listing 1). The results are displayed as both an image (as shown in Fig. 1), and as an objective score.

Listing 1 Script for image evaluation using the Image Assessment tab functionality

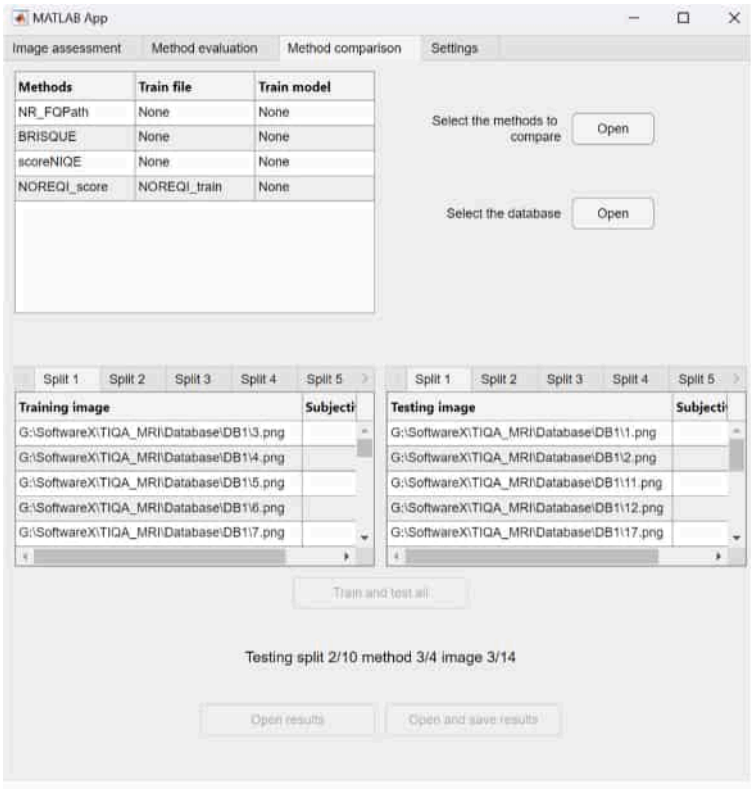
```
% Button pushed function: EvaluateButton
function EvaluateButtonPushed(app, event)
    % Display the 'waiting' message.
    app.Label.Text = 'waiting';
    pause(0.1);
    % Check if a method is selected in the dropdown.
        if isempty(app.SelectthemethodDropDown.Value)
        % If no method is selected, show an error message
            and exit the function.
            app.Label.Text = 'Please select a method from
                the dropdown.';
            return;
        end
        % Determine the function to call based on the
            dropdown values.
        try
            if isempty(app.SelectmodelDropDown_2.Value)
            % Construct and evaluate the function call
                without a model.
                methodCall = strcat(app.
                    SelectthemethodDropDown.Value, '(app.
                    imageProcessed)');
            else
             % Construct and evaluate the function call with
                 a model.
                methodCall = strcat(app.
                    SelectthemethodDropDown.Value, ...
                '(app.imageProcessed, app.
                    SelectmodelDropDown_2.Value)');
            end
            % Evaluate the function call and convert the
                result to a string.
           result = num2str(abs(eval(methodCall)));
           app.Label.Text = result;
           catch ME
           % If there is an error during the evaluation,
               display an error message.
              app.Label.Text = 'Error during evaluation.
                  Please check the selected method and model
                  .';
           % Optionally, display the error message for
               debugging purposes.
              disp(['Error: ', ME.message]);
    end
end
```

3.2. The comparison of IQA methods

The first experiment (Listing 2) that can be conducted using TIQA-MRI involves evaluating a selected image database through the IQA methods and three commonly used evaluation criteria: SRCC, KRCC and PLCC (Fig. 4(a)). Higher values of SRCC, KRCC, and PLCC indicate a better agreement between the results of the proposed approach and the subjective opinions of specialists. The results are displayed as the median of all splits or in tables with individual values for each split (Listing 3). Additionally, individual images can be assessed, as shown in Fig. 4(b). In this experiment, the scores of each image from the



(a) (b)



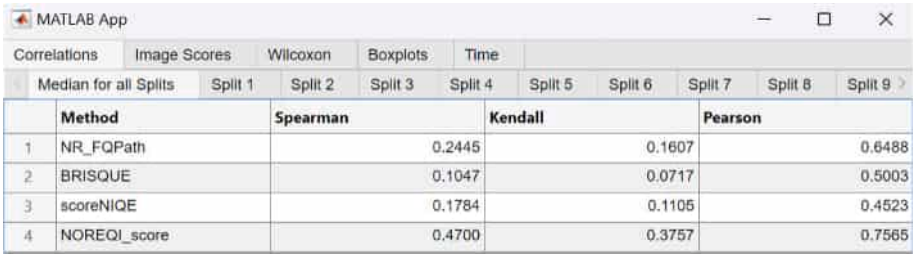
(c)

Fig. 3. *Method comparison* functionality, (a) method selection, (b) database selection, (c) training and testing processes.

evaluated database, assessed using the selected IQA method, can be compared with subjective ratings.

Listing 2 Script to calculate image scores using SRCC, KRCC, and PLCC criteria

```
function [tmpScore, tmpScore1] = computeScores(~, data,
    tmp, numSplits)
  % Compute the Spearman, Kendall, and Pearson scores
      for each method and split
  tmpScore = cell(size(tmp, 1), 1);
  for j = 1:numSplits
     for i = 1:size(tmp, 1)
     % Extract subjective scores for the current split
     subjectiveScores = data.TabGroup2_2.Children(j).
         Children.Children.Data{:, 2};
     % Extract objective scores for the current method
         and split
     objectiveScores = data.scoreAll{i}{j};
     % Compute Spearman, Kendall, and Pearson
         correlation coefficients
     [spearman, kendall, pearson] = metric_evaluation(
         subjectiveScores, objectiveScores);
     % Store the absolute values of the scores in
         tmpScore for the current method and split
     tmpScore{i}(j, :) = [abs(spearman), abs(kendall),
         abs(pearson)];
    end
  end
```



	Method	Spearman	Kendall	Pearson
1	NR_FQPath	0.2445	0.1607	0.6488
2	BRISQUE	0.1047	0.0717	0.5003
3	scoreNIQE	0.1784	0.1105	0.4523
4	NOREQI_score	0.4700	0.3757	0.7565

(a)

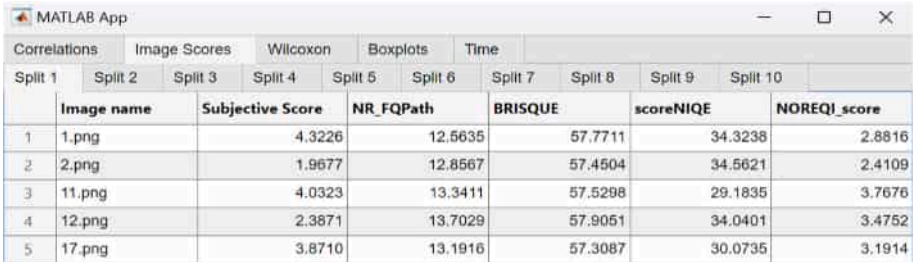


	Image name	Subjective Score	NR_FQPath	BRISQUE	scoreNIQE	NOREQI_score
1	1.png	4.3226	12.5635	57.7711	34.3238	2.8816
2	2.png	1.9677	12.8567	57.4504	34.5621	2.4109
3	11.png	4.0323	13.3411	57.5298	29.1835	3.7676
4	12.png	2.3871	13.7029	57.9051	34.0401	3.4752
5	17.png	3.8710	13.1916	57.3087	30.0735	3.1914
6	18.png	3.4839	13.1273	57.4879	31.6535	3.0467
7	29.png	4.3226	12.0357	58.2450	20.1978	3.2680
8	30.png	3.0968	12.0729	57.8579	22.8272	3.0179
9	47.png	3.5161	13.6950	57.7403	26.4400	3.4038
10	48.png	2.0000	13.9692	58.6275	25.6499	1.3394
11	59.png	3.4839	11.6593	57.9078	34.9283	3.5680
12	60.png	3.3226	11.7691	57.3979	28.9274	3.1645
13	69.png	3.1290	10.9506	58.9717	41.3453	3.7663
14	70.png	4.0000	10.9241	59.0578	38.6414	3.4011

(b)

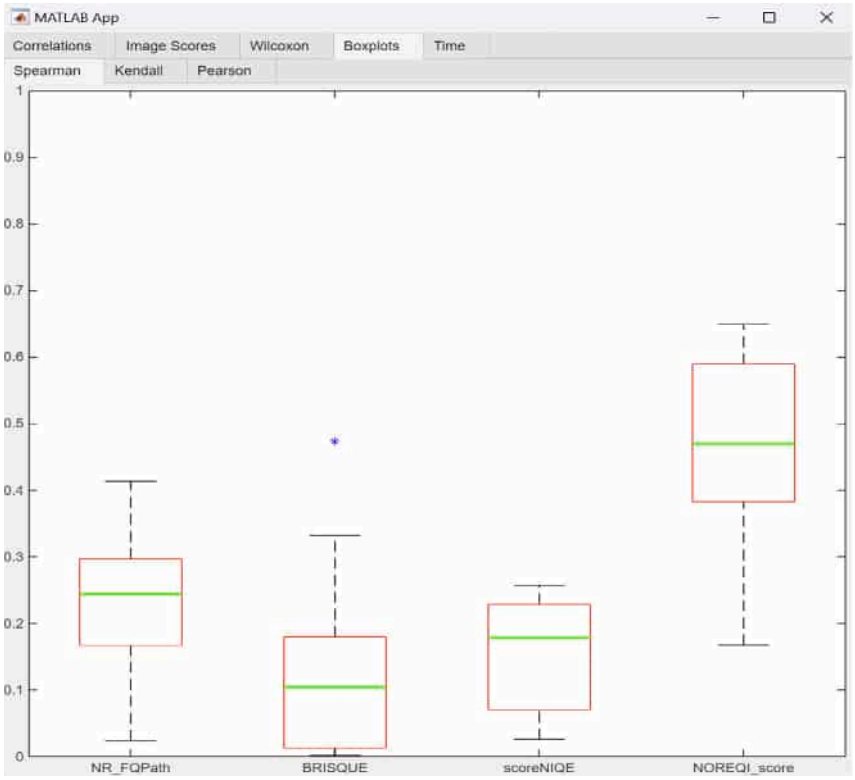
Fig. 4. (a) Selected MR images, (b) MR images evaluated by IQA methods along with their subjective scores.

```
    % Compute the median of the scores across splits for
        each method
    tmpScore1 = cellfun(@(x) median(x, 1), tmpScore, '
        UniformOutput', false);
end
```

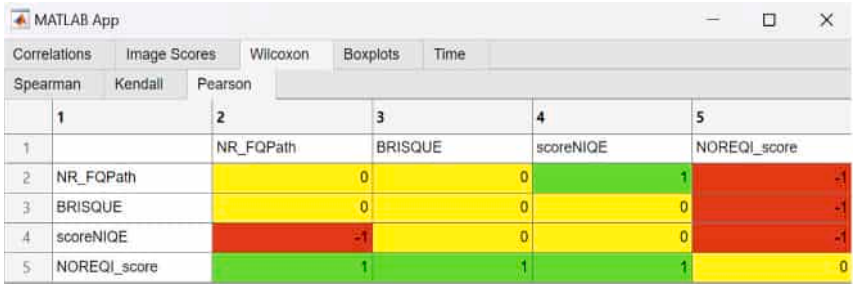
Listing 3 Script to evaluate MR images using IQA methods

```
function imageData = extractImageData(~, data)
    % Check if the input data is empty
    if isempty(data)
        imageData = {}; % Return an empty cell array if
            data is empty
    else
        % Extract the last part of each file path
        parts = regexp(data{:, 1}, filesep, 'split');
        imageData = cellfun(@(x) x{end}, parts, '
            UniformOutput', false);
    end
end
function createImageScoreTabs(app, data, tmp, numSplits)
    % Create tabs for displaying image scores
    app.TabGroupImScore = uitabgroup(app.GridLayout5);
    app.TabGroupImScore.Layout.Row = [1 17];
    app.TabGroupImScore.Layout.Column = [1 8];

    % Adjust column names based on whether tmp is empty
    if isempty(tmp)
        columnNames = {'Image name', 'Subjective Score'};
    else
        columnNames = ['Image name', 'Subjective Score',
            tmp(:, 1)'];
    end

    for i = 1:numSplits
        % Create each split tab and the corresponding
            grid layout and table
        app.tabsImScore{i} = uitab(app.TabGroupImScore, "
            Title", ['Split ' num2str(i)]);
        app.gridImScore{i} = createGridLayout(app, app.
            tabsImScore{i});
        app.tablesTabsImScore{i} = createUITable(app, app
            .gridImScore{i}, columnNames);

        % Extract image data and subjective scores
        imageData = extractImageData(app, data.
            TabGroup2_2.Children(i).Children.Children.
            Data);

        % Check if subjectiveScores is empty and adjust
            if needed
        if isempty(data.TabGroup2_2.Children(i).Children.
            Children.Data)
            subjectiveScores = nan(size(imageData)); %
                Use NaN to indicate missing scores
        else
            subjectiveScores = data.TabGroup2_2.Children(
                i).Children.Children.Data{:, 2};
        end
        % Handle objective scores if tmp is not empty
        if ~isempty(tmp)
            objectiveScores = cellfun(@(x) num2cell(x{i})
                , data.scoreAll(1:end), 'UniformOutput',
                false);
            app.tablesTabsImScore{i}.Data = [imageData,
                num2cell(subjectiveScores),
                objectiveScores{:}];
        else
            % If tmp is empty, only include image names
                and subjective scores
            app.tablesTabsImScore{i}.Data = {'No data
                available for image scores'};
        end
    end
end
```

Additional experimental findings, accessible through the application (see Listings 4 and 5), are illustrated in Fig. 5. The results were based on the typical database division of the images into training-testing



(a)



	1	2	3	4	5
1		NR_FQPath	BRISQUE	scoreNIQE	NOREQI_score
2	NR_FQPath	0	0	1	-1
3	BRISQUE	0	0	0	-1
4	scoreNIQE	-1	0	0	-1
5	NOREQI_score	1	1	1	0

(b)

Fig. 5. (a) Boxplot presenting the resulting values from the entire database, (b) Summary of Wilcoxon's statistical significance tests.

subsets. The tab presents the values for three evaluation criteria, along with boxplots and a summary of the pairwise Wilcoxon tests assessing the statistical significance between the selected methods. The boxplot indicates the median values of the tested samples, calculated at a 5% significance level.

Listing 4 Script presenting the function creating Boxplots

```
function createBoxplotTabs(app, tmpScore, tmp)
    % Create boxplot tabs for each correlation metric
    titles = {'Spearman', 'Kendall', 'Pearson'};
    for j = 1:3
       pause(0.1)
       app.tabsBoxplots{j} = uitab(app.TabBoxplotsTab,"
            Title", titles{j});
       app.gridBoxplots{j} = createGridLayout(app, app.
            tabsBoxplots{j});
       app.tabsUIAxes{j} = uiaxes(app.gridBoxplots{j});
       % Check if tmp is empty
       if isempty(tmp) || size(tmpScore{1}, 1) == 1
         % Display a message when there is no data
         text(app.tabsUIAxes{j}, 0.5, 0.5, 'No data
              available for plotting', '
              HorizontalAlignment', 'center', '
              VerticalAlignment', 'middle', 'FontSize',
              12, 'Color', 'red');
         app.tabsUIAxes{j}.XTick = []; % Remove X-ticks
              for clarity
         app.tabsUIAxes{j}.YTick = []; % Remove Y-ticks
              for clarity
       else
         % Prepare data for boxplot when tmp is not empty
         for i = 1:size(tmpScore, 1)
         % Using the original method to populate tmpTable
             tmpTable(:, i) = tmpScore{i}(:, j);
         end
         % Create the box plot and get the handle
         h = boxplot(app.tabsUIAxes{j}, tmpTable, "Labels
              ", tmp(:, 1));
         % Set the Y-axis limits to be between 0 and 1
         ylim(app.tabsUIAxes{j}, [0 1]);
         % Customize the boxplot appearance
         customizeBoxplotAppearance(app, h);
       end
    end
end
```

Listing 5 Wilcoxon's statistical significance tests function script

```
function createWilcoxonTabs(app, tmpScore, tmp)
    % Create tabs for Wilcoxon test results
    app.TabGroupWilcoxon = uitabgroup(app.GridLayout2);
```

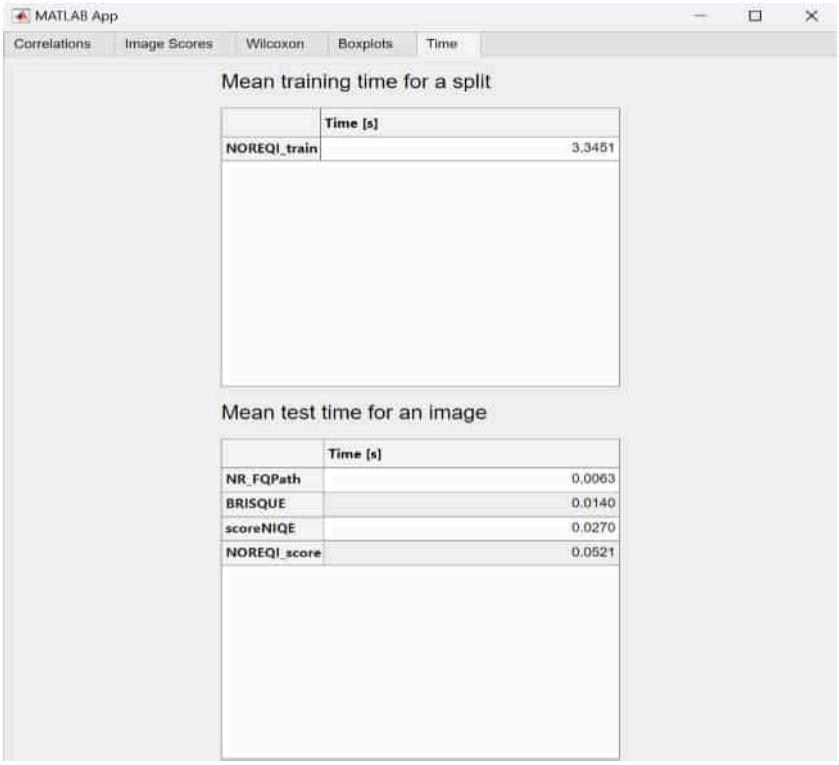


Fig. 6. Run-time comparison for one split or one image.

```
    app.TabGroupWilcoxon.Layout.Row = [1 17];
    app.TabGroupWilcoxon.Layout.Column = [1 8];
    titles = {'Spearman', 'Kendall', 'Pearson'};
    for i = 1:numel(titles)
        % Create the tab for each title
      app.tabsWilcoxon{i} = uitab(app.TabGroupWilcoxon, "
           Title", titles{i});
      app.gridWilcoxon{i} = createGridLayout(app, app.
           tabsWilcoxon{i});
      app.tablesTabsWilcoxon{i} = createUITable(app, app.
           gridWilcoxon{i});
      % Insert data into Wilcoxon table only if tmpScore
           is not empty
      if ~isempty(tmpScore)
          results = wilcoxonFun(app, tmpScore, i, tmp);
          app.tablesTabsWilcoxon{i}.Data = results;
      else
% If tmpScore is empty, you can provide a message or
    handle it differently
          app.tablesTabsWilcoxon{i}.Data = {'No data
               available for Wilcoxon test'};
      end
      % Apply cell styling based on Wilcoxon test results
      applyCellStyling(app, app.tablesTabsWilcoxon{i},
           app.tablesTabsWilcoxon{i}.Data);
    end
end
```

The computational time of the methods (see Listing 6), indicated by the average time required to evaluate an image from the dataset can also be analyzed. The application displays the average computational time for both the training sets and the test images separately. Since not all methods require training, only applicable methods will be shown on the screen (see Fig. 6).

Listing 6 Computational time function script

```
function createComplexityTables(app, tmpScore)
% Create tables for displaying training and testing time
    complexity
   app.tablesComplexityTab{1} = createUITable(app, app.
       GridLayout4, {'Time [s]'}, app.MainApp.
       nameTrainMethod);
   app.tablesComplexityTab{2} = createUITable(app, app.
       GridLayout4, {'Time [s]'}, app.MainApp.
       nameTestMethod);
   app.tablesComplexityTab{1}.Layout.Row = [2 8];
   app.tablesComplexityTab{1}.Layout.Column = [3 6];
   app.tablesComplexityTab{2}.Layout.Row = [10 17];
   app.tablesComplexityTab{2}.Layout.Column = [3 6];
   if ~isempty(tmpScore)
      if ~isempty(app.MainApp.timeTrain)
         app.tablesComplexityTab{1}.Data = (cellfun(@mean
              , app.MainApp.timeTrain))';
      end
         app.tablesComplexityTab{2}.Data = (cellfun(@(x)
              mean(x, 'all'), app.MainApp.timeTest))';
      else
        app.tablesComplexityTab{1}.Data = {'No data
             available for computational complexity'};
         app.tablesComplexityTab{2}.Data = {'No data
             available for computational complexity'};
   end
end
```

4. Impact

The increasing demand for IQA methods developed for MRI has led to the emergence of new techniques and algorithms. To address this need and facilitate the comparison of newly evaluated MR methods with state-of-the-art approaches in MR-IQA (or general IQA), the TIQA-MRI has been developed. Users introducing new MR-IQA techniques can utilize TIQA to incorporate their methods and validate their performance against nine other IQA algorithms. The toolbox includes standard comparison protocols such as IQA performance quality indices, boxplots, Wilcoxon's rank sum test, and a summary of computational time performance.

5. Conclusions

This paper presents TIQA-MRI, a MATLAB-based toolbox specifically designed for the perceptual image quality assessment of MR

images. The toolbox incorporates a range of IQA methods, including advanced perceptual techniques that simulate the human visual system. These methods can be trained using subjective scores to produce objective scores that closely correlate with human perception. TIQA-MRI facilitates comprehensive comparative evaluations of different IQA methods while enabling objective assessments of MR images. By simplifying the process of image quality evaluation, TIQA-MRI empowers researchers to refine algorithms and improve the reliability of their findings.

Future work will focus on expanding the toolbox by introducing additional MATLAB-based IQA methods and developing complementary MR approaches in a Python toolbox. This expansion aims to further enhance accessibility and usability for the research community, promoting ongoing advancements in MRI image analysis.

The code of TIQA-MRI is available on GitHub under the MIT license.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All software code and data used in this article is available on Github.

References

[1] Kathiravan S, Kanakaraj J. A review on potential issues and challenges in mr imaging. Sci World J 2013;2013(1):783715.
[2] Huang SY, Seethamraju RT, Patel P, Hahn PF, Kirsch JE, Guimaraes AR. Body mr imaging: artifacts, k-space, and solutions. Radiographics 2015;35(5):1439–60.
[3] Backhausen LL, Herting MM, Buse J, Roessner V, Smolka MN, Vetter NC. Quality control of structural MRI images applied using FreeSurfer-a hands-on workflow to rate motion artifacts. Front Neurosci 2016;10:558. http://dx.doi.org/10.3389/fnins.2016.00558.
[4] Saba L. Image principles, neck, and the brain. CRC Press; p. 2016.
[5] Stępień I, Oszust M. A brief survey on no-reference image quality assessment methods for magnetic resonance images. J Imaging 2022;8(6). http://dx.doi.org/10.3390/jimaging8060160.
[6] Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ. MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. PLoS One 2017;12(9):1–21. http://dx.doi.org/10.1371/journal.pone.0184661.
[7] Gunawan R, Tran Y, Zheng J, Nguyen H, Chai R. Implementing natural image quality evaluator for performance indicator on noise artefacts recovery in ct scan. In: 2023 45th annual international conference of the IEEE engineering in medicine and biology society. IEEE; 2023, p. 1–4. http://dx.doi.org/10.1109/EMBC40787.2023.10340822.
[8] Huynh-Thu Q, Garcia M-N, Speranza F, Corriveau P, Raake A. Study of rating scales for subjective quality assessment of high-definition video. IEEE Trans Broadcast 2010;57(1):1–14.
[9] Küstner T, Wolf P, Schwartz M, Liebgott A, Schick F, Gatidis S, et al. An easy-to-use image labeling platform for automatic magnetic resonance image quality assessment. In: 2017 IEEE 14th international symposium on biomedical imaging. IEEE; 2017, p. 754–7.
[10] White T, Jansen PR, Muetzel RL, Sudre G, El Marroun H, Tiemeier H, et al. Automated quality assessment of structural magnetic resonance images in children: Comparison with visual inspection and surface-based reconstruction. Hum Brain Mapp 2018;39(3):1218–31.
[11] Gedamu EL, Collins DL, Arnold DL. Automated quality control of brain mr images. J Magn Reson Imaging: Off J Int Soc Magn Reson Med 2008;28(2):308–19.
[12] Woodard JP, Carley-Spencer MP. No-reference image quality metrics for structural mri. Neuroinformatics 2006;4:243–62.
[13] Chow LS, Rajagopal H. Modified-BRISQUE as no reference image quality assessment for structural MR images. Magn Reson Imaging 2017;43:74–87. http://dx.doi.org/10.1016/j.mri.2017.07.016.
[14] Jang J, Bang K, Jang H, Hwang D. Quality evaluation of no-reference mr images using multidirectional filters and image statistics. Magn Reson Med 2018;80(3):914–24. http://dx.doi.org/10.1002/mrm.27084.
[15] Yu S, Dai G, Wang Z, Li L, Wei X, Xie Y. A consistency evaluation of signal-to-noise ratio in the quality assessment of human brain magnetic resonance images. BMC Med Imaging 2018;18(1):17. http://dx.doi.org/10.1186/s12880-018-0256-6.
[16] Qi K, Li H, Rong C, Gong Y, Li C, Zheng H, et al. Blind image quality assessment for MRI with a deep three-dimensional content- adaptive hyper-network. 2021, arXiv:2107.06888.
[17] Oszust M, Piorkowski A, Obuchowicz R. No-reference image quality assessment of magnetic resonance images with high-boost filtering and local features. Magn Reson Med 2020;84(3):1648–60. http://dx.doi.org/10.1002/mrm.28201.
[18] Obuchowicz R, Oszust M, Bielecka M, Bielecki A, Piorkowski A. Magnetic resonance image quality assessment by using non-maximum suppression and entropy analysis. Entropy 2020;22(2):220. http://dx.doi.org/10.3390/e22020220.
[19] Stepien I, Obuchowicz R, Piorkowski A, Oszust M. Fusion of deep convolutional neural networks for no-reference magnetic resonance image quality assessment. Sensors 2021;21(4):1043. http://dx.doi.org/10.3390/s21041043.
[20] Kustner T, Gatidis S, Liebgott A, Schwartz M, Mauch L, Martirosian P, et al. A machine-learning framework for automatic reference-free quality assessment in MRI. Magn Reson Imaging 2018;53:134–47. http://dx.doi.org/10.1016/j.mri.2018.07.003.
[21] Liu S, Thung K-H, Lin W, Shen D, Yap P-T. Hierarchical nonlocal residual networks for image quality assessment of pediatric diffusion MRI with limited and noisy annotations. IEEE Trans Med Imaging 2020;39(11):3691–702. http://dx.doi.org/10.1109/TMI.2020.3002708.
[22] Oszust M, Piórkowski A, Obuchowicz R. No-reference image quality assessment of magnetic resonance images with high-boost filtering and local features. Magn Reson Med 2020;84(3):1648–60.
[23] Brixtel R, Bougleux S, Lézoray O, Caillot Y, Lemoine B, Fontaine M, et al. Whole slide image quality in digital pathology: review and perspectives. IEEE Access 2022;10:131005–35.
[24] Hosseini MS, Plataniotis KN. Image sharpness metric based on maxpol convolution kernels. In: 2018 25th IEEE international conference on image processing. IEEE; 2018, p. 296–300.
[25] Hosseini MS, Zhang Y, Plataniotis KN. Encoding visual sensitivity by maxpol convolution filters for image sharpness assessment. IEEE Trans Image Process 2019;28(9):4510–25.
[26] Leclaire A, Moisan L. Blind deblurring using a simplified sharpness index. In: Scale space and variational methods in computer vision: 4th international conference, SSVM 2013, Schloss Seggau, Leibnitz, Austria, June (2013) 2-6. Proceedings 4. Springer; 2013, p. 86–97.
[27] Chen F, Fu H, Yu H, Chu Y. Using hvs dua-pathway and contrast sensitivity to blindly assess image quality. Sensors 2023;23(10). http://dx.doi.org/10.3390/s23104974.
[28] Simi V, Edla DR, Joseph J. A no-reference metric to assess quality of denoising for magnetic resonance images. Biomed Signal Process Control 2021;70:102962. http://dx.doi.org/10.1016/j.bspc.2021.102962.
[29] Treder MS, Codrai R, Tsvetanov KA. Quality assessment of anatomical mri images from generative adversarial networks: Human assessment and image quality metrics. J Neurosci Methods 2022;109579.
[30] Stępień I, Oszust M. No-reference image quality assessment of magnetic resonance images with multi-level and multi-model representations based on fusion of deep architectures. Eng Appl Artif Intell 2023;123:106283. http://dx.doi.org/10.1016/j.engappai.2023.106283.
[31] Masoudi S, Harmon S, Mehralivand S, Lay N, Bagci U, Wood BJ, et al. No-reference image quality assessment of t2-weighted magnetic resonance images in prostate cancer patients. In: 2021 IEEE 18th international symposium on biomedical imaging. 2021, p. 1201–5. http://dx.doi.org/10.1109/ISBI48211.2021.9434027.
[32] Stępień I, Obuchowicz R, Piórkowski A, Oszust M. Fusion of deep convolutional neural networks for no-reference magnetic resonance image quality assessment. Sensors 2021;21(4). http://dx.doi.org/10.3390/s21041043.
[33] Kastryulin S, Zakirov J, Pezzotti N, Dylov DV. Image quality assessment for magnetic resonance imaging. IEEE Access 2023;11:14154–68.
[34] Gupta A, Sadri AR, Viswanath SE, Tiwari P. Quality assessment of brain MRI scans using a dense neural network model and image metrics. In: Chen G-H, Bosmans H, editors. Medical imaging 2020: physics of medical imaging, vol. 11312. International Society for Optics and Photonics, SPIE; 2020, p. 214–21.
[35] Gao, Yu, Zhu, Huang, Tian. Blind image quality prediction by exploiting multi-level deep representations. Pattern Recognit 2018;81:432–42. http://dx.doi.org/10.1016/j.patcog.2018.04.016.
[36] Mittal A, Moorthy AK, Bovik AC. No-reference image quality assessment in the spatial domain. IEEE Trans Image Process 2012;21(12):4695–708. http://dx.doi.org/10.1109/TIP.2012.2214050.
[37] Moorthy AK, Bovik AC. Blind image quality assessment: From natural scene statistics to perceptual quality. IEEE Trans Image Process 2011;20(12):3350–64. http://dx.doi.org/10.1109/TIP.2011.2147325.

[38] Mittal A, Soundararajan R, Bovik AC. Making a "completely blind" image quality analyzer. IEEE Signal Process Lett 2013;20(3):209–12. http://dx.doi.org/10.1109/LSP.2012.2227726.

[39] Liu L, Liu B, Huang H, Bovik AC. No-reference image quality assessment based on spatial and spectral entropies. Signal Process, Image Commun 2014;29(8):856–63. http://dx.doi.org/10.1016/j.image.2014.06.006.

[40] Bosse S, Maniry D, Müller K-R, Wiegand T, Samek W. Deep neural networks for no-reference and full-reference image quality assessment. IEEE Trans Image Process 2017;27(1):206–19.

[41] Zhang Z, Dai G, Liang X, Yu S, Li L, Xie Y. Can signal-to-noise ratio perform as a baseline indicator for medical image quality assessment. IEEE Access 2018;6:11534–43. http://dx.doi.org/10.1109/ACCESS.2018.2796632.

[42] Hosseini MS, Brawley-Hayes JA, Zhang Y, Chan L, Plataniotis KN, Damaskinos S. Focus quality assessment of high-throughput whole slide imaging in digital pathology. IEEE Trans Med Imaging 2019;39(1):62–74.

[43] Oszust M. No-reference image quality assessment using image statistics and robust feature descriptors. IEEE Signal Process Lett 2017;24(11):1656–60.

[44] Li Q, Lin W, Fang Y. No-reference quality assessment for multiply-distorted images in gradient domain. IEEE Signal Process Lett 2016;23(4):541–5.

[45] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2016, p. 2818–26. http://dx.doi.org/10.1109/CVPR.2016.308.

[46] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015, p. 1–9.

remote sensing

MDPI

Article

No-Reference Quality Assessment of Pan-Sharpening Images with Multi-Level Deep Image Representations

Igor Stępień [1] and Mariusz Oszust [2,*]

[1] Doctoral School of Engineering and Technical Sciences, Rzeszow University of Technology, al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland; igorkrzysztofstepien@gmail.com
[2] Department of Computer and Control Engineering, Rzeszow University of Technology, Wincentego Pola 2, 35-959 Rzeszow, Poland
* Correspondence: marosz@kia.prz.edu.pl

Abstract: The Pan-Sharpening (PS) techniques provide a better visualization of a multi-band image using the high-resolution single-band image. To support their development and evaluation, in this paper, a novel, accurate, and automatic No-Reference (NR) PS Image Quality Assessment (IQA) method is proposed. In the method, responses of two complementary network architectures in a form of extracted multi-level representations of PS images are employed as quality-aware information. Specifically, high-dimensional data are separately extracted from the layers of the networks and further processed with the Kernel Principal Component Analysis (KPCA) to obtain features used to create a PS quality model. Extensive experimental comparison of the method on the large database of PS images against the state-of-the-art techniques, including popular NR methods adapted in this study to the PS IQA, indicates its superiority in terms of typical criteria.

Keywords: pan-sharpening; image quality assessment; remote sensing; multi-level features; network fusion

Citation: Stępień, I.; Oszust, M. No-Reference Quality Assessment of Pan-Sharpening Images with Multi-Level Deep Image Representations. *Remote Sens.* **2022**, *14*, 1119. https://doi.org/10.3390/rs14051119

Academic Editors: Francesca Giannone and Valerio Baiocchi

Received: 13 January 2022
Accepted: 23 February 2022
Published: 24 February 2022

1. Introduction

Pan-sharpening (PS) is an approach to combine spatial details of a high-resolution panchromatic (PAN) image and low-resolution multi-spectral (MS) information of the same region, aiming to produce a high-resolution MS image through the sharpening of the MS bands [1]. PS methods improve the ability of human viewers to interpret satellite imagery. The basic idea of sharpening is to simultaneously preserve the spectral characteristics and the spatial resolution of the image in the obtained object. The acquired image quality differs depending on the used algorithms, as they provide different image sharpening qualities [2]. They can be divided into several categories based on the usage of component substitution (CS) [3,4], multiresolution analysis (MRA) [5], variational optimization (VO) [6], or deep-learning (DL) [7]. Among the PS approaches, the Hue Saturation Value (HSV) leads to the transformation of the R, G, and B bands of an MS image into HSV components. This process replaces a value of the component with a panchromatic image and performs an inverse transformation to gain an MS image with high spatial resolution [8]. One of the most common fusion techniques used for sharpening is the Intensity-Hue-Saturation (IHS) technique [4] that converts a color image to the IHS color space, replaces intensity information with PAN image, and returns to the RGB color space. In another algorithm, Ehlers Fusion (EF), image fusion is based on filtering in the Fourier domain [9]. The method aims to preserve the spectral characteristics of the lower resolution of MS images. In that work, PAN images are fused with Landsat TM and IKONOS multi-spectral data. The algorithm is based on the IHS transform and can be applied to sharpen hyperspectral images without changing their spectral behavior. The High Pass Filter (HPF) resolution [5] creates a PS image with great attention to detail and an accurate depiction of the spectral content of the original MS image. Here, the PAN image is convoluted using a high-pass filter. In further

steps, it is combined with lower-resolution MS imagery. This technique is mostly applied for a large discrepancy in the pixel ratio between the PAN and MS images. In the PS method of Jing et al. [10], an image is synthesized of an image with minimum spectral distortion, considering haze. The method modifies several PAN modulation fusion approaches and generates high-quality synthetic outputs. The main goal of the study of Laben et al. [11] was to create a method that processes any number of bands at the same time. Additionally, it preserves spectral characteristics of the lower spatial resolution MS data in the higher spatial resolution by the Gram-Schmidt transformation on the simulated lower spatial resolution PAN image. The simulated lower spatial resolution image is employed as the first band in the Gram-Schmidt transformation. Another image fusion method that allows the use of any number of bands is the Principal Component Analysis (PCA) [3]. Its standard version is often used for dynamic analysis of multi-source or multi-temporal remote sensing data. Alparone et al. [12] introduced Quality with No Reference (QNR) in which complementary spatial and spectral distortion indices are fused. In its recent version, Hybrid Quality with No Reference (HQNR) method [13], the overall image quality is determined using the DS component of the QNR and spectral distortion metric [14]. A Universal Image Quality Index (Qq) is created by modeling an image distortion as a combination of loss of correlation, distortion of luminance, and contrast [15]. The Spectral Angle Mapper (SAM) technique is used in MS image analysis [16]. It operates on a spectral component and is used to compute the average variation of its angles. This technique has become a common tool for image color analysis or improvement of spatial resolution. In the method, spectral information is reflected by the hue and saturation and is slightly disturbed by a change of intensity. The method proposed by Alcaras et al. [17] considers automatic the PS process of VHR satellite images and the selection of the best of them. The approach of Zhang et al. [18], Object-based Area-To-Point Regression Kriging (OATPRK), fuses the MS and PAN images at the object-based scale. It is composed of image segmentation, object-based regression, and residual downscaling stages. An IQA method to support the visual qualitative analysis of pan-sharpened images by using the Natural Scene Statistics (NSS) is presented by Agudelo-Medina et al. [1]. In the approach, six PS methods are analyzed in the presence of blur and white noise. Since the method requires training a quality model, its development was preceded by the creation of a large PS image database with subjective scores assigned in tests with human observers.

Considering FR PS quality evaluation, the Root Mean Square Error (RMSE) is widely used for this purpose. It measures similarity between bands of original and combined images [19]. Erreur Relative Globale Adimensionalle de Synthèse (ERGAS) [20], in turn, takes into account the number of spectral bands, spatial resolutions of PAN and MS images, and RMSE between fused and original bands. The Edge-based image Fusion Metric (EFM) assesses the edge behavior of PS images and compares the obtained results with the input versions of PAN and MS images [21].

The quality assessment of PS images is a subject of open debate among researchers [7,22]. However, the IQA of natural or medical images is represented by a large diversity of approaches which, as shown in this study, can be adapted to the PS image domain. An NR or Blind Image Quality Assessment (BIQA) approach does not require access to the pristine reference image, which is beneficial since, in most applications, reference images are not available. Among IQA methods devoted to natural images, the BPRI uses as a reference a pseudo-reference image (PRI) and a PRI-based BIQA framework [23], estimating blockiness, sharpness, and noise. A CurveletQA, in turn, operates under a two-stage distortion classification, followed by an evaluation of the quality with a support vector machine (SVM) technique. GWH-GLBP is the NR-IQA method focused on predicting the quality of multiply distorted images, with the help of the weighted local binary pattern (LBP) histogram, calculated based on the gradient map [24]. In the deep learning-based MEON method, the learning process is divided into two stages, i.e., pre-training of the distortion identification subnetwork and quality prediction sub-network training, where the activation function is selected by generalized divisive normalization (GDN) [25]. Popular

Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) extracts statistics of the local luminance signals and measures the naturalness of the image based on the distortion information [26]. A method inspired by the human visual system (HVS), NFERM, extracts image features and uses support vector regression (SVR) to predict image quality [27]. Among deep learning approaches, Blinder [28] extracts features from network architecture and uses the minimum and maximum values of feature maps as a feature vector for quality prediction with the SVR, while the approach of Stępień et al. [29] to IQA of magnetic resonance scans employs jointly trained several networks. In an approach to the IQA of remote sensing images presented by Ieremeiev et al. [30], a set of FR measures designed for natural images are combined using a neural network.

In this paper, a novel NR PS IQA method, Multi-Level Pan-Sharpening Images Evaluator (MLPSIE) technique, is introduced. The method, contrary to other approaches to the PS image evaluation uses deep learning to obtain quality scores correlated with human judgment. To the best knowledge of the authors, it is the first technique that uses deep learning architectures for assessing PS images. Also, contrary to other deep learning methods devoted to the assessment of images from other domains, it takes two complementary deep learning architectures and separately extracts high-dimensional features from their layers, performs layer-wise dimensionality reduction, and creates quality-aware multi-level image representations used to build the quality model.

Contributions of this study are as follows: (1) Application of deep learning to IQA of PS images, (2) Separate extraction and reduction of high-dimensional data from each layer of the networks to provide features for training a quality model, (3) Successful adaptation of IQA methods from different domains to perform the quality evaluation of PS images, (4) Conducting extensive experiments on a large PS image database.

The remainder of this paper is organized as follows. In Section 2, the method is introduced. Then, in Section 3, it is experimentally compared against related IQA methods, and the obtained results are reported and discussed. Finally, in Section 4, conclusions and possible directions of future work are presented.

2. Proposed Method

The proposed MLPSIE uses the two deep learning networks, ResNet18 [31] and VGG19 [32]. However, as it is shown in Section 3.5 (Ablation Tests) the proposed processing of multi-level data allows for obtaining features sensitive to distortions which can be applied to other network pairs or even single architectures, leading to acceptable results. It is worth noticing that the networks are not trained due to the size of the image database and the obtained promising performance of the approach. However, if needed, it is assumed that the released source code of the MLPSIE can be adapted to perform a fine-tuning of the networks to capture image characteristics of a specific problem. The source code is available at http://marosz.kia.prz.edu.pl/MLPSIE.html, accessed on 13 January 2022. As presented in Figure 1 with the block diagram, the PS image composed of the RGB and near-infrared (NIR) bands is used to create the true color (TC) RGB and pseudocolor (PC) NIR + RG inputs to the network pair. Then, high-dimensional network responses at each level are extracted (blue rectangles in the figure) and reduced using the Kernel PCA (KPCA) approach (brown rectangles) [33]. The reduction takes place for concatenated TC and PC information, represented by dashed lines. Finally, the reduced features are concatenated (longer brown block) and used by the quality model obtained with the SVR to predict the quality of the PS image (green circle).

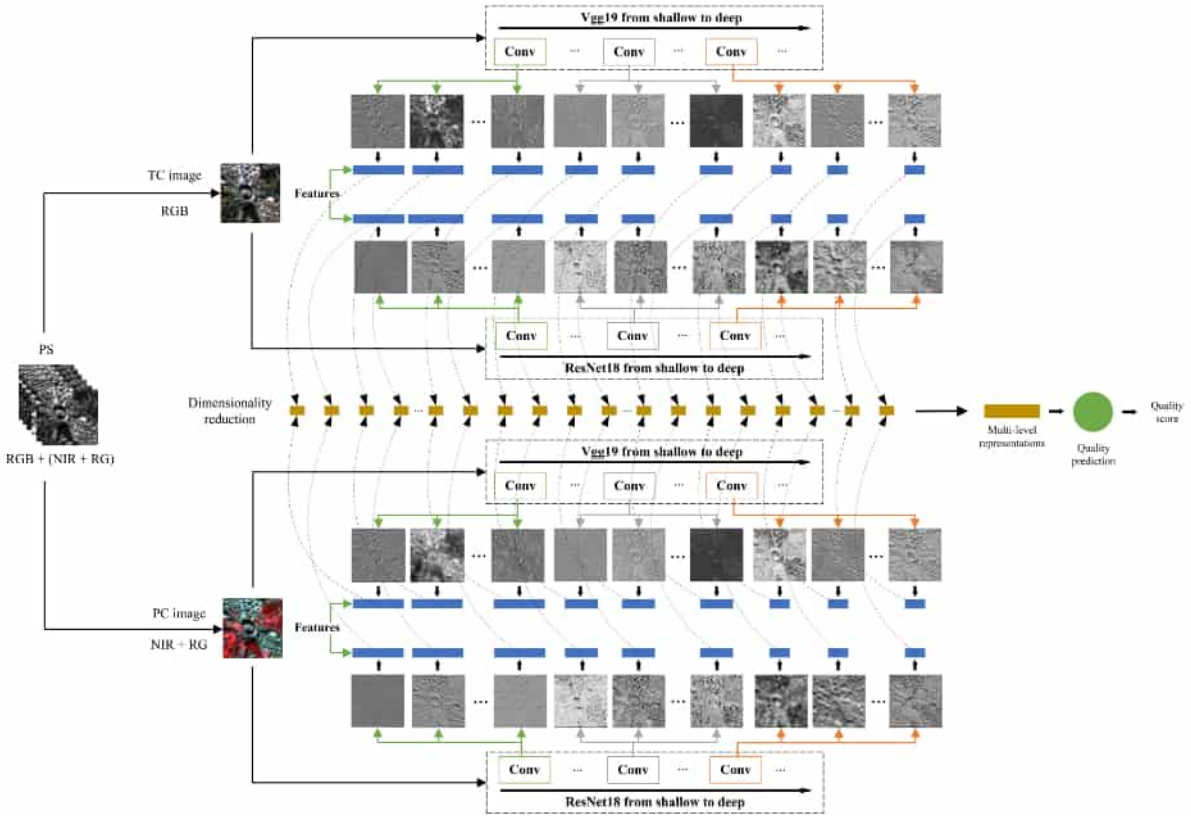


Figure 1. Block diagram of the method. Features extracted from both networks for the PC and TC images of input PS image (represented by blue rectangles) are reduced using KPCA (brown blocks) and concatenated for quality prediction by the trained SVR (green circle).

2.1. Network Architectures

In this paper, ResNet18 and Vgg19 architectures are used for the PS IQA. The Visual Geometry Group Network (VGGNet) is a deep learning algorithm with a multi-layered operation [34]. It consists of 16 convolution layers and three fully-connected layers, where 3×3 convolutional layers are placed on the top to increase with depth level. In the first two convolutional layers, 64 kernels (3×3 filter size) and the same padding are included. In this network architecture, the input is of a fixed size of 224×224. The pre-processing is done by the subtraction of the mean value from each pixel and is calculated for the entire training set. Moreover, max-pooling is performed over a 2×2 pixel window. In the set of fully connected layers, the first two are of size 4096 and the third layer consists of 1000 channels, while the final layer is a SoftMax function. In ResNet18, to avoid two or three layers containing ReLU and batch normalization, the architecture uses shortcut connections. Additionally, it solves the problem of vanishing gradients which increase the training error with a growing number of layers. The shortcut connections that allow skipping the layers allow for the training of deeper networks. At an early stage, the architecture performs the convolution (7×7) and max pooling (3×3). As the last layers, the average pooling and fully-connected layer are used [31,35].

2.2. Multi-Level Features

Since networks are designed to work with three-band RGB images and the proposed approach should be able to produce a quality score based on two three-band images, the feature vector extracted from the l-th layer of the n-th network can be written as V_l^n, where $l = 1, 2, \ldots, L_n$, and L_n is the number of convolutional layers in the network. Hence, multi-level data extracted from the network can be written as $V_n = [V_1^n, V_2^n, \ldots, V_{L_n}^n]$, and taking into account PC and TC images and both networks used in this study ($n = 1$ for ResNet18 and $n = 2$ for VGG19), the resulted representation of PS image is $V =$

$[V_1(RGB), V_1(NIR+RG), V_2(RGB), V_2(NIR+RG)]$. Note that for example, the first layers of ResNet18 and VGG19 contain 802816 and 3211264 values, respectively. Therefore, to create quality models without discarding important information that is stored at various levels of the networks, in this study, each layer is processed independently by the KPCA to produce a compact and distinctive quality-aware vector. Since two networks of each deep learning backbone are used to extract features from the TC and PC images, they are concatenated together. Finally, the vector $\mathcal{V} = [KPCA([V_1^1(RGB), V_1^1(NIR+RG)], \ldots, KPCA([V_{L_1}^1(RGB), V_{L_1}^1(NIR+RG)]), KPCA([V_1^2(RGB), V_1^2(NIR+RG)], \ldots, KPCA([V_{L_2}^2(RGB), V_{L_2}^2(NIR+RG)])]$.

The KPCA implements classical PCA but it can be also used for non-linear problems or problems in which the number of components should be determined automatically [33,36]. It is employed in this work as it provides satisfactory output with ease of implementation.

Once feature vectors characterizing training images are obtained with the proposed method, a quality model can be trained. Here, the SVR is used due to its popularity and dominant position among similar solutions in the IQA literature [28]. The used ε-SVR maps feature vector for an image ($\mathcal{V}$) into its subjective score (S). Given the training data $(\mathcal{V}, S)$, where $\mathcal{V}$ denotes feature vectors of M training images $(\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_M)$ and S contains their subjective scores, i.e., Differential Mean Opinion Scores (DMOS), $(S_1, S_2, \ldots, S_M)$, a function $\mathfrak{f}(\mathcal{V}) = \langle \omega, \mathcal{V} \rangle + b$ is determined in which $\langle \cdot, \cdot \rangle$, ω, and b, are the inner product, weight vector, and a bias parameter, respectively. Once the slack variables ξ_m and ξ_m^* are introduced, the ω and b are the solution of the following optimization problem:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|\omega\|^2 + C\sum_{m=1}^{M}(\xi_m + \xi_m^*) \\ &subject\ to \begin{cases} \langle \omega, \mathcal{V}_m \rangle - (S_m - b) \leq \varepsilon + \xi_m \\ S_n - b - \langle \omega, \mathcal{V}_m \rangle \leq \varepsilon + \xi_m^* \\ \xi_m, \xi_m^* \geq 0, \end{cases} \end{aligned} \tag{1}$$

where C balances ω, ξ_m, and ξ_m^*. The $\omega = \sum_{m=1}^{M} t_m V_m$, where t_m is a combination coefficient. The $\mathcal{V}$ is mapped into $\Phi(\mathcal{V})$,

$$\begin{aligned} \mathfrak{f}(\mathcal{V}) &= \left\langle \sum_{m=1}^{M} t_m \Phi(\mathcal{V}_m), \Phi(\mathcal{V}) \right\rangle + b \\ &= \sum_{m=1}^{M} t_m \langle \Phi(\mathcal{V}_m), \Phi(\mathcal{V}) \rangle + b. \end{aligned} \tag{2}$$

For the RBF kernel,

$$\mathfrak{f}(\mathcal{V}) = \sum_{m=1}^{M} t_m exp(-\gamma(|\mathcal{V}_m - V|)^2) + b, \tag{3}$$

where, the γ is the precision parameter.

3. Results

3.1. Ps Image Database

In experiments, an image database originated from the IKONOS satellite images and assessed by human observers in the study of Agudelo-Medina et al. [1] is used. In that work, five regions of interest were used. However, this study effectively uses only four of them since in its shared version only full sets of their images are associated with subjective scores. It is worth noticing that this dataset is the largest image collection of PS images assessed by human observers and can be employed to thoroughly compare PS methods as well as techniques used for their assessment. The dataset contains 171 PS images (TC and PC pairs) obtained from four reference images using six PS methods (IHS [4]—28,

BDSD [37]—24, PCA [3]—28, MTF-GLP-CBD [38]—28, HPF [5]—28, and ATWT-M2 [39]—31) and one interpolation method (EXT [40]—4). Additionally, images are distorted with blur and additive white Gaussian noise [1]. The following regions are considered: Coliseum, Road, Urban, River, and Villa. However, due to low number of released Villa images, they are only used in the training subsets. Each region of 256 × 256 × 4 pixels for MS and 1024 × 1024 pixels for PAN was extracted from the image of the city of Rome. Subjective scores are assigned to the TC and PC images. Hence, to use both values for the training of a method on the PS images, their geometric mean is employed (DMOS-GM). The dataset also contains four undistorted PS images of the extracted IKONOS scenes. Exemplary undistorted and blurred TC and PC images of the IKONOS Roma Urban scene are presented in Figure 2.

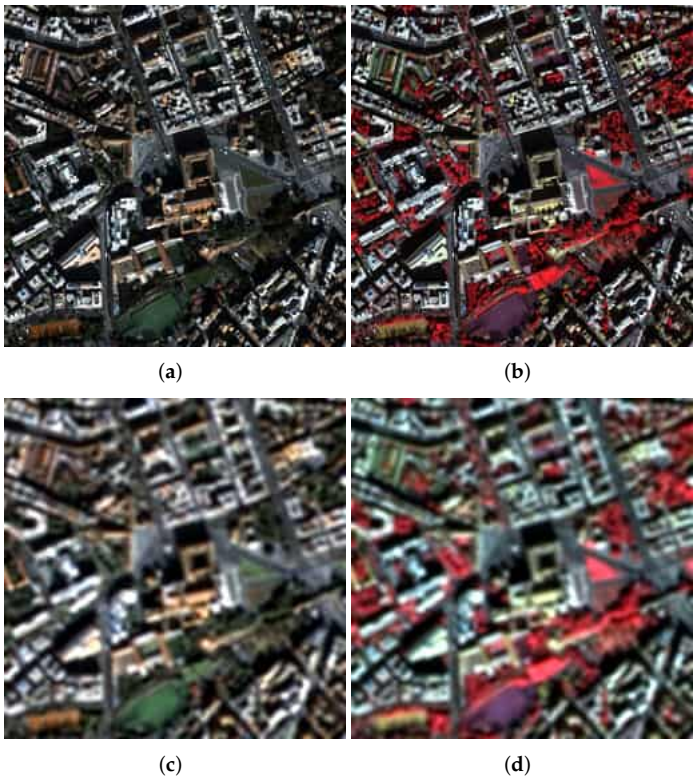


Figure 2. True color (**a**,**c**) and pseudo color (**b**,**d**) images of undistorted (**a**,**b**) and blurred (**c**,**d**) IKONOS Roma Urban scene.

3.2. Experimental Protocol

The proposed method is evaluated using a typical protocol used for the comparison of IQA approaches. In the protocol, four evaluation criteria are used: Spearman Rank-order Correlation Coefficient (SRCC), Kendall Rank Order Correlation Coefficient (KRCC), Pearson Linear Correlation Coefficient (PLCC), and Root Mean Square Error (RMSE) [41]. The criteria are calculated between predicted scores returned by a method and DMOS-GM. The higher correlation value and lower RMSE denote a better IQA method. It is worth noticing that the RMSE is often used to evaluate quality PS images, employed as an FR method. However, similarly to other IQA studies, it is used in this work to assess prediction accuracy, together with the PLCC. The SRCC and KRCC evaluate prediction monotonicity [41]. Since the MLPSIE requires training, images from the database are divided into training and testing subsets, and the evaluation criteria are reported as medians calculated for the testing subsets. Consequently, the methods that do not require training are only using test images in the experiments. Various experimental scenarios are considered, taking into account the number of reference images, random division of examples, or distortion

types. To support the results, statistical significance tests are reported as well as a discussion on the capability of the best methods to sort outputs of PS approaches in comparison to quality scores of human observers (DMOS-GM).

3.3. Comparison of Nr Methods

The proposed method is experimentally compared with 17 state-of-the-art approaches with available source codes: BRISQUE [26], CurveletQA [42], FRIQUEE [43], GMLOG [44], GWH-GLBP [24], NFERM [27], NOREQI [45], Oracle [46], SCORER [47], SISBLIM [48], BPRI [23], SINDEX [49], dipIQ [50], Blinder [28], Q_{oa} [1], ERGAS [20], and R50GR18 [29]. Since the Q_{oa} is reported to outperform other PS IQA methods, the comparison covers also the FR ERGAS technique that yields promising results in many experiments. The Q_{oa} is a training-based technique, similarly to MLPSIE, which allows for the comparison of the IQA capabilities of features used in both solutions. Among the remaining methods, R50GR15, dipIQ, and BLINDER are deep learning-based measures. However, only implementations of the R50GR15 and BLINDER can be trained while the dipIQ does not offer such functionality. Hence, it is evaluated on the testing images as other methods that do not require such a step (CurveletQA, BPRI, SINDEX, and ERGAS). All approaches that originate from the IQA of natural or medical images extract features from PC and TC images and after their concatenation appropriate regression models are trained to provide a quality prediction. Hence, in this work, popular IQA methods are adapted to the PS IQA. In the cases of CurveletQA, BPRI, SINDEX, and ERGAS, TC and PC images are evaluated and their scores are averaged to provide overall quality scores for PS images. The approaches are run in Matlab R2021a, Windows 10, on a PC with an i9-12900k CPU, 128 GB RAM, and an RTX 3090 graphic card. In the MLPSIE, the GPU extracts the features from the networks, while the CPU determines the multi-level image representations and predicts image quality. The SVR parameters of relevant methods are obtained using grid search.

Since there are four base PS scenes, i.e., Coliseum, Road, Urban, and River, in the first experimental scenario, all images that belong to one scene are used for testing while the remaining images train the NR methods. The median evaluation criteria from four tests are reported in Table 1. As presented, the introduced MLPSIE outperforms other approaches by a large margin. It is followed by CurvletQA and NFERM. Interestingly, the CurvletQA does not require training which can be seen as an additional advantage. Taking into account IQA methods designed for the PS images, ERGAS outperforms Q_{oa} for all evaluation criteria and is on par with other deep learning approaches trained on PS images (Blinder and R50GR18).

To provide a more thorough examination, similar to the scenario proposed by Agudelo-Medina et al. [1] along with the dataset, the entire image collection is randomly divided into disjoint training and testing samples (80%:20%), disregarding the scene of origin. The evaluation criteria are reported as the median values resulting from 1000 such divisions. The results for the compared methods can be seen in Table 2. In this scenario, the MLPSIE is the leading technique, with Q_{oa} as the second approach, followed by NOREQI and BRISQUE. This experiment favors methods with powerful features and the capability of creating a quality model. Hence, a simpler ERGAS or deep learning model with pre-trained implementation (dipIQ) obtain inferior results in these tests. To provide additional insight and test whether the indicated differences among results are statistically significant, the Wilcoxon rank-sum test is conducted. This test is considered with a 5% significance level and measures the equivalence of the median value of the samples [51]. In the experiment, a method with a significantly higher SRCC median obtained a score of "1", the worse "−1", and indistinguishable "0". Finally, the results are added to highlight the best approaches (Figure 3). As presented, the statistical significance tests confirm the results shown in Table 2, indicating the best performance of the MLPSIE and promising results for Q_{oa}.

Table 1. Comparison of performance of approaches for different scenes.

Method	SRCC	KRCC	PLCC	RMSE
BRISQUE	0.9007	0.7473	0.9655	1.9523
CurveletQA	**0.9443**	**0.8098**	**0.9720**	**1.6473**
FRIQUEE	0.8183	0.6336	0.9124	3.2830
GMLOG	0.9200	0.7698	0.9653	2.1162
GWH-GLBP	0.8766	0.7121	0.9377	2.7140
NFERM	**0.9373**	**0.7981**	**0.9698**	**1.8099**
NOREQI	0.8779	0.7110	0.9511	2.5011
Oracle	0.7882	0.5879	0.4137	7.1081
SCORER	0.7955	0.5857	0.7728	4.6842
SISBLIM	0.8026	0.6135	0.8808	3.8421
BPRI	0.4599	0.2957	0.6107	5.8090
SINDEX	0.8749	0.6848	0.9118	3.1256
dipIQ	0.6204	0.4762	0.6821	5.6095
Blinder	0.9115	0.7647	0.9650	1.8550
R50GR18	0.8003	0.6257	0.8136	4.0225
Q_{oa} *	0.8971	0.7307	0.9381	2.7947
ERGAS *°	0.9265	0.7796	0.9658	1.9560
MLPSIE *	**0.9548**	**0.8295**	**0.9803**	**1.5797**

Note: Three best results for each criterion are written in bold; Approach with '*' is designed for the PS IQA, while ° denotes a FR-IQA method.

Table 2. Comparison of methods using 1000 random database splits into training and testing samples.

Method	SRCC	KRCC	PLCC	RMSE
BRISQUE	0.9646	0.8610	**0.9844**	**1.3989**
CurveletQA	0.9092	0.7576	0.9488	2.5022
FRIQUEE	0.8921	0.7255	0.9105	3.2761
GMLOG	0.9326	0.7968	0.9556	2.3363
GWH-GLBP	0.9429	0.8146	0.9564	2.3137
NFERM	0.8946	0.7433	0.9435	2.6367
NOREQI	**0.9670**	**0.8645**	**0.9826**	1.4676
Oracle	0.7385	0.5401	0.6025	6.6669
SCORER	0.7328	0.5259	0.7706	5.0609
SISBLIM	0.6796	0.4866	0.7926	4.7989
BPRI	0.3563	0.2442	0.4792	6.9266
SINDEX	0.7792	0.5936	0.8426	4.2549
dipIQ	0.5568	0.4421	0.6565	5.9519
Blinder	0.9426	0.8075	0.9451	2.5781
R50GR18	0.9610	0.8431	0.9693	1.9243
Q_{oa} *	**0.9688**	**0.8681**	0.9823	**1.4568**
ERGAS *°	0.6321	0.4688	0.7482	5.2766
MLPSIE *	**0.9749**	**0.8895**	**0.9878**	**1.2359**

Note: Three best results for each criterion are written in bold. Approach with '*' is designed for the PS IQA, while ° denotes a FR-IQA method.

The third experimental scenario considers the division of images based on distortion types, i.e., using images affected by two distortion types for training and the remaining distortion for testing. The results reported in Table 3 evidence that MLPSIE is among the three best methods in terms of evaluation criteria for undistorted and blurred images. However, its performance for images distorted with Additive White Gaussian Noise (AWGN) is on par with Q_{oa}, behind FRIQUEE, Blinder, or NFERM which by design take into account this distortion type, since it is common in natural images.

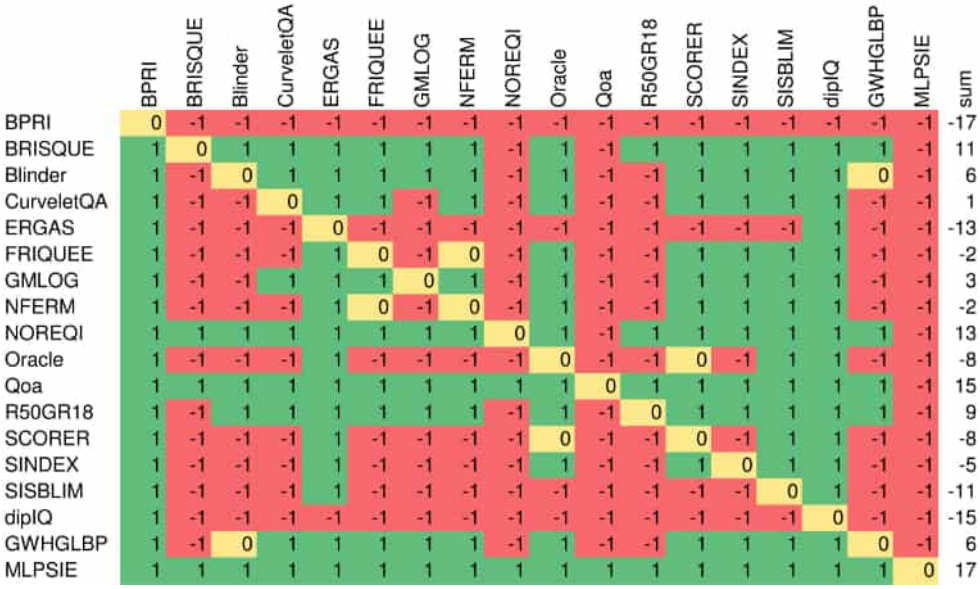


Figure 3. Results of statistical significance tests.

Table 3. Comparison of approaches considering different distortion types.

	UD				BLR				AWGN			
Method	**SRCC**	**KRCC**	**PLCC**	**RMSE**	**SRCC**	**KRCC**	**PLCC**	**RMSE**	**SRCC**	**KRCC**	**PLCC**	**RMSE**
BRISQUE	0.9217	0.7619	0.9796	2.3142	**0.8760**	**0.7062**	**0.9008**	**2.3757**	0.7162	0.5321	0.8245	3.2687
CurveletQA	0.9398	0.8042	0.9787	2.3672	0.7885	0.5992	0.8490	2.8896	0.4229	0.3005	0.4579	5.1348
FRIQUEE	0.8331	0.6402	0.8953	5.1334	0.7508	0.5541	0.8179	3.1469	**0.7809**	**0.6049**	**0.8799**	**2.7445**
GMLOG	0.8610	0.7143	0.9640	3.0649	**0.8610**	**0.6724**	**0.8913**	**2.4798**	0.6262	0.4476	0.7821	3.6134
GWH-GLBP	0.9414	0.7989	**0.9826**	**2.1427**	0.7388	0.5445	0.8284	3.0633	0.4497	0.3130	0.5005	5.0004
NFERM	0.9119	0.7566	0.9649	3.0271	0.8203	0.6370	0.8559	2.8283	**0.7691**	**0.5790**	**0.8584**	**2.9625**
NOREQI	**0.9562**	**0.8413**	**0.9847**	**2.0071**	0.6737	0.4897	0.7672	3.5079	0.4743	0.3419	0.5609	4.7816
Oracle	0.5534	0.3810	0.4637	10.4433	0.5824	0.4196	0.6402	4.4801	0.4453	0.3492	0.1295	5.7758
SCORER	0.7203	0.5397	0.8650	5.7804	0.5395	0.4197	0.3034	5.2115	0.6417	0.4563	0.7331	3.9340
SISBLIM	0.6273	0.5026	0.9216	4.4724	0.7201	0.5155	0.7963	3.3087	0.4966	0.2887	0.6241	4.5131
BPRI	0.7493	0.5291	0.8227	6.5507	0.3381	0.2306	0.3901	5.0361	0.2397	0.1541	0.3680	5.3705
SINDEX	0.6678	0.5238	0.8563	5.9500	0.8389	0.6539	0.8465	2.9120	0.7285	0.5290	0.7894	3.5453
dipIQ	0.1779	0.1429	0.8000	6.9126	0.8445	0.6628	0.8777	2.6213	0.2110	0.1886	0.6949	4.1536
Blinder	0.9294	0.7778	0.9623	3.1339	0.7914	0.6072	0.8347	3.0119	**0.8045**	**0.6369**	**0.8974**	**2.5483**
R50GR18	0.9223	0.7884	0.9818	2.1880	0.7931	0.6129	0.8669	2.7264	0.7116	0.5290	0.7751	3.6495
Q_{oa} *	**0.9502**	**0.8307**	0.9765	2.4809	0.8353	0.6636	0.8688	2.7112	0.7207	0.5266	0.8198	3.3079
ERGAS *°	0.5742	0.4286	0.8348	6.3432	0.5026	0.3561	0.6305	4.2450	0.3960	0.2838	0.6227	4.5193
MLPSIE *	**0.9425**	**0.8307**	**0.9876**	**1.8124**	**0.9393**	**0.7891**	**0.9488**	**1.7280**	0.7176	0.5313	0.7877	3.5586

Note: Three best results for each criterion are written in bold. Approach with '*' is designed for the PS IQA, while ° denotes a FR-IQA method.

3.4. Computational Complexity

Table 4 shows the computational complexity measured in terms of the average running time needed for image quality assessment of an image in the database. It can be seen that the MLPSIE is of moderate complexity. The most computationally demanding step in MLPSIE is associated with dimensionality reduction. Note that some of its steps can be performed in parallel to further reduce its computation time. In this experiment, ERGAS is the leading technique, followed by SINDEX or GMLOG. However, despite shorter running times, their IQA efficiency is far behind the proposed approach.

Table 4. Run-time comparison.

Method	Time(s)
BRISQUE	0.0057
CurveletQA	0.2457
FRIQUEE	1.7476
GMLOG	0.0053
GWH-GLBP	0.0076
NFERM	2.7032
NOREQI	0.0266
Oracle	0.0722
SCORER	0.0721
SISBLIM	0.1563
BPRI	0.1096
SINDEX	0.0048
dipIQ	0.7501
Blinder	0.6605
R50GR18	0.0079
Q_{oa}	0.2898
ERGAS	0.0006
MLPSIE	0.7117

3.5. Ablation Tests

Since in the MLPSIE the KPCA reduces the high-dimensional vectors, the influence of the dimensionality of the resulted feature vector on the performance of the method should be examined. Therefore, the experiment in which PS images are randomly split into training and testing samples 1000 times (see Table 2) is conducted with the dimensionality of vectors for networks' layers ranging from 5 to 20 with the step of 5. As reported in Figure 4, the performance of the MLPSIE is stable for different values, and the employed dimensionality of 15 seems a reasonable choice.

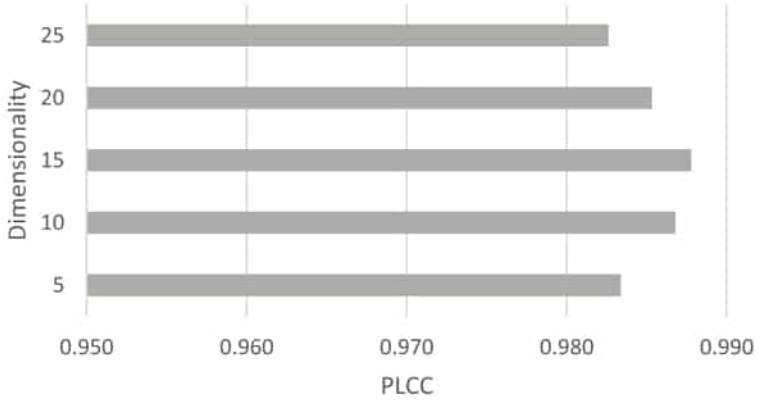


Figure 4. Median PLCC values for the MLPSIE with different dimensionality of reduced vectors.

The same experiment is also employed to show the distinctiveness of convolutional layers in exemplary network architecture. As presented in Figures 5 and 6, most data extracted from layers and reduced by the KPCA is of high importance for the performance of the MLPSIE. The layers with lower SRCC values are likely to be appropriately weighted by the SVR while training the model. Nevertheless, even with single short vectors for layers, the method outperforms many compared techniques (cmp. Table 2). The usage of both networks in the fusion of their such multi-level image representations is responsible for their outstanding performance.

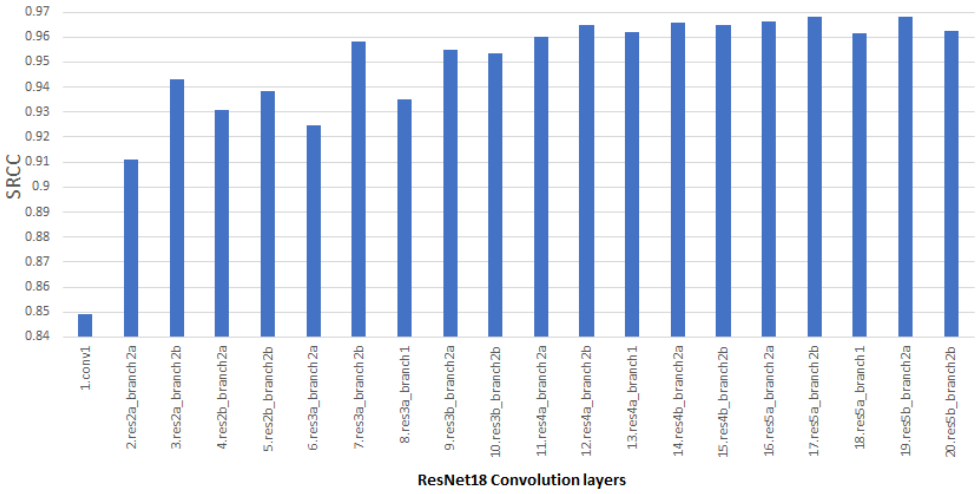


Figure 5. Performance of features in convolutional layers of the ResNet18 network in terms of the SRCC criterion.

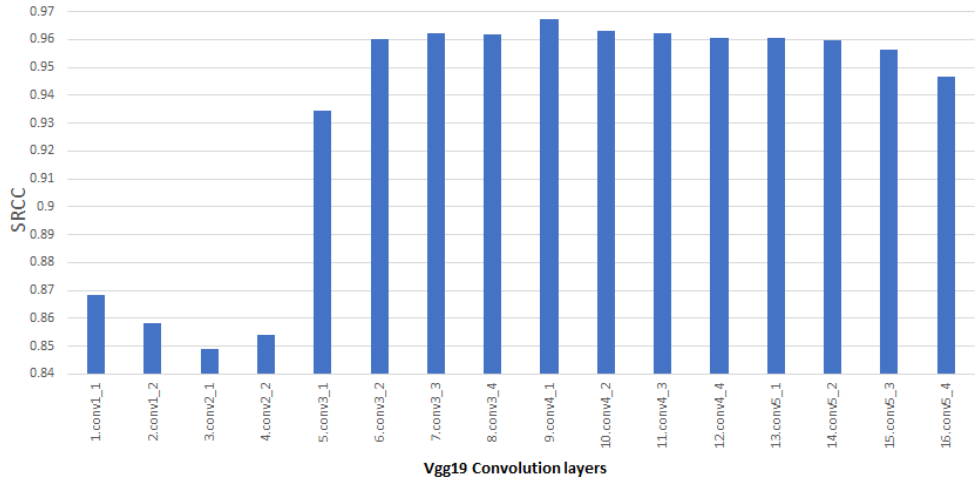


Figure 6. Performance of features in convolutional layers of the VGG19 network in terms of the SRCC criterion.

Since the MLPSIE uses two networks, their complementarity should be compared with those of other network alternatives. Therefore, the experiment with random dataset division is performed considering single networks of reasonably low complexity (VGG19, ResNet18, Alexnet, and SqueezeNet) and their combinations. As shown in Table 5, the employed fusion of the VGG19 and ResNet18 is the most beneficial, in terms of almost all evaluation criteria. However, other network pairs or even single networks also exhibit promising performance. Interestingly, as the networks represent different approaches to deep-learning-based image processing and offer different features, their results are similar, justifying the introduced in this work way of creating and using multi-level image representations as reduced high-dimensional feature vectors extracted from the network layers. Such representations, disregarding the difference in architectures of deep-learning methods that are used to extract features, lead to superior results when compared with other IQA approaches used for the quality prediction of PS images.

Table 5. Comparison of the performance of single and fused networks.

Method	SRCC	KRCC	PLCC	RMSE
VGG19	0.9742	0.8859	0.9852	1.3420
ResNet18	0.9759	0.8877	0.9860	1.2944
AlexNet	0.9697	0.8681	0.9779	1.6430
SqueezeNet	0.9728	0.8824	0.9863	1.3099
AlexNet + ResNet18	0.9746	0.8859	0.9866	1.2839
AlexNet + SqueezeNet	0.9740	0.8859	0.9866	1.2859
MLPSIE (ResNet18 + VGG19)	0.9749	0.8895	0.9878	1.2359

The employed channel configuration in MLPSIE assumes that a PS image is transformed into PC and TC image pair, each characterized by three channels since the considered network architectures are devoted to processing RGB images. The used configuration is proposed by Alparone et al. [52] and used in subjective tests with human observers in the work of Agudelo-Medina et al. [1]. However, as presented in Table 6, the MLPSIE can be successfully used with other channel configurations. Here, the combination of the NIR channel with any of two RGB components provides promising results, while creating input composed of three copies of the NIR channel decreases the performance. Consequently, it can be assumed that the quality assessment of images that contain more than channels would be possible once the RGB components are mixed with non-RGB channels of such an image. This would also require the addition of more VGG19 and ResNet18 pairs. The promising results obtained for different combinations of channels, as well as various deep learning backbones (Table 5), allow assuming that the proposed would be suitable, directly or after adaptation and fine-tuning, for applications that require the processing of more channels or involve a fusion of multi-sensor data [53,54].

Table 6. Influence of input channel configuration on the quality prediction performance of the MLPSIE. The pair (RGB, NIR + RG) is the default MLPSIE input configuration.

Channel Config.	SRCC	KRCC	PLCC	RMSE
RGB, NIR + RG	0.9749	0.8895	0.9878	1.2359
RGB, NIR + GB	0.9756	0.8895	0.9878	1.2273
RGB, NIR + RB	0.9756	0.8895	0.9879	1.2278
RGB, NIR + NIR + NIR	0.9743	0.8859	0.9871	1.2503

To show the capability of obtained multi-level image representations to distinguish images of different quality, two-dimensional t-SNE embeddings [55] of features are shown in Figure 7. The figure contains multi-level image representations of images in the dataset for separate (Figure 7b,c) and jointly (Figure 7a) considered networks. To facilitate the visualization with a limited palette of colors, the DMOS-GM scores are scaled. Presented scatter plots for all networks evidence that features allow for the distinction of images that belong to different locations (Coliseum, Road, Urban, and River). At this point in the image processing pipeline, image quality is not considered in the method, since the SVR responsible for the quality prediction is using these features. Hence, the clusters of images of similar quality, reflected by dots of the same or close color in the plots, confirm that the multi-level image representations are sensitive to image distortions and can be used for quality prediction. The t-SNE embeddings for the ResNet and VGG18 are different, and despite irregular cluster boundaries, in most cases, they can be easily differentiated. For the comparison, feature vectors of well-performing Q_{oa} and BRISQUE are also presented. As reported, they have difficulties in clustering images of similar quality. However, in the visualization for the Q_{oa}, image clusters of different quality are better distinguished than it can be seen for the BRISQUE.

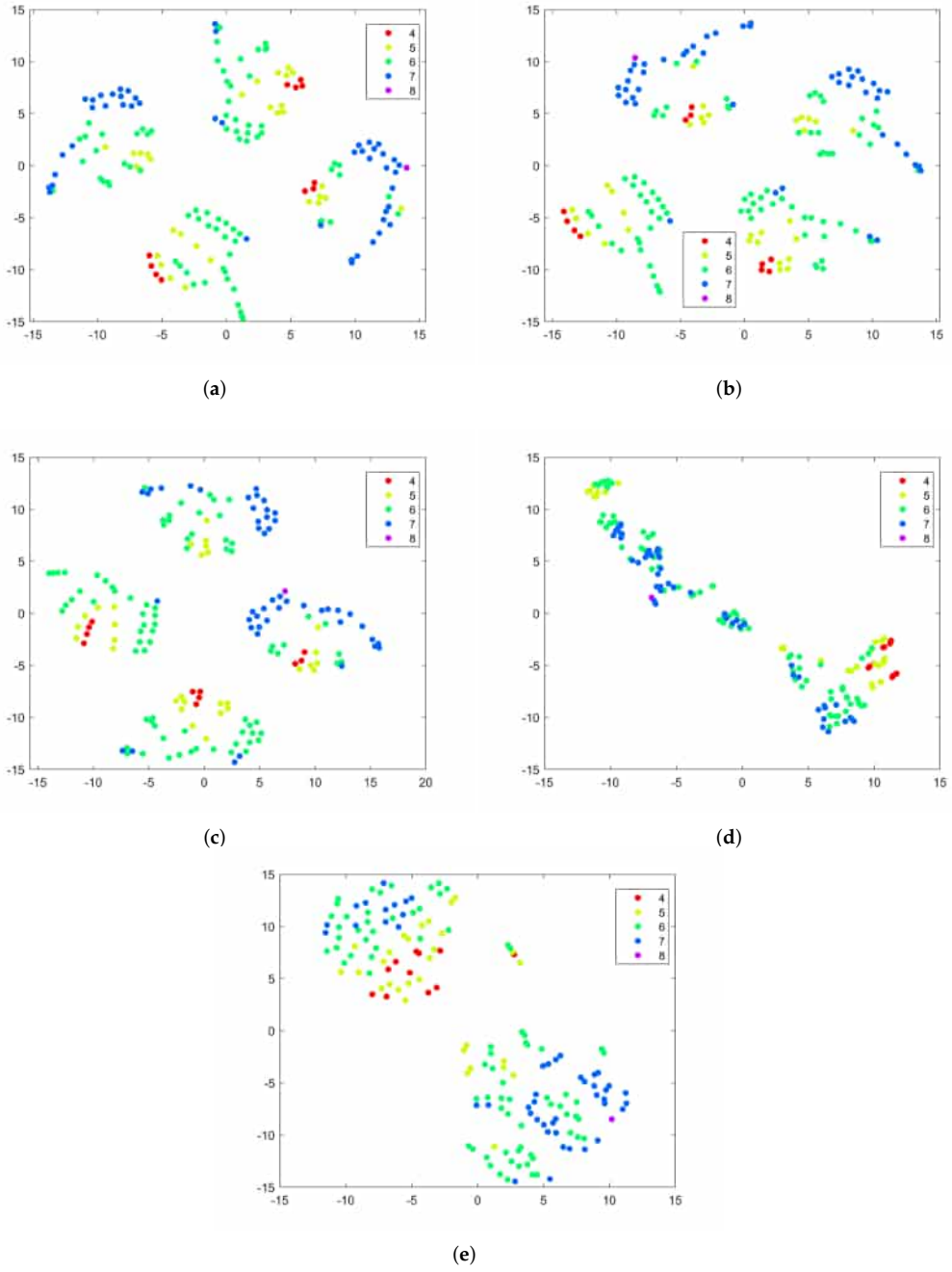


Figure 7. Scatter plots with two-dimensional t-SNE embeddings of MLPSIE multi-level features for the PS images of four categories in the dataset (**a**), VGG19 (**b**), and ResNet18 (**c**). Embeddings for the BRISQUE (**d**) and Q_{oa} (**e**) are also shown. Colors indicate scaled DMOS-GM scores for images.

3.6. Ranking of Ps Methods

Previous tests consider large data quantities to determine the general performance of the methods. However, to support such tests, the ability of the best approaches to select the best PS method and rank their outputs similarly to human observers is also considered. In Tables 7–10, scores returned by the best methods indicated in previous experiments, as well as subjective scores (DMOS-GM) and the resultant quality precedence are reported. Four distorted images are separately considered in this experiment. Blinder, MLPSIE, and Q_{oa} were trained on three remaining scenes and their distorted equivalents. The precedence of images produced by different PS methods is shown and written in bold

to facilitate analysis. The tables also contain the number of consistent scores obtained by IQA methods with subjective scores. It can be seen from tables that the MLPSIE correctly assessed the precedence of images placing them 18 out of 28 times. This test is particularly challenging since many images are similar according to the DMOS scores and their differentiation requires powerful quality-aware features and a quality model. Here, the Q_{oa} is the second-best technique with 16 correctly placed images. The remaining methods, ERGAS and Blinder, were able to determine the correct position of 15 images. Interestingly, the MLPSIE and Q_{oa} determined the best PS image only two times, and all techniques similarly identified the worst images.

To support the analysis, Table 9 is accompanied by TC images for IKONOS Rome River scene (Figure 8), with magnified selected parts of the images.

Table 7. Subjective (DMOS-GM) and objective scores for IKONOS Rome Urban scene. The lower part of the table contains precedence images produced by PS methods.

	MLPSIE	**Blinder**	Q_{OA}	**ERGAS**	**DMOS-GM**
PCA	51.69	54.22	47.40	27.16	38.90
MTF-GLP-CBD	53.28	54.21	48.00	26.91	39.84
BDSD	53.1	54.13	49.08	27.34	39.93
IHS	51.96	54.24	49.41	26.12	40.69
HPF	54.55	54.96	51.33	34.70	49.84
ATWT-M2	57.82	55.83	49.81	45.74	60.21
EXP	63.02	56.64	58.97	55.40	71.38
PCA	**PCA**	BDSD	**PCA**	IHS	
MTF-GLP-CBD	IHS	**MTF-GLP-CBD**	BDSD	**MTF-GLP-CBD**	
BDSD	**BDSD**	PCA	ATWT-M2	PCA	
IHS	MTF-GLP-CBD	**IHS**	**IHS**	BDSD	
HPF	**HPF**	**HPF**	MTF-GLP-CBD	**HPF**	
ATWT-M2	**ATWT-M2**	**ATWT-M2**	HPF	**ATWT-M2**	
EXP	**EXP**	**EXP**	**EXP**	**EXP**	
Number of scores consistent with DMOS	5	5	3	4	

Table 8. Subjective (DMOS-GM) and objective scores for IKONOS Rome Road scene. The lower part of the table contains precedence images produced by PS methods.

	MLPSIE	**Blinder**	Q_{OA}	**ERGAS**	**DMOS-GM**
IHS	54.48	56.46	53.14	26.67	31.97
BDSD	55.06	55.56	49.04	25.93	32.53
PCA	54.09	55.92	53.01	27.89	33.19
MTF-GLP-CBD	55.44	55.73	48.97	25.78	35.50
HPF	56.89	56.36	55.36	32.06	42.37
ATWT-M2	60.31	57.45	60.97	42.40	53.83
EXP	64.47	58.13	66.90	52.90	64.98
IHS	PCA	BDSD	MTF-GLP-CBD	MTF-GLP-CBD	
BDSD	IHS	MTF-GLP-CBD	**BDSD**	**BDSD**	
PCA	BDSD	**PCA**	**PCA**	IHS	
MTF-GLP-CBD	**MTF-GLP-CBD**	HPF	IHS	PCA	
HPF	**HPF**	IHS	**HPF**	**HPF**	
ATWT-M2	**ATWT-M2**	**ATWT-M2**	**ATWT-M2**	**ATWT-M2**	
EXP	**EXP**	**EXP**	**EXP**	**EXP**	
Number of scores consistent with DMOS	4	3	5	4	

Table 9. Subjective (DMOS-GM) and objective scores for IKONOS Rome River scene. The lower part of the table contains precedence images produced by PS methods.

	MLPSIE	**Blinder**	Q_{OA}	**ERGAS**	**DMOS-GM**
PCA	55.36	53.95	34.89	25.14	40.14
BDSD	56.19	53.66	38.43	24.27	40.37
IHS	55.41	54.25	33.36	25.08	40.63
MTF-GLP-CBD	56.28	53.78	35.57	24.87	41.83
HPF	57.89	54.74	42.93	33.62	48.82
ATWT-M2	61.07	56.10	46.43	44.78	56.96
EXP	64.66	56.62	59.31	53.23	68.50
PCA	**PCA**	BDSD	IHS	BDSD	
BDSD	IHS	MTF-GLP-CBD	PCA	MTF-GLP-CBD	
IHS	BDSD	PCA	MTF-GLP-CBD	**IHS**	
MTF-GLP-CBD	**MTF-GLP-CBD**	IHS	BDSD	PCA	
HPF	**HPF**	**HPF**	**HPF**	**HPF**	
ATWT-M2	**ATWT-M2**	**ATWT-M2**	**ATWT-M2**	**ATWT-M2**	
EXP	**EXP**	**EXP**	**EXP**	**EXP**	
Number of scores consistent with DMOS	5	3	3	4	

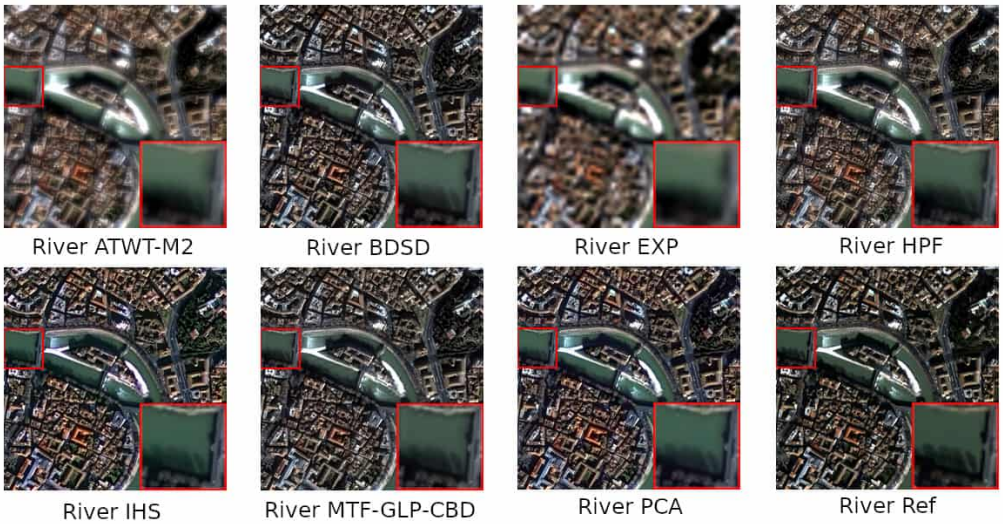


Figure 8. TC images obtained with PS methods compared with the reference (Ref) for IKONOS Rome River scene.

Table 10. Subjective (DMOS-GM) and objective scores for IKONOS Rome Coliseum scene. The lower part of the table contains precedence images produced by PS methods.

	MLPSIE	**Blinder**	Q_{OA}	**ERGAS**	**DMOS-GM**
BDSD	60.43	55.99	49.29	35.33	34.07
PCA	59.17	56.54	56.30	47.13	34.26
IHS	59.70	56.67	53.10	37.25	34.74
MTF-GLP-CBD	60.64	56.06	51.64	37.29	34.81
HPF	61.28	57.09	56.50	40.35	44.65
ATWT-M2	63.64	58.47	60.34	50.00	53.64
EXP	66.82	59.16	63.67	56.96	62.91

Table 10. *Cont.*

	MLPSIE	Blinder	Q_{OA}	ERGAS	DMOS-GM
BDSD	PCA	**BDSD**	**BDSD**	**BDSD**	
PCA	IHS	MTF-GLP-CBD	MTF-GLP-CBD	IHS	
IHS	BDSD	PCA	**IHS**	MTF-GLP-CBD	
MTF-GLP-CBD	**MTF-GLP-CBD**	IHS	PCA	HPF	
HPF	**HPF**	**HPF**	**HPF**	PCA	
ATWT-M2	**ATWT-M2**	**ATWT-M2**	**ATWT-M2**	**ATWT-M2**	
EXP	**EXP**	**EXP**	**EXP**	**EXP**	
Number of scores consistent with DMOS	4	4	5	3	

4. Conclusions

In this paper, a novel NR IQA method is proposed aimed at quality prediction of PS images. In the method, multi-level representations of PS images offered by two deep learning network architectures are employed as quality-aware features to provide a successful quality model. Since the extracted features from the networks are high-dimensional, they are reduced using the KPCA technique in a layer-wise manner, taking into account the joint reduction of information that describes TC and PC images. As the extensive experimental comparison with 17 approaches reveals, the proposed approach outperforms related methods as well as IQA approaches adapted in this work to the PS domain and can be used to rank PS techniques based on the quality of fused images.

Future work will be focused on the combination of deep learning and classical IQA methods for the IQA of PS images. An organization of subjective tests that include outputs of more PS methods compared to using several high-resolution satellite images, such as QuickBird or WorldView, is also considered.

The code of the introduced MLPSIE is available at http://marosz.kia.prz.edu.pl/MLPSIE.html, accessed on 13 January 2022.

Author Contributions: Conceptualization, I.S. and M.O.; methodology, M.O.; software, I.S.; validation, I.S. and M.O.; investigation, I.S. and M.O.; writing and editing, I.S. and M.O.; supervision: M.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank Oscar A. Agudelo-Medina, Hernan Dario Benitez-Restrepo, Gemine Vivone, and Alan Bovik for sharing their dataset.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Agudelo-Medina, O.A.; Benitez-Restrepo, H.D.; Vivone, G.; Bovik, A. Perceptual Quality Assessment of Pan-Sharpened Images. *Remote Sens.* **2019**, *11*, 877. [CrossRef]
2. Govind, N.R.; Rishikeshan, C.A.; Ramesh, H. Comparison of Different Pan Sharpening Techniques using Landsat 8 Imagery. In Proceedings of the 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 29–31 March 2019; pp. 1–4. [CrossRef]
3. Du, Q.; Gungor, O.; Shan, J. Performance evaluation for pan-sharpening techniques. In Proceedings of the 2005 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seoul, Korea, 29–29 July 2005; pp. 4264–4266. [CrossRef]
4. Te-Ming.; Huang.; Chung-Ling.; Chien-Ping. A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery. *IEEE Geosci. Remote Sens. Lett.* **2004**, *1*, 309–312. [CrossRef]
5. Jat, M.; Garg, P.; Dahiya, S. A comparative study of various pixel based image fusion techniques as applied to an urban environment. *Int. J. Image Data Fusion* **2013**, *4*, 197–213. [CrossRef]
6. Ballester, C.; Caselles, V.; Verdera, J.; Rouge, B. A Variational Model for P+XS Image Fusion. *Int. J. Comput. Vis.* **2006**, *69*, 43–58. [CrossRef]
7. Huang, W.; Xiao, L.; Wei, Z.; Liu, H.; Tang, S. A new pan-sharpening method with deep neural networks. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1037–1041. [CrossRef]

8. Kau, L.J.; Lee, T.L. An HSV Model-Based Approach for the Sharpening of Color Images. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 13–16 October 2013; pp. 150–155. [CrossRef]
9. Jawak, S.D.; Luis, A.J. A Comprehensive Evaluation of PAN-Sharpening Algorithms Coupled with Resampling Methods for Image Synthesis of Very High Resolution Remotely Sensed Satellite Data. *Adv. Remote Sens.* **2013**, *2013*, 332–344. [CrossRef]
10. Jing, L.; Cheng, Q. Two improvement schemes of PAN modulation fusion methods for spectral distortion minimization. *Int. J. Remote Sens.* **2009**, *30*, 2119–2131. [CrossRef]
11. Laben, C.A.; Brower, B.V. Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpening. U.S. Patent 6,011,875, 4 January 2000.
12. Alparone, L.; Aiazzi, B.; Baronti, S.; Garzelli, A.; Nencini, F.; Selva, M. Multispectral and Panchromatic Data Fusion Assessment Without Reference. *ASPRS J. Photogramm. Eng. Remote Sens.* **2008**, *74*, 193–200. [CrossRef]
13. Vivone, G.; Dalla Mura, M.; Garzelli, A.; Restaino, R.; Scarpa, G.; Ulfarsson, M.O.; Alparone, L.; Chanussot, J. A New Benchmark Based on Recent Advances in Multispectral Pansharpening: Revisiting Pansharpening With Classical and Emerging Pansharpening Methods. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 53–81. [CrossRef]
14. Khan, M.; Alparone, L.; Chanussot, J. Pansharpening Quality Assessment Using the Modulation Transfer Functions of Instruments. *Geosci. Remote Sens. IEEE Trans.* **2009**, *47*, 3880–3891. [CrossRef]
15. Zhou.; Bovik. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]
16. Vivone, G.; Alparone, L.; Chanussot, J.; Dalla Mura, M.; Garzelli, A.; Licciardi, G.A.; Restaino, R.; Wald, L. A Critical Comparison Among Pansharpening Algorithms. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2565–2586. 2361734. [CrossRef]
17. Alcaras, E.; Parente, C.; Vallario, A. Automation of Pan-Sharpening Methods for Pleiades Images Using GIS Basic Functions. *Remote Sens.* **2021**, *13*, 1550. [CrossRef]
18. Zhang, Y.; Atkinson, P.M.; Ling, F.; Foody, G.M.; Wang, Q.; Ge, Y.; Li, X.; Du, Y. Object-Based Area-to-Point Regression Kriging for Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 8599–8614. [CrossRef]
19. Sarp, G. Spectral and spatial quality analysis of pan-sharpening algorithms: A case study in Istanbul. *Eur. J. Remote Sens.* **2014**, *47*, 19–28. [CrossRef]
20. Kim, M.; Holt, J.; Madden, M. Comparison of Global- and Local-scale Pansharpening for Rapid Assessment of Humanitarian Emergencies. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 51–63. [CrossRef]
21. Javan, F.D.; Samadzadegan, F.; Reinartz, P. Spatial Quality Assessment of Pan-Sharpened High Resolution Satellite Imagery Based on an Automatically Estimated Edge Based Metric. *Remote Sens.* **2013**, *5*, 6539–6559. [CrossRef]
22. Alimuddin, I.; Sumantyo, J.T.S.; Kuze, H. Assessment of pan-sharpening methods applied to image fusion of remotely sensed multi-band data. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 165–175.
23. Min, X.; Gu, K.; Zhai, G.; Liu, J.; Yang, X.; Chen, C.W. Blind quality assessment based on pseudo-reference image. *IEEE Trans. Multimed.* **2017**, *20*, 2049–2062. [CrossRef]
24. Li, Q.; Lin, W.; Fang, Y. No-Reference Quality Assessment for Multiply-Distorted Images in Gradient Domain. *IEEE Signal Process. Lett.* **2016**, *23*, 541–545. [CrossRef]
25. Ma, K.; Liu, W.; Zhang, K.; Duanmu, Z.; Wang, Z.; Zuo, W. End-to-End Blind Image Quality Assessment Using Deep Neural Networks. *IEEE Trans. Image Process.* **2018**, *27*, 1202–1213. [CrossRef] [PubMed]
26. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [CrossRef] [PubMed]
27. Gu, K.; Lin, W.; Zhai, G.; Yang, X.; Zhang, W.; Chen, C.W. No-reference quality metric of contrast-distorted images based on information maximization. *IEEE Trans. Cybern.* **2016**, *47*, 4559–4565. [CrossRef]
28. Gao.; Yu.; Zhu.; Huang.; Tian. Blind image quality prediction by exploiting multi-level deep representations. *Pattern Recognit.* **2018**, *81*, 432–442. [CrossRef]
29. Stępień, I.; Obuchowicz, R.; Piórkowski, A.; Oszust, M. Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment. *Sensors* **2021**, *21*, 1043. [CrossRef]
30. Ieremeiev, O.; Lukin, V.; Okarma, K.; Egiazarian, K. Full-Reference Quality Metric Based on Neural Network to Assess the Visual Quality of Remote Sensing Images. *Remote Sens.* **2020**, *12*, 2349. [CrossRef]
31. Napoletano, P.; Piccoli, F.; Schettini, R. Anomaly Detection in Nanofibrous Materials by CNN-Based Self-Similarity. *Sensors* **2018**, *18*, 209. [CrossRef]
32. Pires de Lima, R.; Marfurt, K. Convolutional Neural Network for Remote-Sensing Scene Classification: Transfer Learning Analysis. *Remote Sens.* **2020**, *12*, 86. [CrossRef]
33. Wang, W.; Zhang, M.; Wang, D.; Jiang, Y. Kernel PCA feature extraction and the SVM classification algorithm for multiple-status, through-wall, human being detection. *EURASIP J. Wirel. Commun. Netw.* **2017**, *2017*, 151. [CrossRef]
34. Kemker, R.; Salvaggio, C.; Kanan, Ch. Algorithms for Semantic Segmentation of Multispectral Remote Sensing Imagery using Deep Learning. *arXiv* **2018**, arXiv:1703.06452.
35. Xie.; Girshick.; Dollar.; Tu.; He. Aggregated Residual Transformations for Deep Neural Networks. *arXiv* **2017**, arXiv:1611.05431.
36. Licciardi, G.; Vivone, G.; Dalla Mura, M.; Restaino, R.; Chanussot, J. Multi-resolution analysis techniques and nonlinear PCA for hybrid pansharpening applications. *Multidimens. Syst. Signal Process.* **2016**, *27*, 807–830. [CrossRef]

37. Vivone, G. Robust Band-Dependent Spatial-Detail Approaches for Panchromatic Sharpening. *IEEE Trans. Geosci. Remote. Sens.* **2019**, *9*, 6421–6433. [CrossRef]
38. Jiao, J.; Wu, L.; Qian, K. A Segmentation-Cooperated Pansharpening Method Using Local Adaptive Spectral Modulation. *Electronics* **2019**, *8*, 685. [CrossRef]
39. Li, J.; Hu, Q.; Ai, M. Multispectral and panchromatic image fusion based on spatial consistency. *Int. J. Remote Sens.* **2018**, *39*, 1017–1041. [CrossRef]
40. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2300–2312. [CrossRef]
41. Sheikh, H.; Sabir, M.; Bovik, A. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Trans. Image Process.* **2006**, *15*, 3440–3451. [CrossRef]
42. Ahmed, I.T.; Der, C.S. Enhancement of no-reference image quality assessment for contrast-distorted images using natural scene statistics features in Curvelet domain. In Proceedings of the 2017 7th IEEE International Conference on System Engineering and Technology (ICSET), Shah Alam, Malaysia, 2–3 October 2017; pp. 128–133. [CrossRef]
43. Ghadiyaram, D.; Bovik, A.C. Perceptual Quality Prediction on Authentically Distorted Images Using a Bag of Features Approach. *arXiv* 2016, arXiv:1609.04757.
44. Abdul, R. New Learning Frameworks for Blind Image Quality Assessment Model. Ph.D. Thesis, University of Sheffield, Sheffield, UK, 2018.
45. Oszust, M. No-reference image quality assessment using image statistics and robust feature descriptors. *IEEE Signal Process. Lett.* **2017**, *24*, 1656–1660. [CrossRef]
46. Babonneau, F.; Beltran, C.; Haurie, A.; Tadonki, C.; Vial, J.P. Proximal-ACCPM: A versatile oracle based optimisation method. In *Optimisation, Econometric and Financial Analysis*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 67–89._4 [CrossRef]
47. Liu, T.J.; Liu, K.H. No-Reference Image Quality Assessment by Wide-Perceptual-Domain Scorer Ensemble Method. *IEEE Trans. Image Process.* **2018**, *27*, 1138–1151. [CrossRef]
48. Gu, K.; Zhai, G.; Yang, X.; Zhang, W. Hybrid no-reference quality metric for singly and multiply distorted images. *IEEE Trans. Broadcast.* **2014**, *60*, 555–567. [CrossRef]
49. Leclaire, A.; Moisan, L. No-reference image quality assessment and blind deblurring with sharpness metrics exploiting fourier phase information. *J. Math. Imaging Vis.* **2015**, *52*, 145–172. [CrossRef]
50. Ma, K.; Liu, W.; Liu, T.; Wang, Z.; Tao, D. dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs. *IEEE Trans. Image Process.* **2017**, *26*, 3951–3964. [CrossRef] [PubMed]
51. de Barros, R.S.M.; Hidalgo, J.I.G.; de Lima Cabral, D.R. Wilcoxon rank sum test drift detector. *Neurocomputing* **2018**, *275*, 1954–1963. [CrossRef]
52. Alparone, L.; Wald, L.; Chanussot, J.; Thomas, C.; Gamba, P.; Bruce, L.M. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3012–3021. [CrossRef]
53. Ehlers, M.; Klonus, S.; Astrand, P.J.; Rosso, P. Multi-sensor image fusion for pansharpening in remote sensing. *Int. J. Image Data Fusion* **2010**, *1*, 25–45. [CrossRef]
54. Kizel, F.; Benediktsson, J.A. Spatially Enhanced Spectral Unmixing Through Data Fusion of Spectral and Visible Images from Different Sensors. *Remote Sens.* **2020**, *12*, 1255. [CrossRef]
55. Hinton, G.; Roweis, S. Stochastic Neighbor Embedding. In Proceedings of the 15th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 9–14 December 2002; MIT Press: Cambridge, MA, USA, 2002; pp. 857–864.

Engineering Applications of Artificial Intelligence 139 (2025) 109594

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Research paper

Three-branch neural network for No-Reference Quality assessment of Pan-Sharpened Images

Igor Stępień, Mariusz Oszust *

Doctoral School of the Rzeszow University of Technology, al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland
Department of Computer and Control Engineering, Rzeszow University of Technology, Wincentego Pola 2, 35-959 Rzeszow, Poland

ARTICLE INFO

Keywords:
Remote sensing
Pan-sharpening
Image quality assessment
Neural networks
Three-branch neural network
Pan-sharpening quality prediction

ABSTRACT

Pan-Sharpening (PS) techniques aim to enhance the spatial resolution of low-resolution multispectral images by leveraging data from high-resolution panchromatic images. Their comparison typically relies on the quality assessment of the resulting Full-Resolution (FS) pan-sharpened images. However, in the absence of a reference image, a dedicated No-Reference (NR) method must be employed. Therefore, this paper introduces a novel approach called the Three-Branch Neural Network for No-Reference Quality Assessment of Pan-Sharpened Images (TBN-PSI). The network consists of three subnetworks designed for perceptual processing of image channels, featuring shared extraction of low-level features and high-level semantics. Extensive experimental evaluation demonstrates the superiority of the approach over the state-of-the-art NR PS image quality assessment methods, using six datasets containing diverse satellite images that span urban areas, green vegetation, and water scenarios. Specifically, TBN-PSI outperforms the compared methods by 4% to 9% in terms of Spearman's Rank-Order Correlation Coefficient (SRCC), Pearson's Linear Correlation Coefficient (PLCC), and Kendall's Rank Correlation Coefficient (KRCC) between the obtained scores and those of three representative full-reference methods.

1. Introduction

Pan-Sharpening (PS) aims to enhance the spatial resolution of a low-spatial-resolution (LR) multispectral (MS) image by leveraging a high-spatial-resolution (HR) panchromatic (PAN) image, resulting in a higher-resolution MS image (Huang et al., 2015). Due to incoming radiation energy, onboard storage capacity, or data transmission constraints, a compromise between the spatial and spectral resolutions of the images is often necessary (Bao et al., 2022; Nia et al., 2015). The state-of-the-art PS methods are typically categorized into Multiresolution Analysis (MRA), Component Substitution (CS), Color-Based (CB), Deep Learning (DL), Variational Optimization (VO), and hybrid approaches (Rahmani et al., 2010; Serifoglu Yilmaz et al., 2022). However, it is important to note that there is no standardized categorization of these methods, as different studies apply varying classification criteria. CS-based methods represent the most widely adopted pan-sharpening techniques, marking a shift from traditional to more universal approaches (Tu et al., 2001). In the conventional paradigm, the MS bands are transformed into a new space via spectral manipulation, and one component is replaced with the high-resolution PAN image. An inverse projection is then performed to obtain the fused image.

Methods based on Multiresolution Analysis (MRA) typically follow a three-step process. First, the source images are decomposed into multiple scale levels using wavelet, curvelet, or pyramid transforms. Next, spectral and spatial features are combined at different decomposition levels using a fusion rule. Finally, an inverse transform is applied to generate the PS data within the original color domain (Ghassemian, 2016; Wang et al., 2023). CB methods, on the other hand, employ specific procedures to preserve the color information inherent in the source MS data (Yilmaz, 2021). VO-based PS methods are grounded in variational theory, where the primary process involves optimizing an energy function (Meng et al., 2021). In DL-based pan-sharpening methods, a network architecture is typically proposed and trained (Vivone et al., 2021; Wang et al., 2024; Zhou et al., 2023), while hybrid methods integrate diverse techniques (Dadrass Javan et al., 2018; Yilmaz, 2023).

Among popular approaches, the Intensity-Hue-Saturation (IHS) method (Rahmani et al., 2010) replaces the intensity component of the MS image with that of the PAN image. Other fusion methods, such as Principal Component Analysis (PCA) (Shah et al., 2008), the Gram–Schmidt (GS) method (Dalla Mura et al., 2015), and the Brovey method (Gharbia et al., 2014), are based on different transformation

* Corresponding author.
E-mail addresses: istepien@kia.prz.edu.pl (I. Stępień), marosz@kia.prz.edu.pl (M. Oszust).

https://doi.org/10.1016/j.engappai.2024.109594
Received 28 June 2024; Received in revised form 20 October 2024; Accepted 30 October 2024
Available online 12 November 2024

techniques. Adaptive methods, including Band-Dependent Spatial Detail (BDSD) (Garzelli et al., 2007) and Partial Replacement Adaptive Component Substitution (PRACS) (Choi et al., 2010), are also widely discussed in the literature. The Additive Wavelet Transform Model 2 (ATWT-M2) (Ranchin and Wald, 2000) focuses on decomposing the source PAN data to extract its spatial structure content, which is then integrated into the MS image. Additionally, Modulation Transfer Function (MTF)-based pan-sharpening methods have garnered significant interest due to their promising performance (Alparone et al., 2007).

The objective of Image Quality Assessment (IQA) methods is to provide automated, consistent, and precise evaluation of images, aiming to replace human observer-based assessments (Li et al., 2022; Rubel et al., 2022; Sheikh et al., 2006). Over the past few decades, substantial research has been dedicated to this field, resulting in the development of novel algorithms across various facets of IQA (Gabarda et al., 2018). IQA is typically divided into subjective and objective quality evaluation (Opozda and Sochan, 2014). Subjective evaluation depends on human perception, where evaluators rate the overall quality of images, considering factors like blurriness or distortions. The Mean Opinion Score (MOS) (Sheikh et al., 2006) is then obtained based on these assessments. However, this approach is time-consuming, as it requires human evaluators to manually assess images.

In contrast, objective quality evaluation uses algorithms and automated processes to assess image quality, predicting perceptual quality while eliminating subjectivity. Objective IQA methods are classified into three categories based on the availability of reference images: Full-Reference (FR) approaches, which require access to a source image for comparison (Athar and Wang, 2023); Reduced-Reference (RR) methods, which need only partial reference information; and No-Reference (NR) or "blind" approaches, which are the most challenging, as they predict image quality without any reference image (Muthusamy and Sathyamoorthy, 2023). In FR and RR approaches, accurate evaluation of pan-sharpened images has significant implications for secure image processing, particularly in cases where high-resolution data must be enhanced without compromising the integrity or authenticity of the original data (Koziel et al., 2016). NR-IQA methods frequently attempt to replicate characteristics of the human visual system or employ features sensitive to image distortions to assess quality (Agudelo-Medina et al., 2019; Firmansyah et al., 2024; Singh Kushwah et al., 2022; Wang et al., 2004).

In this paper, a novel NR PS IQA method is proposed. The Three-Branch Neural Network for No-Reference Quality Assessment of Pan-Sharpened Images (TBN-PSI) is designed with three subnetworks dedicated to processing PS images. These subnetworks share information while extracting both low-level features and high-level semantic information. The evaluation of PS IQA methods is crucial for monitoring environmental changes, such as deforestation, urban expansion, and water quality. Developing robust pan-sharpened image quality assessment techniques enhances decision-making processes related to environmental protection and management. High-quality satellite imagery is particularly essential for observing natural disasters like hurricanes, floods, and earthquakes, enabling more accurate damage assessments and effective disaster response strategies. Beyond environmental applications, reliable satellite IQA plays a critical role in AI-driven decision support systems, ensuring the production of high-quality images, thus improving the operational efficiency of dependent systems.

The key contributions of this study are:

- The development of a network architecture for PS IQA that incorporates subnetworks sharing information during training via an additional middle branch.
- The adaptation of various IQA methods, including efficient deep learning techniques, to effectively assess the quality of PS images.
- Comprehensive experiments conducted on six datasets of PS images to evaluate and compare the performance of the proposed approach against state-of-the-art methods.

The remainder of this paper is organized as follows: Section 2 reviews related works, Section 3 introduces the proposed method, and Section 4 presents an experimental comparison with state-of-the-art IQA methods. Finally, Section 5 summarizes the findings and outlines future research directions.

2. Related works

In the literature, quality metrics such as Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS) (Wald, 2002), Spectral Angle Mapper (SAM) (Yuhas et al., 1992), and $Q2^n$ (Garzelli and Nencini, 2009) are commonly used to evaluate PS images. These methods require a high-resolution multispectral reference image and can serve as substitutes for human scores when such evaluations are not available. ERGAS is calculated as a weighted sum of the Root of the Mean Square Error (RMSE):

$$ERGAS = 100\frac{g}{k}\sqrt{\frac{1}{S}\sum_{s=1}^{S}(\frac{RMSE(X_s, Y_s)}{E(X_s)})^2}, \tag{1}$$

where X denotes the predicted fused images, Y is the reference image, S is the number of image channels, g denotes spatial resolutions of the PAN image k is the spatial resolution of MS image, and

$$RMSE(X, Y) = \sqrt{E[(X - Y)^2]}, \tag{2}$$

where $E[\cdot]$ is the average value of all pixels. SAM measures the angle between the reference and processed vectors of a specific pixel within the spectral feature space (Yuhas et al., 1992):

$$SAM(MS, F) = Arccos(\frac{\sum_{i=1}^{S} MS_i F_i}{\sqrt{\sum_{i=1}^{S} MS_i MS_i}\sqrt{\sum_{i=1}^{S} F_i F_i}}), \tag{3}$$

where MS_i is the spectral vector of the multispectral images, F_i is the spectral vector of the fused images, and S denotes the number of image bands. SCC extracts the high-frequency details of PAN and fused images using a high-pass 2-D filter. Subsequently, it employs a correlation coefficient to assess the spatial similarity between the PAN and fused images (Pushparaj and Hegde, 2017; Sulaiman et al., 2020). It is calculated as follows (Chang and Wang, 2011):

$$SCC(Y, X) = \frac{\sum\sum(Y_i - \gamma_Y)^2(X_i - \gamma_X)^2}{\sqrt{\sum(Y_i - \gamma_Y)^2 \sum(X_i - \gamma_X)^2}}, \tag{4}$$

where Y is the PAN image, X is the fused image, and γ_Y with γ_X are their means. SAM and ERGAS are used for spectral evaluations, while SCC assesses the spatial quality validation of pan-sharpened images.

The requirement for a reference image presents a significant limitation in quality assessment applications, especially when a high-resolution reference image is unavailable. This challenge has driven the development of NR-IQA techniques for pan-sharpened images. For example, the Universal Image Quality Index (Qq) models image distortion as a combination of loss of correlation, luminance distortion, and contrast changes (Wang and Bovik, 2002). In the Opinion-Aware Analyzer (Q_{oa}), a regression module is employed to map a quality-aware feature space of scores, thereby generating an opinion-aware quality model (Agudelo-Medina et al., 2019). Recently, Scarpa and Ciotola (2022) introduced two NR full-resolution assessment methods: the Reprojection Protocol for Spectral Accuracy Assessment, which uses Khan's index and reprojection-based indexes with embedded alignment, and the Correlation-Based Spatial Consistency Index, which complements the reprojection-based protocol by focusing on spatial consistency evaluation.

In the field of deep learning methods, Badal et al. (2022) proposed Deep Pan-Sharpened IQA (DPIQA), utilizing an NR deep learning approach to predict $Q2^n$ and SAM values. Another approach, SCC, optimizes weights in both the intensity image calculation and the approximation images derived from intensity and histogram-matched PAN images (Zhang, 2019). The Quality Without Reference (QNR)

Table 1
Characteristics of datasets used in the experiments.

Satellite sensors		Spatial dimension	Spectral dimension	Dimension size	Number of thematic scene images				
					Urban	Green vegetation	Water scenario	Unlabeled	Data volume
IKONOS	PAN MS	1 m 4 m	1 band 4 bands	1024 × 1024 256 × 256 × 4	60	8	6	126	200
QuickBird	PAN MS	0.61 m 2.44 m	1 band 4 bands	1024 × 1024 256 × 256 × 4	150	20	42	288	500
Gaofen-1	PAN MS	2 m 8 m	1 band 4 bands	1024 × 1024 256 × 256 × 4	5	90	10	305	410
WorldView-4	PAN MS	0.31 m 1.24 m	1 band 4 bands	1024 × 1024 256 × 256 × 4	90	85	95	230	500
WorldView-2	PAN MS	0.5 m 2 m	1 band 8 bands	1024 × 1024 256 × 256 × 8	150	35	145	170	500
WorldView-3	PAN MS	0.31 m 1.24 m	1 band 8 bands	1024 × 1024 256 × 256 × 8	55	20	20	65	160

index evaluates spectral and spatial distortions independently, combining these results for a comprehensive assessment (Khan et al., 2009). Similarly, the Hybrid Quality with No Reference (HQNR) (Aiazzi et al., 2014) primarily relies on the $Q2^n$ index for full-resolution assessment, incorporating spectral and spatial distortion information to evaluate pan-sharpened image quality. Research on $Q2^n$ has expanded to full-scale assessments via multiscale extrapolations (Carla et al., 2015). Kwan et al. (2017) introduced the GQNR index for assessing fused WorldView-3 images. Zhou et al. (2019) proposed a new method that considers both spatial and spectral distortions in fused images, integrating typical information indices such as the Normalized Difference Water Index (NDWI) and the Normalized Difference Vegetation Index (NDVI) to form a quality-aware feature representation. Mehravar et al. (2022) focused on spatial quality assessment of PS images by assigning weight factors to pixels based on their spatial information content. In the authors' previous approach, features from multiple layers of hybrid neural networks were extracted and reduced using Kernel Principal Component Analysis (KPCA), followed by concatenation and training of Support Vector Regression (SVR) for quality prediction (Stepien and Oszust, 2022). While this approach, which extracts information from multiple layers similar to classic hand-crafted methods, provides valuable insights, it is time and memory-intensive due to the large number of transformed features, thus limiting its applicability to relatively small datasets.

In the literature, numerous methods have been developed for the IQA of natural images, which can also be applied to perform NR PS IQA. Among them, the popular Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) extracts statistics from local luminance signals to assess the naturalness of an image based on distortion information (Mittal et al., 2012a). Another approach, the Natural Image Quality Evaluator (NIQE), builds upon a 'completely blind' IQA model (Mittal et al., 2012b). NIQE relies solely on measurable deviations from expected statistical regularities observed in high-quality natural images, thus eliminating the need for training on human-rated distorted images and avoiding exposure to or training on distorted images. With the advent of modern neural network-based approaches, an increasing number of methods leverage their capabilities. For instance, the transformer-based vision model (ViT) (Dosovitskiy et al., 2021) has been utilized by You and Korhonen (2021). In a similar model, a transformer with a multilayer perceptron head is employed, and the resulting feature maps are input into a shallow transformer encoder to address blind IQA tasks (Tang et al., 2023). Overall, the limited number of recent approaches to NR PS IQA, along with their availability and complexity, underscores the necessity for the development of new promising methods. In contrast to previously introduced methods in the literature, TBN-PSI does not focus solely on evaluating spatial consistency. Instead, it is designed to achieve effective sharpening by addressing both spectral and spatial consistency, particularly for complex pan-sharpened images. To facilitate the development of a robust architecture, several large satellite datasets were employed for its evaluation, as methods that perform efficiently on small datasets may fail to scale effectively to larger ones. Unlike many existing approaches that were initially developed for natural images and later adapted for PS images, TBN-PSI was specifically designed from the outset to accommodate the unique characteristics of PS imagery.

3. Proposed method

3.1. Architecture

The proposed TBN-PSI uses a predefined number of blocks of layers from the deep learning backbone network to create three connected subnetworks. Specifically, the proposed architecture is based on the first 87 layers of the 315 layers in the Inception-v3 network. The set of useful layers, gathered into blocks, has been determined experimentally, taking into account the image quality assessment perspective (see ablation tests in Section 4). Sets of blocks have been duplicated to handle two inputs of the network designated for RGB and NIR + GB channels of 299 × 299 size, as specified by the Inception-v3 input size. Furthermore, an additional branch has been added to process information from both inputs after each block of layers of the lower and upper branches, as shown in Fig. 1. This approach allows for fusing features of different granularity, from low-level information to high-level semantics. The data from the three branches is combined using the Depth Concatenation layer and subsequently processed through the Fully Connected, Global Average Pool, and regression layers.

The proposed TBN-PSI utilizes a predefined number of blocks of layers from the deep learning backbone network to create three interconnected subnetworks. Specifically, the architecture is based on the first 87 layers of the 315 layers in the Inception-v3 network. The selection of useful layers, organized into blocks, was determined experimentally with consideration for the image quality assessment perspective (see ablation tests in Section 4). The sets of blocks have been duplicated to accommodate two inputs designated for the RGB and NIR + GB channels, each of size 299 × 299, as specified by the Inception-v3 input size. Furthermore, an additional branch has been introduced to process information from both inputs after each block of layers in the lower and upper branches, as depicted in Fig. 1. This approach facilitates the fusion of features with varying granularity, ranging from low-level information to high-level semantics. The data from the three branches is combined using the Depth Concatenation layer and subsequently processed through the Fully Connected, Global Average Pool, and regression layers.

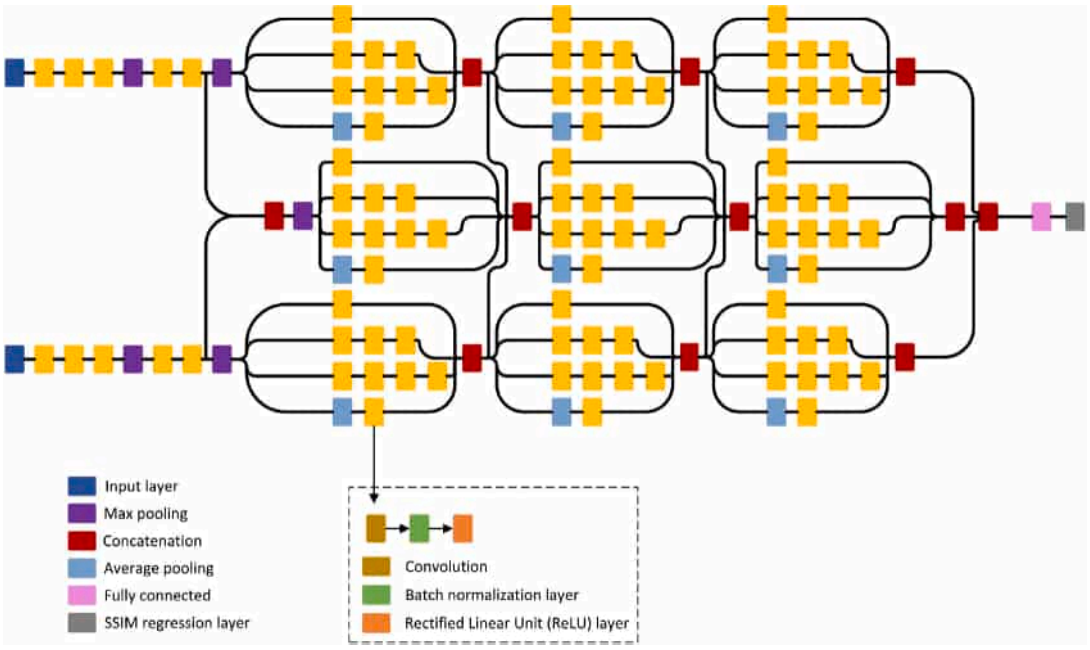


Fig. 1. Proposed network architecture with three-branch subnetworks and connections between branches.

3.2. Loss function

In this work, the regression layer also incorporates information on quality, as the Structural Similarity Index Measurement (SSIM) (Qi et al., 2020; Wang et al., 2004) is utilized as the loss function instead of the Mean Square Error (MSE). This application constitutes one of the contributions of this study. The SSIM index, calculated between two images x and y, is defined as follows:

$$SSIM(x,y) = \frac{(2v_x v_y + w_1)(2\rho_{xy} + w_2)}{(v_x^2 + v_y^2 + w_1)(\rho_x^2 + \rho_y^2 + w_2)}, \quad (5)$$

where v_x and v_y are the averages of pixel intensities, ρ_x^2 and ρ_y^2 their standard deviations, ρ_{xy} is their covariance, $w_1 = (k_1 L)^2, w_2 = (k_2 L)^2$ are two variables to stabilize the division with weak denomination, $L = 1$, $k_1 = 0.01$, and $k_2 = 0.01$.

4. Experiments and results

4.1. PS image datasets

The proposed approach is evaluated on the NBU PansharpRSData benchmark (Meng et al., 2021), which consists of six satellite datasets (see Fig. 2): IKONOS (Agudelo-Medina et al., 2019), Quickbird (Toutin and Cheng, 2002), Gaofen (Chen et al., 2022), WorldView-4 (WV4) (Sefercik et al., 2021), WorldView-3 (WV3) (Longbotham et al., 2015), and WorldView-2 (WV2) (Padwick et al., 2010). The distinct features of each satellite dataset included in the NBU PansharpRSData are presented in Table 1. In the experiments, popular PS methods were applied to pairs of PAN and MS images, including ATWT_M2 (Ranchin and Wald, 2000), BDSD (Garzelli et al., 2007), GS (Dalla Mura et al., 2015), IHS (Rahmani et al., 2010), MTF-GLP-CBD (Alparone et al., 2007), and PCA (Shah et al., 2008). Exemplary images of these methods are presented in Fig. 3.

Finally, for the IKONOS dataset, 1200 pan-sharpened images were created; for QuickBird, 3000 images; for Gaofen, 2460 images; for WorldView-4, 3000 images; for WorldView-2, 3000 images; and for WorldView-3, 960 images. Since subjective scores for the created PS images are not available, the SAM, ERGAS, and SCC metrics are employed to validate the image quality assessment approaches.

4.2. Experimental protocol

Three popular metrics are employed to evaluate the performance of the method: Spearman's Rank-Order Correlation Coefficient (SRCC), Pearson's Linear Correlation Coefficient (PLCC), and Kendall's Rank Correlation Coefficient (KRCC) (Sheikh et al., 2006). These coefficients are calculated between the predicted scores and the scores of the utilized spectral or spatial full-reference quality metrics (ERGAS, SAM, and SCC). A superior IQA metric yields higher values for these correlation coefficients.

In the experiments, 80% of the reference images, along with their corresponding distorted equivalents, are selected for the training set, while the remaining 20% of the images are allocated for the testing set (Min et al., 2017; Ye et al., 2012). The results are reported as the medians obtained from ten random splits.

4.3. Comparison with state-of-the-art methods

The proposed TBN-PSI is compared against nine approaches: $Q2^n$ (Garzelli and Nencini, 2009), Q (Wang and Bovik, 2002), Inception-v3 (Szegedy et al., 2016), ResNet50 (He et al., 2016), NIQE (Mittal et al., 2012b), BRISQUE (Mittal et al., 2012a), GoogleNet (Szegedy et al., 2015), Vision Transformer (ViT) (Pushparaj and Hegde, 2017), and Q_{oa} (Agudelo-Medina et al., 2019). $Q2^n$, Q, and Q_{oa} are image quality indexes (QIs) frequently applied in PS IQA. NIQE and BRISQUE are NR-IQA approaches designed to evaluate natural images and trained to assess PS images, while Inception-v3, ResNet50, GoogleNet, and ViT are CNN architectures finetuned for IQA of PS images by the authors of this work. The source code of the proposed method is available online (Stepien and Oszust, 2024).

The upper and lower branches in the proposed architecture utilize weights identical to those of the Inception-v3 layers trained on ImageNet, serving as a starting point for training the entire network. In contrast, the weights in the middle branch are initialized randomly due to the additional connections from the lower and upper branches. The network is trained using the Adaptive Moment Estimation (ADAM) optimizer (Kingma and Ba, 2014), with an initial learning rate of 10e−3, a mini-batch size of 32, and a total of 50 epochs. Training of the TBN-PSI is conducted without freezing any layers. The approaches are executed in Matlab R2023b on a Windows 10 system, utilizing a PC

Fig. 2. Exemplary PS images from NBU PansharpRSData benchmark.

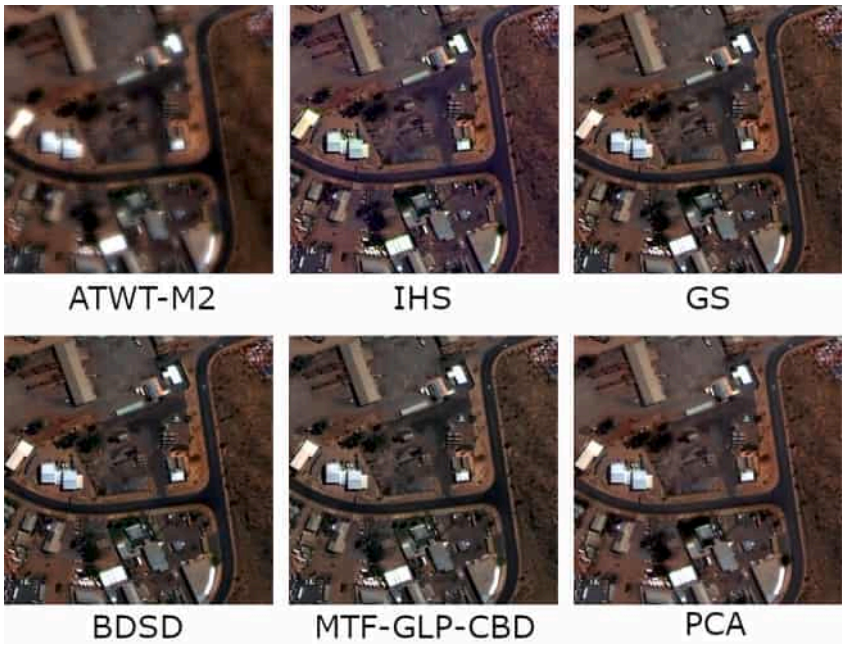


Fig. 3. Exemplary images obtained with PS methods.

equipped with an i9-12900k CPU, 128 GB of RAM, and an RTX 3090 graphics card. With this configuration, training the proposed TBN-PSI on the PS images based on WV3 takes approximately 820 s.

The performance of the proposed method compared with other approaches across six IQA datasets is presented in Table 2. For the SRCC criterion, TBN-PSI demonstrates superior performance on the IKONOS, Quickbird, WorldView-2, and WorldView-3 datasets. On the Quickbird and WorldView-2 datasets, it achieves the second-best performance, ranking between Inception-v3 and GoogleNet. When considering the KRCC, TBN-PSI outperforms the other methods twice and ranks as the second-best approach three times, following Inception-v3. Experiments conducted using the PLCC criterion further highlight the superiority of TBN-PSI for the IKONOS and WV3 datasets. For the remaining datasets, it ranks either second or fourth, with Inception-v3 as the leading approach. Focusing on the widely used IKONOS data, TBN-PSI achieves better SRCC results than the second-best method (Inception-v3, Q_{oa}) by 7%, and by 6% and 11% for KRCC and PLCC, respectively. Overall, the average capability of learning to mimic ERGAS scores is reflected in the ranks, which facilitate comparisons between approaches. As illustrated, TBN-PSI and GoogleNet are ranked second and third, respectively, following Inception-v3. However, the difference between the two leading approaches for PS IQA is approximately half a rank.

Since the methods demonstrated similar performances in many instances, the Wilcoxon rank-sum test is employed to compare the SRCC median values of independent samples at a significance level of 5%. For each dataset, a significantly better result for the method in the row compared to the method in the column receives a score of '1', while worse and indistinguishable results receive '−1' and '0', respectively. The scores across all six datasets are subsequently summed. Table 3

Table 2
Performance comparison of nine IQA algorithms and proposed TBN-PSI calculated with the ERGAS metric. The best three results for each dataset are bolded.

Criterion	Dataset	$Q2^n$	Q	Inception-v3	ResNet50	NIQE	BRISQUE	GoogleNet	ViT	Q_{oa}	TBN-PSI
SRCC	**IKONOS**	0.3400	0.3617	**0.8517**	0.7844	0.1052	0.7970	0.8159	0.3557	**0.8584**	**0.9198**
	Quickbird	0.3235	0.3036	**0.8848**	0.8407	0.2106	0.7741	**0.8528**	0.3012	0.8392	**0.8818**
	Gaofen	0.0435	0.2212	0.9620	**0.9716**	0.6933	**0.9724**	0.9554	0.7451	**0.9747**	0.9587
	WV4	0.0241	0.0190	**0.9024**	0.8815	0.0271	0.8138	**0.8898**	0.7206	0.8648	**0.9022**
	WV2	0.0847	0.0695	**0.9257**	0.9010	0.1428	0.7749	**0.9171**	0.7017	0.8795	**0.9215**
	WV3	0.0417	0.0307	**0.8657**	0.8510	0.1963	0.7376	**0.8727**	0.6253	0.8190	**0.9217**
KRCC	**IKONOS**	0.2574	0.2770	**0.6612**	0.5829	0.0797	0.5969	0.6235	0.2470	**0.6711**	**0.7542**
	Quickbird	0.2398	0.2244	**0.7031**	0.6405	0.1491	0.5742	**0.6604**	0.2048	0.6413	**0.6931**
	Gaofen	0.0562	0.2099	**0.9224**	**0.9273**	0.5647	0.9164	0.8747	0.5688	**0.9380**	0.8708
	WV4	0.0057	0.0011	**0.7446**	0.7118	0.0168	0.6228	**0.7257**	0.5308	0.6791	**0.7358**
	WV2	0.0315	0.0259	**0.7711**	0.7353	0.0921	0.5839	**0.7564**	0.4890	0.7058	**0.7667**
	WV3	0.0608	0.0552	**0.6961**	0.6859	0.1396	0.5659	**0.7072**	0.4512	0.6616	**0.7756**
PLCC	**IKONOS**	0.4144	0.4444	**0.8417**	0.7788	0.3102	0.7808	0.8173	0.3822	**0.8526**	**0.9076**
	Quickbird	0.4121	0.3990	**0.8685**	0.8223	0.3198	0.7838	0.8415	0.3376	**0.8500**	**0.8521**
	Gaofen	0.4165	0.5087	**0.9516**	0.9316	0.8486	0.9498	**0.9505**	0.8040	**0.9522**	0.8655
	WV4	0.2671	0.1164	**0.9498**	**0.9294**	0.1315	0.8475	**0.9378**	0.8129	0.9130	0.9039
	WV2	0.3653	0.1965	**0.9388**	**0.9135**	0.2595	0.7276	**0.9196**	0.6391	0.9025	0.9071
	WV3	0.3755	0.3812	**0.8414**	0.8374	0.4337	0.7739	**0.8575**	0.6025	0.7743	**0.9225**
Average rank	All	8.67	8.94	2.00	4.22	8.72	5.39	3.22	7.67	3.61	2.56

Table 3
Summary of statistical significance tests over six datasets for the ERGAS metric.

	BRISQUE	GoogleNet	Inception-v3	NIQE	$Q2^n$	Q	Q_{oa}	ResNet50	ViT	TBN-PSI	Sum
BRISQUE	0	−4	−5	6	6	6	−5	−4	6	−5	1
GoogleNet	4	0	−3	6	6	6	2	0	6	−3	24
Inception-v3	5	3	0	6	6	6	4	3	6	−2	37
NIQE	−6	−6	−6	0	−3	1	−6	−6	−6	−6	−44
$Q2^n$	−6	−6	−6	3	0	3	−6	−6	−4	−6	−34
Q	−6	−6	−6	−1	−3	0	−6	−6	−6	−6	−46
Q_{oa}	5	−2	−4	6	6	6	0	−1	6	−5	17
ResNet50	4	0	−3	6	6	6	1	0	6	−4	22
ViT	−6	−6	−6	6	4	6	−6	−6	0	−6	−20
TBN-PSI	5	3	2	6	6	6	5	4	6	0	**43**

presents the results for the ERGAS metric. As illustrated, the introduced TBN-PSI exhibits superior performance relative to other methods based on the overall number of statistically better results in pairwise comparisons, with Inception-v3 ranking as the second-best approach. Interestingly, although Inception-v3 is frequently indicated as the best method for the ERGAS metric (see Table 2), it produces scores that are comparable to those of other approaches. Consequently, TBN-PSI is statistically superior to the compared methods six times more often than Inception-v3.

The results for the SAM metric, as presented in Table 4, indicate significantly better performance for the proposed approach, with superior values across all three criteria in 14 instances. Consequently, TBN-PSI achieves the best average rank value (1.56), while Inception-v3 and Q_{oa} rank second and third on average, respectively. The four instances in which the proposed approach performed worse involve the PLCC criterion and the Gaofen, WV4, and WV2 datasets. For the widely used IKONOS data, TBN-PSI outperforms the second-best approach by 3.9% in terms of SRCC, and by 3.3% and 2% for KRCC and PLCC, respectively.

Table 5 summarizes the statistical significance tests for the SAM metric, supporting the superior performance of the proposed approach, which achieves a total sum of 48. In this context, Inception-v3 ranks as the second most effective method, with a total sum of 32.

Table 6 presents the performance of TBN-PSI in comparison to other methods based on the SCC metric. For the SRCC and KRCC criteria, TBN-PSI demonstrates superior performance across all evaluated datasets. Additionally, it yields promising results for the PLCC criterion in the IKONOS, Quickbird, Worldview-4, and Worldview-3 datasets. However, for the Gaofen dataset, it produces a slightly lower value than BRISQUE, Q*oa*, and Inception-v3. Furthermore, on the Worldview-2 dataset, TBN-PSI ranks second-best, positioned between Inception-v3 and Q*oa*. Considering the average ranks obtained by the various approaches, TBN-PSI outperforms state-of-the-art methods with an average rank of 1.22, followed by Q*oa* (2.0) and BRISQUE (2.89). Interestingly, Inception-v3, which exhibited strong results for the ERGAS and SAM metrics, is ranked fourth. For the widely used IKONOS data, TBN-PSI achieves better SRCC results than the second-best approach by 6.9%, and by 7.6% and 2.6% for KRCC and PLCC, respectively. The statistical significance tests (see Table 7) clearly indicate the superior performance of TBN-PSI, which obtains a sum of 52, followed by Q*oa* (40), BRISQUE (27), and Inception-v3 (13).

Overall, the results affirm the capability of the proposed approach to effectively mimic the scores of popular full-reference metrics, highlighting its potential utility in reflecting human quality perception.

4.4. Cross-dataset evaluation

An experiment was conducted across datasets to evaluate the generalization capability of the different approaches. In this experiment, the methods were trained on one dataset and subsequently tested on another. The PLCC, SRCC, and KRCC values are presented in Tables 8 and 9, with the best results highlighted in bold.

In the experiments, the approaches requiring training were trained on Worldview-3 and tested on Quickbird. The results, based on the SAM metric, highlight the performance of the proposed method. For most criteria (PLCC, SRCC, and KRCC), TBN-PSI consistently achieved either the best or second-best results, typically ranking between Inception-v3 and ResNet50. For both the ERGAS and SCC metrics, TBN-PSI excelled, particularly in the SRCC and KRCC criteria. Regarding the SAM metric, the proposed method outperforms all others across every criterion. However, the results for PLCC were slightly lower on the WorldView-3 dataset, which can be attributed to its relatively small size.

In the case in which Quickbird is used for the training, as outlined in Table 9, TBN-PSI outperforms all other methods across three criteria.

Table 4
Performance comparison of nine IQA algorithms and proposed TBN-PSI calculated with the SAM metric. The best three results for each dataset are bolded.

	Dataset	$Q2^n$	Q	Inception-v3	ResNet50	NIQE	BRISQUE	GoogleNet	ViT	Q_{oa}	TBN-PSI
SRCC	**IKONOS**	0.2127	0.2422	**0.8798**	0.8092	0.0891	0.8244	0.8409	0.4544	**0.9027**	**0.9400**
	Quickbird	0.2389	0.2240	**0.8756**	0.8325	0.0479	0.8114	**0.8614**	0.3342	0.8592	**0.9092**
	Gaofen	0.0014	0.2014	0.9568	0.9553	0.6863	**0.9687**	0.8890	0.3743	**0.9747**	**0.9797**
	WV4	0.3169	0.1722	**0.9440**	0.9241	0.3066	0.8736	**0.9261**	0.7755	0.9195	**0.9460**
	WV2	0.6087	0.5765	**0.9527**	0.9335	0.6334	0.8886	0.9438	0.7607	**0.9513**	**0.9685**
	WV3	0.1025	0.0851	**0.9089**	0.8826	0.1703	0.7638	0.8735	0.7008	**0.8843**	**0.9240**
KRCC	**IKONOS**	0.1589	0.1780	**0.7019**	0.6332	0.0671	0.6265	0.6567	0.3119	**0.7217**	**0.7899**
	Quickbird	0.2062	0.2028	**0.6932**	0.6376	0.0427	0.6147	0.6660	0.2266	**0.6749**	**0.7340**
	Gaofen	0.0475	0.1985	0.9098	0.8868	0.5511	**0.9335**	0.7711	0.1657	**0.9380**	**0.9132**
	WV4	0.2163	0.1214	**0.8076**	0.7780	0.2048	0.6997	**0.7861**	0.5834	0.7695	**0.8124**
	WV2	0.4019	0.3829	**0.8182**	0.7803	0.4231	0.7111	0.7984	0.5544	**0.8173**	**0.8544**
	WV3	0.0321	0.0303	**0.7515**	0.7183	0.1004	0.5862	0.6986	0.5098	**0.7211**	**0.7807**
PLCC	**IKONOS**	0.3600	0.3817	**0.8842**	0.8366	0.3429	0.8231	0.8512	0.4810	**0.9065**	**0.9301**
	Quickbird	0.4206	0.4127	**0.8759**	0.8140	0.2957	0.8086	0.8299	0.3463	**0.8663**	**0.8845**
	Gaofen	0.6066	0.4550	**0.9647**	**0.9530**	0.9007	**0.9539**	0.9530	0.8410	0.9522	0.9101
	WV4	0.2399	0.1706	**0.9527**	**0.9360**	0.2980	0.8625	0.9344	0.8121	0.9240	**0.9393**
	WV2	0.7509	0.7481	**0.9638**	0.9488	0.7138	0.9195	0.9532	0.7907	**0.9674**	**0.9637**
	WV3	0.6751	0.6779	**0.8970**	0.8860	0.6059	0.8692	0.8861	0.7121	**0.8943**	**0.9338**
Average rank	All	8.72	9.11	2.28	4.64	8.83	5.33	4.14	7.33	3.06	1.56

Table 5
Summary of statistical significance tests over six datasets for the SAM metric.

	BRISQUE	GoogleNet	Inception-v3	NIQE	$Q2^n$	Q	Q_{oa}	ResNet50	ViT	TBN-PSI	Sum
BRISQUE	0	−3	−5	6	6	6	−5	−3	5	−5	2
GoogleNet	3	0	−2	6	6	6	−2	−1	6	−6	16
Inception-v3	5	2	0	6	6	6	1	3	6	−3	32
NIQE	−6	−6	−6	0	−2	−2	−6	−6	−6	−6	−46
$Q2^n$	−6	−6	−6	2	0	2	−6	−6	−6	−6	−38
Q	−6	−6	−6	2	−2	0	−6	−6	−6	−6	−42
Q_{oa}	5	2	−1	6	6	6	0	2	6	−5	27
ResNet50	3	1	−3	6	6	6	−2	0	6	−5	18
ViT	−5	−6	−6	6	6	6	−6	−6	0	−6	−17
TBN-PSI	5	6	3	6	6	6	5	5	6	0	**48**

Table 6
Performance comparison of nine IQA algorithms and proposed TBN-PSI calculated with the SCC metric. The best three results for each dataset are bolded.

Criterion	Dataset	$Q2^n$	Q	Inception-v3	ResNet50	NIQE	BRISQUE	GoogleNet	ViT	Q_{oa}	TBN-PSI
SRCC	**IKONOS**	0.5930	0.6235	0.5609	0.3222	0.0742	**0.7410**	0.1746	0.1104	**0.8234**	**0.8842**
	Quickbird	0.3741	0.3696	0.6774	0.4431	0.1299	**0.8207**	0.4901	0.1424	**0.8539**	**0.9052**
	Gaofen	0.3313	0.1488	0.6657	0.6586	0.5056	**0.7962**	0.7366	0.4922	**0.7886**	**0.9239**
	WV4	0.0783	0.0722	0.7159	0.5278	0.1360	**0.7978**	0.6331	0.1997	**0.8409**	**0.8875**
	WV2	0.0724	0.0381	0.8367	0.6888	0.4335	**0.8587**	0.7791	0.0931	**0.9247**	**0.9388**
	WV3	0.4495	0.4554	0.5757	0.3177	0.1848	**0.7706**	0.3634	0.3623	**0.8901**	**0.9168**
KRCC	**IKONOS**	0.4700	0.4985	0.4012	0.2232	0.0552	**0.5642**	0.1135	0.0622	**0.6540**	**0.7079**
	Quickbird	0.2948	0.2986	0.4812	0.3073	0.1003	**0.6158**	0.3388	0.0933	**0.6648**	**0.7268**
	Gaofen	0.2397	0.0889	0.5640	0.5000	0.4126	**0.6529**	0.5577	0.3136	**0.6653**	**0.8128**
	WV4	0.0455	0.0567	0.5193	0.3665	0.0906	**0.5994**	0.4363	0.1336	**0.6445**	**0.7057**
	WV2	0.0047	0.0193	0.6438	0.5036	0.2966	**0.6636**	0.5850	0.0627	**0.7671**	**0.7930**
	WV3	0.3443	0.3501	0.4046	0.2221	0.1222	**0.5932**	0.2420	0.2663	**0.7239**	**0.7586**
PLCC	**IKONOS**	0.6541	0.6560	0.4780	0.3328	0.2825	**0.7544**	0.1848	0.1439	**0.8505**	**0.8734**
	Quickbird	0.5108	0.5676	0.6724	0.4413	0.2908	**0.8437**	0.4692	0.1433	**0.8616**	**0.9017**
	Gaofen	0.5308	0.4505	**0.9127**	0.6721	0.8079	**0.9590**	0.8598	0.6788	**0.9334**	0.8743
	WV4	0.1572	0.1827	0.6676	0.5119	0.2428	**0.7724**	0.5877	0.1818	**0.8330**	**0.8725**
	WV2	0.5351	0.5350	**0.8801**	0.7497	0.3810	0.8376	0.7909	0.0856	**0.9404**	**0.9060**
	WV3	0.6685	0.6992	0.5085	0.3598	0.2130	**0.7652**	0.3712	0.3118	**0.8978**	**0.9028**
Average rank	All	7.50	7.22	4.39	6.83	8.44	2.89	6.00	8.50	2.00	1.22

Overall, the proposed method demonstrates high robustness and generalization capabilities, exhibiting consistent and stable performance across various types of satellite data from different satellite platforms.

4.5. Ablation study

To investigate the relationship between the performance of the approach and design choices, ablation experiments were conducted to assess the impact of varying the network structure and applying different loss functions on the performance of TBN-PSI. In these experiments, the lengths of the Inception-V3 backbone were varied from 2 to 4 blocks, as shown in Table 10. The results clearly indicate that reducing the length of the Inception-V3 architecture positively influences performance, with benefits observed for networks comprising two or four blocks. Additionally, the findings highlight the advantage of using SSIM as a loss function over MSE.

To demonstrate the capability of network branches in distinguishing images of varying quality, a visualization of activations from selected layers in the branches, is presented, utilizing an input image characterized by differing quality levels. Each layer consists of 64 channels; however, only the first channel is employed for visualization purposes. A colormap is applied to enhance the visual identification of

Table 7
Summary of statistical significance tests over six datasets for the SCC metric.

	BRISQUE	GoogleNet	Inception-v3	NIQE	$Q2^n$	Q	Q_{oa}	ResNet50	ViT	TBN-PSI	Sum
BRISQUE	0	5	4	6	6	6	−5	5	6	−6	27
GoogleNet	−5	0	−4	5	3	3	−5	1	4	−6	−4
Inception-v3	−4	4	0	6	4	4	−5	5	5	−6	13
NIQE	−6	−5	−6	0	−4	−5	−6	−4	−3	−6	−45
$Q2^n$	−6	−3	−4	4	0	−2	−6	−3	−1	−6	−27
Q	−6	−3	−4	5	2	0	−6	−3	−1	−6	−22
Q_{oa}	5	5	5	6	6	6	0	5	6	−4	40
ResNet50	−5	−1	−5	4	3	3	−5	0	3	−6	−9
ViT	−6	−4	−5	3	1	1	−6	−3	0	−6	−25
TBN-PSI	6	6	6	6	6	6	4	6	6	0	**52**

Table 8
Evaluation results of models trained on Wordview-3 and tested on Quickbird datasets.

	SAM			ERGAS			SCC		
	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC
$Q2^n$	0.1033	0.0963	0.3562	0.1848	0.1348	**0.3231**	**0.2553**	**0.1977**	**0.4330**
Q	0.0828	0.0824	0.3571	0.1607	0.1197	**0.3215**	0.2430	**0.1926**	**0.4877**
Inception-v3	**0.4395**	**0.3073**	0.4240	**0.2298**	**0.1516**	0.2566	0.1832	0.1239	0.1987
ResNet50	**0.4230**	**0.2959**	**0.4293**	0.1914	0.1272	0.2393	0.1220	0.0815	0.1202
NIQE	0.0479	0.0427	0.2957	**0.2106**	**0.1491**	0.3198	0.1299	0.1003	0.2908
BRISQUE	0.0775	0.0400	**0.5761**	0.0721	0.0514	**0.3747**	**0.2650**	0.1820	0.3030
GoogleNet	0.3539	0.2448	0.3708	0.1420	0.0939	0.2610	0.1364	0.0904	0.1235
ViT	0.3686	0.2537	0.3300	0.2094	0.1427	0.2047	0.2406	0.1647	0.2079
Q_{oa}	0.3317	0.2278	0.3785	0.1131	0.0807	0.0595	0.1388	0.0981	**0.3253**
TBN-PSI	**0.4346**	**0.3084**	**0.4811**	**0.2148**	**0.1495**	0.2670	**0.2638**	**0.1885**	0.2439

Table 9
Evaluation results of models trained on Quickbird and tested on Wordview-3 datasets.

	SAM			ERGAS			SCC		
	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC
$Q2_n$	0.3334	0.2166	**0.6723**	0.0520	0.0095	0.3376	0.4510	**0.3496**	**0.6319**
Q	0.3304	0.2146	**0.6871**	0.0479	0.0071	0.3270	**0.4553**	**0.3565**	**0.6604**
Inception-v3	0.3849	**0.2685**	0.4815	0.1638	0.1081	0.4234	0.2236	0.1482	0.2004
ResNet50	**0.4236**	**0.3086**	0.4662	0.1326	0.0927	0.3768	0.1886	0.1230	0.1355
NIQE	0.1703	0.1004	0.6059	0.1963	0.1396	**0.4337**	0.1848	0.1222	0.2130
BRISQUE	0.0458	0.0178	0.4623	0.0857	0.0585	0.2012	0.0318	0.0267	0.3979
GoogleNet	**0.3958**	0.2666	0.4745	**0.2646**	**0.1816**	**0.4400**	0.2893	0.1945	0.2325
ViT	0.1390	0.1010	0.0879	**0.3314**	**0.2376**	0.3417	0.3794	0.2779	0.3292
Q_{oa}	0.1253	0.0753	0.5019	0.1134	0.0765	0.2839	**0.4557**	0.3137	0.5335
TBN-PSI	**0.6725**	**0.4945**	**0.8115**	**0.6348**	**0.4655**	**0.6256**	**0.7438**	**0.5435**	**0.7265**

Table 10
Ablation study of the proposed TBN-PSI on Worldview-3 dataset and the SAM metric.

Network variation, Loss Function	SRCC	KRCC	PLCC
Inception-v3, MSE	0.9089	0.7515	0.8970
Inception-v3, 2 blocks, MSE	0.4551	0.3221	0.5632
Inception-v3, 4 blocks, MSE	0.3950	0.2876	0.6100
Inception-v3, 2 blocks, SSIM	0.9129	**0.7639**	0.9286
Inception-v3, 3 blocks, SSIM	**0.9168**	0.7611	**0.9304**
Inception-v3, 4 blocks, SSIM	**0.9152**	**0.7669**	**0.9342**
TBN-PSI, MSE	0.4919	0.3462	0.5563
TBN-PSI, SSIM	**0.9240**	**0.7807**	**0.9338**

distinct regions within the images. As illustrated in Fig. 4, all branches, irrespective of the quality of the input image, exhibit layers with subtly different activations, underscoring the network's ability to extract rich information from the image. Moreover, the variations in activations within the same branches for images of different quality further demonstrate the network's sensitivity to image degradation. Consequently, the integrated processing of the extensive information gathered across all branches, combined with their sensitivity to distortions, positions this network as a promising solution for assessing the quality of pan-sharpened images, contributing to its superior performance.

Table 11
The comparison of computational load of the CNN-based methods.

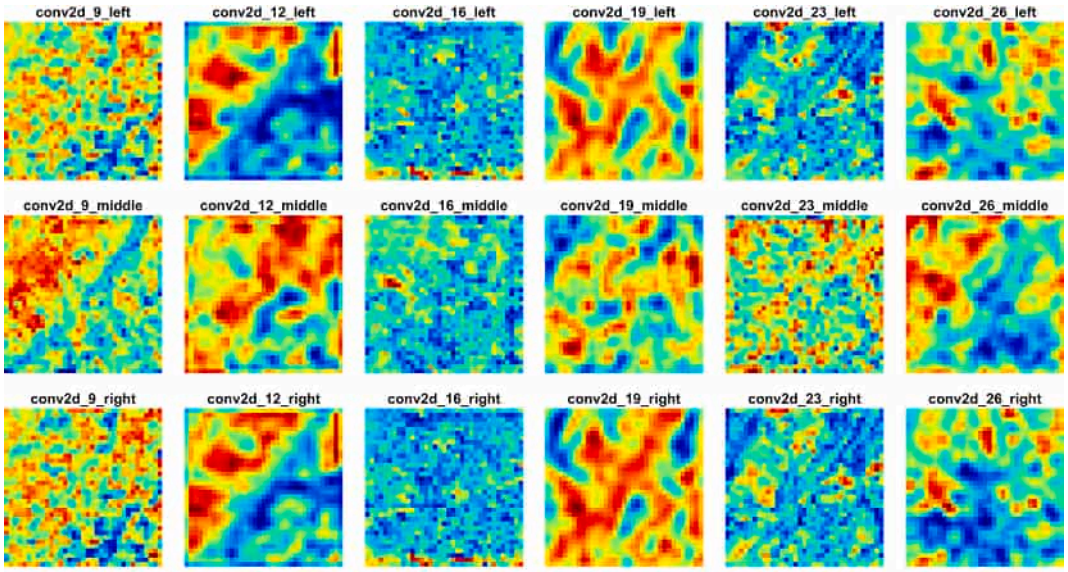
Method	Params [M]
ResNet50	23.5
Inception-v3	21.8
GoogleNet	5.9
ViT	5.5
TBN-PSI	4.1

4.6. Computational complexity

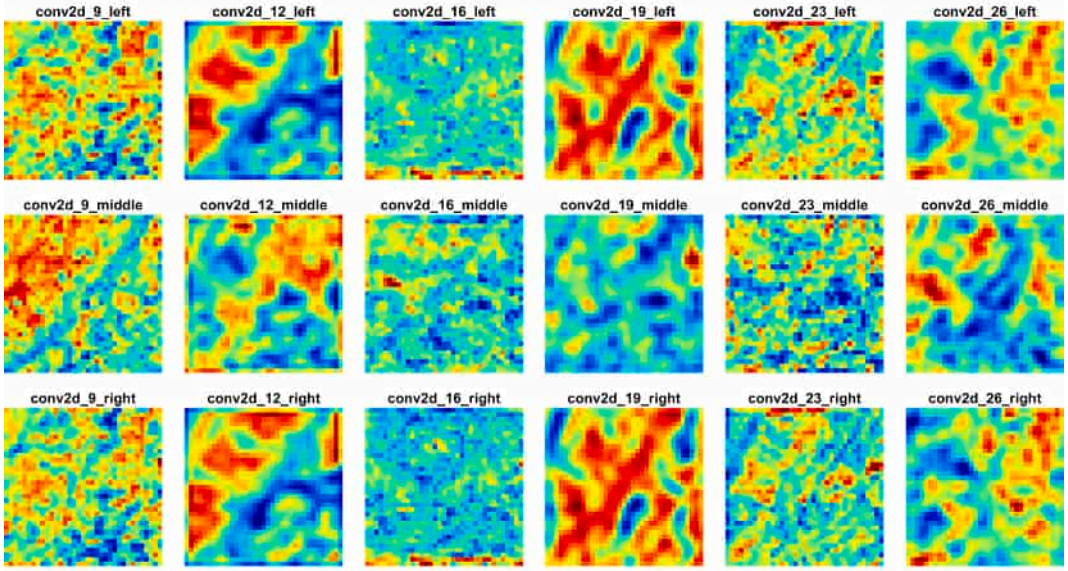
The computational complexity of the proposed architecture is compared in Table 11 with several CNN-based methods, including ResNet50, Inception-v3, GoogleNet, and ViT. As reported, TBN-PSI demonstrates a more lightweight design than these CNN methods; despite comprising three branches, it possesses nearly five times fewer parameters than its backbone (Inception-v3), while still delivering superior performance in image quality prediction.

4.7. Limitations of the study

Although the proposed approach is novel, it is not without limitations pertaining to image processing and data retrieval. Image processing shares conceptual parallels with cryptographic practices, where



(a) Activations of selected layers for a high-quality PS image; each row represents one branch.



(b) Activations of selected layers for a lower-quality PS image; each row represents one branch.

Fig. 4. Visualization of activations of selected layers in three branches, based on two PS images of the same region.

the objective is to establish a reliable, fault-resistant mechanism—specifically for assessing image quality in this context. This domain necessitates sophisticated algorithms and extensive testing to ensure resilience against failures or attacks, ultimately contributing to the overall integrity and trustworthiness of the applications in which they are deployed. In the context of pan-sharpening techniques, fault detection can be likened to identifying anomalies or discrepancies in the quality assessment of acquired full-resolution images (Koziel et al., 2017).

Pan-sharpening techniques are susceptible to security threats, particularly poisoning attacks. These attacks involve malicious modifications to the training data, which can significantly compromise the integrity of the neural network's performance. For instance, if adversarial data were introduced into the training set, the TBN-PSI model could be manipulated to generate inaccurate estimates of the quality of pan-sharpened images. Such alterations could lead to erroneous conclusions regarding the effectiveness of pan-sharpening methods, thereby undermining the reliability of the image enhancement process—akin to how poisoning attacks in healthcare can ultimately result in misdiagnoses. Addressing these security threats is crucial to ensuring the robustness and reliability of pan-sharpening techniques and their applications (Mozaffari-Kermani et al., 2015; Nia et al., 2015).

5. Conclusions

In this study, a new low-complexity NR IQA method for predicting the quality of pan-sharpened images was introduced. The method employs a three-branch neural network architecture specifically designed for assessing the quality of pan-sharpened images. Selected blocks of layers from a backbone network were utilized, with two branches processing the input images—one dedicated to the RGB channels and the other to the NIR + GB channels. A central branch integrates information from both inputs, capturing features at varying levels of granularity. Experimental results demonstrate that the proposed method outperforms related approaches, consistently ranking pan-sharpening techniques

more effectively based on the quality of their fused images. TBN-PSI achieved exceptional performance in quality prediction as indicated by metrics such as SAM, ERGAS, and SCC. This approach advances the field by providing a more accurate and efficient tool for evaluating the quality of pan-sharpened images, which may lead to significant improvements in remote sensing and image fusion techniques.

Future work will focus on:

- Integrating CNN-based approaches with hand-crafted methods for IQA of PS images to enhance overall performance.
- Leveraging the low complexity and reduced computational requirements of hand-crafted solutions to increase efficiency.
- Developing new network architectures based on recently introduced Transformer networks or a combination of various architectures to further improve accuracy and adaptability.
- Expanding dataset diversity and enhancing robustness against potential adversarial attacks to ensure the reliability of the proposed model across various real-world applications.

CRediT authorship contribution statement

Igor Stępień: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Mariusz Oszust:** Writing – original draft, Methodology, Supervision, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Agudelo-Medina, O.A., Benitez-Restrepo, H.D., Vivone, G., Bovik, A., 2019. Perceptual quality assessment of pan-sharpened images. Remote Sens. 11 (7), 877.

Aiazzi, B., Alparone, L., Baronti, S., Carlà, R., Garzelli, A., Santurri, L., 2014. Full-scale assessment of pansharpening methods and data products. In: Image and Signal Processing for Remote Sensing XX. Vol. 9244, SPIE, 924402.

Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P., Bruce, L.M., 2007. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-s data-fusion contest. IEEE Trans. Geosci. Remote Sens. 45 (10), 3012–3021.

Athar, S., Wang, Z., 2023. Degraded reference image quality assessment. IEEE Trans. Image Process. 32, 822–837.

Badal, N., Soundararajan, R., Garg, A., Patil, A., 2022. No reference pansharpened image quality assessment through deep feature similarity. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 7235–7247. http://dx.doi.org/10.1109/JSTARS.2022.3199446.

Bao, K., Meng, X., Chai, X., Shao, F., 2022. A blind full resolution assessment method for pansharpened images based on multistream collaborative learning. IEEE Trans. Geosci. Remote Sens. 60, 1–11. http://dx.doi.org/10.1109/TGRS.2022.3177607.

Carla, R., Santurri, L., Aiazzi, B., Baronti, S., 2015. Full-scale assessment of pansharpening through polynomial fitting of multiscale measurements. IEEE Trans. Geosci. Remote Sens. 53 (12), 6344–6355.

Chang, H.W., Wang, M.H., 2011. Sparse correlation coefficient for objective image quality assessment. Signal Process., Image Commun. 26 (10), 577–588. http://dx.doi.org/10.1016/j.image.2011.07.003.

Chen, L., Letu, H., Fan, M., Shang, H., Tao, J., Wu, L., Zhang, Y., Yu, C., Gu, J., Zhang, N., Hong, J., Wang, Z., Zhang, T., 2022. An introduction to the Chinese high-resolution earth observation system: Gaofen-17 civilian satellites. J. Remote Sens. 2022, 9769536. http://dx.doi.org/10.34133/2022/9769536.

Choi, J., Yu, K., Kim, Y., 2010. A new adaptive component-substitution-based satellite image fusion by using partial replacement. IEEE Trans. Geosci. Remote Sens. 49 (1), 295–309.

Dadrass Javan, F., Samadzadegan, F., Fathollahi, F., 2018. Spectral and spatial quality assessment of IHS and wavelet based pan-sharpening techniques for high resolution satellite imagery. Adv. Image Video Process. 6, http://dx.doi.org/10.14738/aivp.62.4226.

Dalla Mura, M., Vivone, G., Restaino, R., Addesso, P., Chanussot, J., 2015. Global and local gram-Schmidt methods for hyperspectral pansharpening. In: 2015 IEEE International Geoscience and Remote Sensing Symposium. IGARSS, pp. 37–40. http://dx.doi.org/10.1109/IGARSS.2015.7325691.

Dosovitskiy, Beyer, K., Weissenborn, Z., Unterthiner, D., Minderer, H., Gelly, U., Houlsby, 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021.

Firmansyah, H.R., Sarli, P.W., Twinanda, A.P., Santoso, D., Imran, I., 2024. Building typology classification using convolutional neural networks utilizing multiple ground-level image process for city-scale rapid seismic vulnerability assessment. Eng. Appl. Artif. Intell. 131, 107824.

Gabarda, S., Cristóbal, G., Goel, N., 2018. Anisotropic blind image quality assessment: Survey and analysis with current methods. J. Vis. Commun. Image Represent. 52, 101–105. http://dx.doi.org/10.1016/j.jvcir.2018.02.008.

Garzelli, A., Nencini, F., 2009. Hypercomplex quality assessment of multi/hyperspectral images. IEEE Geosci. Remote Sens. Lett. 6 (4), 662–665.

Garzelli, A., Nencini, F., Capobianco, L., 2007. Optimal MMSE pan sharpening of very high resolution multispectral images. IEEE Trans. Geosci. Remote Sens. 46 (1), 228–236.

Gharbia, R., El Baz, A.H., Hassanien, A.E., Tolba, M.F., 2014. Remote sensing image fusion approach based on brovey and wavelets transforms. In: Kömer, P., Abraham, A., Sná˘ sel, V. (Eds.), Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014. pp. 311–321.

Ghassemian, H., 2016. A review of remote sensing image fusion methods. Inf. Fusion 32, 75–89.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Huang, W., Xiao, L., Wei, Z., Liu, H., Tang, S., 2015. A new pan-sharpening method with deep neural networks. IEEE Geosci. Remote Sens. Lett. 12 (5), 1037–1041. http://dx.doi.org/10.1109/LGRS.2014.2376034.

Khan, M.M., Alparone, L., Chanussot, J., 2009. Pansharpening quality assessment using the modulation transfer functions of instruments. IEEE Trans. Geosci. Remote Sens. 47 (11), 3880–3891. http://dx.doi.org/10.1109/TGRS.2009.2029094.

Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. CoRR arXiv:1412.6980.

Koziel, B., Azarderakhsh, R., Mozaffari Kermani, M., Jao, D., 2017. Post-quantum cryptography on FPGA based on isogenies on elliptic curves. IEEE Trans. Circuits Syst. I. Regul. Pap. 64 (1), 86–99. http://dx.doi.org/10.1109/TCSI.2016.2611561.

Koziel, B., Jalali, A., Azarderakhsh, R., Jao, D., Mozaffari-Kermani, M., 2016. NEON-SIDH: Efficient implementation of supersingular isogeny diffie-hellman key exchange protocol on ARM. In: Cryptology and Network Security: 15th International Conference, CANS 2016, Milan, Italy, November 14-16, 2016, Proceedings 15. Springer, pp. 88–103.

Kwan, C., Budavari, B., Bovik, A.C., Marchisio, G., 2017. Blind quality assessment of fused worldview-3 images by using the combinations of pansharpening and hypersharpening paradigms. IEEE Geosci. Remote Sens. Lett. 14 (10), 1835–1839.

Li, F., Lukin, V., Ieremeiev, O., Okarma, K., 2022. Quality control for the BPG lossy compression of three-channel remote sensing images. Remote Sens. 14 (8), http://dx.doi.org/10.3390/rs14081824.

Longbotham, N., Pacifici, F., Malitz, S., Baugh, W., Camps-Valls, G., 2015. Measuring the spatial and spectral performance of WorldView-3. In: Fourier Transform Spectroscopy and Hyperspectral Imaging and Sounding of the Environment. Optica Publishing Group, p. HW3B.2. http://dx.doi.org/10.1364/HISE.2015.HW3B.2.

Mehravar, S., Dadrass Javan, F., Samadzadegan, F., Toosi, A., Moghimi, A., Khatami, R., Stein, A., 2022. Varying weighted spatial quality assessment for high resolution satellite image pan-sharpening. Int. J. Image Data Fusion 13 (1), 44–70.

Meng, X., Xiong, Y., Shao, F., Shen, H., Sun, W., Yang, G., Yuan, Q., Fu, R., Zhang, H., 2021. A large-scale benchmark data set for evaluating pansharpening performance: Overview and implementation. IEEE Geosci. Remote Sens. Mag. 9 (1), 18–52. http://dx.doi.org/10.1109/MGRS.2020.2976696.

Min, X., Gu, K., Zhai, G., Liu, J., Yang, X., Chen, C.W., 2017. Blind quality assessment based on pseudo-reference image. IEEE Trans. Multimed. 20 (8), 2049–2062.

Mittal, A., Moorthy, A.K., Bovik, A.C., 2012a. No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. 21 (12), 4695–4708. http://dx.doi.org/10.1109/TIP.2012.2214050.

Mittal, A., Soundararajan, R., Bovik, A.C., 2012b. Making a "completely blind" image quality analyzer. IEEE Signal Process. Lett. 20 (3), 209–212.

Mozaffari-Kermani, M., Sur-Kolay, S., Raghunathan, A., Jha, N.K., 2015. Systematic poisoning attacks on and defenses for machine learning in healthcare. IEEE J. Biomed. Health Inf. 19 (6), 1893–1905. http://dx.doi.org/10.1109/JBHI.2014.2344095.

Muthusamy, D., Sathyamoorthy, S., 2023. Feature sampling based on multilayer perceptive neural network for image quality assessment. Eng. Appl. Artif. Intell. 121, 106015. http://dx.doi.org/10.1016/j.engappai.2023.106015.

Nia, A.M., Mozaffari-Kermani, M., Sur-Kolay, S., Raghunathan, A., Jha, N.K., 2015. Energy-efficient long-term continuous personal health monitoring. IEEE Trans. Multi-Scale Comput. Syst. 1 (2), 85–98. http://dx.doi.org/10.1109/TMSCS.2015.2494021.

Opozda, S., Sochan, A., 2014. The survey of subjective and objective methods for quality assessment of 2D and 3D images. Theor. Appl. Inform. 26 (1-2), 39–67.

Padwick, C., Deskevich, M., Pacifici, F., Smallwood, S., 2010. WorldView-2 pan-sharpening. In: Proceedings of the ASPRS 2010 Annual Conference, San Diego, CA, USA. Vol. 2630, pp. 1–14.

Pushparaj, J., Hegde, A.V., 2017. Evaluation of pan-sharpening methods for spatial and spectral quality. Appl. Geomatics 9, 1–12.

Qi, J., Du, J., Siniscalchi, S.M., Ma, X., Lee, C.-H., 2020. On mean absolute error for deep neural network based vector-to-vector regression. IEEE Signal Process. Lett. 27, 1485–1489.

Rahmani, S., Strait, M., Merkurjev, D., Moeller, M., Wittman, T., 2010. An adaptive IHS pan-sharpening method. IEEE Geosci. Remote Sens. Lett. 7 (4), 746–750. http://dx.doi.org/10.1109/LGRS.2010.2046715.

Ranchin, T., Wald, L., 2000. Fusion of high spatial and spectral resolution images: The arsis concept and its implementation. Photogramm. Eng. Remote Sens. 66 (1), 49–61.

Rubel, A., Ieremeiev, O., Lukin, V., Fastowicz, J., Okarma, K., 2022. Combined no-reference image quality metrics for visual quality assessment optimized for remote sensing images. Appl. Sci. 12 (4), http://dx.doi.org/10.3390/app12041986.

Scarpa, G., Ciotola, M., 2022. Full-resolution quality assessment for pansharpening. Remote Sens. 14 (8), http://dx.doi.org/10.3390/rs14081808.

Sefercik, U.G., Alkan, M., Jacobsen, K., Atalay, C., Buyuksalih, G., 2021. Quality analysis of worldview-4 DSMs generated by least squares matching and semiglobal matching. J. Appl. Remote Sens. 15 (3), 034515.

Serifoglu Yilmaz, C., Yilmaz, V., Gungor, O., 2022. A theoretical and practical survey of image fusion methods for multispectral pansharpening. Inf. Fusion 79, 1–43. http://dx.doi.org/10.1016/j.inffus.2021.10.001.

Shah, V.P., Younan, N.H., King, R.L., 2008. An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets. IEEE Trans. Geosci. Remote Sens. 46 (5), 1323–1335. http://dx.doi.org/10.1109/TGRS.2008.916211.

Sheikh, H., Sabir, M., Bovik, A., 2006. A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Process. 15 (11), 3440–3451. http://dx.doi.org/10.1109/TIP.2006.881959.

Singh Kushwah, J., Kumar, A., Patel, S., Soni, R., Gawande, A., Gupta, S., 2022. Comparative study of regressor and classifier with decision tree using modern tools. Mater. Today: Proc. 56, 3571–3576. http://dx.doi.org/10.1016/j.matpr.2021.11.635.

Stepien, I., Oszust, M., 2022. No-reference quality assessment of pan-sharpening images with multi-level deep image representations. Remote Sens. 14 (5), http://dx.doi.org/10.3390/rs14051119.

Stepien, I., Oszust, M., 2024. TBN-PSI: Three-branch neural network for no-reference quality assessment of pan-sharpening images. https://github.com/Igorles/TBN_PSI.

Sulaiman, A.G., Elashmawi, W.H., El-Tawel, G.S., 2020. A robust pan-sharpening scheme for improving resolution of satellite images in the domain of the nonsubsampled shearlet transform. Sens. Imaging 21, 1–27.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 2818–2826. http://dx.doi.org/10.1109/CVPR.2016.308.

Tang, Z., Chen, Z., Li, Z., Zhong, B., Zhang, X., Zhang, X., 2023. Unifying dual-attention and siamese transformer network for full-reference image quality assessment. ACM Trans. Multimedia Comput. Commun. Appl. 19 (6), 24. http://dx.doi.org/10.1145/3597434.

Toutin, T., Cheng, P., 2002. QuickBird–a milestone for high resolution mapping. Earth Obs. Mag. 11 (4), 14–18.

Tu, T.-M., Su, S.-C., Shyu, H.-C., Huang, P., 2001. A new look at IHS-like image fusion methods. Inf. Fusion 2, 177–186. http://dx.doi.org/10.1016/S1566-2535(01)00036-7.

Vivone, G., Dalla Mura, M., Garzelli, A., Restaino, R., Scarpa, G., Ulfarsson, M., Alparone, L., Chanussot, J., 2021. A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods. IEEE Geosci. Remote Sens. Mag. 9, 53–81. http://dx.doi.org/10.1109/MGRS.2020.3019315.

Wald, L., 2002. Data Fusion. Definitions and Architectures - Fusion of Images of Different Spatial Resolutions. Presses de l'Ecole, Ecole des Mines de Paris, Paris, France, p. 200, ISBN 2-911762-38-X.

Wang, Z., Bovik, A., 2002. A universal image quality index. IEEE Signal Process. Lett. 9, 81–84. http://dx.doi.org/10.1109/97.995823.

Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E., 2004. Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. 13, 600–612. http://dx.doi.org/10.1109/TIP.2003.819861.

Wang, J., Lu, T., Huang, X., Zhang, R., Feng, X., 2024. Pan-sharpening via conditional invertible neural network. Inf. Fusion 101, 101980. http://dx.doi.org/10.1016/j.inffus.2023.101980.

Wang, P., Yao, H., Huang, B., Leung, H., Liu, P., 2023. Multiresolution analysis pansharpening based on variation factor for multispectral and panchromatic images from different times. IEEE Trans. Geosci. Remote Sens. 61, 1–17. http://dx.doi.org/10.1109/TGRS.2023.3252001.

Ye, P., Kumar, J., Kang, L., Doermann, D., 2012. Unsupervised feature learning framework for no-reference image quality assessment. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1098–1105.

Yilmaz, V., 2021. A non-dominated sorting genetic algorithm-II-based approach to optimize the spectral and spatial quality of component substitution-based pansharpened images. Concurr. Comput.: Pract. Exper. 33, e6030. http://dx.doi.org/10.1002/cpe.6030.

Yilmaz, V., 2023. Adaptive hybrid pansharpening: a novel approach for combining two methods to achieve superior pansharpening performance. Int. J. Remote Sens. 44 (14), 4301–4325. http://dx.doi.org/10.1080/01431161.2023.2234095.

You, J., Korhonen, J., 2021. Transformer for image quality assessment. In: 2021 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 1389–1393.

Yuhas, R.H., Goetz, A.F., Boardman, J.W., 1992. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In: JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop.

Zhang, J., 2019. Spatial correlation coefficient estimator for frequency selective MIMO channels. Electron. Lett. 55 (5), 290–292.

Zhou, M., Huang, J., Hong, D., Zhao, F., Li, C., Chanussot, J., 2023. Rethinking pan-sharpening in closed-loop regularization. IEEE Trans. Neural Netw. Learn. Syst. 1–15. http://dx.doi.org/10.1109/TNNLS.2023.3279931.

Zhou, B., Shao, F., Meng, X., Fu, R., Ho, Y.-S., 2019. No-reference quality assessment for pansharpened images via opinion-unaware learning. IEEE Access 7, 40388–40401.

SoftwareX 23 (2023) 101494

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original software publication

TIQA-PSI: Toolbox for perceptual Image Quality Assessment of Pan-Sharpened Images

Igor Stępień [a], Mariusz Oszust [b,*]

[a] Doctoral School of Engineering and Technical Sciences at the Rzeszow University of Technology, al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland

[b] Department of Computer and Control Engineering, Rzeszow University of Technology, Wincentego Pola 2, 35-959 Rzeszow, Poland

ARTICLE INFO

Article history:
Received 23 April 2023
Received in revised form 31 July 2023
Accepted 1 August 2023

Keywords:
Image Quality Assessment
Hyperspectral Imaging
Pan-sharpening
Toolbox
Perceptual image quality
MATLAB

ABSTRACT

A Pan-Sharpening (PS) technique integrates spatial details of a high-resolution panchromatic (PAN) image with spectral information of a low-resolution multi-spectral (MS) image, creating a single high-resolution color image. Since different PS methods produce images of a different quality, they should be compared using Image Quality Assessment (IQA) approaches that mimics human visual perception. In this paper, a MATLAB-based Toolbox for perceptual Image Quality Assessment of Pan-Sharpened Images (TIQA-PSI) is presented. TIQA-PSI aims to stimulate the development and facilitate the evaluation of new PS or PS-IQA approaches as it contains a large number of state-of-the-art IQA methods and images generated by representative PS techniques with corresponding subjective scores. They can be used to assess the output of any new PS algorithm or to serve as baselines for the comparison with new PS-IQA methods. Furthermore, apart from popular PS-IQA methods, the toolbox offers an opportunity to employ perceptual IQA approaches that can be effectively trained on scores obtained in tests with human subjects for superior image evaluation. The code of TIQA-PSI is available on GitHub.

Code metadata

Current code version	v0.1
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-23-00263
Code Ocean compute capsule	none
Legal Code License	MIT
Code versioning system used	git
Software code languages, tools, and services used	MATLAB 2022b, libsvm, MatlabPyrTools
Compilation requirements, operating environments & dependencies	MATLAB toolbox: Image Processing, Statistics and Machine Learning, Deep Learning, curve fitting, Computer Vision, Parallel Computing, Signal Processing, Deep Learning Toolbox Model: VGG-16, VGG-19, ResNet-18
If available Link to developer documentation/manual	https://github.com/Igorles/TIQA_PSI
Support email for questions	istepien@kia.prz.edu.pl

1. Motivation and significance

The success of remote sensing research has relied upon the creation of accurate imagery along with effective interpretation and analysis [1–5]. Spectral imaging is used in satellite image studies to enhance selected components of captured objects [6]. Unlike conventional imaging methods based on three-channel color imaging, spectral imaging collects image data at specific wavelength intervals over the electromagnetic spectrum as the spectral dimension can be used e.g., for the identification of the materials in the scene. Despite the higher spectral resolution, the lower spatial resolution affects the hyperspectral (HS) data processing implementation [7]. The approach for integrating the spatial details of a high-resolution panchromatic (PAN) image and the spectral information of a low-resolution multi-spectral (MS) image is named Pan-Sharpening (PS). The integration is obtained at the same time, over the same region, resulting in

* Corresponding author.
E-mail addresses: istepien@kia.prz.edu.pl (Igor Stępień), marosz@kia.prz.edu.pl (Mariusz Oszust).

https://doi.org/10.1016/j.softx.2023.101494

a high-resolution MS image [8]. HS data are suitable to evaluate the spectral behavior of PS approaches since they can provide a smooth spectral curve of an object by using a set of higher-spectral-resolution detectors.

The PAN and MS images can be generated using numerous combinations of HS data [9]. The PS methods improve the analysis of satellite images due to spatial and spectral enhancements. Moreover, PS is one of the steps for improving images prior change detection [10], object recognition [11], visual image analysis, or scene interpretation [12]. The quality of achieved images depends on the employed algorithms, as different algorithms create different PS outputs [13]. Furthermore, a PS method may provide a perfect spectral consistency along with a poor spatial consistency. To make a reliable assessment of fusion results, the evaluation of both spectral and spatial consistency should be implemented.

The literature categorizes classical PS methods into component substitution (CS), multiresolution analysis (MRA), also known as high-pass filtering (HPF), and model-based methods. CS methods exchange the spatial information in the original MS image with spatial details contained in a PAN image. This replacement can lead to visually-appealing PS images robust against small misregistration errors. The MRA methods extract PAN details via spatial filtering while preserving spectral information and providing results robust against temporal misalignments [14]. Model-based approaches can be divided into compressed sensing-based [15] and sparse matrix factorization technique-based methods [16].

Since the human eye is the fundamental evaluator of visual experience, consequently the modeling of the human visual system (HVS) is the main issue for objective Image Quality Assessment (IQA). Consequently, both the development of IQA for PS images and PS approaches require a pool of state-of-the-art methods for comparison. Furthermore, new PS methods cannot be created without an automatic assessment of their results and an opportunity to compare the results with related techniques using the same IQA algorithm. This also indicates that the development of IQA methods is of a great interest to the field since approaches returning scores for PS images that are better correlated with human scores are desired.

The IQA methods are divided into three groups, depending on the type of used images: Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR) techniques. The FR approaches compare an image with its distortion-free version, RR methods employ only a part of the distortion-free image, and NR methods assess an image without any access to the undistorted image. Hence, the NR methods are the most challenging and most desirable in practice. Taking into account the specificity of PS images and the lack of their distortion-free counterparts, the usability of NR images requires the creation of software packages with recent approaches along with the widely-accepted protocols for comparison of PS approaches, or, in the case of the introduction of a new IQA method, its validation using a large number of PS images with subjective scores.

In the work of Meng et al. [17], a dataset with very high resolution MS and PAN image pairs has been presented in which dated and state-of-the-art methods are compared and ranked through classical evaluation indices. The approach is designed to train ML-based approaches, without focusing on the assessment of the performance of new PS methods. Vivone et al. [8], in turn, introduced the PAirMax benchmark for evaluating the performance of PS algorithms using nine performance indices. It is composed of 14 PAN and MS pairs collected by different satellites. The previous benchmark of Vivone et al. [18], introduced in 2014, defines performance protocols for the assessment of PS images, and uses three IQA methods and 19 PS methods. However, the benchmarks or software toolboxes in the literature lack the opportunity to compare PS images using a large number of classical, state-of-the-art, or perceptual IQA approaches. Also, the number of used PS and IQA methods is relatively small. Furthermore, they cannot be used for the development of new PS-IQA techniques as they are designed for the comparison of PS methods.

In this paper, a new toolbox (called TIQA-PSI, Image Quality Assessment of Pan-Sharpened Images toolbox) that includes 30 IQA methods and 16 PS techniques is introduced. It is created to support the development of new PS techniques that can be added to the benchmark and assessed with recent methods. Similarly, due to the large number of IQA methods in the toolbox, the addition of a new IQA method and its comparison using a typical procedure is simplified. Moreover, since the database with PS images and subjective scores is used, the training-based IQA methods are also included to provide often superior IQA of PS images. Consequently, the development of new PS methods can be assisted with perceptual IQA approaches that reflect the subjectivity of the image assessment process. The main contributions of the MATLAB-based software toolbox introduced in this paper are as follows. (1) Simplification of training and comparison of new IQA approaches for evaluating PS images with 30 state-of-the-art methods, (2) Evaluation of selected PS methods using IQA approaches, incorporating superior perceptual methods trained on images assessed by human observers, (3) Automated generation of statistical significance test results, evaluation criteria values, and images for qualitative tests.

2. Software description

TIQA-PSI is used to test algorithms for assessing the quality of images created by PS algorithms. A new method can be relatively easily added to the set of state-of-the-art methods and tested. This supports the creation of both new IQA and PS approaches due to facilitated comparison and a large competitor pool. The folder contains three main MATLAB scripts.

TIQA_PSI_FullDB.m script trains IQA models for training-based approaches. It uses the image collection of PS images with subjective scores obtained in tests with humans [6]. Here, the machine learning-based methods are trained on the entire Agudelo-Medina et al. database with subjective scores and is tested on additional images [6]. The training can be organized using a different database, ensuring the availability of subjective scores [19].

As a result, the script displays the objective scores returned by IQA methods and the rankings of test images.

In *TIQA_PSI_Train.m* script, the images are randomly divided into disjoint training and testing subsets, i.e., based on the popular 80:20 ratio, and then, the quality assessment methods are run a predefined number of times. Finally, the median values of the correlation coefficients between returned objective scores and human assessments are calculated. The results are presented in a table with Spearman's Rank Correlation Coefficient (SRCC), Kendall's Order Correlation Coefficient (KRCC), and Pearson's Linear Correlation Coefficient (PLCC), as well as statistical tests in the form of a boxplot and the results of the Wilcoxon's rank sum test. In the Wilcoxon's test, the method with significantly better PLCC median scores *"1"*, while the worse and indistinguishable approaches are assigned with *"−1"* and *"0"*, respectively.

The *TIQA_PSI_PAN.m* script runs selected PS methods and assesses their results using IQA methods. Perceptual machine-learning IQA methods can be trained using the entire Agudelo-Medina et al. [6]. It is advisable to use the method that exhibits superior or acceptable performance in the evaluation provided by the *TIQA_PSI_Train.m* script or use a method that satisfactorily assesses user-provided images considered in *TIQA_PSI_FullDB.m* script. The results are presented in the form of tables that include objective IQA scores and the ranking of the run PS methods.

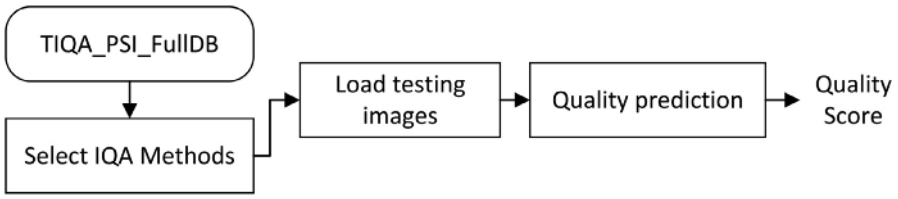


Fig. 1. Flowchart of the *TIQA_PSI_FullDB.m* functionality in which IQA methods are trained or used to assess images.

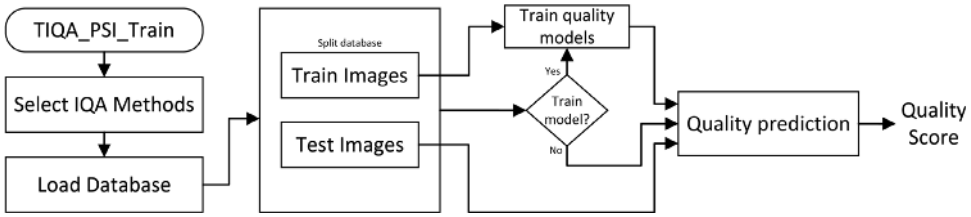


Fig. 2. Flowchart of the *TIQA_PSI_Train.m* script in which quality models are trained on a part of the database and tested on the other part.

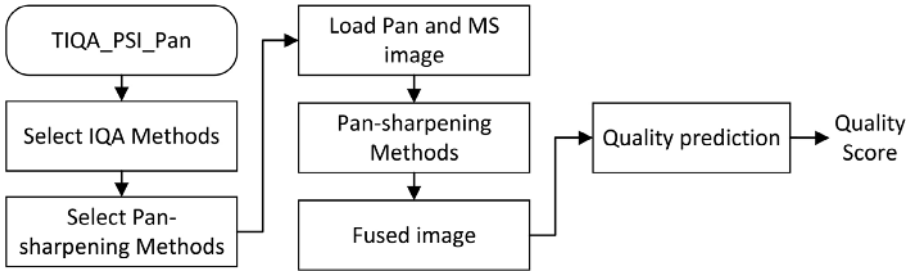


Fig. 3. Flowchart of the *TIQA_PSI_PAN.m* functionality of TIQA-PSI toolbox, in which PAN and MS images are processed by PS algorithms and evaluated by IQA techniques.

2.1. Database

In experiments, a database with PS images that originate from the IKONOS satellite images is used. The images were assessed by human observers in the study of Agudelo-Medina et al. [6]. It contains 171 PS images from the input of four reference images using seven PS methods. The availability of a reference image indicates that the study employed the reduced-resolution images. The images of four regions were distorted with blur and additive white Gaussian noise. Each Rome region (Coliseum, Road, Urban, and River) of $256 \times 256 \times 4$ pixels for MS and 1024×1024 pixels for PAN was extracted [6].

2.2. Software architecture

The TIQA-PSI toolbox is created on the MATLAB platform, a widely used programming language for data analysis. The toolbox is launched by selecting one of the three scripts available in the main folder. Depending on the needs of a user of the toolbox, codes of new IQA, PS-IQA, or PS methods can be following the already included state-of-the-art approaches gathered in TIQA-PSI. All methods and scripts often use MATLAB toolboxes for image processing, computer vision, machine learning, deep learning, and signal processing.

2.3. Software functionalities

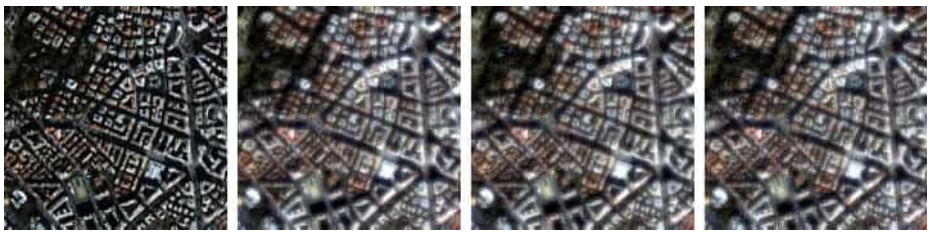
The TIQA-PSI contains scripts that perform three functionalities. As it is presented in Fig. 1, the first functionality uses the entire Agudelo-Medina et al. [6] database to train methods that require creating quality models. These methods and the remaining techniques are then used to assess the quality of user-provided testing images. Most of them use Support Vector Regression (SVR) technique for this purpose. However, depending on the method, the toolbox does not restricts the addition of other quality models or learning them in the end-to-end fashion.

The second functionality, presented in Fig. 2, offers the training of the IQA methods and applies the widely-accepted protocol in which performance criteria are computed to compare them. Finally, user is presented with comparative evaluation of selected IQA or PS-IQA methods of different types (FR and NR) or fields of origin (e.g., IQA of natural images, PS-IQA). In the case of the selection of several IQA methods, the quality prediction is performed by them all and a set of objective scores is returned for analysis.

In the third functionality, quality models are created for methods that require them. Then, PAN and MS images are loaded and processed by selected set of PS algorithms. Finally, all generated PS images are evaluated by IQA techniques (Fig. 3).

2.4. Methods

TIQA-PSI contains 30 IQA methods. The full-reference approaches designed for the assessment of PS images are ERGAS [20], SAM [21], Q [22], and Q2n [8], while the NR techniques are represented by MLPSIE [23]. Here, only MLPSIE is representative of perceptual methods with the MIT license. To facilitate the application of more IQA approaches based on natural image processing the FR (DISTS [24], MDSI [25], HaarPSI [26], SCC [27], IFC [28], SUMMER [9], MSUNIQUE [29], ReSIFT [30], UNIQUE [31], VIF [26], SSIM [26], and PerSIM [32]) and NR (BLINDER [33], BRISQUE [34], CEIQ [35], DIIVINE [34], FRIQUEE [34], GWH-GLBP-BIQA [36], NIQE [34], SSEQ [37], SINDEX [38], SR-metric [39],



(a) Image 1 (b) Image 2 (c) Image 3 (d) Image 4

	Image1	Image2	Image3	Image4
BRISQUE	42.5384	53.0602	60.4086	62.8865
MLPSIE	58.5943	66.5632	67.6644	69.1528
NOREQI	31.1623	61.0162	61.0663	63.3561
SSEQ	55.2963	58.0442	57.9266	57.9552
SVD	0.0000	54.6332	54.7707	54.6196
ERGAS	0.0000	17.5381	17.7219	18.3092
Q2n	1.0000	0.7943	0.7844	0.7649
Q	1.0000	0.8330	0.8362	0.8266
SCC	1.0000	0.9294	0.9204	0.9041

(e)

	Image1	Image2	Image3	Image4
BRISQUE	1	2	3	4
MLPSIE	1	2	3	4
NOREQI	1	2	3	4
SSEQ	1	4	2	3
SVD	1	3	4	2
ERGAS	1	2	3	4
Q2n	1	2	3	4
Q	1	3	2	4
SCC	1	2	3	4
Ranking	1	2	3	4

(f)

Fig. 4. (a–d) Exemplary images, (e–f) results of image assessment of selected IQA methods.

and NOREQI [40]) methods are included. Among them, FR perceptual approaches are represented by MULTISSIM and SVD, while BRISQUE, BLINDER, FRIQUEE, GWH-GLBP-BIQA, SSEQ, and NOREQI belong to such NR methods.

The toolbox contains the following 16 PS algorithms: EXP [41], SR-D [42], FE-HPM [43], MF [44], BDSD, BDSD_PC, C_BDSD, BT-H, C-GSA, MTF-GLP, MTF-GLP-CBD, MTF-GLP-FS, MTF-GLP-HPM, MTF-GLP-HPM-R, MTF-GLP-HPM-H, and PRACS [8].

3. Illustrative examples

3.1. Training of IQA methods and assessment of PS images

In *TIQA_PSI_FullDB.m* script, learning-based (perceptual) IQA methods are trained on the entire database and then, together with the remaining methods, used to assess test images. The script, exemplary PS images, and the obtained results are shown in Listing 1 and Fig. 4.

The image quality is presented either through objective scores obtained from the used IQA methods (Fig. 4e) or as a ranking of methods, with 1 indicating the highest-rated image and 4 indicating the lowest-rated image (Fig. 4f).

3.2. Evaluation of IQA methods

The *TIQA_PSI_Train.m* (Listing 2) is responsible for the evaluation of IQA methods. As an illustrative example, the results obtained on the database based on the typical division of its images into training-testing samples are shown in Fig. 5. Apart from the values of four evaluation criteria displayed in the command window (see Table 1), boxplot and summary of the pairwise Wilcoxon's statistical significance tests between selected methods are shown. For the boxplot, the median values of the tested samples are measured with a *5%* significance level. From the example, it can be seen, that MPLSIE outperforms the remaining methods on the database, in terms of SRCC, KRCC, and PLCC. However, its PLCC results are not significantly better than those of BRISQUE. Consequently, it is advised to use one of them for the perceptual evaluation of PS images.

Table 1
The results of the comparative evaluation of selected IQA methods.

Method	SRCC	KRCC	PLCC
BRISQUE	0.9642	0.8592	0.9836
MLPSIE	0.9717	0.8788	0.9854
SINDEX	0.7901	0.6043	0.8483
DIIVINE	0.6112	0.4635	0.7000
SSIM	0.7384	0.5419	0.7820
ERGAS	0.6582	0.5098	0.6817
Q2n	0.6504	0.4795	0.6753
Q	0.4280	0.2888	0.3539

3.3. Evaluation of PS methods

TIQA_PSI_PAN.m displays PS images and their evaluation by IQA methods (Listing 3). To facilitate the visual comparison of PS methods, parts of the images are marked and magnified (see Fig. 6). Also, the reference image is presented in the visualization. Furthermore, the assessment of images made by selected IQA methods is presented in the table. In the image, the FR methods return 0 or 1 while assessing the reference image depending on the established scale, i.e., they return greater or smaller values for images of quality closer to the quality of the reference image.

4. Impact

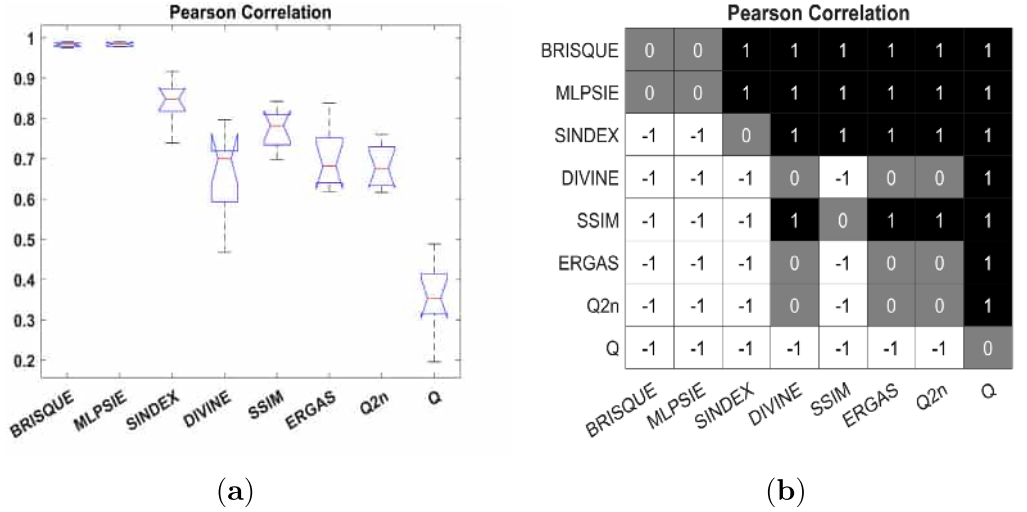
Recently, a growing demand for IQA methods concerning IQA satellite images can be observed. Therefore, to facilitate the comparison of new PS methods with state-of-the-art approaches based on the assessment of recent PS-IQA (or general IQA), the TIQA has been designed. Furthermore, with TIQA, the researchers creating new PS-IQA techniques gain unique opportunity to add their method to the toolbox and validate its performance against other 30 IQA algorithms. The toolbox is equipped with typically employed comparison protocols, including IQA performance quality indices, boxplots, Wilcoxon's rank sum test, and visualization of PS images with magnified regions.

Listing 1: *TIQA_PSI_FullDB.m* script.

```
addpath(genpath('Tools'));
load Mos_imgs.mat
IQA = {... Full-Reference methods 18
    'SVD','ERGAS','Q2n','Q','SCC',...
    ...'DISTS', 'HaarPSI', 'IFC', 'MDSI', 'MS-UNIQUE', 'ReSIFT', 'SAM', 'SUMMER', 'VIF', 'SSIM', 'MULTISSIM', 'PerSIM', 'UNIQUE',...
    ... No-Reference methods 12
    'BRISQUE','MLPSIE','NOREQI','SSEQ','SINDEX'
    ...'BLINDER', 'CEIQ', 'DIVINE', 'FRIQUEE', 'GWH-GLBP-BIQA', 'NIQE', 'SR-metric'
    };
PC_img = strcat(pwd,filesep,'Database',filesep,'SS',filesep,PC_img);
TC_img = strcat(pwd,filesep,'Database',filesep,'SS',filesep,TC_img);
%Create Svr models for IQA methods, where 0 is the svr model name
Create_Svr_Models(PC_img,TC_img,DMOS_PC_TC,0,IQA)
%% Inputs:
% images : true colour image (R+G+B), imagesRef : true colour reference image (R+G+B)
% images2 : pseudo colour image (NIR+R+G), imagesRef : pseudo colour reference image (NIR+R+G)
images ={strcat(pwd,filesep,'Database\SS\Test\TC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\TC_Test_ATWT-M2_AWGN_Level-1.tif'); ...
        strcat(pwd,filesep,'Database\SS\Test\TC_Test_ATWT-M2_AWGN_Level-2.tif'); strcat(pwd,filesep,'Database\SS\Test\TC_Test_ATWT-M2_AWGN_Level-3.tif')};
imagesRef ={strcat(pwd,filesep,'Database\SS\Test\TC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\TC_Test_REF_UD_UD.tif'); ...
        strcat(pwd,filesep,'Database\SS\Test\TC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\TC_Test_REF_UD_UD.tif')};
images2 ={strcat(pwd,filesep,'Database\SS\Test\PC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\PC_Test_ATWT-M2_AWGN_Level-1.tif'); ...
        strcat(pwd,filesep,'Database\SS\Test\PC_Test_ATWT-M2_AWGN_Level-2.tif'); strcat(pwd,filesep,'Database\SS\Test\PC_Test_ATWT-M2_AWGN_Level-3.tif')};
imagesRef2 ={strcat(pwd,filesep,'Database\SS\Test\PC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\PC_Test_REF_UD_UD.tif'); ...
        strcat(pwd,filesep,'Database\SS\Test\PC_Test_REF_UD_UD.tif'); strcat(pwd,filesep,'Database\SS\Test\PC_Test_REF_UD_UD.tif')};
%% USAGE
% If you only have true colour image iqa_metrics(images,[],[],[],IQA,0), where 0 is the svr model name
% If you have true colour and reference images iqa_metrics(images,imagesRef,[],[],IQA,0)
% If you have true colour and pseudo colour images iqa_metrics(images,[],images2,[],IQA,0)
% If you have true colour and pseudo colour images and reference images iqa_metrics(images,imagesRef,images2,imagesRef2,IQA,0)
score = iqa_metrics(images,imagesRef,images2,imagesRef2,IQA,0);
tableData = prepareScoreResult(score,[]);
```

Listing 2: *TIQA_PSI_Train.m* script.

```
rand('seed',0);
addpath(genpath('Tools'));
load Mos_imgs_2.mat
IQA = {... Full-Reference methods 18
    'Q','Q2n','SSIM','ERGAS',...
    ...'DISTS','HaarPSI','IFC','MDSI','MS-UNIQUE',
    ...'ReSIFT','SAM','SCC','SUMMER','SVD','VIF','MULTISSIM','PerSIM','UNIQUE',...
    ... No-Reference methods 12
    'MLPSIE','SINDEX','DIVINE','BRISQUE'
    ... 'BLINDER','CEIQ','FRIQUEE','GWH-GLBP-BIQA',
    ... 'NIQE','NOREQI','SR-metric','SSEQ'
    };
tiqapsitrainHelper % script preparing the database for the popular 80:20
result = createTableResult(resultTmp,names)
createBoxPlot(resultTmp,result)
wilcoxonTable(resultTmp,names)
```



(a) (b)

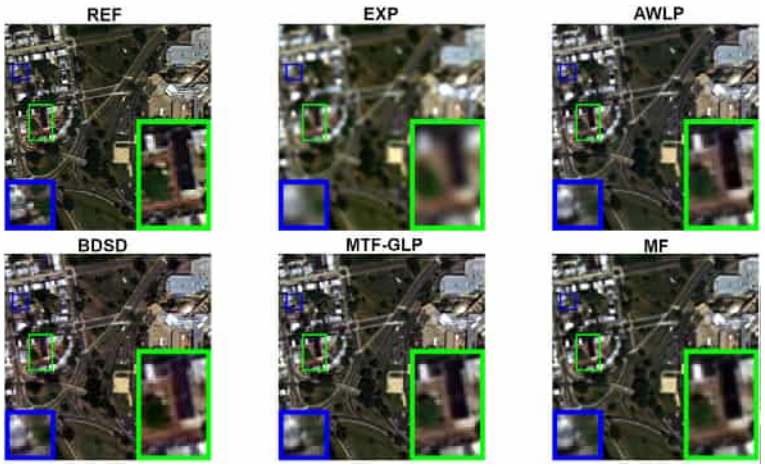
Fig. 5. (a) Boxplot based on the obtained PLCC values for selected methods (higher values denote better performance), (b) summary of the Wilcoxon's statistical significance tests in which significantly better, worse, or indistinguishable PLCC results of method in the row than those of the method in the column obtains '1', '−1', or '0', respectively.

5. Conclusions

This paper introduces the MATLAB-based toolbox TIQA-PSI: Toolbox for Perceptual Image Quality Assessment of Pan-sharpened Images. It provides access to a large number of IQA methods, including rarely employed in this field perceptual methods that simulate the HVS and can be trained on subjective scores to provide well-correlated objective scores. The software allows for the comparative evaluation of IQA methods and the evaluation of PS images produced by different PS methods.

Listing 3: *TIQA_PSI_PAN.m* script script.

```
addpath(genpath('Tools'));
load Mos_imgs.mat
load(strcat(pwd,filesep,'Database',filesep,'Dataset',filesep,'Test(HxWxC)_wv2_data1.mat'))
IQA = { %Full-Reference methods 18
    'VIF','SSIM','SAM','HaarPSI','MDSI',...
    ... 'DISTS', 'ERGAS', 'IFC', 'MS-UNIQUE', 'Q', 'Q2n', 'ReSIFT', 'SCC', 'SUMMER', 'SVD', 'MULTISSIM', 'PerSIM', 'UNIQUE',...
    ... No-Reference methods 12
    'BLINDER','DIVINE','MLPSIE'
    ... 'BRISQUE', 'CEIQ', 'FRIQUEE', 'GWH-GLBP-BIQA', 'NIQE', 'NOREQI', 'SINDEX', 'SR-metric', 'SSEQ',
    };
PC_img = strcat(pwd,filesep,'Database',filesep,'SS',filesep,PC_img);
TC_img = strcat(pwd,filesep,'Database',filesep,'SS',filesep,TC_img);
%Create Svr models for IQA methods, where 0 is the svr model name
Create_Svr_Models(PC_img,TC_img,DMOS_PC_TC,0,IQA)
%Pan-Sharpening algorithms
algorithms = {... Require REF
    'REF','EXP','AWLP','BDSD','MTF-GLP','MF'
    ... 'C-BDSD', 'BDSD_PC', 'BT-H', 'C-GSA', 'SR-D', 'MTF-GLP-FS', 'MTF-GLP-HPM', 'MTF-GLP-HPM-R', 'MTF-GLP-CBD', 'FE-HPM', 'PRACS','MTF-GLP-HPM-H'
    };
ratio = 4; % Resize Factor
sensor = 'WV2'; % Sensor
[imageFused,imagesRef,VisibleImage,algorithms] = Pan_sharpening_algorithms(gt,lms,ms,pan,ratio,algorithms,sensor);
%% Inputs:
% ImageFused : true colour fused image (R+G+B)
% imagesRef : true colour reference image (R+G+B)
score = iqa_metrics(imageFused,imagesRef,[],[],IQA,0);
tableData = prepareScoreResult(score,algorithms);
```



(a)

	REF	EXP	AWLP	BDSD	MTF-GLP	MF
BLINDER	55.0180	55.0181	55.0180	55.0180	55.0180	55.0180
MLPSIE	61.4612	61.5020	61.4941	61.4835	61.4902	61.4996
DIVINE	34.2820	72.8712	39.0960	36.3976	36.3737	39.7636
VIF	1.0000	0.5917	0.8906	0.9369	0.9508	0.9184
SSIM	1.0000	0.5141	0.7705	0.7872	0.8063	0.7745
MDSI	0.0000	0.0720	0.0530	0.0490	0.0497	0.0538
HaarPSI	1.0000	0.9994	0.9998	0.9999	0.9999	0.9998
SAM	0.0000	8.5468	9.4497	9.4434	8.5322	8.4170

(b)

Fig. 6. (a) PS images with magnified regions, (b) provided assessment by IQA methods.

Future work will be focused on either adding more MATLAB-based IQA and PS approaches or creating a version of the toolbox dedicated to methods implemented in Python.

The code of TIQA-PSI is available on GitHub under the MIT license.

CRediT authorship contribution statement

Igor Stępień: Conceptualization, Software, Validation, Writing, Visualization, Investigation. **Mariusz Oszust:** Conceptualization, Writing, Investigation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Hong D, Gao L, Yao J, Zhang B, Plaza A, Chanussot J. Graph convolutional networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens 2020;59(7):5966–78.
[2] Hong D, Gao L, Yokoya N, Yao J, Chanussot J, Du Q, Zhang B. More diverse means better: Multimodal deep learning meets remote-sensing imagery classification. IEEE Trans Geosci Remote Sens 2020;59(5):4340–54.
[3] Yao J, Zhang B, Li C, Hong D, Chanussot J. Extended vision transformer (exvit) for land use and land cover classification: A multimodal deep learning framework. IEEE Trans Geosci Remote Sens 2023;61:1–15.
[4] Roy SK, Deria A, Hong D, Rasti B, Plaza A, Chanussot J. Multimodal fusion transformer for remote sensing image classification. IEEE Trans Geosci Remote Sens 2023;61:1–20.
[5] Li C, Zhang B, Hong D, Yao J, Chanussot J. Lrr-net: An interpretable deep unfolding network for hyperspectral anomaly detection. IEEE Trans Geosci Remote Sens 2023;61:1–12.
[6] Agudelo-Medina OA, Benitez-Restrepo HD, Vivone G, Bovik A. Perceptual quality assessment of pan-sharpened images. Remote Sens 2019;11(7). http://dx.doi.org/10.3390/rs11070877.
[7] Irmak H, Akar GB, uksel SEY. Image fusion for hyperspectral image super-resolution. In: 2018 9th Workshop on hyperspectral image and signal processing: evolution in remote sensing. 2018, p. 1–5. http://dx.doi.org/10.1109/WHISPERS.2018.8747231.
[8] Vivone G, Dalla Mura M, Garzelli A, Restaino R, Scarpa G, Ulfarsson MO, Alparone L, Chanussot J. A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods. IEEE Geosci Remote Sens Mag 2021;9(1):53–81. http://dx.doi.org/10.1109/MGRS.2020.3019315.
[9] Temel D, AlRegib G. Perceptual image quality assessment through spectral analysis of error representations. Signal Process, Image Commun 2019;70:37–46. http://dx.doi.org/10.1016/j.image.2018.09.005.
[10] Souza C, Firestone L, Silva LM, Roberts D. Mapping forest degradation in the eastern amazon from spot 4 through spectral mixture models. Remote Sens Environ 2003;87(4):494–506. http://dx.doi.org/10.1016/j.rse.2002.08.002, large Scale Biosphere Atmosphere Experiment in Amazonia.
[11] Mohammadzadeh A, Tavakoli A, Valadan Zoej MJ. Road extraction based on fuzzy logic and mathematical morphology from pan-sharpened ikonos images. Photogrammetric Rec 2006;21(113):44–60. http://dx.doi.org/10.1111/j.1477-9730.2006.00353.x.
[12] Laporterie-Déjean F, de Boissezon H, Flouzat G, Lefèvre-Fonollosa M-J. Thematic and statistical evaluations of five panchromatic/multispectral fusion methods on simulated pleiades-hr images. Inf Fusion 2005;6(3):193–212. http://dx.doi.org/10.1016/j.inffus.2004.06.006, fusion of Remotely Sensed Data over Urban Areas.
[13] Yang S, Wang M, Jiao L. Fusion of multispectral and panchromatic images based on support value transform and adaptive principal component analysis. Inf Fusion 2012;13(3):177–84. http://dx.doi.org/10.1016/j.inffus.2010.09.003.
[14] Baronti S, Aiazzi B, Selva M, Garzelli A, Alparone L. A theoretical analysis of the effects of aliasing and misregistration on pansharpened imagery. IEEE J Sel Top Sign Proces 2011;5(3):446–53. http://dx.doi.org/10.1109/JSTSP.2011.2104938.
[15] Li S, Yang B. A new pan-sharpening method using a compressed sensing technique. IEEE Trans Geosci Remote Sens 2011;49(2):738–46. http://dx.doi.org/10.1109/TGRS.2010.2067219.
[16] Huang B, Song H, Cui H, Peng J, Xu Z. Spatial and spectral image fusion using sparse matrix factorization. IEEE Trans Geosci Remote Sens 2014;52(3):1693–704. http://dx.doi.org/10.1109/TGRS.2013.2253612.
[17] Meng X, Xiong Y, Shao F, Shen H, Sun W, Yang G, Yuan Q, Fu R, Zhang H. A large-scale benchmark data set for evaluating pansharpening performance: Overview and implementation. IEEE Geosci Remote Sens Mag 2021;9(1):18–52. http://dx.doi.org/10.1109/MGRS.2020.2976696.
[18] Vivone G, Alparone L, Chanussot J, Dalla Mura M, Garzelli A, Licciardi GA, Restaino R, Wald L. A critical comparison among pansharpening algorithms. IEEE Trans Geosci Remote Sens 2015;53(5):2565–86. http://dx.doi.org/10.1109/TGRS.2014.2361734.
[19] Xiong Y, Shao F, Meng X, Jiang Q, Sun W, Fu R, Ho Y-S. A large-scale remote sensing database for subjective and objective quality assessment of pansharpened images. J Vis Commun Image Represent 2020;73:102947. http://dx.doi.org/10.1016/j.jvcir.2020.102947.
[20] Wald L. Data fusion. definitions and architectures - fusion of images of different spatial resolutions. In: Presses de l'Ecole, Ecole des Mines de Paris, Paris, France. 2002, iSBN 2-911762-38-X.
[21] Vivone G, Restaino R, Chanussot J. Full scale regression-based injection coefficients for panchromatic sharpening. IEEE Trans Image Process 2018;27(7):3418–31. http://dx.doi.org/10.1109/TIP.2018.2819501.
[22] Wang Z, Bovik A. A universal image quality index. IEEE Signal Process Lett 2002;9(3):81–4. http://dx.doi.org/10.1109/97.995823.
[23] Stępień I, Oszust M. No-reference quality assessment of pan-sharpening images with multi-level deep image representations. Remote Sens 2022;14(5). http://dx.doi.org/10.3390/rs14051119.
[24] Ding K, Ma K, Wang S, Simoncelli EP. Image quality assessment: Unifying structure and texture similarity. IEEE Trans Pattern Anal Mach Intell 2022;44(5):2567–81. http://dx.doi.org/10.1109/TPAMI.2020.3045810.
[25] Ziaei Nafchi H, Shahkolaei A, Hedjam R, Cheriet M. Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator. IEEE Access 2016;4:5579–90. http://dx.doi.org/10.1109/ACCESS.2016.2604042.
[26] Okarma K. Current trends and advances in image quality assessment. Elektron Elektrotech 2019;25:77–84. http://dx.doi.org/10.5755/j01.eie.25.3.23681.
[27] wen Chang H, hui Wang M. Sparse correlation coefficient for objective image quality assessment. Signal Process, Image Commun 2011;26(10):577–88. http://dx.doi.org/10.1016/j.image.2011.07.003.
[28] Sheikh H, Bovik A, de Veciana G. An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Trans Image Process 2005;14(12):2117–28. http://dx.doi.org/10.1109/TIP.2005.859389.
[29] Prabhushankar M, Temel D, AlRegib G. Ms-unique: Multi-model and sharpness-weighted unsupervised image quality estimation. Electron Imaging 2017.
[30] Temel D, AlRegib G. Resift: reliability-weighted sift-based image quality assessment. In: 2016 IEEE International Conference on Image Processing. 2016, p. 2047–51. http://dx.doi.org/10.1109/ICIP.2016.7532718.
[31] Zhang W, Ma K, Zhai G, Yang X. Uncertainty-aware blind image quality assessment in the laboratory and wild. IEEE Trans Image Process 2021;30:3474–86. http://dx.doi.org/10.1109/TIP.2021.3061932.
[32] Temel D, AlRegib G. Persim: Multi-resolution image quality assessment in the perceptually uniform color domain. In: 2015 IEEE International Conference on Image Processing. ICIP, 2015, p. 1682–6. http://dx.doi.org/10.1109/ICIP.2015.7351087.
[33] Gao Yu, Zhu Huang, Tian. Blind image quality prediction by exploiting multi-level deep representations. Pattern Recognit 2018;81:432–42. http://dx.doi.org/10.1016/j.patcog.2018.04.016.
[34] Bosse S, Maniry D, Müller K-R, Wiegand T, Samek W. Deep neural networks for no-reference and full-reference image quality assessment. IEEE Trans Image Process 2018;27(1):206–19. http://dx.doi.org/10.1109/TIP.2017.2760518.
[35] Xu M, Wang Z. No-reference quality assessment of contrast-distorted images. In: 2016 IEEE International Conference on Signal and Image Processing. ICSIP, 2016, p. 362–7. http://dx.doi.org/10.1109/SIPROCESS.2016.7888285.
[36] Li Q, Lin W, Fang Y. No-reference quality assessment for multiply-distorted images in gradient domain. IEEE Signal Process Lett 2016;23(4):541–5. http://dx.doi.org/10.1109/LSP.2016.2537321.
[37] Liu L, Liu B, Huang H, Bovik AC. No-reference image quality assessment based on spatial and spectral entropies. Signal Process, Image Commun 2014;29(8):856–63. http://dx.doi.org/10.1016/j.image.2014.06.006.
[38] Leclaire A, Moisan L. No-reference image quality assessment and blind deblurring with sharpness metrics exploiting fourier phase information. J. Math. Imaging Vis. 2015;52(1):145–72. http://dx.doi.org/10.1007/s10851-015-0560-5.
[39] Ma C, Yang C-Y, Yang X, Yang M-H. Learning a no-reference quality metric for single-image super-resolution. Comput Vis Image Underst 2017;158:1–16. http://dx.doi.org/10.1016/j.cviu.2016.12.009.
[40] Oszust M. No-reference image quality assessment using image statistics and robust feature descriptors. IEEE Signal Process Lett 2017;24(11):1656–60.
[41] Aiazzi B, Alparone L, Baronti S, Garzelli A. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. IEEE Trans Geosci Remote Sens 2002;40(10):2300–12. http://dx.doi.org/10.1109/TGRS.2002.803623.
[42] Vicinanza MR, Restaino R, Vivone G, Dalla Mura M, Chanussot J. A pansharpening method based on the sparse representation of injected details. IEEE Geosci Remote Sens Lett 2015;12(1):180–4. http://dx.doi.org/10.1109/LGRS.2014.2331291.
[43] Vivone G, Simes M, Dalla Mura M, Restaino R, Bioucas-Dias JM, Licciardi GA, Chanussot J. Pansharpening based on semiblind deconvolution. IEEE Trans Geosci Remote Sens 2015;53(4):1997–2010. http://dx.doi.org/10.1109/TGRS.2014.2351754.
[44] Restaino R, Vivone G, Dalla Mura M, Chanussot J. Fusion of multispectral and panchromatic images based on morphological operators. IEEE Trans Image Process 2016;25(6):2882–95. http://dx.doi.org/10.1109/TIP.2016.2556944.

Streszczenie w języku polskim

Niniejsza rozprawa doktorska jest monotematycznym cyklem publikacji naukowych poświęconych percepcyjnej ocenie jakości obrazów cyfrowych bez referencji z zastosowaniem metod sztucznej inteligencji, ze szczególnym uwzględnieniem głębokich sieci neuronowych. Praca ma charakter interdyscyplinarny, łącząc osiągnięcia z zakresu Informatyki Technicznej i Telekomunikacji z zastosowaniami w medycynie i teledetekcji. Koncentruje się na dwóch głównych obszarach badawczych: ocenie jakości obrazów rezonansu magnetycznego (MRI, ang. *Magnetic Resonance Imaging*) oraz wyostrzonych obrazów satelitarnych (PS, ang. *Pansharpened*). Dla obrazów rezonansu magnetycznego opracowano metody oceny jakości, które umożliwiają analizę wpływu różnorodnych zniekształceń bez konieczności ich jednoznacznej klasyfikacji. W przypadku obrazów PS zaproponowano podejścia oparte na wielopoziomowych reprezentacjach danych oraz modyfikacjach architektur sieci neuronowych, dostosowanych do percepcyjnego przetwarzania kanałów obrazów wielospektralnych. Uwzględniając powyższe, w pracy sformułowano następującą hipotezę badawczą:

> *Zastosowanie metod sztucznej inteligencji, w szczególności różnorodnych architektur sieci neuronowych, umożliwia ocenę jakości obrazów satelitarnych oraz obrazów rezonansu magnetycznego bez referencji, zapewniając wysoką zgodność z subiektywnymi ocenami ekspertów. Opracowane metody nie ustępują pod względem skuteczności predykcji jakości rozwiązaniom znanym z literatury.*

W świetle uzyskanych wyników można stwierdzić, że efektywność zaproponowanych rozwiązań jest porównywalna lub lepsza w porównaniu z istniejącymi metodami z literatury. Ich skuteczność została w znacznym stopniu potwierdzona w badanich eksperymentalnych na wielu zróznicowanych zbiorach danych, poprzez realizację następujących zadań badawczych:

1) Przeprowadzenie analizy literatury dotyczącej najnowszych metod oceny jakości obrazów rezonansu magnetycznego, z uwzględnieniem typowych zniekształceń, protokołów zastosowań oraz wykorzystywanych baz danych.

2) Opracowanie narzędzi do percepcyjnej oceny jakości wyostrzonych obrazów satelitarnych oraz platformy do oceny jakości obrazów rezonansu magnetycznego.

3) Zastosowanie głębokich sieci neuronowych w ocenie jakości obrazów satelitarnych i medycznych bez referencji, w celu opracowania metod zgodnych z oceną subiektywną i konkurencyjnych względem rozwiązań literaturowych, z wykorzystaniem:

   a) zewnętrznego połączenia architektur głębokich splotowych sieci neuronowych,

   b) wewnętrznego połączenia dwóch architektur sieci,

c) struktury sieci złożonej z trzech podsieci,

d) informacji wyodrębnionych z warstw sieci.

4) Przeprowadzenie eksperymentów z wykorzystaniem opracowanych metod, porównanie wyników z metodami z literatury oraz analiza i interpretacja rezultatów.

Zadanie 1. Przeprowadzenie analizy literatury dotyczącej najnowszych metod oceny jakości obrazów rezonansu magnetycznego, z uwzględnieniem typowych zniekształceń, protokołów zastosowań oraz wykorzystywanych baz danych.

Zadanie zostało zrealizowane poprzez analizę literatury dotyczącej najnowszych metod oceny jakości obrazów rezonansu magnetycznego, uwzględniając typy zniekształceń najczęściej występujących w tym rodzaju obrazowania, protokoły zastosowań tych metod oraz najczęściej używane bazy obrazów. W przypadku obrazów rezonansu magnetycznego przeanalizowano zniekształcenia związane z szumem termicznym, artefakty ruchowe, zniekształcenia geometryczne oraz efekty heterogeniczności pola magnetycznego, wskazując na ich wpływ na jakość i interpretację wyników. Przedstawiono też przegląd baz danych używanych w tym kontekście, wskazując najczęściej stosowane zestawy danych w badaniach obrazów MRI. Uzyskano pełny obraz aktualnych trendów i metod w ocenie jakości obrazowania MRI, umożliwiając wskazanie dalszych kierunków badań oraz potencjalnych wyzwań w tej dziedzinie.

Zadanie 2. Opracowanie narzędzi do percepcyjnej oceny jakości wyostrzonych obrazów satelitarnych oraz platformy do oceny jakości obrazów rezonansu magnetycznego.

Zadanie zostało zrealizowane poprzez opracowanie dwóch dedykowanych narzędzi: TIQA-MRI oraz TIQA-PSI. TIQA-MRI wspiera ocenę jakości obrazów MRI, oferując zaawansowane funkcje, takie jak dziewięć metod IQA, elastyczne ustawienia eksperymentalne oraz testy istotności statystycznej. TIQA-PSI umożliwia ocenę jakości wyostrzonych obrazów satelitarnych, wykorzystując 30 metod, w tym 16 zaprojektowanych z myślą o technikach wyostrzania. Oba narzędzia integrują kompleksowe analizy statystyczne, takie jak test istotności statystycznej Wilcoxona i obliczenia współczynników korelacji, co pozwala na precyzyjną i wszechstronną ocenę jakości obrazów.

Zadanie 3. Zastosowanie głębokich sieci neuronowych w ocenie jakości obrazów satelitarnych i medycznych bez referencji, w celu opracowania metod zgodnych z oceną subiektywną i konkurencyjnych względem rozwiązań literaturowych.

a) Zadanie zostało zrealizowane poprzez opracowanie metody oceny jakości obrazów MRI opartej na fuzji głębokich splotowych sieci neuronowych oraz transfer

learningu. Rozwiązanie to pozwoliło na skuteczną ocenę jakości obrazów MRI. W przypadku obrazów satelitarnych opracowano metodę łączącą architektury ResNet-18 i VGG-19, umożliwiającą ocenę jakości obrazów PS bez potrzeby korzystania z oryginalnych obrazów referencyjnych.

b) W ramach tego etapu zaproponowano metodę fuzji wielowymiarowych danych pochodzących z warstw sieci takich jak ResNet-18 i ResNet-50. Wykorzystano Kernel PCA do redukcji wymiarowości oraz model SVR do precyzyjnego przewidywania jakości obrazów. Dzięki ponownemu trenowaniu sieci uzyskano znaczną poprawę zdolności do różnicowania obrazów rezonansu magnetycznego pod kątem jakości percepcyjnej.

c) Zrealizowano metodę łączącą dwie komplementarne architektury sieci neuronowych, takie jak ResNet-18 i ResNet-50. Dzięki zastosowaniu transfer learningu oraz odpowiednio dostosowanego procesu trenowania uzyskano wyższą precyzję w ocenie jakości obrazów MRI.

d) Opracowano trójgałęziową sieć neuronową przeznaczoną do percepcyjnej oceny jakości obrazów satelitarnych, zoptymalizowaną pod kątem analizy cech przestrzennych i spektralnych. Sieć ta została zoptymalizowana pod kątem analizy specyficznych cech przestrzennych i spektralnych. Dzięki zastosowaniu metryk takich jak SSIM model uzyskał znaczące przewagi nad istniejącymi metodami oceny jakości PS, co zostało potwierdzone w testach na obszernych zbiorach danych.

Zadanie 4. Przeprowadzenie eksperymentów z wykorzystaniem opracowanych metod, porównanie wyników z metodami z literatury oraz analiza i interpretacja rezultatów.

Zadanie zostało zrealizowane poprzez przeprowadzenie eksperymentów z wykorzystaniem opracowanych metod oceny jakości obrazów rezonansu magnetycznego, polegających na zastosowaniu architektur głębokich sieci neuronowych oraz ich fuzji w celu uzyskania reprezentacji cech charakteryzujących jakość na wielu poziomach przetwarzania obrazu w artykule artykułów [A-3]. W podejściu tym wykorzystano zarówno cechy uzyskane z pojedynczych modeli ResNet18 i ResNet50, jak i ich wewnętrznej fuzji, co umożliwiło uchwycenie różnorodnych aspektów percepcyjnych obrazów MRI. Również w publikacjach [A-3] i [A-2] uzyskane wyniki zostały porównane z wynikami otrzymanymi przez metody, takie jak BRISQUE, GWH-GLBP, CurveletQA, czy MEON, a także przez inne klasycze podejścia oparte na analizie cech lokalnych i globalnych. Ponadto odniesiono się do skuteczności metod zaprojektowanych pierwotnie dla naturalnych lub satelitarnych obrazów (m.in. metoda MLPSIE) i dostosowanych do obrazów medycznych w artykułach [A-5] i [A-2]. Analiza porównawcza przeprowadzona została na bazach obrazów MR zawierających obrazy ocenione przez doświadczonych radiologów, uwzględniając subiektywne opinie ekspertów [A-3] i [A-2]. Interpretacja rezultatów wy-

kazała przewagę zaproponowanych podejść nad istniejącymi rozwiązaniami, szczególnie w kontekście zdolności do predykcji subiektywnej jakości obrazów bez konieczności posiadania obrazu referencyjnego.

Słowa kluczowe: ocena jakości obrazów, głębokie sieci neuronowe, rezonans magnetyczny, obrazy satelitarne, wyostrzanie, fuzja sieci głębokich.

Streszczenie w języku angielskim

The dissertation is a monothematic cycle of scientific publications devoted to the no-reference perceptual image quality assessment using artificial intelligence methods, with particular emphasis on deep neural networks. The work is interdisciplinary in nature, combining achievements in technical computer science and telecommunications with applications in medicine and remote sensing. It focuses on two main research areas: the quality assessment of magnetic resonance imaging (MRI) and pansharpened satellite images (PS).

For MRI images, methods have been developed that allow for the analysis of various distortions without the need for their explicit classification. In the case of PS images, the dissertation proposes approaches based on multi-level data representations and modifications of neural network architectures, adapted to the perceptual processing of multispectral image channels.

Considering the above, the following research hypothesis was formulated:

> *The application of artificial intelligence methods, particularly various architectures of neural networks, enables no-reference image quality assessment of satellite and magnetic resonance images, ensuring high consistency with subjective expert evaluations. The developed methods are not inferior in predictive performance to the state-of-the-art approaches found in the literature.*

In light of the obtained results, it can be stated that the performance of the proposed methods is comparable to or better than existing solutions. Their effectiveness has been largely confirmed through experimental studies on various datasets, by accomplishing the following research tasks:

1) A literature review of the latest methods for MRI image quality assessment, including typical distortions, application protocols, and utilized databases.

2) Development of tools for perceptual quality assessment of pansharpened satellite images and a platform for MRI image quality evaluation.

3) Application of deep neural networks for no-reference quality assessment of satellite and medical images to develop methods consistent with subjective assessments and competitive with literature-based solutions, using:

   a) external fusion of deep convolutional neural network architectures,

   b) internal fusion of two network architectures,

   c) a network structure composed of three sub-networks,

   d) features extracted from network layers.

4) Conducting experiments using the developed methods, comparing the results with literature-based methods, and analyzing and interpreting the outcomes.

Task 1. Literature review of the latest methods for assessing the quality of magnetic resonance images, including typical distortions, application protocols, and utilized databases.

The task was completed by reviewing the literature on state-of-the-art MRI quality assessment methods, considering the most common types of distortions in this imaging modality, such as thermal noise, motion artifacts, geometric deformations, and magnetic field inhomogeneities. Their impact on image quality and diagnostic interpretability was discussed. A review of the most widely used datasets was also presented, providing an overview of current trends and identifying directions for future research and potential challenges in this domain.

Task 2. Development of tools for perceptual quality assessment of pansharpened satellite images and a platform for MRI image quality assessment.

The task was carried out by developing two dedicated tools: TIQA-MRI and TIQA-PSI. TIQA-MRI supports the quality assessment of MRI images, offering advanced features such as nine IQA methods, flexible experimental configurations, and statistical significance testing. TIQA-PSI allows for the assessment of pansharpened satellite image quality using 30 methods, including 16 specifically designed for sharpening techniques. Both tools integrate comprehensive statistical analyses, such as the Wilcoxon significance test and correlation coefficient calculations, enabling accurate and robust image quality evaluation.

Task 3. Application of deep neural networks for no-reference image quality assessment of satellite and medical images, in order to develop methods aligned with subjective evaluations and competitive with literature-based approaches.

a) The task was accomplished by developing an MRI image quality assessment method based on the fusion of deep convolutional neural networks and transfer learning, enabling effective evaluation of MRI quality. For satellite images, a method combining ResNet-18 and VGG-19 architectures was developed, enabling PS image quality assessment without the need for reference images.

b) A method was proposed for fusing multidimensional features extracted from network layers such as ResNet-18 and ResNet-50. Kernel PCA was used for dimensionality reduction, and an SVR model was used for precise quality prediction. Retraining the networks significantly improved their ability to distinguish perceptual quality in MRI images.

c) A method combining two complementary network architectures (ResNet-18 and ResNet-50) was implemented. Using transfer learning and a dedicated training procedure, high precision in MRI quality assessment was achieved.

d) A three-branch neural network was developed for perceptual quality assessment of pansharpened satellite images, optimized for the analysis of spatial and spectral

features. The model demonstrated significant advantages over existing PS quality assessment methods using metrics such as SSIM, confirmed by extensive testing on large datasets.

Task 4. Conducting experiments using the developed methods, comparing results with literature methods, and analyzing and interpreting the outcomes.

The task was completed by conducting experiments using the developed methods for assessing the quality of MRI images. This included the application and fusion of deep neural network architectures to extract multi-level representations of image quality features, as described in publication [A-3]. Both individual ResNet18 and ResNet50 models, as well as their internal fusion, were used to capture various perceptual aspects of MRI images. Furthermore, in publications [A-3] and [A-2], the obtained results were compared with existing methods such as BRISQUE, GWH-GLBP, CurveletQA, MEON, and other traditional techniques based on local and global feature analysis. Additionally, the effectiveness of methods originally designed for natural or satellite images (e.g., MLPSIE) and adapted to medical images was discussed in publications [A-5] and [A-2]. The comparative analysis was conducted using MRI datasets evaluated by experienced radiologists, incorporating subjective expert assessments [A-3], [A-2]. The interpretation of the results showed the superiority of the proposed methods, especially in predicting subjective image quality without the need for reference images.

Keywords: image quality assessment, deep neural networks, magnetic resonance imaging, satellite images, pansharpening, deep network fusion.

Oświadczenia współautorów

Poniższy rozdział zawiera oświadczenia dotyczące indywidualnego wkładu merytorycznego autora rozprawy oraz współautorów w przygotowanie, przeprowadzenie i opracowanie badań oraz przedstawienie prac w formie publikacji, a także informację procentowym wkładzie autorskim. Oświadczenia dotyczą kolejno następujących artykułów:

[A-1] **Stępień, I.**, Oszust, M. *A Brief Survey on No-Reference Image Quality Assessment Methods for Magnetic Resonance Images.* J. Imaging 2022, 8(6), 160. https://doi.org/10.3390/jimaging8060160, IF_2022 = 3,2; CS: 5,9; wkład 50%; liczba punktów: 20, (str. 166).

[A-2] **Stępień, I.**, Obuchowicz, R., Piórkowski, A., Oszust, M. *Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment.* Sensors 2021, 21(4), 1043. https://doi.org/10.3390/s21041043, IF_2021 = 3,847; CS: 7,3; wkład 25%; liczba punktów: 100, (str. 168).

[A-3] **Stępień, I.**, Oszust, M. *No-Reference Image Quality Assessment of Magnetic Resonance Images with Multi-Level and Multi-Model Representations Based on Fusion of Deep Architectures.* Engineering Applications of Artificial Intelligence 2023, 123, 106283. https://doi.org/10.1016/j.engappai.2023.106283, IF_2023 = 7,5; CS: 9,6; wkład 50%; liczba punktów: 140, (str. 172).

[A-5] **Stępień, I.**, Oszust, M. *No-Reference Quality Assessment of Pan-Sharpening Images with Multi-Level Deep Image Representations.* Remote Sens. 2022, 14(5), 1119. https://doi.org/10.3390/rs14051119, IF_2022 = 5; CS: 8.3; wkład 50%; liczba punktów: 100, (str. 174).

[A-6] **Stępień, I.**, Oszust, M. *Three-Branch Neural Network for No-Reference Quality Assessment of Pan-Sharpened Images.* Engineering Applications of Artificial Intelligence 2025, 139(B), 109594. https://doi.org/10.1016/j.engappai.2024.109594, IF_2023 = 7,5; CS: 9,6; wkład 50%; liczba punktów: 140, (str. 176).

[A-7] **Stępień, I.**, Oszust, M. *TIQA-PSI: Toolbox for Perceptual Image Quality Assessment of Pan-Sharpened Images.* SoftwareX 2023, 23, 101494, https://doi.org/10.016/j.softx.2023.101494, IF_2023 =2,4; CS: 4.1; wkład 50%; liczba punktów: 200, (str. 178).

Rzeszów, 26.06.2025

miejscowość, data

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Oszust Mariusz: „A Brief Survey on No-Reference Image Quality Assessment Methods for Magnetic Resonance Images", Journal of Imaging 2022, 8(6), 160.

Mój udział polegał na:

- Analiza literatury,
- Współredagowaniu publikacji,
- Współopracowaniu recenzji i końcowej wersji publikacji.

Mój udział oceniam na 50%.

Igor Stępień

podpis

Podwyższ udział procentowy : merytoryczny dodatkowo

Rzeszów 23.01.2025

miejscowość, data

dr hab. inż. Mariusz Oszust, prof. PRz.

Katedra Informatyki i Automatyki

Politechnika Rzeszowska im. Ignacego Łukasiewicza

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Oszust Mariusz: „A Brief Survey on No-Reference Image Quality Assessment Methods for Magnetic Resonance Images", Journal of Imaging 2022, 8(6), 160.

Mój udział polegał na:

- Sformułowaniu koncepcji publikacji,
- Współredagowaniu publikacji,
- Współopracowaniu recenzji i końcowej wersji publikacji,
- Opiece merytorycznej.

Mój udział oceniam na 50%.

podpis

Rzeszów, 26.06.2025

miejscowość, data

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Obuchowicz Rafał, Piórkowski Adam, Oszust Mariusz: „Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment", Sensors 2021, 21(4), 104.

Mój udział polegał na:

- Współopracowaniu koncepcji publikacji,
- Opracowaniu architektury metody,
- Współprowadzenie eksperymentów,
- Współredagowaniu publikacji.

Mój udział oceniam na 25%.

Igor Stępień

podpis

Potwierdzam udział procentowy i merytoryczny doktoranta

Potwierdzam udział procentowy i merytoryczny doktoranta

Potwierdzam niniejszy udział procentowy i merytoryczny doktoranta

Rzeszów 23.06.2025

miejscowość, data

dr hab. inż. Mariusz Oszust, prof. PRz.

Katedra Informatyki i Automatyki

Politechnika Rzeszowska im. Ignacego Łukasiewicza

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Obuchowicz Rafał, Piórkowski Adam, Oszust Mariusz: „Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment", Sensors 2021, 21(4), 104.

Mój udział polegał na:

- Współopracowaniu koncepcji publikacji,
- Wyborze metodyki badań,
- Współopracowaniu i implementacji oprogramowania,
- Współplanowaniu i przeprowadzeniu eksperymentów,
- Współredagowaniu publikacji,
- Opiece merytorycznej.

Mój udział oceniam na 25%.

podpis

Kraków 29/06/25

miejscowość, data

Prof. dr hab. n. med. Rafał Obuchowicz

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Obuchowicz Rafał, Piórkowski Adam, Oszust Mariusz: „Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment", Sensors 2021, 21(4), 104.

Mój udział polegał na:

- Współopracowaniu koncepcji publikacji,
- Weryfikacji poprawności zastosowanej metody badawczej,
- Współplanowaniu i przeprowadzeniu eksperymentów,
- Współredagowaniu publikacji,
- Opiece merytorycznej.

Mój udział oceniam na 25%.

Rafał Obuchowicz

podpis

Kraków 26.06.2025

miejscowość, data

Prof. dr hab. inż. Adam Piórkowski

Katedra Biocybernetyki i Inżynierii Biomedycznej

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Obuchowicz Rafał, Piórkowski Adam, Oszust Mariusz: „Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment", Sensors 2021, 21(4), 104.

Mój udział polegał na:

- współopracowaniu koncepcji publikacji,
- weryfikacji poprawności zastosowanej metody badawczej,
- współredagowaniu publikacji,
- opiece merytorycznej.

Mój udział oceniam na 25%.

Adam Piórkowski

podpis

Rzeszów, 26.06.2025

miejscowość, data

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Oszust Mariusz: „No-Reference Image Quality Assessment of Magnetic Resonance Images with Multi-Level and Multi-Model Representations Based on Fusion of Deep Architectures", Engineering Applications of Artificial Intelligence 2023, 123, 10628.

Mój udział polegał na:

- Współopracowaniu koncepcji publikacji,
- Opracowaniu architektury metody,
- Przeprowadzeniu eksperymentów,
- Współredagowaniu publikacji.

Mój udział oceniam na 50%.

Igor Stępień

podpis

Podział udziału procentowy i merytoryczny zgodny

Rzeszów 13.06.2025

miejscowość, data

dr hab. inż. Mariusz Oszust, prof. PRz.

Katedra Informatyki i Automatyki

Politechnika Rzeszowska im. Ignacego Łukasiewicza

Oświadczenie

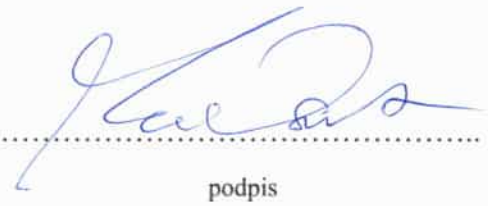
Oświadczam, że w pracy:

Stępień Igor, Oszust Mariusz: „No-Reference Image Quality Assessment of Magnetic Resonance Images with Multi-Level and Multi-Model Representations Based on Fusion of Deep Architectures", Engineering Applications of Artificial Intelligence 2023, 123, 10628.

Mój udział polegał na:

- Opracowaniu koncepcji publikacji,
- Wyborze metodyki badań,
- Współplanowaniu i przeprowadzeniu eksperymentów,
- Współredagowaniu publikacji,
- Opiece merytorycznej.

Mój udział oceniam na 50%.



podpis

Rzeszów, 26.06.2025

miejscowość, data

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Oszust Mariusz: „No-Reference Quality Assessment of Pan-Sharpening Images with Multi-Level Deep Image Representations" Remote Sensing 2022, 14(5), 1119.

Mój udział polegał na:

- Współautorstwie koncepcji publikacji,
- Współwyborze metodyki badań,
- Opracowaniu architektury metody,
- Przeprowadzenie eksperymentów,
- Współredagowaniu publikacji.

Mój udział oceniam na 50%.

…/p. Stępień/……

podpis

Potwierdzam udział pracy i merytoryczny

Rzeszów 23.06.2025

miejscowość, data

dr hab. inż. Mariusz Oszust, prof. PRz.

Katedra Informatyki i Automatyki

Politechnika Rzeszowska im. Ignacego Łukasiewicza

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Oszust Mariusz: „No-Reference Quality Assessment of Pan-Sharpening Images with Multi-Level Deep Image Representations" Remote Sensing 2022, 14(5), 1119.

Mój udział polegał na:

- Opracowaniu koncepcji publikacji,
- Wyborze metodyki badań,
- Współplanowaniu i przeprowadzeniu eksperymentów,
- Współredagowaniu publikacji,
- Opiece merytorycznej.

Mój udział oceniam na 50%.

podpis

Rzeszów, 26.06.2025
miejscowość, data

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Oszust Mariusz: „Three-Branch Neural Network for No-Reference Quality Assessment of Pan-Sharpened Images", Engineering Applications of Artificial Intelligence 2025, 139(B), 109594.

Mój udział polegał na:

- Współautorstwie koncepcji publikacji,
- Współwyborze metodyki badań,
- Opracowaniu architektury metody,
- Przeprowadzenie eksperymentów,
- Współredagowaniu publikacji.

Mój udział oceniam na 50%.

Igor Stępień
podpis

Potwierdzam udział procentowy i merytoryczny

Rzeszów 23.06.2025

miejscowość, data

dr hab. inż. Mariusz Oszust, prof. PRz.

Katedra Informatyki i Automatyki

Politechnika Rzeszowska im. Ignacego Łukasiewicza

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Oszust Mariusz: „Three-Branch Neural Network for No-Reference Quality Assessment of Pan-Sharpened Images", Engineering Applications of Artificial Intelligence 2025, 139(B), 109594.

Mój udział polegał na:

- Opracowaniu koncepcji publikacji,
- Wyborze metodyki badań,
- Współplanowaniu eksperymentów,
- Współredagowaniu publikacji,
- Opiece merytorycznej.

Mój udział oceniam na 50%.

podpis

Rzeszów, 26.06.2025

miejscowość, data

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Oszust Mariusz: „TIQA-PSI: Toolbox for Perceptual Image Quality Assessment of Pan-Sharpened Images", SoftwareX 2023, 23, 10149.

Mój udział polegał na:

- Współopracowaniu koncepcji publikacji,
- Zaprojektowanie architektury aplikacji,
- Przeprowadzeniu eksperymentów,
- Współredagowaniu publikacji.

Mój udział oceniam na 50%.

Igor Stępień

podpis

Potwierdzam wkład procentowy i merytoryczny doktoranta

……Rzeszów, 23.06.2025……..

miejscowość, data

dr hab. inż. Mariusz Oszust, prof. PRz

Katedra Informatyki i Automatyki

Politechnika Rzeszowska im. Ignacego Łukasiewicza

Oświadczenie

Oświadczam, że w pracy:

Stępień Igor, Oszust Mariusz: „TIQA-PSI: Toolbox for Perceptual Image Quality Assessment of Pan-Sharpened Images", SoftwareX 2023, 23, 10149.

Mój udział polegał na:

- Współopracowaniu koncepcji publikacji,
- Współplanowaniu i przeprowadzeniu eksperymentów,
- Współredagowaniu publikacji,
- Opiece merytorycznej.

Mój udział oceniam na 50%.

……..

podpis