



ROZPRAWA DOKTORSKA

Rekonstrukcja nadrozdzielcza
obrazów cyfrowych
z wykorzystaniem
głębokich konwolucyjnych
sieci neuronowych

mgr inż. Paweł Benecki

promotor: dr hab. inż. Michał Kawulok, prof. PŚ
promotor pomocniczy: dr inż. Daniel Kostrzewa

dyscyplina główna: Informatyka Techniczna
i Telekomunikacja

Gliwice, 22 września 2023



**Rzeczpospolita
Polska**

Unia Europejska
Europejski Fundusz Społeczny



Niniejsza praca doktorska powstała w ramach projektu:

*CyPhiS – Projekt nowoczesnych studiów
doktoranckich w dziedzinie systemów
cyber–fizycznych*

Projekt ten był dofinansowany z Funduszy Europejskich
w ramach Programu Operacyjnego Wiedza Edukacja Rozwój 2014-2020
przez Wydział Automatyki, Elektroniki i Informatyki Politechniki Śląskiej
w ramach Działania 3.2 Studia doktoranckie
na podstawie umowy nr POWR.03.02.00-00-I007/17-00

Spis treści

1	Wstęp	1
1.1	Wprowadzenie	1
1.2	Tezy rozprawy	4
1.3	Cele rozprawy	5
1.4	Opublikowane artykuły naukowe	6
1.5	Struktura rozprawy	9
2	Przegląd literatury	11
2.1	Metody rekonstrukcji nadrozdzielczej	11
2.1.1	Metody jednoobrazowe	13
2.1.2	Metody wieloobrazowe	17
2.1.3	Rekonstrukcja wieloobrazowa realizowana za pomocą sieci głębokich	22
2.2	Miary używane w ocenie jakości obrazów	26
2.2.1	Miary analityczne	26
2.2.2	Miary bezreferencyjne	30
2.2.3	Miary oparte na uczeniu maszynowym	31
2.2.4	Wady znanych metryk analitycznych	34
2.3	Zbiory danych	34
2.3.1	Zbiór PROBA-V	35
2.3.2	Zbiór danych symulowanych z Sentinel-2	36
2.3.3	Zbiór MuS2	40
2.3.4	Zbiór LIVE	41
2.3.5	Zbiór B4MultiSR	42
2.3.6	Zbiór WorldStrat	43
3	Opracowane metody	45
3.1	Połączenie klasycznych metod rekonstrukcji nadrozdzielczej oraz sieci głębokich	45
3.1.1	Ewolucyjna optymalizacja hiperparametrów metod wieloobra- zowych – EvoIM	45

3.1.2	Metoda EvoNet	47
3.2	Miary podobieństwa obrazów	50
3.2.1	Miary oparte na detektorach punktów kluczowych	52
3.2.2	Miary oparte na uczeniu maszynowym	55
3.3	Użycie miary opartej na cechach głębokich w rekonstrukcji wielo- obrazowej	56
4	Walidacja eksperymentalna	59
4.1	Połączenie metod klasycznych z sieciami głębokimi	59
4.2	Badanie cech metryk pod kątem rekonstrukcji nadrozdzielczej	68
4.3	Sieci głębokie jako funkcja straty	73
4.3.1	Dobór funkcji straty	73
4.3.2	Wyniki dla zbioru MuS2	79
4.3.3	Ocena wyników przez obserwatorów	86
4.4	Dyskusja wyników i wnioski	91
5	Podsumowanie i możliwe kierunki dalszych badań	101
5.1	Podsumowanie	101
5.2	Wnioski	102
5.3	Możliwe kierunki dalszych badań	103
	Bibliografia	107

Spis rysunków

2.1	Porównanie działania interpolacji oraz rekonstrukcji nadrozdzielczej	12
2.2	Zasada działania metody median shift and add	21
2.3	Schemat sieci DeepSum	23
2.4	Schemat sieci RAMS	24
2.5	Schemat sieci PIUnet	25
2.6	Przykład nieoczekiwanego działania miary L2 w porównywaniu obrazów	35
2.7	Przykład nieoczekiwanego działania miary SSIM w porównywaniu obrazów	36
2.8	Przykładowe obrazy ze zbioru danych PROBA-V	37
2.9	Przykładowe obrazy ze zbioru danych symulowanych Sentinel-2	39
2.10	Przykładowe obrazy ze zbioru danych MuS2	40
2.11	Przykłady obrazów ze zbioru LIVE	41
3.1	Sposób kodowania jąder splotu w metodzie EvoIM	47
3.2	Obliczanie wartości funkcji przystosowania osobnika w algorytmie EvoIM	48
3.3	Schemat działania metody EvoNet	49
3.4	Porównanie wyglądu obrazów z różnych źródeł	51
3.5	Działanie klasycznych miar - symulacja zmian histogramu obrazów	52
3.6	Działanie detektorów punktów kluczowych	53
3.7	Sposób treningu sieci głębokiej z użyciem miar opartych na sieciach głębokich	57
4.1	Przykład obrazów HR oraz LR z różnych satelitów	61
4.2	Przykłady wyników EvoNet na zbiorze obrazów sztucznie degradowanych	62
4.3	Przykłady wyników EvoNet na zbiorze obrazów rzeczywistych	63
4.4	Przykładowe pytania z ankiety dotyczącej wyników metody EvoNet. Poniżej nazwy metody podano liczbę oddanych głosów.	67
4.5	Przykłady degradacji obrazów w badaniu czułości metryk	71

4.6	Wyniki rekonstrukcji nadrozdzielczej w badaniu czułości metryk . . .	72
4.7	Pasma spektralne satelitów Sentinel-2 oraz WorldView-2	77
4.8	Przykładowe wyniki dla wybranych funkcji straty na zbiorze testowym MuS2	80
4.9	Przykładowe wyniki dla wybranych funkcji straty na zbiorze testowym MuS2	81
4.10	Przykładowe wyniki dla wybranych funkcji straty na zbiorze testowym MuS2	82
4.11	Przykładowe wyniki dla wybranych funkcji straty na zbiorze testowym MuS2	83
4.12	Przykładowe pytania z ankiety 2AFC	94
4.13	Obrazy wybrane do ankiety	95
4.14	Porównanie obrazów SR L1+LPIPS z L1+DKFS _{MIX} , cz. 1	96
4.15	Porównanie obrazów SR L1+LPIPS z L1+DKFS _{MIX} , cz. 2	97
4.16	Porównanie obrazów SR L1+LPIPS z L1+DKFS _{MIX} , cz. 3	98
4.17	Porównanie obrazów SR L1+LPIPS z L1+DKFS _{MIX} , powiększenie - artefakty	99

Spis tabel

4.1	Wyniki działania metody EvoNet	64
4.2	Wyniki działania metody EvoNet – opinie ankietowanych	65
4.3	Wyniki działania metody EvoNet – podsumowanie opinii ankietowanych	66
4.4	Wyniki analizy rozpatrywanych miar	73
4.5	Wyniki działania modeli na zbiorze MuS2	84
4.6	Testy statystyczne na zbiorze MuS2 dla wartości cPSNR	85
4.7	Testy statystyczne na zbiorze MuS2 dla wartości cSSIM	86
4.8	Testy statystyczne na zbiorze MuS2 dla wartości LPIPS	87
4.9	Ranking modeli względem różnych miar na zbiorze MuS2	88
4.10	Wyniki działania modeli na zbiorze symulowanym Sentinel-2	89
4.11	Wyniki działania modeli na zbiorze PROBA-V RED	90
4.12	Wyniki działania modeli na zbiorze PROBA-V NIR	91
4.13	Wyniki ankiety dla metod rekonstrukcji	91
4.14	Wyniki ankiety dla poszczególnych kombinacji metod	92

Rozdział 1

Wstęp

1.1 Wprowadzenie

Rozwój technik komputerowych doprowadził do tego, że możliwe jest przetwarzanie i używanie obrazów o coraz wyższych rozdzielczościach, rozumianych w kontekście: przestrzennym (wymiary obrazu) [1], [2], czasowym (np. częstotliwość odwiedzin danego miejsca na powierzchni Ziemi przez satelitę)[1], [3] lub spektralnej (liczba pasm spektralnych) [4]. W niniejszej rozprawie rozpatrywany jest wyłącznie aspekt rozdzielczości przestrzennej. Niska jakość danych obrazowych utrudnia możliwości ich analizy, obniżając jakość działania systemów wizyjnych. Często zaś dochody firm lub kwestie bezpieczeństwa bywają uzależnione od jakości działania wspomnianych systemów [5]. W wielu praktycznych scenariuszach uzyskanie obrazów o wysokiej rozdzielczości (ang. *high-resolution*, HR) jest trudne lub kosztowne. Przyczyny mogą mieć charakter techniczny, fizyczny (optyka) lub, jak w przypadku obserwacji satelitarnych, wynikający z ograniczonego pasma transmisji danych. Często potencjalna wymiana aparatury do akwizycji obrazu na lepszą jest kosztowna, albo jak w wypadku satelitów do obserwacji Ziemi, wręcz niemożliwa, co wymusza dłuższe korzystanie z instrumentów o słabszej charakterystyce.

Techniki rekonstrukcji nadrozdzielczej (ang. *super-resolution*, SR) mają na celu przewyciężyć to ograniczenie przez rekonstrukcję obrazów o wysokiej rozdzielczości przestrzennej z obserwacji o niskiej rozdzielczości (ang. *low-resolution*, LR). Uzyskane w ten sposób obrazy o wysokiej rozdzielczości mogą dostarczyć cennych informacji do analizy, podejmowania decyzji i wizualizacji [6]. W świetle tych informacji rekonstrukcja nadrozdzielcza staje się jednym z podstawowych problemów wizji komputerowej i przetwarzania obrazów. Celem metod SR jest stworzenie obrazu o wyższej rozdzielczości z ogólnie lepszą jakością wizualną i zawierającego więcej szczegółów. Ten obszar badań zyskał znaczną uwagę ze względu na jego szerokie zastosowanie w różnych dziedzinach, w tym w obrazowaniu me-

dycznym [7], systemach inspekcji wizyjnej i kryminalistyce [8], teledetekcji [9], [10] i fotografii cyfrowej [5], [6], [11]–[13]. Próbując opisać w najprostszy sposób co jest zadaniem rekonstrukcji nadrozdzielczej, można powiedzieć, że łączy ona cele realizowane zwykle przez interpolację oraz rekonstrukcję, a zatem: zwiększanie rozdzielczości wraz z uzupełnianiem szczegółów obrazu.

Istnieją dwie główne grupy metod SR: rekonstrukcja jednoobrazowa (ang. *single-image SR*, SISR) i rekonstrukcja wieloobrazowa (ang. *multi-image SR*, MISR) [5], [12], [13]. Różnią się one postacią danych wejściowych i w większości wypadków w bardziej fundamentalny sposób – celami.

Celem większości metod SISR jest poprawa wyglądu obrazu, czyli tak naprawdę generacja obrazu wyglądającego w sposób przypominający podobne obrazy w wyższej rozdzielczości [5]. Istnieją co prawda metody SISR używane np. w teledetekcji [14]–[16], które próbują zrekonstruować faktyczną informację, ale nie jest to główny nurt SISR [13]. W typowym wypadku SISR brakująca informacja jest uzupełniana na podstawie innych podobnych obrazów, a zgodność z prawdziwym wyglądem obserwowanej sceny nie jest krytyczna [17]. Dlatego też możliwe są duże powiększenia – cztero-, ośmiokrotne lub nawet większe. Chociaż poczyniono znaczne postępy w SISR, nadal istnieje wiele wyzwań i ograniczeń. Jednym z głównych jest odtworzenie szczegółów i ostrych krawędzi, co jest trudne ze względu na ograniczoną ilość informacji dostępnych w pojedynczym obrazie [5]. Kompromis między dokładnością rekonstrukcji a kosztem obliczeniowym jest kolejnym krytycznym wyzwaniem. Często algorytmy powinny generować wysokiej jakości wyniki, będąc jednocześnie wykonalnymi dla aplikacji działających w czasie rzeczywistym lub zbliżonym do rzeczywistego [18], [19]. Dodatkowo należy rozważyć niwelowanie artefaktów wprowadzanych podczas procesu rekonstrukcji oraz radzenie sobie z ograniczeniami samego systemu obrazowania, takimi jak zakłócenia w obrazie wejściowym oraz szum czujnika i rozmycie wynikające z ruchu obiektu fotografowanego względem instrumentu obrazującego.

Dla metod MISR natomiast głównym celem jest pozyskanie faktycznych danych w wyższej rozdzielczości. Brakująca informacja jest zatem rekonstruowana, a nie uzupełniana. Zachodzi tutaj fuzja informacji z obrazów niskiej rozdzielczości, a możliwość powiększania zależy bezpośrednio od ilości i jakości danych LR. Brakująca informacja wysokiej rozdzielczości jest uzupełniana na podstawie przesunięć o odległości mniejsze od piksela, czyli subpikselowe [20]. W warunkach rzeczywistych, gdy obrazy pochodzą z akwizycji dokonanych w różnym czasie, urządzenie fotografujące oraz obiekt fotografowany są zawsze minimalnie przesunięte względem siebie, co powoduje, że każdy piksel w obrazie LR zawiera informację z nieco różnych obszarów obrazu HR [20]. Natura danych, które na przykład w wypadku obrazowania satelitarne pochodzą z urządzeń o różnej charakterystyce i są poddawane dodatkowej wymaganej obróbce po transferze z satelity,

powoduje problemy spotykane w trakcie realizacji zadania MISR. Wyzwania te można sklasyfikować następująco [21]–[23]:

- 1) Właściwe wyznaczenie przesunięć między obserwacjami LR, czyli dokładna rejestracja na poziomie subpikselowym.
- 2) Uwzględnianie okluzji takich jak: chmury na obrazach satelitarnych, cienie wysokich budynków, uszkodzone piksele itp.
- 3) Zmienność czasowa – dla obrazów satelitarnych np. zmiany zabudowy, wegetacji, szczególnie widoczne w zakresie bliskiej podczerwieni itp.
- 4) Skalowalność do obsługi zmiennej liczby obrazów wejściowych i ignorowanie kolejności ich prezentowania metodzie realizującej rekonstrukcję.
- 5) Skuteczna fuzja informacji z wielu obrazów przy jednoczesnym unikaniu artefaktów powstających w wyniku działania metod rekonstrukcji.
- 6) Lokalne zmiany charakterystyki obrazów takie jak miejscowe wariacje oświetlenia, cienie itp. – zarówno między obrazami HR i LR, jak i pomiędzy poszczególnymi obserwacjami w niskiej rozdzielczości.
- 7) Różnice w charakterystyce akwizycji obrazów między sensorami wykonującymi obrazy wysokiej i niskiej rozdzielczości (na przykład dwa różne instrumenty satelitarne), które są następnie używane w treningu i ewaluacji metod. Chodzi tutaj zarówno o różnice wynikające z właściwości samych sensorów, jak i różnic w przetwarzaniu końcowym (ang. *post-processing*) surowych obrazów z różnych źródeł.
- 8) Problem oceny podobieństwa obrazów w kontekście tradycyjnie stosowanych miar i zmienności obrazów HR i LR wspomnianych w punktach 6 i 7.

Wspólnym wyzwaniem dla obu grup metod – SISR i MISR – jest to, że są to zagadnienia źle postawione [20], czyli dla obrazu/obrazów LR istnieje wiele rozwiązań w postaci obrazu zrekonstruowanego SR. Rozprawa skupia się na trzech ostatnich punktach ze zdefiniowanej wcześniej listy wyzwań MISR (punkty 6–8).

Badania, które są opisane w rozdziałach 4.1 (ewolucyjna optymalizacja metod MISR z zastosowaniem przetwarzania wstępnego za pomocą SISR realizowane przez głębokie sieci konwolucyjne) oraz 4.2 (miary podobieństwa obrazów) były prowadzone w ramach projektów finansowanych przez Europejską Agencję Kosmiczną: SISPARE¹ w latach 2017-2018 oraz SuperDeep² w latach 2018-2020.

¹Satellite Image SPAtial Resolution Enhancement, <https://better.future-processing.com/projects/sispare>

²Super-resolution reconstruction of satellite images using deep convolutional neural networks

1.2 Tezy rozprawy

Jak wspomniano, w niniejszej pracy został podjęty temat rozwoju metod rekonstrukcji nadrozdzielczej wieloobrazowej. Szczególna uwaga została poświęcona pracy na danych pochodzących z satelitów przeznaczonych do obserwacji powierzchni Ziemi, w których uwidaczniają się problemy opisane wcześniej [21]–[23]. Zwykle współczesne metody rozwiązujące problem rekonstrukcji nadrozdzielczej oparte są na uczeniu maszynowym [5], [23], w szczególności konwolucyjnych sieciach neuronowych (ang. *convolutional neural networks*, CNN). Zbadane zostały możliwości połączenia klasycznych metod MISR z algorytmami opartymi na sztucznych sieciach neuronowych. W tym zakresie sprawdzono możliwość optymalizacji hiperparametrów metod za pomocą podejścia ewolucyjnego ze szczególnym zwróceniem uwagi na dane rzeczywiste pochodzące z satelitarnych obserwacji Ziemi. Jak pokazano w pracach [24]–[26] operowanie na realnych zdjęciach satelitarnych pochodzących z różnych instrumentów obserwacyjnych wymaga odpowiedniego sposobu oceny jakości obrazu wynikowego, tak by proces treningu przebiegał w odpowiednim kierunku. Z tego względu w rozprawie dokonano oceny miar spotykanych w literaturze pod kątem cech oczekiwanych w rekonstrukcji nadrozdzielczej. Skonstruowano również miary oparte na punktach charakterystycznych obrazu, które posiadają pożądaną w SR charakterystykę. Dalszym kierunkiem prac było sprawdzenie, czy oparte na głębokich sieciach konwolucyjnych (splotowych) miary podobieństwa, polegające na wykrywaniu punktów charakterystycznych obrazu – inne niż tradycyjnie stosowane L1 i L2 – mogą zostać efektywnie zastosowane w zadaniu MISR operującym na tego typu danych wejściowych.

W niniejszej rozprawie zostały zatem postawione dwie tezy:

- T1 Połączenie klasycznych metod wieloobrazowej rekonstrukcji nadrozdzielczej z technikami jednoobrazowymi wykorzystującymi głębokie splotowe sieci neuronowe, pozwala na uzyskanie lepszych jakościowo wyników niż zastosowanie którejkolwiek z metod bazowych z osobna.**

- T2 W wieloobrazowej rekonstrukcji nadrozdzielczej obrazów satelitarnych, realizowanej za pomocą głębokich splotowych sieci neuronowych, możliwe jest uzyskanie poprawy wyników poprzez zastosowanie przy obliczaniu funkcji straty metryki opartej na cechach głębokich służących do wykrywania punktów charakterystycznych w obrazach.**

Aby zweryfikować poprawność tezy T1, zostały wykonane eksperymenty polegające na wstępnym przetwarzaniu danych wejściowych za pomocą metod SISR

opartych na sztucznych sieciach neuronowych, a następnie fuzji informacji za pomocą klasycznych metod MISR z uwzględnieniem optymalizacji ich hiperparametrów. Wykonano eksperymenty na danych symulowanych, które potwierdziły ilościowo poprawę jakości rekonstrukcji. W przypadku danych symulowanych i rzeczywistych poprawa jakości została potwierdzona badaniem ankietowym wśród ludzkich obserwatorów.

Teza T2 została zweryfikowana z użyciem metody MISR, jednej z wiodących zgodnie z najnowszym stanem wiedzy. Zgodnie z wynikami badań z literatury [27], metryki typowo stosowane w SR, takie jak L1 i L2 są odpowiednie dla obrazów symulowanych, dla obrazów rzeczywistych należy znaleźć zaś odpowiedniejsze rozwiązanie. Zaproponowano tutaj autorską miarę, która za pomocą sieci głębokiej realizuje porównanie obrazów oparte na podobieństwie punktów charakterystycznych obrazu. Ta miara została użyta jako funkcja straty w trakcie treningu wspomnianej metody MISR. Zostały wykonane eksperymenty polegające na treningu wspomnianej metody MISR na danych satelitarnych z użyciem tak jej oryginalnej funkcji straty, jak i miar opartych na głębokich sieciach neuronowych. Poprawa jakości wyników została potwierdzona numerycznie za pomocą uznanych w literaturze metryk oraz za pomocą analizy subiektywnej oceny obserwatorów w postaci ankiety.

1.3 Cele rozprawy

Aby zweryfikować tezy postawione wcześniej, zdefiniowano następujące główne cele rozprawy:

1. Zaprojektować i zaimplementować sposób połączenia metod rekonstrukcji nadrozdzielczej opartych na sieciach głębokich z metodami klasycznymi.
2. Zweryfikować eksperymentalnie poprawę jakości dla danych symulowanych oraz rzeczywistych obrazów satelitarnych z użyciem zdefiniowanego zaproponowanego podejścia.
3. Zaprojektować i zaimplementować miary podobieństwa obrazu, które posiadają cechy pożądane w SR.
4. Zweryfikować eksperymentalnie przydatność wspomnianych miar podobieństwa obrazu w rekonstrukcji nadrozdzielczej w porównaniu do metryk spotykanych w literaturze.
5. Zaprojektować sposób, w jaki zaprojektowane miary podobieństwa mogą zostać użyte w treningu algorytmu MISR opartego na sieci głębokiej.

6. Zweryfikować eksperymentalnie poprawę jakości rekonstrukcji z wykorzystaniem zaproponowanej metody na rzeczywistych danych satelitarnych.

1.4 Opublikowane artykuły naukowe

Część z badań wykonanych w ramach niniejszej pracy, a związanych z tematyką rekonstrukcji nadrozdzielczej została już opublikowana w czasopismach oraz w materiałach z recenzowanych konferencji naukowych. Lista artykułów jest wymieniona wraz z punktacją MEiN (na rok publikacji rozprawy) oraz liczbą cytowań (Google Scholar, wraz z autocytowaniami) na dzień 1. września 2023.

Artykuły opublikowane w czasopismach

1. Kawulok Michał, **Benecki Paweł**, Piechaczek Szymon, Hrynczenko Krzysztof, Kostrzewa Daniel, Nalepa Jakub: Deep learning for multiple-image super-resolution, IEEE Geoscience and Remote Sensing Letters, 2020, vol. 17, nr 6, s.1062-1066. DOI:10.1109/LGRS.2019.2940483, **140 punktów MEiN, 83 cytowania.**

Autorska część: współudział w pracach koncepcyjnych, współudział w pracach implementacyjnych i walidacja eksperymentalna.

2. **Benecki Paweł**, Kawulok Michał, Kostrzewa Daniel, Skonieczny Łukasz: Evaluating super-resolution reconstruction of satellite images, Acta Astronautica, 2018, vol. 153, s.15-25. DOI:10.1016/j.actaastro. 2018.07.035, **35 punktów MEiN, 27 cytowań.**

Autorska część: współudział w opracowaniu koncepcji miar opartych na punktach charakterystycznych, implementacja i walidacja eksperymentalna.

Materiały konferencyjne

1. Kawulok Michał, Kostrzewa Daniel, **Benecki Paweł**, Skonieczny Łukasz: Evolving imaging model for super-resolution reconstruction, W: Proceedings of the 2018 Genetic and Evolutionary Computation Conference Companion : GECCO'18 Companion, Kyoto, Japan, 15-19 July 2018 / Aguirre H. (red.), 2018, Association for Computing Machinery, s.284-285, ISBN 978-1-4503-5764-7. DOI:10.1145/3205651.3205676, **140 punktów MEiN, 12 cytowań**

Autorska część: współudział w pracach koncepcyjnych, współudział w pracach implementacyjnych oraz współudział w walidacji eksperymentalnej.

2. Kawulok Michał, **Benecki Paweł**, Kostrzewa Daniel, Skonieczny Łukasz: Towards evolutionary super-resolution, W: Applications of evolutionary computation : 21st International conference, EvoApplications 2018, Parma, Italy,

April 4-6, 2018. Proceedings / Sim K., Kaufmann P. (red.), Lecture Notes In Computer Science, 2018, vol. 10784, Springer, s.480-496, ISBN 978-3-319-77537-1. DOI:10.1007/978-3-319-77538-8_33, **70 punktów MEiN, 6 cytowań**

Autorska część: współudział w pracach koncepcyjnych, współudział w pracach implementacyjnych oraz współudział w walidacji eksperymentalnej.

3. Kawulok Michał, **Benecki Paweł**, Nalepa Jakub, Kostrzewa Daniel: Evaluating super-resolution of satellite images: a Proba-V case study, W: 2020 IEEE International Geoscience and Remote Sensing Symposium : IGARSS 2020, 26 September - 2 October 2020, Virtual, Waikoloa, United States. Proceedings, 2020, Institute of Electrical and Electronics Engineers, s.641-644, ISBN 978-1-7281-6374-1.

DOI:10.1109/IGARSS39084.2020.9323908, **20 punktów MEiN**

Autorska część: współudział w pracach koncepcyjnych, implementacja, walidacja eksperymentalna.

4. Kawulok Michał, **Benecki Paweł**, Hrynczenko Krzysztof, Kostrzewa Daniel, Piechaczek Szymon, Nalepa Jakub, Smółka Bogdan: Deep learning for fast super-resolution reconstruction from multiple images, W: Real-Time Image Processing and Deep Learning 2019, Baltimore, Maryland, United States, 15-16 April 2019 / Kehtarnavaz N., Carlsohn M. (red.), Proceedings of SPIE: Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments, 2019, vol. 10996, Bellingham, SPIE - The International Society for Optics and Photonics, s.109960B-1 - 109960B-8, DOI:10.1117/12.2519579, **20 punktów MEiN, 10 cytowań**

Autorska część: współudział w pracach koncepcyjnych, współudział w pracach implementacyjnych oraz współudział w walidacji eksperymentalnej.

5. Kostrzewa Daniel, Piechaczek Szymon, Hrynczenko Krzysztof, **Benecki Paweł**, Nalepa Jakub, Kawulok Michał: Super-resolution reconstruction using deep learning: should we go deeper?, W: Beyond databases, architectures and structures : Paving the road to smart data processing and analysis. 15th International conference, BDAS 2019, Ustroń, Poland, May 28-31, 2019. Proceedings / Kozielski Stanisław [i in.] (red.), Communications in Computer and Information Science, 2019, vol. 1018, Springer, s.204-216, ISBN 978-3-030-19092-7. DOI:10.1007/978-3-030-19093-4_16, **20 punktów MEiN**

Autorska część: współudział w walidacji eksperymentalnej.

6. Kawulok Michał, Piechaczek Szymon, Hrynczenko Krzysztof, **Benecki Paweł**, Kostrzewa Daniel, Nalepa Jakub: On training deep networks for satellite image super-resolution, W: 2019 IEEE International Geoscience &

Remote Sensing Symposium : IGARSS 2019, July 28 - August 2, 2019, Yokohama, Japan. Proceedings, 2019, Institute of Electrical and Electronics Engineers, s.3125-3128, ISBN 978-1-5386-9155-7. DOI:10.1109/IGARSS.2019.8899098, **20 punktów MEiN, 12 cytowań**

Autorska część: współudział w pracach implementacyjnych i współudział w walidacji eksperymentalnej.

7. Kawulok Michał, **Benecki Paweł**, Nalepa Jakub, Kostrzewa Daniel, Skonieczny Łukasz: Towards robust evaluation of super-resolution satellite image reconstruction, W: Intelligent information and database systems : 10th Asian Conference, ACIIDS 2018, Dong Hoi City, Vietnam, March 19-21, 2018. Proceedings / Thanh Nguyen N. [i in.] (red.), Lecture Notes In Computer Science, 2018, Springer, s.476-486, ISBN 978-3-319-75416-1. DOI:10.1007/978-3-319-75417-8_45, **20 punktów MEiN, 3 cytowania**

Autorska część: walidacja eksperymentalna.

8. Kostrzewa Daniel, Skonieczny Łukasz, **Benecki Paweł**, Kawulok Michał: B4MultiSR: a benchmark for multiple-image super-resolution reconstruction, W: Beyond databases, architectures and structures : Facing the challenges of data proliferation and growing variety. 14th International conference, BDAS 2018 held at the 24th IFIP World Computer Congress, WCC 2018, Poznan, Poland, September 18-20, 2018. Proceedings / Kozielski Stanisław [i in.] (red.), Communications in Computer and Information Science, 2018, Springer, s.365-375, ISBN 978-3-319-99986-9. DOI:10.1007/978-3-319-99987-6_28, **20 punktów MEiN, 5 cytowań**

Autorska część: współudział w walidacji eksperymentalnej.

9. **Paweł Benecki**, Daniel Kostrzewa: Michal Kawulok, Perceptual loss for training multi-image super-resolution, W: 2023 IEEE International Geoscience and Remote Sensing Symposium : IGARSS 2023, July 16-21 2023, Pasadena, United States. **[w druku] 20 punktów MEiN**

Autorska część: współudział w pracach koncepcyjnych, implementacja i walidacja eksperymentalna.

1.5 Struktura rozprawy

Dalsza część pracy zawiera:

1. Przegląd spotykanych w literaturze algorytmów SR (rozdział 2.1), metod porównywania obrazów (rozdział 2.2) oraz zbiorów danych stosowanych w trakcie badań nad SR (rozdział 2.3).
2. Opis algorytmów opracowanych w ramach badań związanych z niniejszą rozprawą: zastosowanie optymalizacji ewolucyjnej w SR (rozdział 3.1), propozycje miar podobieństwa obrazów (rozdział 3.2), wraz z miarami podobieństwa opartymi na sieciach głębokich (rozdział 3.2.2) oraz sposób użycia tego typu miar w treningu modeli SR (rozdział 3.3).
3. Walidację eksperymentalną wymienionych wcześniej metod i podejść (rozdział 4) wraz z dyskusją wyników (rozdział 4.4).
4. Podsumowanie i możliwe kierunki dalszych badań (rozdział 5).

Rozdział 2

Przegląd literatury

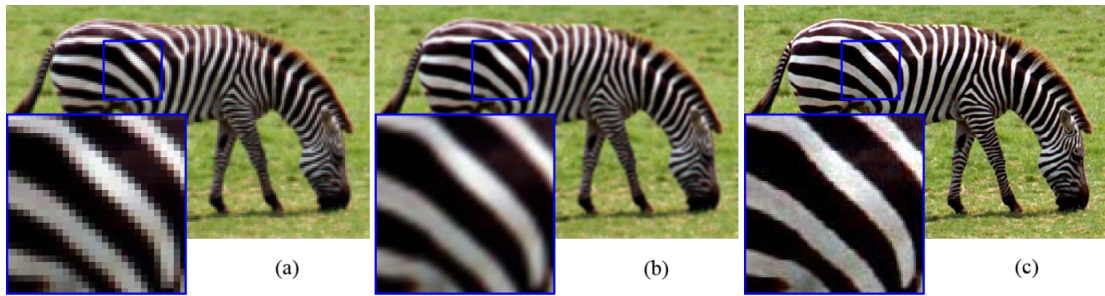
W niniejszym rozdziale został przedstawiony przegląd literatury związanej z tematyką SR. Opisano zarówno obecnie wiodące metody SR, jak i te, które stanowią podstawę do budowy nowych podejść MISR opisanych w dalszej części pracy (rozdział 2.1). Szczególny nacisk położono na pokazanie, jakie metryki są stosowane do oceny jakości obrazów – jakie są ich zalety i wady w zastosowaniu w MISR (rozdział 2.2). Równie ważnym elementem badań nad SR są dane wejściowe, dlatego też w dalszej części tego rozdziału przedstawiono zbiory (rozdział 2.3), które są używane w badaniach opisanych w rozdziale 4.

2.1 Metody rekonstrukcji nadrozdzielczej

Rozdzielczość obrazu cyfrowego jest jedną z głównych wartości charakteryzujących jakość obrazu. Większa rozdzielczość jest pożądana, a często wymagana w wielu zastosowaniach [5], ponieważ często oznacza więcej użytecznych szczegółów obrazu. Możliwe jest uzyskanie wyższej rozdzielczości przy pomocy lepszego sensora lub optyki, jest to jednak rozwiązanie kosztowne, a w przypadku wielu urządzeń takich jak na przykład już istniejące satelity, podmiana aparatury jest niemożliwa. Częstość ograniczenia wynikają również z tego, że instrument o lepszych parametrach cechowałby się większymi rozmiarami i wagą. Stąd metody pozwalające na podwyższenie rozdzielczości obrazów są bardzo pożądane w dziedzinie przetwarzania obrazów cyfrowych. Dlatego też od wielu lat jest to aktywny problem badawczy [5].

Rekonstrukcja nadrozdzielcza obrazów cyfrowych jest pojęciem obejmującym metody generujące zrekonstruowane obrazy wyższej rozdzielczości $\mathcal{I}^{(sr)}$ – z jednego lub więcej obrazów niższej rozdzielczości $\mathbf{I}^{(L)}$. Przykład działania rekonstrukcji i porównanie z interpolacją są pokazane na rys. 2.1.

W niniejszej rozprawie skupiono się na zastosowaniu w teledetekcji, w szczegól-



Rysunek 2.1: Porównanie działania interpolacji oraz rekonstrukcji nadrozdzielczej: a) obraz niskiej rozdzielczości, powiększony dla celów wizualizacji interpolacją NN, b) interpolacja dwuliniowa, c) obraz poddany rekonstrukcji nadrozdzielczej jedno-obrazowej. Rysunek pochodzi z pracy [28].

ności na zwiększaniu rozdzielczości obrazów satelitarnych. Inne obszary, w których SR ma zastosowanie to m.in:

- Obrazowanie medyczne: SR może mieć zastosowanie w obrazowaniu medycznym do poprawy rozdzielczości obrazów takich jak obrazy rentgenowskie, tomografia komputerowa i rezonans magnetyczny [29], [30]. W tej dziedzinie SR pozwala lekarzom dostrzegać mniejsze szczegóły i nieprawidłowości, tym samym pozwalając na stawianie dokładniejszych diagnoz. Przykładem praktycznej implementacji jest *Philips MicroFlow Imaging*¹ używany w obrazowaniu ultrasonograficznym.
- Teledetekcja: SR użyte w aplikacjach teledetekcji do zwiększenia rozdzielczości obrazów satelitarnych i lotniczych może pomóc lepiej zrozumieć i monitorować wykorzystanie terenu, wzorce pogodowe i inne czynniki środowiskowe [13], [21].
- Systemy nadzoru oraz kryminalistyka: rekonstrukcja nadrozdzielcza może być wykorzystywana w aplikacjach nadzoru do poprawy jakości nagrań z kamer bezpieczeństwa o niskiej rozdzielczości. Dzięki temu śledczy mogą mieć możliwość lepszej identyfikacji podejrzanych oraz skuteczniejszego gromadzenia dowodów [31], [32]. Komercyjne systemy tej klasy to m.in. *Amped FIVE*² lub *Cognitech Video Investigator*³

¹<https://www.usa.philips.com/healthcare/resources/feature-detail/ultrasound-microflow-imaging>

²<https://ampedsoftware.com/five>

³<https://cognitech.com/cognitech-video-investigator-64/>

- Fotografia: technika ta może być stosowana w fotografii do zwiększenia rozdzielczości obrazów o niskiej rozdzielczości wykonanych telefonem komórkowym lub innym aparatem [33]–[35]. Przykładem praktycznego wdrożenia jest mechanizm *Samsung Scene Optimiser*⁴. Do obróbki obrazów już wykonanych można wykorzystać metody SR dostarczane z Adobe Photoshop⁵.
- Przetwarzanie wideo: SR jest używane w aplikacjach do przetwarzania wideo, aby poprawić jakość wideo o niskiej rozdzielczości [36]–[38] – przykłady komercyjnych wdrożeń to: *Infognition* – dodatek do Adobe After Effects⁶, *Topaz Video Enhance AI*⁷, czy też *DaVinci Resolve*⁸.
- Systemy mikroskopowe – dodawanie szczegółów o wymiarach niższych niż limit wyznaczony długością fali świetlnej, przykładowo zaimplementowane w systemach *Hamamatsu*⁹.

Rekonstrukcja jednoobrazowa polega na modelowaniu relacji między obrazami wysokiej i niskiej rozdzielczości. Zasadniczo tego typu przetwarzanie generuje obraz, który wygląda lepiej niż klasyczna interpolacja. Dodawana informacja pochodzi z pewnych założeń a priori lub też wiedzy wyuczonej na zbiorze treningowym. Zwykle jednak dodane szczegóły są takie jakie byłyby oczekiwane w podobnej scenie w docelowej rozdzielczości, a niekoniecznie odzwierciedlają faktyczny wygląd sceny w wysokiej rozdzielczości [5]. Z kolei metody wieloobrazowe używają informacji wyekstrahowanej z obrazów w trakcie rekonstrukcji, stąd są formą fuzji informacji. Dlatego też dodane detale mogą być faktyczną rekonstrukcją wyglądu obiektu lub obszaru w wysokiej rozdzielczości. Obrazy LR tej samej sceny zwykle są przesunięte względem siebie na poziomie subpikselowym. Dzięki temu każde z ujęć LR zawiera nieco inne szczegóły, które mogą zostać uwytatnione w obrazie HR w trakcie rekonstrukcji [13], [21].

2.1.1 Metody jednoobrazowe

Kodowanie rzadkie

Dużą grupę algorytmów rekonstrukcji nadrozdzielczej stanowią metody oparte na kodowaniu rzadkim (ang. *sparse coding*) [39], [40]. Jest to jedna z metod uczenia

⁴<https://www.samsung.com/ie/support/mobile-devices/how-galaxy-cameras-combine-super-resolution-technologies-with-ai-to-produce-high-quality-images-of-the-moon/>

⁵<https://www.adobe.com/products/photoshop.html>

⁶http://www.infognition.com/super_resolution/

⁷<https://www.topazlabs.com/topaz-video-ai>

⁸<https://www.blackmagicdesign.com/pl/products/davinciresolve>

⁹<https://camera.hamamatsu.com/content/dam/hamamatsu-photonics/sites/static/system/en/documents/Using%20Super-Resolution%20Nanorulers%20to%20study%20the%20Capabilities%20of%20EM-CCD%20and%20sCMOS.pdf>

maszynowego oparta na słowniku funkcji bazowych (atomów). Wynikiem działania algorytmów tej klasy, w tym wypadku obrazem rekonstruowanym, jest kombinacja liniowa atomów. W rekonstrukcji nadrozdzielczej słownik atomów jest uczony z zestawu obrazów o wysokiej rozdzielczości. Niektóre metody działają bezpośrednio na obrazie, inne w trakcie rekonstrukcji operują w dziedzinie falek [41] lub innych transformatach obrazu.

Jedną z popularnych metod rekonstrukcji nadrozdzielczej z wykorzystaniem kodowania rzadkiego jest podejście oparte na reprezentacji rzadkiej (ang. *sparse representation-based approach*, SRBA) [42]. Algorytm SRBA uczy się słownika funkcji bazowych ze zbioru obrazów o wysokiej rozdzielczości, a następnie wykorzystuje ten słownik do rekonstrukcji obrazu o niskiej rozdzielczości. Rekonstrukcja jest osiągnięta przez rozwiązanie problemu optymalizacyjnego, który ma na celu znalezienie reprezentacji obrazu LR w kategoriach wyuczonego słownika.

Inną metodą wykorzystującą kodowanie rzadkie jest CSBA (ang. *compressed-sensing based approach*) [43]. Algorytm CSBA opiera się na tym, że sygnał może być zrekonstruowany z niewielkiej liczby pomiarów, o ile ma on reprezentację rzadką w jakiejś bazie. Algorytm CSBA najpierw pozyskuje zestaw pomiarów o niskiej rozdzielczości, a następnie wykorzystuje kodowanie rzadkie do rekonstrukcji obrazu o wysokiej rozdzielczości.

Wadą wielu podejść opartych na kodowaniu rzadkim jest to, że operują na poziomie fragmentów obrazów (ang. *patch*), przetwarzając każdy z nich osobno. Co prawda fragmenty zwykle zachodzą na siebie, ale wzajemne relacje między nimi są ignorowane [44]. Idea kodowania rzadkiego bywa również używana w innych podejściach takich jak sieci konwolucyjne oparte na kodowaniu rzadkim (ang. *convolutional sparse coding*) [44], [45].

Uczenie głębokie

W ostatnich latach nastąpił duży postęp w wykorzystaniu uczenia głębokiego w dziedzinie przetwarzania obrazów. Nie pozostało to bez wpływu na dziedzinę rekonstrukcji nadrozdzielczej, powodując znaczny rozwój metod, w szczególności opisanych dalej. Rozwój metod SISR jest też stymulowany przez konkursy dotyczące SR, takie jak NTIRE (ang. *New Trends in Image Restoration and Enhancement*) lub opisywany w dalszej części rozprawy (rozdział 2.1.3) konkurs PROBA-V.

Sieci SRCNN i FSRCNN

Pierwszą opublikowaną metodą był algorytm SRCNN (ang. *super-resolution CNN*) [46]. Metoda ta była inspirowana podejściem sparse coding, które zaimplementowano za pomocą sieci składającej się z zestawu warstw konwolucyjnych. Wejściem do sieci jest obraz $\mathcal{I}^{(l)}$, powiększony wstępnie do pożądanego rozmiaru za pomocą

interpolacji. Zadaniem sieci jest przeprowadzenie odpowiedniej filtracji tego obrazu w celu poprawy jakości odwzorowania szczegółów. Parametry procesu filtracji są reprezentowane przez wagi sieci, których wartości są wyznaczane w procesie treningu. Ci sami autorzy zaprojektowali model szybszy i dający lepszą jakość rekonstrukcji (*ang. fast SRCNN*, FSRCNN) [47]. Tutaj z kolei obraz nie był powiększany na wejściu, ale dopiero po kilku warstwach konwolucyjnych za pomocą warstwy dekonwolucyjnej.

Użycie przeciwstawnych sieci generatywnych – SRGAN

Innym podejściem, opartym o sieci generatywne (*ang. generative adversarial networks*, GAN) jest SRGAN [48] z 2017 roku. Architektura GAN [49] składa się z dwóch sieci: generatora i dyskryminatora. Pierwsza z nich służy do generowania wyniku, druga natomiast pełni rolę klasyfikatora oceniającego czy otrzymane na jej wejściu dane pochodzą z generatora, czy też są danymi rzeczywistymi. Proces uczenia GAN przebiega naprzemiennie w dwóch fazach. Jedną z nich jest trening dyskryminatora, który jest uczony na rzeczywistych danych ze zbioru treningowego oraz na danych wygenerowanych przez generator. Drugą fazę stanowi trening generatora, w trakcie którego minimalizowana jest zdolność dyskryminatora do poprawnego klasyfikowania wygenerowanych danych. W trakcie iteracji wspomnianych faz, generator generuje dane, które są dla dyskryminatora coraz trudniejsze do odróżnienia od prawdziwych.

W przypadku SRGAN, generator ma za zadanie zwiększyć rozdzielczość obrazów, podczas gdy dyskryminator próbuje odróżnić je od prawdziwych obrazów wysokiej jakości. Celem jest doprowadzenie do stanu, gdy dyskryminator nie jest w stanie odróżnić obrazów generowanych od realnych. Architektura SRGAN składa się z generatora, zawierającego warstwy służące do ekstrakcji cech obrazu w postaci sieci ResNet (*ang. residual network*, sieć z połączeniami rezydualnymi) oraz dyskryminatora. Generowany obraz jest na bieżąco oceniany przez dyskryminator, co pozwala na poprawę jakości obrazu w trakcie procesu uczenia. SRGAN umożliwia uzyskanie wyraźnych i naturalnie wyglądających szczegółów w obrazach o niskiej rozdzielczości. Wadami tego rozwiązania są spotykane niekiedy w rozwiązaniach typu GAN artefakty – halucynacje, czyli obiekty, których nie ma w rzeczywistości oraz zjawisko nazywane *mode collapse*, polegające na generowaniu mało zróżnicowanych wyników pomimo zróżnicowanego zbioru treningowego [48], [50], [51].

EDSR

SRGAN stał się podstawą dalszych badań, co doprowadziło do powstania następnym metod SR. Wskazywano na słabości podejścia GAN wymienione wcześniej

oraz wysokie wymagania pamięciowe oraz obecność warstw normalizacji wsadowej (ang. *batch normalization*), które w tym zastosowaniu mogły redukować wydajność modelu. Jedną z nich jest metoda EDSR (ang. *Enhanced Deep Residual Networks for Single Image Super-Resolution*) (EDSR) [52]. EDSR wykorzystuje podobnie jak SRGAN głęboką architekturę ResNet (bez wspomnianych warstw normalizacji wsadowej) z zastosowaniem warstw dekonwolucyjnych, jednak nie jest oparty na sposobie treningu GAN.

Inne metody oparte na GAN

Kolejnym podejściem jest generowanie SR w GAN przez warunkowanie na podstawie treści (ang. *Content-Conditioned GAN for Super-Resolution*) z 2018 roku [53]. Metoda ta opiera się na wykorzystaniu autoenkodera do wydobycia cech obrazu oraz sieci dyskryminatora, która ma za zadanie ocenić jakość wygenerowanego obrazu. Innym ciekawym podejściem jest High-Resolution Image Synthesis via Progressive Growing of GANs (ProGAN) [54]. Metoda ta wykorzystuje stopniowe zwiększanie rozdzielczości generowanych obrazów podczas procesu uczenia. SISR trenowane w schemacie GAN – OpTiGAN zostało użyte również do rekonstrukcji obrazów satelitarnych [55], umożliwiając zwiększenie rozdzielczości z GSD 31 cm do 9 cm, czyli 3,69-krotnie.

LapSRN

W pracy [56] zaproponowano model oparty na głębokich sieciach neuronowych, wykorzystujący piramidy Laplace’a (ang. *Laplacian pyramids*). Model przyjmuje obraz LR jako wejście i stopniowo generuje dodatkową informację na różnych poziomach.

Model składa się z dwóch głównych odgałęzień: ekstrakcji cech i rekonstrukcji obrazu. Gałąź ekstrakcji cech składa się z warstw konwolucyjnych i transponowanych warstw konwolucyjnych. Na każdym poziomie wydobywane są cechy, które są następnie poddawane operacji zwiększania rozdzielczości za pomocą transponowanej warstwy konwolucyjnej. Wyjście z tej warstwy jest połączone z dwoma oddzielnymi warstwami konwolucyjnymi: jedną służącą do odtwarzania obrazu resztkowego na bieżącym poziomie, a drugą do ekstrakcji cech na kolejnym, bardziej szczegółowym poziomie. Sama ekstrakcja cech odbywa się w przestrzeni LR, co redukuje złożoność obliczeniową. Współdzielone reprezentacje cech z niższych poziomów zwiększają nieliniowość sieci, umożliwiając naukę złożonych odwzorowań na bardziej szczegółowych poziomach.

Gałąź rekonstrukcji obrazu wykorzystuje transponowaną warstwę konwolucyjną do zwiększenia rozdzielczości obrazu wejściowego dwukrotnie na każdym poziomie. Początkowym przybliżeniem obrazu SR jest interpolacja dwuliniowa. Tak otrzy-

mane przybliżenie obrazu jest sumowane z danymi z gałęzi ekstrakcji cech, co daje jako rezultat obraz o wysokiej rozdzielczości. Wygenerowany obraz na danym poziomie jest używany jako wejście do gałęzi rekonstrukcji obrazu na poziomie następnym. Cała sieć składa się z kaskady sieci konwolucyjnych o podobnej strukturze na każdym poziomie i umożliwia powiększenia będące potęgami liczby 2.

2.1.2 Metody wieloobrazowe

Proces rejestracji obrazów

Rekonstrukcja wieloobrazowa wykorzystuje proces fuzji informacji z wielu obrazów LR. Ponieważ obrazy te mogą się różnić w zakresie położenia obiektów i obszarów, które prezentują, należy zapewnić dopasowanie przestrzenne tych obrazów. Robi się to za pomocą techniki zwanej rejestracją obrazów. Jest to jeden z podstawowych procesów w dziedzinie przetwarzaniu obrazów. Polega na wyrównaniu wielu obrazów tej samej sceny wykonanych w różnym czasie, z różnych punktów widzenia lub za pomocą różnych czujników. Celem rejestracji obrazu jest doprowadzenie tych obrazów do wspólnego układu współrzędnych, wyrównując je przestrzennie, tak aby odpowiadające sobie punkty w obrazach, prezentujące te same obiekty, dokładnie do siebie pasowały. Rejestracja obrazu polega na znalezieniu transformacji (takiej jak translacja, obrót, skalowanie lub nieliniowe deformacje), która doprowadzi obrazy do wyrównania. Transformacja ta jest szacowana na podstawie odpowiednich cech lub punktów na obrazach. Cechy te mogą być oznaczane ręcznie lub wykrywane automatycznie przy użyciu technik takich jak wykrywanie krawędzi, wykrywanie narożników lub metody wykrywania punktów kluczowych. W przypadku rejestracji subpikselowej są stosowane metody pozwalające na znalezienie transformacji dopasowujących obrazy na poziomie odległości poniżej pojedynczego piksela. Oba rodzaje rejestracji są stosowane w rekonstrukcji nadrozdzielczej wieloobrazowej.

Model obrazowania

Podczas akwizycji danych obrazowych zachodzą procesy wynikające z budowy optyki i układów elektronicznych, które wprowadzają do obrazu niepożądane efekty takie jak rozmycie i szumy. Są one związane z konkretnym systemem obrazującym – różne układy mają różną charakterystykę tych degradacji. W pracach związanych z rekonstrukcją nadrozdzielczą pojawia się pojęcie *modelu obrazowania* (ang. *imaging model*, IM), który formalnie opisuje wspomniane efekty [5], [11], [20] modelując transformacje takie jak m.in.: rozmycie, przesunięcia i obroty, zmniejszanie rozdzielczości przestrzennej oraz szumy [5]. Procesy zachodzące w metodach wieloobrazowych można opisać jako próbę odwrócenia procesu obrazowania opisanego

za pomocą IM . Mamy zatem degradację hipotetycznego obrazu HR $\mathcal{I}_i^{(h)}$ w zbiór N obrazów LR: $I^{(L)} = \left\{ \mathcal{I}_i^{(l)} : i \in [1..N] \right\}$.

Metody rekonstrukcji wieloobrazowej opierają się na założeniu, że obserwacje LR ($\mathcal{I}^{(l)}$) są otrzymywane z hipotetycznego obrazu HR ($\mathcal{I}^{(h)}$) w następujący sposób [33]:

$$\mathcal{I}_i^{(l)} = \mathbf{D}_i \mathbf{B}_i \mathbf{W}_i \mathcal{I}^{(h)} + n_i, \quad (2.1)$$

gdzie:

- \mathbf{D} jest macierzą opisującą zmniejszanie rozdzielczości,
- \mathbf{B} jest macierzą opisującą operację rozmycia,
- \mathbf{W} jest macierzą opisującą przekształcenie geometryczne, zawierającą przesunięcie i obrót,
- n oznacza szum dodany do obrazu LR.

Gdy akwizycja obrazów jest dokonywana za pomocą tego samego sensora, zakłada się, że \mathbf{B} oraz \mathbf{D} są wspólne dla całej serii wytworzonych obrazów, ponieważ wynikają one z charakterystyki, w szczególności optycznej instrumentu. Stąd można uznać, że różnice między obserwacjami LR wynikają tylko z przekształceń geometrycznych (obrotu i przesunięć) oraz z szumu.

Odwrócenie transformacji modelu obrazowania

Rekonstrukcja nadrozdzielcza jest zagadnieniem źle postawionym (ang. *ill-posed problem*), posiadającym wiele rozwiązań. Tego typu zadania są rozwiązywane za pomocą metod Bayesowskich lub opartych na optymalizacji gradientowej z dodaniem składnika regularyzacji by zapewnić naturalny wygląd obrazu wynikowego $\mathcal{I}^{(sr)}$. Możliwymi celami regularyzacji, w zależności od użytej funkcji, są m.in.: redukcja szumów, zachowanie krawędzi oraz wygładzanie obrazu [20]. Regularyzacja zmienia problem z postaci *maksymalnego prawdopodobieństwa* (ang. *maximum likelihood*) w estymację typu *maximum a posteriori* [57]. Pierwsze z podejść skupia się na znalezieniu wartości parametrów, które maksymalizują prawdopodobieństwo obserwowanych danych dla tych parametrów, bez uwzględniania wcześniejszych informacji. Drugie – łączy maksymalizację prawdopodobieństwa danych z informacjami wstępnymi (ang. *prior information*).

Historycznie, w jednych z pierwszych prac SR była używana iteracyjna projekcja wsteczna (ang. *iterative back projection*, IBP) [58], metoda szeroko stosowana również m.in. w obrazowaniu medycznym. Jako dane wejściowe podawane były obrazy LR poddane rejestracji z dokładnością subpikselową.

Z kolei w metodach opartych na rzutowaniu gradientu (ang. *gradient projection algorithm*, GPA) [59] używana była hierarchiczna estymacja przesunięć sub-pikselowych, połączona z metodami Bayesowskimi. W algorytmach używanych do rekonstrukcji sekwencji wideo do modelu obrazowania dodatkowo dodane jest rozmycie wynikające z ruchu obiektów (ang. *motion blur*) [60].

Inną metodą optymalizacji stosowaną w SR jest rzutowanie na zbiory wypukłe (ang. *projection onto convex sets*, POCS) [61], która polega na iteracyjnej zmianie estymowanego obrazu HR za pomocą błędu między $\mathcal{I}^{(l)}$ a $\mathcal{I}'^{(l)}$, czyli pomniejszonej do rozmiaru LR wersji obrazu $\mathcal{I}'^{(h)}$, która jest degradowana założonym modelem obrazowania.

Spotyka się również metody działające na podstawie adaptacyjnych filtrów Wienera [62] i ukrytych pól Markowa [63].

SR oparte na największym prawdopodobieństwie *a posteriori*

W algorytmach działających na podstawie teorii największego prawdopodobieństwa *a posteriori* rozwiązanie w postaci obrazu wynikowego $\mathcal{I}^{(sr)}$ wynika z minimalizacji problemu opisanego wyrażeniem:

$$\mathcal{I}'^{(h)} = \arg \min_{\mathcal{X}} \sum_{i=1}^N \rho \left(\mathcal{I}_i^{(l)}, \mathbf{DBW}_i \mathcal{X} \right) + \lambda U(\mathcal{X}) \quad (2.2)$$

gdzie:

- ρ oznacza miarę niepodobieństwa między obrazami,
- U oznacza regularyzację dodaną z współczynnikiem wpływu λ .

Istniejące algorytmy SR używają różnych postaci macierzy degradacji oraz metod regularyzacji. Również strategie optymalizacji bywają definiowane w różny sposób.

Projekcja wsteczna

Iteracyjna projekcja wsteczna została zaproponowana do realizacji SR w pracy [58]. Główną ideą IBP jest iteracyjne minimalizowanie błędu rekonstrukcji między obrazem HR a obserwacjami LR przez naprzemienną realizację dwóch etapów – reprojekcji i projekcji wstecznej:

- Reprojekcja: w tym etapie algorytm szacuje obraz HR, rzutując obserwacje LR na współrzędne HR. Polega to na interpolacji obrazów LR, a następnie zastosowaniu odpowiedniej rejestracji i splocie z funkcją rozmycia punktu (ang. *point spread function*, PSF) w celu uwzględnienia modelu obrazowania.

- Projekcja wsteczna: w etapie projekcji wstecznej bieżące oszacowanie HR jest próbkowane z powrotem do dziedziny LR. Następnie obliczany jest błąd między próbkowanym obrazem HR a oryginalnymi obserwacjami LR, który jest reprojekowany na domenę HR w celu udoskonalenia oszacowania HR.

Oba etapy są powtarzane iteracyjnie do momentu spełnienia określonego kryterium zatrzymania takiego jak maksymalna liczba iteracji lub próg błędu rekonstrukcji.

IBP w rekonstrukcji nadrozdzielczej zostało zastosowane w pracy [58]. Metoda ta wykorzystuje hierarchiczną technikę rejestracji do oszacowania przesunięć subpikselowych między obrazami LR, a następnie te informacje są wprowadzane podczas etapu reprojekcji w celu rekonstrukcji obrazu HR.

Metoda Fast and Robust Super-Resolution

W metodzie Fast and Robust Super Resolution (FRSR) [20] zaproponowano adaptacyjną metodę IBP, opierającą się na estymacji maksymalnego prawdopodobieństwa z uproszczoną regularyzacją w postaci utworzonego przez autorów podejścia *bilateral total variation*. Algorytm ten w trakcie działania operuje od razu na współrzędnych HR (opisane dalej), dzięki czemu unika się kosztownej operacji iteracyjnego zmniejszania / powiększania obrazu. Metoda ta jest bardzo ważna w kontekście niniejszej rozprawy ze względu na to, że bazują na niej algorytmy EvoIM oraz EvoNet opisane w rozdziale 3.1. Ważnym elementem używanym w tym algorytmie jest podejście MSA (ang. *median shift and add*), które pozwala na inicjalizację pierwotnego oszacowania HR w taki sposób, by piksele LR były rzutowane na siatkę HR zgodnie z ich lokalizacją oszacowaną za pomocą rejestracji, co przedstawione jest na rys. 2.2. Tak rzutowane są wszystkie obrazy LR, następnie dla każdego piksela na siatce HR obliczana jest mediana wartości pochodzących z każdego z obrazów wejściowych. Z kolei piksele nieokreślone są następnie uzupełniane za pomocą operatora mediany w otoczeniu dwuwymiarowym.

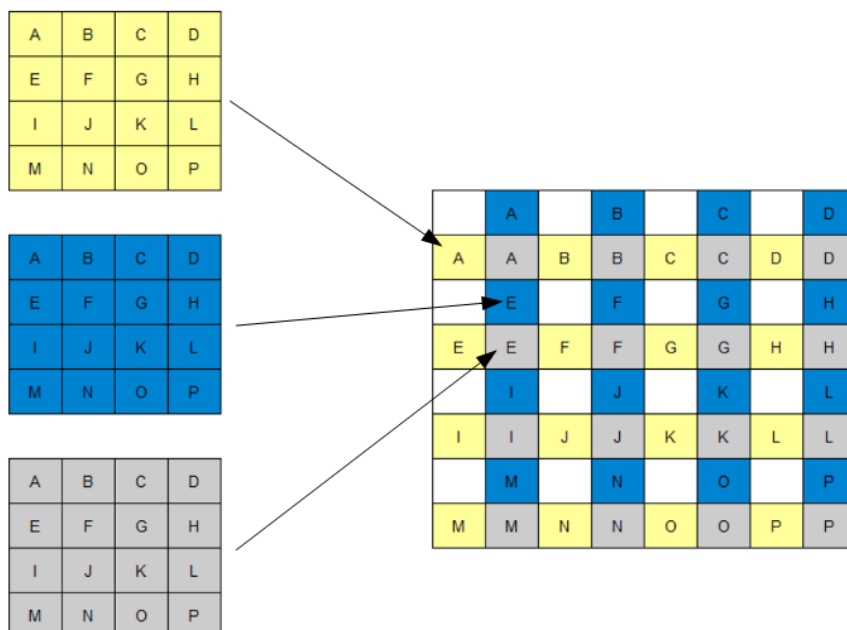
Po wykonaniu operacji MSA otrzymuje się oszacowanie:

$$\hat{Z} = \mathbf{H}\hat{X}, \quad (2.3)$$

gdzie \hat{Z} jest wynikiem operacji MSA, \mathbf{H} jest nieznaną macierzą PSF a \hat{X} jest wynikiem HR. Aby oszacować \hat{X} przeprowadzana jest iteracyjna operacja:

$$\begin{aligned} \hat{X}_{n+1} = & \hat{X}_n - \beta \{ \mathbf{H}^T \mathbf{A}^T \text{signum}(\mathbf{A}\mathbf{H}\hat{X}_n - \mathbf{A}\hat{Z}) + \\ & + \lambda' \sum_{l=-P}^P \sum_{m=0}^P \alpha^{|m|+|l|} [\mathbf{I} - \mathbf{S}_y^{-m} \mathbf{S}_x^{-l}] \text{signum}(\hat{X}_n - \mathbf{S}_x^l \mathbf{S}_y^m \hat{X}_n) \}, \end{aligned} \quad (2.4)$$

gdzie \mathbf{I} jest macierzą jednostkową, a \mathbf{S} jest macierzą przesunięcia, zaś macierz \mathbf{A} jest macierzą diagonalną o wartościach na przekątnej równych pierwiastkowi



Rysunek 2.2: Zasada działania metody MSA, użytej w [20] oraz metodzie EvoIM/EvoNet. Piksele LR są rzutowane na siatkę HR zgodnie z oszacowanymi przesunięciami obrazów LR względem siebie – tutaj obraz niebieski jest przesunięty względem żółtego o wektor $(0,5; 0,5)$, a szary względem żółtego o $(0,5; 0)$. W przypadku, gdy jest więcej niż jedna wartość dla jednego piksela HR, obliczana jest mediana. Nieokreślone piksele na siatce HR (białe pola) są uzupełniane filtrem medianowym z wartości pikseli określonych w otoczeniu takiego punktu.

kwadratowemu z liczby pomiarów, które przyczyniły się do utworzenia każdego elementu macierzy \hat{Z} . α, β, λ są hiperparametrami metody FRSR. W metodach EvoIM / EvoNet osobno optymalizowana jest postać macierzy \mathbf{H} .

Nowsze metody bazujące na IBP

Koncepcja projekcji wstecznej, która stanowi podstawę metody IBP, została również wykorzystana w nowszych rozwiązaniach opartych o splotowe sieci neuronowe, ponieważ zapewnia elastyczny i interpretowalny proces rekonstrukcji.

Jedną z metod, która bazuje na idei podejścia IBP (aczkolwiek jest to metoda SISR) jest głęboka sieć reprojekcji wstecznej (ang. *deep back projection network*, DBPN) [64], która zawiera iteracyjną naturę IBP w architekturze uczenia głębokiego.

Inną ostatnio opublikowaną metodą jest praca definiująca sieć reprojekcji wstecz-

nej z wykorzystaniem metod atencji (AG-BPN) [65], integrująca mechanizm atencji z procesem reprojekcji wstecznej. Umożliwia to AG-BPN przechwytywanie bardziej złożonych zależności między pikselami oraz globalnych informacji kontekstowych, co dodatkowo poprawia wyniki SR.

2.1.3 Rekonstrukcja wieloobrazowa realizowana za pomocą sieci głębokich

Rozwój metod wieloobrazowych SR został przyspieszony po opublikowaniu zbioru danych z satelity PROBA-V: w 2019 Europejska Agencja Kosmiczna ogłosiła konkurs, którego polegał na rekonstrukcji nadrozdzielczej przy użyciu danych z tego satelity¹⁰. Został opublikowany duży zbiór obrazów LR i HR, różniących się rozdzielczością w stosunku 1:3 (obrazy LR mają wielkość 128×128 , HR – 384×384), gdzie dla jednego obrazu HR istnieje co najmniej 9 obrazów LR. Dzięki wysokiej jakości danych wejściowych oraz następującemu w międzyczasie przyrostowi mocy obliczeniowej GPU możliwe stało się trenowanie rozbudowanych modeli.

DeepSUM

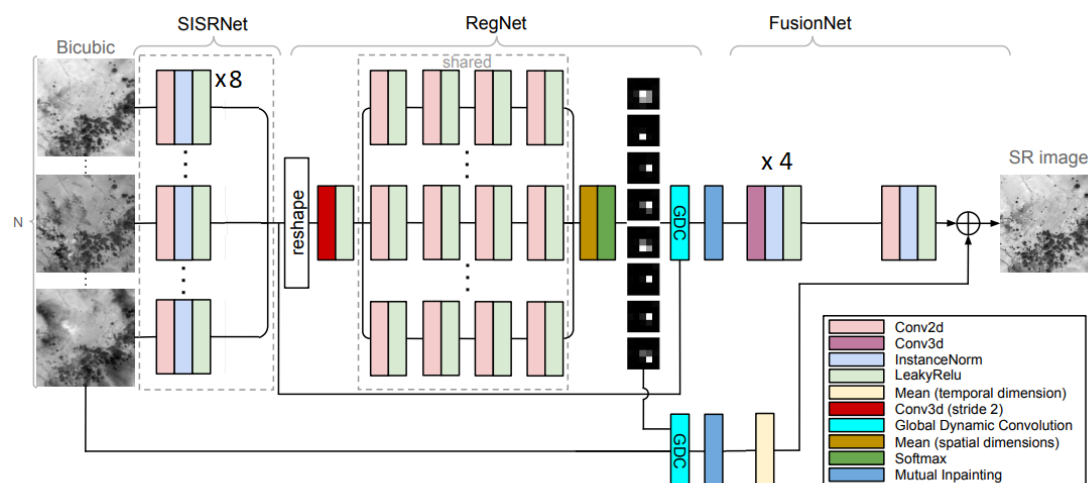
Wszystkie czołowe miejsca w konkursie PROBA-V zajęły metody wykorzystujące metody uczenia głębokiego ([66], [67]). Zwycięski zespół opublikował swoją metodę w pracy [67]. Model DeepSUM wykorzystuje jako dane wejściowe niezarejestrowane obrazy LR. Metoda składa się z trzech głównych składników: modułu ekstrakcji cech, modułu rejestracji obrazu oraz modułu rekonstrukcji (rys. 2.3).

Moduł rejestracji (nazwany przez autorów *RegNet*) obrazu odpowiedzialny jest za dopasowanie obrazów LR na poziomie subpikselowym (pełnopikselowym w rozdzielczości obrazu HR). Ten etap w innych klasycznych metodach [20], [58], [59] był wykonywany w kroku przetwarzania wstępnego za pomocą innych środków, niewykorzystujących sieci głębokich. Jest to istotna nowość wprowadzona w opisywanej metodzie. Dzięki rejestracji, następny krok algorytmu operuje już na obrazach dopasowanych do siebie przestrzennie.

Moduł ekstrakcji cech (SISRNet) wykorzystuje głęboką sieć konwolucyjną do wydobycia cech wysokiego poziomu z obrazów LR. Struktura sieci została zaprojektowana tak, aby uchwycić zarówno zależności przestrzenne, jak między obrazami LR.

Moduł rekonstrukcji (FusionNet) używa uprzednio wyodrębnionych cech i wykorzystując zestaw warstw konwolucyjnych i dekonwolucyjnych generuje zrekonstruowany obraz.

¹⁰<https://kelvins.esa.int/proba-v-super-resolution/>

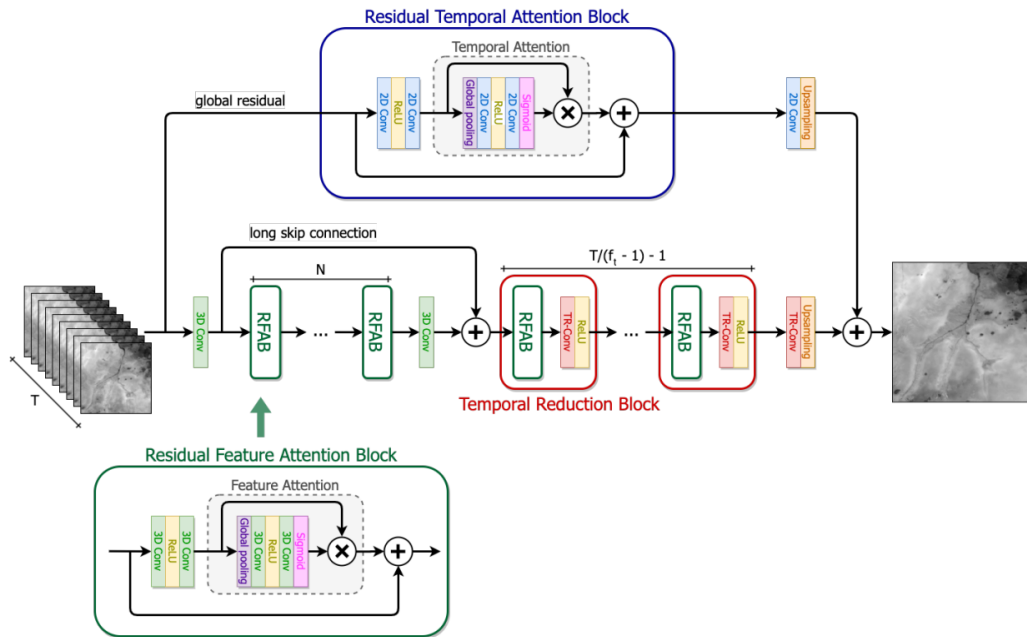


Rysunek 2.3: Schemat sieci DeepSum: moduł ekstrakcji cech (SISRNet), rejestracji (RegNet) oraz rekonstrukcji (FusionNet). Rysunek pochodzi z pracy [67].

Jako funkcja straty w treningu sieci używana jest miara L2 z kompensacją różnic jasności i przesunięć pełnopikselowych – wspomniana wcześniej metoda cPSNR.

RAMS

Opisana w artykule [66] metoda RAMS (ang. *Residual Attention Multi-image Super-resolution*) również powstała w ramach prac nad konkursem PROBA-V. Wykorzystuje sieć neuronową z blokami uwagi rezydualnej (ang. *residual attention*), która efektywnie ekstrahuje cechy wysokiego poziomu z obrazów LR. Mechanizm uwagi pozwala zidentyfikować najbardziej istotne cechy i zwiększyć ich wagę, zmniejszając znaczenie mniej istotnych, tym samym redukując ograniczenia lokalnego zasięgu operacji splotu. Ponadto wykorzystanie połączeń rezydualnych pozwala na przepływ informacji o niskiej częstotliwości z zestawu obrazów LR przez sieć, umożliwiając modelowi skoncentrować się tylko na składnikach o wysokiej częstotliwości. Informacja o wysokiej częstotliwości jest bardziej istotna dla rekonstrukcji obrazu o wyższej rozdzielczości, podczas gdy obrazy LR zawierają dużo informacji o niskiej częstotliwości, które można bezpośrednio przekazać do ostatecznego wyniku. Wszystkie obliczenia w sieci są wykonywane efektywnie w przestrzeni cech LR, wymagając tylko operacji zwiększania rozmiaru (ang. *upsampling*) na końcowym odcinku architektury sieci. Schemat sieci RAMS jest pokazany na rys. 2.4.



Rysunek 2.4: Schemat sieci RAMS. Moduły i ich skrótowe oznaczenia są opisane w tekście. Rysunek pochodzi z pracy [66].

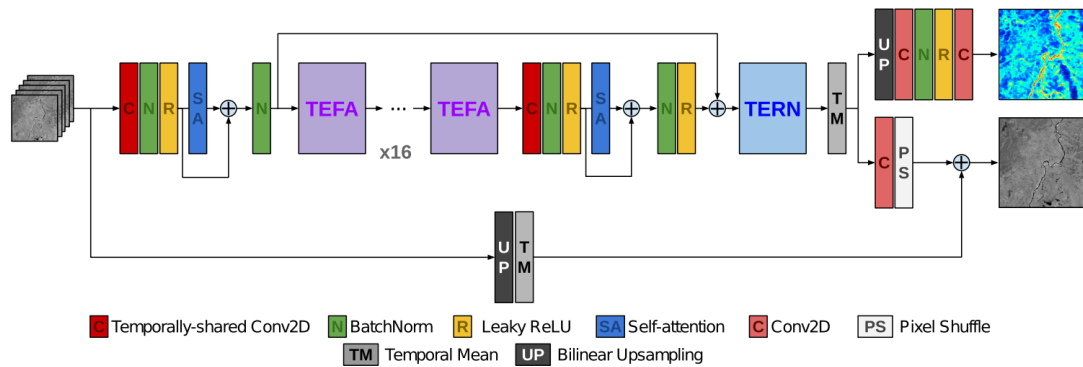
Architektura sieci RAMS składa się z trzech głównych bloków: uwagi dla cech (ang. *residual feature attention*), uwagi dla wymiaru czasowego (ang. *residual temporal attention*) oraz bloku redukującego wymiarowość temporalną (ang. *temporal reduction*) oraz dwóch gałęzi: głównej oraz rezydualnej (ang. *global residual path*). W artykule przedstawiono również proces treningu sieci, gdzie jako funkcji straty używana jest miara L1 obliczana między fragmentami HR oraz rekonstruowanymi – SR.

HighRes-net

W artykule [68] przedstawiono metodę HighRes-net. Architektura sieci składa się z trzech głównych modułów: kodowania, fuzji i dekodowania. W module kodowania, sieć ekstrahuje cechy z obrazów LR, skupiając się na różnicach między różnymi obserwacjami LR. Fuzja odbywa się w kilku etapach, w których liczba stanów niskiej rozdzielczości jest sukcesywnie zmniejszana. Dzięki temu możliwe jest scalenie informacji z różnych obserwacji i uzyskanie spójnej reprezentacji na poziomie wysokiej rozdzielczości. Ostatni blok, czyli dekodowanie, polega na generowaniu rekonstrukcji obrazu o wysokiej rozdzielczości na podstawie danych z modułu fuzji. W treningu sieci użyto danych z konkursu PROBA-V oraz jako funkcji straty

metryki L2.

PIUnet



Rysunek 2.5: Schemat sieci PIUnet. Blok TEFA (ang. *temporally equivariant feature attention*) dokonuje selekcji najbardziej znaczących cech, TERN (ang. *temporally equivariant registration network*) zajmuje się rejestracją obrazów wejściowych. Rysunek pochodzi z pracy [69].

Autorzy [69] zaproponowali rozwiązanie problemu MISR za pomocą sieci neuronowej, nazwanej PIUnet (ang. *Permutation Invariance and Uncertainty*). Wejściem sieci jest zbiór z dowolną liczbą obrazów LR. Kontrastuje to z innymi istniejącymi technikami, które są zaprojektowane dla stałej liczby obrazów. Zakłada się, że obrazy LR są wstępnie dorejestrowane (w dokładności pełnopikselowej) ze sobą w trakcie przetwarzania wstępnego. Kluczową cechą sieci PIUnet jest niezależność od permutacji danych wejściowych LR oraz w szacowaniu zarówno obrazu SR, jak i odpowiadającej mu wartości niepewności rekonstrukcji. Osiąga się to za pomocą dwóch równoległych ścieżek przetwarzania. Schemat sieci jest przedstawiony na rys. 2.5. Sieć generuje dwie informacje wynikowe: obraz zrekonstruowany (na rysunku szary obraz po prawej) oraz mapę niepewności rekonstrukcji (obraz w skali czerwony-niebieski).

Podsumowanie sieci głębokich służących do rekonstrukcji wieloobrazowej oraz sposobu treningu

Wspólnym aspektem wszystkich wspomnianych modeli jest wykorzystanie głębokich sieci neuronowych do ekstrakcji cech wysokiego poziomu z obrazów niskiej rozdzielczości, która to informacja jest używana następnie do generowania obrazów SR. Modele różnią się w wyborze konkretnej funkcji straty (L2 lub L1). Jed-

nak zazwyczaj żadna inna metryka poza wspomnianymi podstawowymi nie jest stosowana w treningu. Zostały opublikowane podejścia do SISR z miarami perceptualnymi ([70], [71]), ale problem MISR jak na razie nie doczekał się tego typu rozwiązań. Zaprezentowane sieci są zarazem stosunkowo złożone i bardzo różnorodne. W chwili rozpoczęcia pisania rozprawy metodą dającą najlepsze wyniki w rankingu opartym na zbiorze PROBA-V [72] była sieć PIUnet, która została wykorzystana do przeprowadzenia eksperymentów opisanych w rozdziale 4.

2.2 Miary używane w ocenie jakości obrazów

Ocena jakości wyniku SR jest zadaniem skomplikowanym, gdyż sposób postrzegania podobieństwa przez ludzki zmysł wzroku do tej pory nie został dokładnie zbadany i opisany. Jest to wciąż aktywny problem badawczy [73], [74]. Różnorodne miary wykorzystywane są w przemyśle filmowym i telewizyjnym, w zakresie kompresji stratnej oraz w metodach rekonstrukcji [75]. Nie istnieje jedna metryka, która miałaby uniwersalne zastosowanie. Istnieją miary, które zostały przyjęte jako standard *de facto* takie jak opisywane dalej proste miary analityczne. Jednak mają one istotne wady, które częściowo niwelują miary bardziej złożone, lub też oparte na bardziej skomplikowanych modelach używających sieci głębokich.

Do oceny jakości można podejść dwójako. Pierwszą możliwością jest zastosowanie metod bezreferencyjnych, czyli oceniających jakość obrazu bez porównywania z obrazem wzorcowym, biorąc pod uwagę wyłącznie właściwości pojedynczego obrazu [76]. Drugą grupą metod jest porównywanie podobieństwa obrazu $\mathcal{I}^{(d)}$ do obrazu wzorcowego – referencyjnego [75] $\mathcal{I}^{(r)}$. W przypadku rekonstrukcji nadrozdzielczej będzie to obraz HR $\mathcal{I}^{(h)}$. W dalszej części rozdziału zostaną opisane metody spotykane w literaturze oraz wskazane zostaną ich typowe wady, ze zwróceniem uwagi na zastosowanie w SR.

2.2.1 Miary analityczne

Istnieje wiele miar opartych na porównaniu obrazów piksel do piksela lub też przy wykorzystaniu metod operujących na otoczeniu lokalnym. Podstawowymi metrykami spotykanymi w literaturze są opisane poniżej:

Średni błąd bezwzględny (ang. *Mean Absolute Error*, MAE) to jedna z najprostszych metryk podobieństwa, opisana jako średnia wartości bezwzględnych różnic między poszczególnymi pikselami porównywanych obrazów [77]:

$$\text{MAE} = \frac{1}{M \cdot N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |\mathcal{I}^{(d)}(i, j) - \mathcal{I}^{(r)}(i, j)|, \quad (2.5)$$

gdzie M oraz N to wymiary obrazów $\mathcal{I}^{(d)}$ i $\mathcal{I}^{(r)}$.

Stosunek sygnału szczytowego do szumu (ang. *peak signal-to-noise ratio*, PSNR), oparty na błędzie średniokwadratowym (ang. *Mean Squared Error*, MSE), obliczany dla wszystkich pikseli obrazów w sposób przedstawiony podanymi dalej wzorami [78]:

$$\text{MSE} = \frac{1}{M \cdot N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [\mathcal{I}^{(d)}(i, j) - \mathcal{I}^{(r)}(i, j)]^2, \quad (2.6)$$

$$\text{PSNR} = 20 \cdot \log_{10} \frac{\mathcal{M}}{\sqrt{\text{MSE}}}, \quad (2.7)$$

gdzie M oraz N to wymiary obrazów $\mathcal{I}^{(d)}$ i $\mathcal{I}^{(r)}$, zaś \mathcal{M} to maksymalna możliwa wartość przyjmowana przez piksele obrazów w stosowanej rozdzielczości bitowej.

PSNR jest mierzony w decybelach (dB), przy czym wyższe wartości PSNR wskazują na większe podobieństwo porównywanych obrazów. Typowo przyjmuje wartości z zakresu 20–50 dB [79] dla obrazów 8-bitowych.. PSNR jest często stosowany do oceny jakości obrazów kompresowanych stratnie [80]. Zaletą stosowania PSNR jako miary jest to, że jest ona łatwa do obliczenia. Jednak PSNR ma istotne ograniczenia, które powodują, że wartość PSNR nie zawsze koreluje z percepcją jakości wizualnej. Spowodowane to jest czułością na przesunięcia (wynikającą z porównywania piksel do piksela) oraz na zmiany jasności, kontrastu lub ogólniej – histogramu obrazu. Dzieje się tak dlatego, że ocena jakości obrazu przez człowieka jest złożona i opiera się na zależnościach bardziej skomplikowanych niż różnice między poszczególnymi punktami obrazu [78].

Miara podobieństwa strukturalnego (ang. *structural similarity index metric*, SSIM) – metryka opracowana w 2004 roku do oceny jakości sygnału wideo w laboratorium LIVE na University of Texas w Austin w Stanach Zjednoczonych [81]. W odróżnieniu do PSNR, ocena jakości jest przeprowadzana na poziomie obszarów (okien) większych niż pojedynczy piksel:

$$\text{SSIM}(i, j) = \frac{(2\mu_i\mu_j + c_1)(2\sigma_{ij} + c_2)}{(\mu_i^2 + \mu_j^2 + c_1)(\sigma_i^2 + \sigma_j^2 + c_2)}, \quad (2.8)$$

gdzie μ_i to średnia, a σ_i^2 wariancja wartości obrazu w oknie i , σ_{ij} zaś to kowariancja między wartościami obrazu w oknach i oraz j . c_1 oraz c_2 to odpowiednio dobrane stałe, które stabilizują numerycznie wynik przy małej wartości mianownika.

SSIM uwzględnia aspekty: luminancji (jasności), kontrastu oraz element strukturalny otoczenia piksela. Jest to miara bezjednostkowa, przyjmująca wartości z zakresu $\langle 0; 1 \rangle$. Wartość 0 oznacza brak podobieństwa między obrazami, 1 z kolei oznacza, że obrazy są identyczne.

Miara wierności informacji wizualnej (ang. *visual information fidelity*, VIF) - kolejna miara opracowana przez zespół tworzący SSIM w roku 2006 [82], [83]. W trakcie jej projektowania brano pod uwagę budowę ludzkiego systemu postrzegania obrazu. Miara oblicza wartości działając na obrazach przetransformowanych do przestrzeni falkowej, wykorzystując strukturę reprezentacji w postaci piramidy (ang. *steerable pyramid*) [84]. VIF opisuje różnice między obrazami w kontekście jasności, koloru i tekstury. Jest to miara bezjednostkowa, gdzie wyższa wartość miary oznacza większe podobieństwo. Przyjmuje wartości z zakresu $\langle 0; 1 \rangle$, gdzie 0 oznacza brak podobieństwa, a 1 – podobieństwo idealne.

Skorygowany PSNR (cPSNR) – miara utworzona na potrzeby konkursu dot. rekonstrukcji nadrozdzielczej obrazów z satelity PROBA-V [26] (ang. *PROBA-V Super Resolution competition*¹¹). Jest oparta na PSNR, ale z częściową minimalizacją wcześniej wspomnianych wad w odniesieniu do obrazów różniących się poziomem jasności. Metryka została zaprojektowana dla zadania rekonstrukcji nadrozdzielczej i dodatkowo uwzględnia realizację następujących wymagań:

- Odpowiadające sobie piksele na obrazie zrekonstruowanym oraz referencyjnym mają wartości zbliżone do siebie, z uwzględnieniem średniego przesunięcia jasności między obrazami (opisane wzorami w dalszej części rozdziału).
- W ocenie powinny być uwzględnione tylko piksele, które w obrazie referencyjnym nie są oznaczone jako błędne lub ignorowane (przykryte przez chmury, uszkodzone piksele).
- Proces oceny powinien uwzględniać możliwe pełnopikselowe przesunięcia między obrazem referencyjnym oraz zrekonstruowanym. Mogą one wynikać z: translacji widoku sceny między obrazami LR oraz HR; przesunięć między obrazami LR, które skutkują przesunięciem widoku obrazu zrekonstruowanego; wyboru obrazu LR, do którego są dorejestrowywane pozostałe obrazy niskiej rozdzielczości; bądź też wreszcie ze sposobu działania rekonstrukcji, która może przesuwać widok obrazu wynikowego względem obrazów wejściowych.

W tym celu wprowadzono obliczenie średniego przesunięcia jasności między obrazami porównywanymi:

$$b = \frac{1}{|clear(\mathcal{I}_{u,v}^{(r)})|} \left(\sum_{x,y \in clear(\mathcal{I}_{u,v}^{(r)})} \mathcal{I}_{u,v}^{(r)}(x,y) - \mathcal{I}^{(d)}(x,y) \right), \quad (2.9)$$

gdzie $\mathcal{I}^{(r)}$ oznacza obraz referencyjny, $\mathcal{I}^{(d)}$ to obraz porównywany (w tym wypadku – zrekonstruowany), $clear(I)$ reprezentuje zbiór pikseli, które są oznaczone

¹¹Strona konkursu dostępna jest pod adresem: <https://kelvins.esa.int/proba-v-super-resolution/home/>

jako poprawne (maska poprawnych pikseli), a u, v oznacza pełnopikselową translację obrazów względem siebie.

Następnie ta wartość jest uwzględniana przy liczeniu skorygowanego błędu średniokwadratowego:

$$\text{cMSE}_{u,v} = \frac{1}{|\text{clear}(\mathcal{I}_{u,v}^{(r)})|} \sum_{x,y \in \text{clear}(\mathcal{I}_{u,v}^{(r)})} (\mathcal{I}_{u,v}^{(r)}(x,y) - (\mathcal{I}^{(d)}(x,y) + b))^2. \quad (2.10)$$

Z kolei obliczana jest skorygowana wartość PSNR dla translacji u, v :

$$\text{cPSNR}_{u,v} = -10 \log_{10}(\text{cMSE}(\mathcal{I}_{u,v}^{(r)}, \mathcal{I}^{(d)})). \quad (2.11)$$

I ostatecznie wynik dla wszystkich możliwych translacji jest definiowany następująco:

$$\text{cPSNR}(\mathcal{I}^{(r)}, \mathcal{I}^{(d)}) = \max_{u,v \in \{-3, \dots, 3\}} (\text{cPSNR}_{u,v}(\mathcal{I}_{u,v}^{(r)}, \mathcal{I}^{(d)})). \quad (2.12)$$

Maksymalny zakres przesunięć w tym wyrażeniu przyjęto dla typowych możliwych przesunięć możliwych w praktyce do wygenerowania za pomocą sieci konwolucyjnej.

Skorygowany SSIM (cSSIM) – analogicznie do cPSNR, możliwe jest również uzyskanie podobnej korekty dla miary SSIM – cSSIM [27], dopasowując obrazy między sobą o przesunięcie jasności b , opisane wzorem 2.9:

$$\text{cSSIM}(\mathcal{I}^{(r)}, \mathcal{I}^{(d)}) = \text{SSIM}(\mathcal{I}^{(r)}, \mathcal{I}^{(d)} + b). \quad (2.13)$$

Inne miary podobieństwa Oprócz wcześniej wspomnianych, najczęściej spotykanych w literaturze, zaprojektowane zostały inne metryki podobieństwa obrazów. Można tutaj wymienić następujące:

HaarPSI (ang. *Haar Perceptual Similarity Index*) [85] - używa transformacji falkowej Haara, co pozwala na analizę szczegółów obrazu na wielu poziomach. Transformacje falkowe są uznawane za reprezentację mającą cechy spotykane w ludzkim układzie postrzegania obrazów ([86]–[89]), co zapewnia pewien poziom perceptualności odzwierciedlony w nazwie metody. Miara może działać na obrazach w skali szarości oraz kolorowych, operując w przestrzeni barw YIQ. Autorzy deklarują, że miara jest bardziej odporna od tradycyjnych metryk na zmiany w jasności i kontraście między obrazami $\mathcal{I}^{(r)}$ oraz $i\mathcal{I}^{(d)}$.

GMSD (ang. *Gradient Magnitude Similarity Deviation*) [90], która jest metryką podobieństwa obrazów opartą na gradiencie obrazu. Jej działanie jest opisane następująco: są obliczone magnitudy gradientu (uzyskane za pomocą filtru Prewitta [91]) obrazów referencyjnego $\mathcal{I}^{(r)}$ oraz porównywanego $\mathcal{I}^{(d)}$ w punkcie i , oznaczone za pomocą $m^r(i)$ oraz $m^d(i)$. Definiowana jest mapa podobieństwa

magnitud gradientów (ang. *gradient magnitude similarity*, GMS):

$$\text{GMS}(i) = \frac{2m^d(i)m^r(i) + c}{(m^d(i))^2 + (m^r(i))^2 + c}, \quad (2.14)$$

gdzie c to odpowiednio dobrana stała stabilizująca wynik numerycznie. Obliczana jest następnie średnia wartość GMS – GMSM:

$$\text{GMSM} = \frac{1}{N} \sum_{i=1}^N \text{GMS}(i), \quad (2.15)$$

gdzie N to liczba pikseli w obrazach. Końcowym wynikiem jest wartość

$$\text{GMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{GMS}(i) - \text{GMSM})^2}. \quad (2.16)$$

Wartość ta reprezentuje podobieństwo między dwoma obrazami. Według autorów metoda ta jest wrażliwa na szczegóły, pozostając jednocześnie mało czuła na zmiany w histogramie obrazu.

2.2.2 Miary bezreferencyjne

Oprócz wspomnianych wcześniej metryk podobieństwa, które uwzględniają różnice między dwoma obrazami, możliwe jest również użycie jako miar wartości generowanych przez podejścia bezreferencyjne, obliczane dla jednego tylko obrazu. Może mieć w szczególności sens obliczenie takich metryk dla dwóch wersji tego samego obrazu, gdzie porównanie wartości takich metryk można czasami zinterpretować jako ocenę jakości obrazu. Do tego typu miar należą:

Entropia - opiera się na rozkładzie prawdopodobieństwa intensywności pikseli w obrazie. Zakładając, że obraz jest w skali szarości, entropię można obliczyć za pomocą następującego wzoru:

$$H = - \sum_i p(i) \log_2(p(i)), \quad (2.17)$$

gdzie: H reprezentuje wartość entropii, $p(i)$ zaś to prawdopodobieństwo wystąpienia określonej intensywności piksela. Aby obliczyć entropię, należy obliczyć rozkład prawdopodobieństwa intensywności pikseli, zliczając częstotliwość występowania każdej wartości intensywności w obrazie.

Miara EME (ang. *measure of enhancement*), wprowadzona w [92] jest narzędziem oceniającym jakość obrazu. Obliczana jest w oknach $w_{k,l}$, na które jest dzielony wejściowy obraz X :

$$\text{EME} = \max_{\Phi \in \{\Phi\}} \left(\frac{1}{k_1 k_2} \sum_{l=1}^{k_2} \sum_{k=1}^{k_1} 20 \log \frac{I_{max;k;l}^w}{I_{min;k;l}^w} \right), \quad (2.18)$$

gdzie $I_{max;k;l}^w$ oraz $I_{min;k;l}^w$ są odpowiednio minimum i maksimum wartości w obrazie X wewnątrz bloku $w_{k,l}$

2.2.3 Miary oparte na uczeniu maszynowym

Procesy zachodzące w ludzkim postrzeganiu obrazu, które pozwalają na ocenę podobieństwa obiektów i obrazów są znacznie bardziej złożone niż przedstawione wcześniej metody mogą uwzględnić [93]. Miary podobieństwa starają się zapewnić zbieżność z odczuciami ludzi, jednak ze względu na trudność w ustaleniu co tak naprawdę gra rolę w postrzeganiu różnic przez człowieka, zdefiniowanie miary perceptualnie zadowalającej jest trudne [81]. Jest to ciągle aktywny problem badawczy.

Miary definiowane analitycznie mają wiele dobrze znanych wad. Przykładowo dla miar L2 takich jak MSE czy PSNR, znany jest fakt, że dla operacji rozmycia zmiana wartości miary jest mała, w przeciwieństwie do tego co odczuwa człowiek – małe zmiany obrazu polegające nawet na niewielkim rozmyciu są bardzo łatwo zauważalne [93].

Metryki takie jak SSIM czy VIF zostały zaprojektowane w ten sposób, by lepiej odzwierciedlać ludzkie postrzeganie różnic między obrazami. Jednak największy postęp nastąpił wraz z nadejściem możliwości trenowania głębokich sieci neuronowych. Podjęte zostały próby [93]–[95] zastosowania tej technologii do celów porównywania obrazów oraz oceny ich jakości. Ocena jakości przez człowieka jest wielopoziomowa i zależy tak od drobnych szczegółów obrazu, jak i od cech takich jak tekstury, kształty i dostrzegalne obiekty [93]. Wszystkie tego typu cechy mogą wyodrębnić i uwzględniać sieci głębokie, które można wytrenować w ten sposób by naśladowały ludzką percepcję.

LPIPS

Jednym z modeli najczęściej używanych w porównywaniu obrazów jest LPIPS (ang. *Learned Perceptual Image Patch Similarity*). Jest to metryka podobieństwa obrazów opracowana w celu oceny perceptualnego podobieństwa między dwoma obrazami. W porównaniu do tradycyjnych metryk analitycznych, LPIPS ma na celu lepsze odzwierciedlenie tego, jak ludzie postrzegają różnice między obrazami. Jej skuteczność powoduje, że wydaje się odpowiedni dla zadań takich jak synteza, kompresja stratna i rekonstrukcja obrazów [96], [97].

LPIPS został zaprojektowany przez naukowców z University of California w Berkeley w Stanach Zjednoczonych i Adobe Research [93]. Model opiera się na sieci konwolucyjnej operującej na wyniku innej, wstępnie wytrenowanej sieci, która służy do ekstrakcji cech obrazu.

Sposób działania LPIPS jest następujący:

- Przetwarzanie wstępne – wejściowe obrazy są przeskalowywane i normalizowane zgodnie z warunkami wymaganymi przez opisywaną dalej sieć ekstrahującą cechy.
- Ekstrakcja cech – wytrenowana osobno sieć, taka jak VGG [98], AlexNet [99] czy SqueezeNet [100], służy do ekstrakcji cech z obu obrazów wejściowych. Przejście w przestrzeń tego typu cech zapewnia interpretację informacji z obrazu na poziomie wyższym niż porównywanie pikseli, czy też ich małego otoczenia.
- Obliczanie odległości - wynik podobieństwa LPIPS obliczany jest jako odległość między głębokimi cechami dwóch obrazów wejściowych. Mniejsza odległość odpowiada wyższemu podobieństwu percepcyjnemu.

W celu przeprowadzenia treningu sieci LPIPS został opracowany przez autorów pracy [93] zbiór obrazów oraz wyników uzyskanych z ankietowania w postaci dwóch rodzajów zapytań:

- Wybór jednej z dwóch alternatyw (ang. *two-alternatives forced choice*, 2AFC), w której dla fragmentu obrazu \mathcal{I} jest on degradowany dwoma procesami, w wyniku których uzyskujemy obrazy \mathcal{I}_1 oraz \mathcal{I}_2 . Następnie ankietowanemu przedstawiane są wszystkie trzy wspomniane obrazy wraz z pytaniem, który z obrazów \mathcal{I}_1 , \mathcal{I}_2 jest bliższy \mathcal{I} [101].
- Tylko zauważalne różnice (ang. *just noticeable differences*, JND). Potencjalną wadą zadania 2AFC jest to, że uczestnicy mogą świadomie wybrać, na których aspektach podobieństwa skupią się podczas wykonywania zadania, co może wprowadzać niepożądane tendencje do ocen. Aby uniknąć tego typu efektów, w tym rodzaju ankiety osobie badanej pokazywane są: obraz referencyjny \mathcal{I} oraz jeden z obrazów degradowanych \mathcal{I}_1 . Pytanie zadawane ankietowanemu brzmi: czy przedstawione obrazy są takie same czy różne [102].

Zbiór stosowanych przy wytwarzaniu danych treningowych LPIPS metod degradacji obrazów jest różnorodny, co według autorów powinno zapewnić wszechstronność metody:

- Zmiany w fotometrii: przesunięcia jasności, przesunięcia w kolorach, zmiany kontrastu i nasycenia.
- Szum: biały szum, biały szum Gaussa, szum różowy i niebieski, szum Gaussa w zakresie kolorów [103], [104], artefakty szachownicy.
- Rozmycie: splot z funkcją Gaussa, filtrowanie bilateralne.

- Zmiany przestrzenne: przesunięcie, niewielkie przekształcenie afiniczne, przekształcenie homograficzne, aberracja chromatyczna.
- Artefakty kompresji metodą JPEG (ang. *Joint Photographic Experts Group*) [105].
- Degradacja za pomocą sieci głębokich (autoenkodery [106]–[108]) wytrenowanych z danymi wejściowymi degradowanymi w sposób opisany wcześniej oraz z architekturą i hiperparametrami zmodyfikowanymi względem oryginalnych wartości.

Zbiór danych zawiera również wiele elementów pochodzących z metod przetwarzania obrazów takich jak: rekonstrukcja nadrozdzielcza (dane z konkursu dotyczącego rekonstrukcji jednoobrazowej – ang. *New Trends in Image Restoration and Enhancement*, NTIRE z 2017 roku [109]), interpolacja, usuwanie rozmycia oraz koloryzacja. Obrazy oryginalne pochodzą ze zbiorów używanych w literaturze do badań nad oceną jakości obrazów: LIVE (zob. rozdział 2.3.4) [110], CSIQ (Categorical Image Quality) [111], TID2008 (Tampere Image Database) [112], TID2013 [113].

Tak otrzymane wyniki są użyte w treningu sieci. Ponieważ dane wejściowe oparte są wyłącznie na wartościach uzyskanych od osób ankietowanych, a zbiór wartości jest różnorodny i poddany różnorodnym degradacjom, autorzy wysuwają tezę, że wartości generowane przez model dobrze naśladują postrzeganie podobieństwa przez ludzki zmysł wzroku.

NIMA

NIMA (ang. *Neural Image Assessment*) [94] to metoda oceny jakości obrazu oparta na sieciach neuronowych, opracowana w 2018 roku. NIMA według autorów pozwala w ocenie podobieństwa uwzględniać zarówno aspekty estetyki, jak i technicznej jakości. Wykorzystuje cechy generowane przez głębokie sieci konwolucyjne w celu nauczenia się cech istotnych dla ludzkiego postrzegania jakości obrazu.

NIMA jest wytrenowane na zbiorach danych zawierających oceny jakości obrazów dokonywanymi przez ludzi: TID2013 [113] oraz AVA (ang. *Aesthetic Visual Analysis*) [114]. W stosunku do LPIPS różni się podejściem do generowania wartości jakości obrazu oraz sposobem treningu. Bazowa sieć (VGG [98], MobileNet, [115] Inception [116]) jest pozbawiona ostatnich warstw dokonujących klasyfikacji tak, by pozostały tylko elementy generujące cechy na wielu poziomach. Następnie dołączona jest warstwa w pełni połączona generująca 10 klas wyników jakości (0-10) zakończona modułem SoftMax [106], który normalizuje wynik. Architektura NIMA jest prosta i skuteczna, jednak jest to miara bezreferencyjna, oceniająca jakość pojedynczego obrazu, a nie klasyczna miara podobieństwa.

PieAPP

Podobnie jak w LPIPS, w pracy [95] autorzy definiują głęboką sieć konwolucyjną, która ekstrahuje cechy obrazu i operując w takiej przestrzeni obliczają podobieństwo między obrazami. Również tutaj kluczowe jest utworzenie dużego zbioru obrazów wraz z ocenami zebranymi od osób ankietowanych. Wykorzystana jest metoda zbierania ocen 2AFC. Podobnie jak LPIPS metryka ta generuje wartości reprezentujące odległość między obrazami.

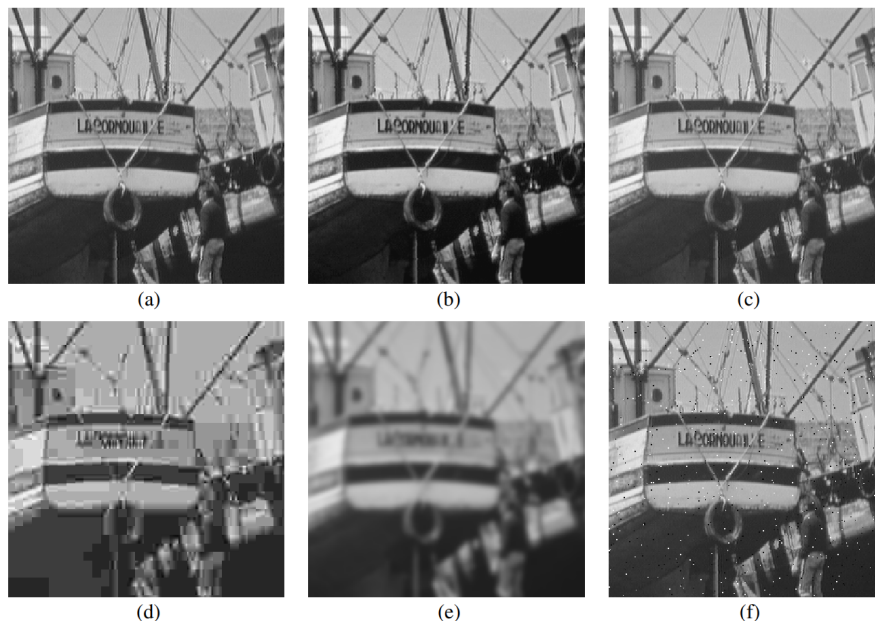
2.2.4 Wady znanych metryk analitycznych

W badaniach nad rekonstrukcją nadrozdzielczą najczęściej stosowanymi miarami oceny są wciąż miary L2 takie jak MSE lub oparte na nim PSNR oraz cPSNR. PSNR pozostaje wciąż najczęściej stosowaną metryką ([117]). Są to miary proste i stąd bardzo szybkie do obliczenia. Były i są stosowane w wielu pracach dotyczących przetwarzania obrazów, co pozwala porównywać wyniki nowych algorytmów z opracowanymi wcześniej. Szczególnie istotną zaletą w kontekście uczenia maszynowego jest różniczkowalność, która umożliwia bezpośrednie użycie tych miar jako funkcji straty w popularnych środowiskach obsługujących sztuczne sieci neuronowe takie jak PyTorch, TensorFlow lub Keras. Wady takich prostych metryk są opisywane od wielu lat ([81]). Wizualny przykład nieoczekiwanego zachowania jest pokazany na rys. 2.6.

Wykazano, że najczęściej używane miary klasy L2 nie są dobrym wyborem dla zadań polegających na rekonstrukcji obrazu [119]. Słabo korelują one z odczuciami ludzkiego zmysłu wzroku. Użycie metryki L2 niejawnie zakłada, że wpływ szumu jest niezależny od lokalnej charakterystyki obrazu, natomiast ludzki wzrok jest czuły na szumy w zależności od lokalnej charakterystyki luminancji, kontrastu i struktury obrazu [81]. Większość tego typu problemów miała zostać usunięta w nowszych miarach takich jak SSIM. Jednak jak się okazało nie jest trudno pokazać przykłady, gdy nawet ta metryka zwraca wyniki sprzeczne z oczekiwaniami ludzkiego obserwatora (rys. 2.7). W pracy [119] wskazano, że w zadaniach: rekonstrukcji nadrozdzielczej, usuwaniu szumu oraz artefaktów blokowych kompresji JPEG [105], dobrym wyborem okazuje się miara podobieństwa w postaci metryki L1.

2.3 Zbiory danych

W niniejszym rozdziale opisano wybrane zbiory danych: tak zawierające dane pozwalające na realizację zadania wieloobrazowej rekonstrukcji nadrozdzielczej, jak i zestaw wykorzystywany przy konstruowaniu metryk podobieństwa obrazów.

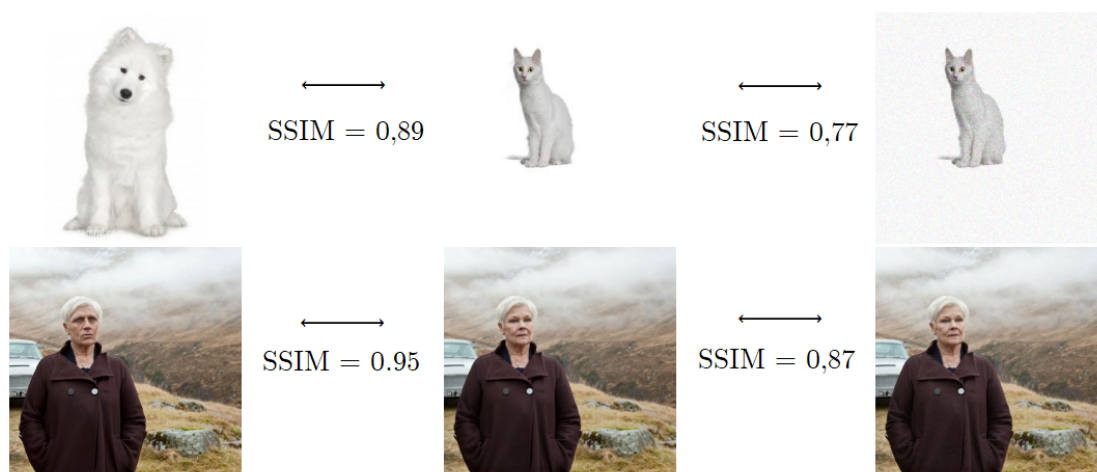


Rysunek 2.6: Nieoczekiwane zachowanie miary L2 w porównywaniu obrazów. a) obraz oryginalny, b) zmiana kontrastu, c) zmiana jasności, d) kompresja JPEG [105], e) rozmycie, f) szum typu sól i pieprz. Dla każdej pary (a, b), (a, c) .. (a, f) MSE wynosi 210. Rysunek pochodzi z pracy [81].

Część z wspomnianych zbiorów danych została użyta podczas walidacji eksperymentalnej.

2.3.1 Zbiór PROBA-V

Zbiór danych PROBA-V jest zbiorem danych satelitarnych, które są pozyskiwane przez satelitę PROBA-V. Satelita ten został wystrzelony w maju 2013 roku przez Europejską Agencję Kosmiczną (ESA) z zamiarem monitorowania powierzchni Ziemi i zbierania danych dotyczących roślinności, poziomu wody oraz temperatury powierzchniowej. Dzięki tym danym naukowcy i badacze mogą badać zmiany klimatyczne oraz środowisko naturalne, a także przeprowadzać badania nad roślinnością i glebą. W kontekście rekonstrukcji nadrozdzielczej PROBA-V zostało użyte w konkursie ogłoszonym przez ESA - utworzono wówczas wspomniany zbiór danych [26]. Dane pochodzą z 74 wyselekcjonowanych ręcznie regionów na całym świecie. W obrazach dokonano korekcji radiometrycznych i geometrycznych. Same dane pochodzą z pasma czerwonego i bliskiej podczerwieni i są zarejestrowane w rozdzielczości 300 m (LR) i 100 m (HR), w odwzorowaniu kartograficznym Plate Carrée. Dane LR są dostarczane jako obrazy o rozmiarze 128×128 pikseli



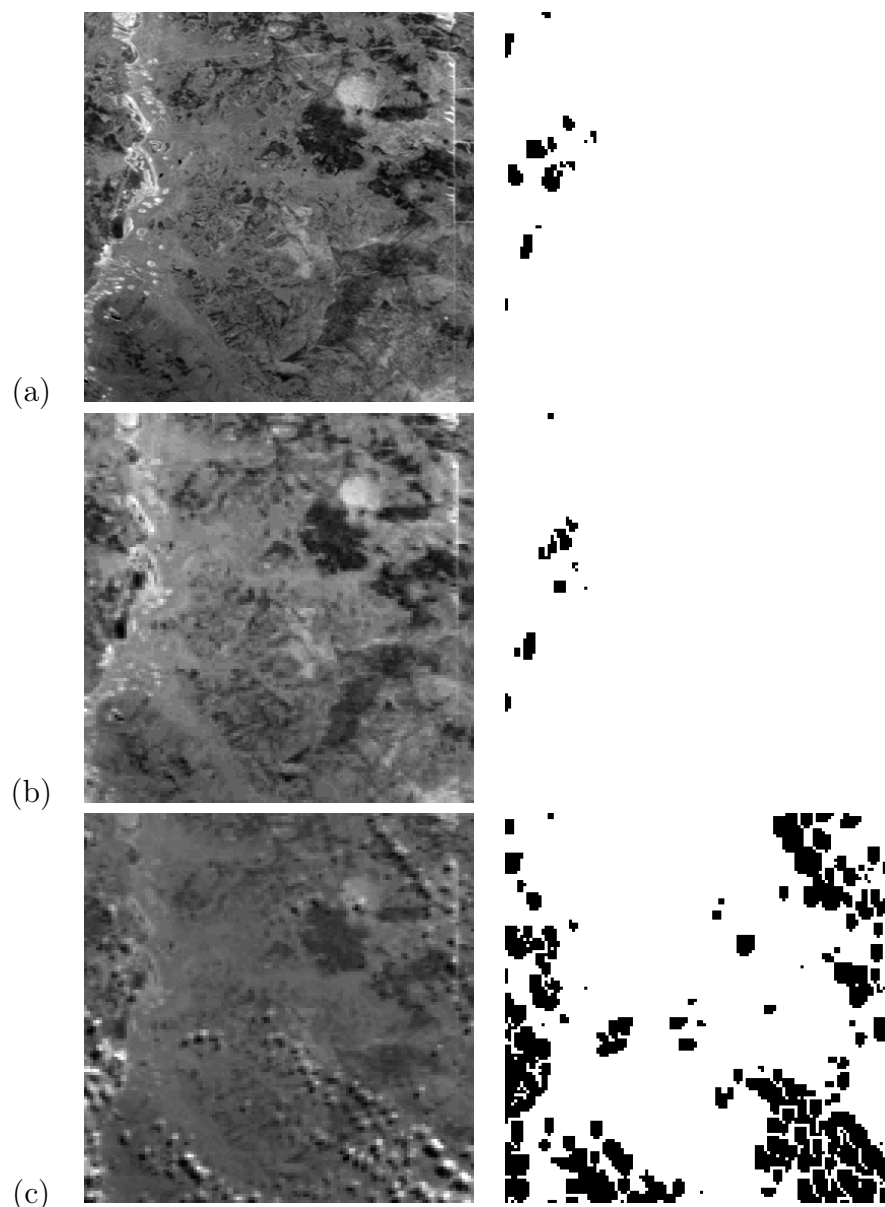
Rysunek 2.7: Nieoczekiwane działanie miary SSIM w porównywaniu obrazów. Po lewej – podmiana obrazu lub istotnego fragmentu (twarzy), w środku – oryginał, po prawej – dodanie jednorodnego szumu. Dla rysunków zawierających te same lub bardziej zbliżone obiekty oczekiwane byłyby wartości powinny być większe. U góry widać również wpływ perceptualnie "pustych" fragmentów obrazu na zawyżenie wartości miary. Rysunek na podstawie pracy [118]

w skali szarości, a dane HR – 384×384 . Głębia bitowa obrazów wynosi 14. Każdy obraz posiada maskę jakości, wskazującą piksele, które są ukryte (np. chmury, cienie chmur, uszkodzenia). Minimalna wartość pikseli zobrazowanych jako jasne, potrzebna do uwzględnienia danego obrazu w zestawie danych, wynosi 75% dla obrazów o rozdzielczości HR i 60% dla obrazów o rozdzielczości LR. Dla każdej sceny istnieje jeden obraz o rozdzielczości 100 m i przeciętnie 19 (co najmniej 9) obrazów o rozdzielczości 300 m z tego samego miejsca. W sumie zestaw danych zawiera 1450 scen: 1160 w zbiorze treningowym oraz 290 scen testowych, dla których nie są ujawnione dane HR. Przykładowe obrazy z pojedynczej sceny ze zbioru treningowego są pokazane na rys. 2.8.

Dane zostały zarejestrowane w ciągu 30 dni, co oznacza niewielkie różnice w obrazach, głównie ze względu na zmiany w vegetacji. Obrazy dostarczone przez PROBA-V zostały zarejestrowane z dokładnością pikselową, pozostawiając różnice subpikselowe, co czyni je przydatnymi w zadaniu MISR.

2.3.2 Zbiór danych symulowanych z Sentinel-2

W pracy [120] pokazano, że niewystarczające dane treningowe są powszechnym, ale mało badanym problemem w dziedzinie SR. Wkładem tej pracy jest zapropo-



Rysunek 2.8: Przykładowe obrazy ze zbioru danych PROBA-V: pierwsza kolumna prezentuje obrazy, druga - maski poprawnych pikseli (czarne piksele na masce to miejsca, które zostały zidentyfikowane jako niepoprawne: chmury, cienie chmur, uszkodzenia matrycy itp.). a) - obraz HR, b-c) - przykładowe obrazy LR przeskalowane do rozmiaru HR.

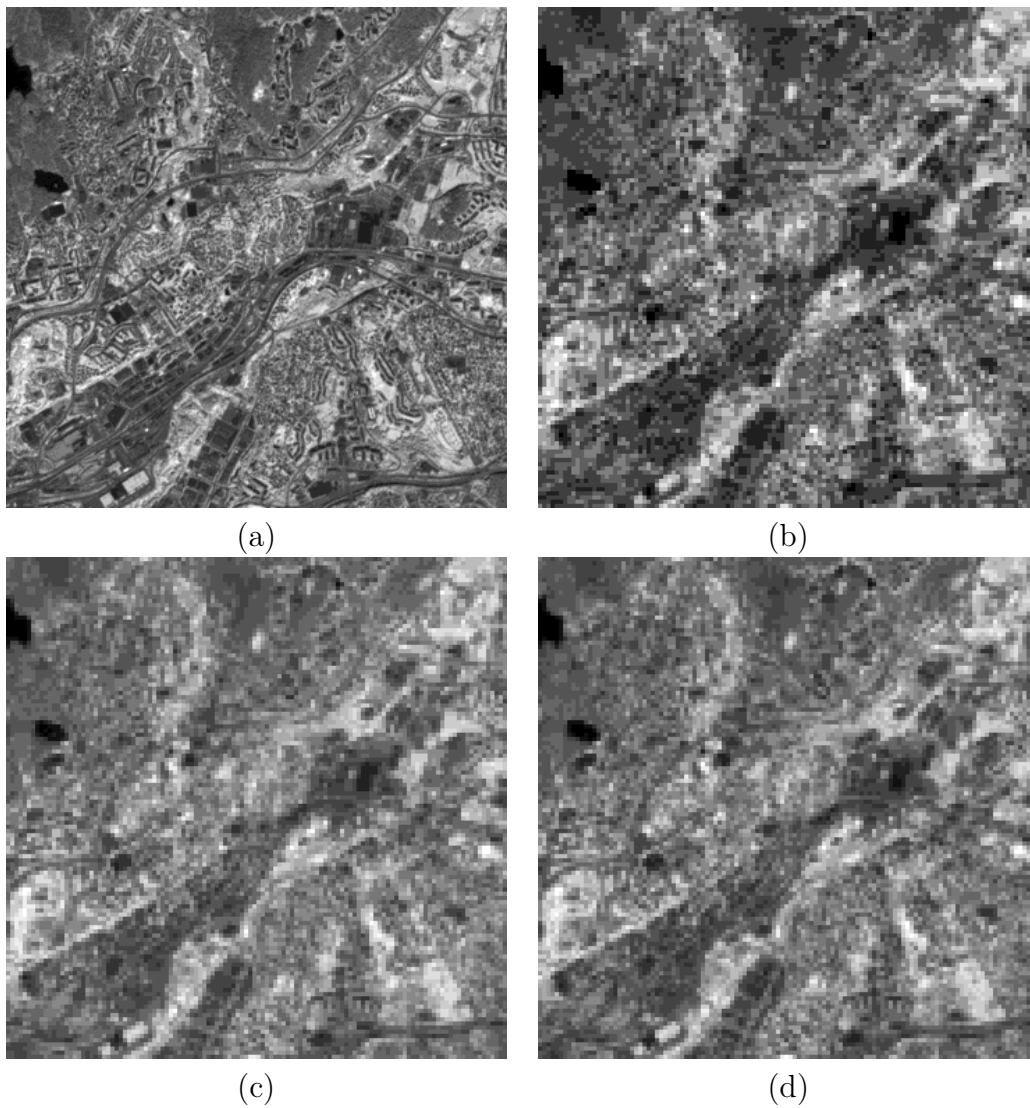
nowanie nowej metody generowania częściowo symulowanych obrazów LR, które przypominają autentyczne obserwacje. Aby stworzyć pojedynczy przykład, wyma-

gane jest zebranie wielu obserwacji jednej sceny, a następnie ich pełnopikselowa rejestracja. Rozdzielczość obrazów jest następnie zmniejszana, aby dopasować się do rozdzielczości LR, pozostawiając jeden obraz bez zmian, aby służył jako referencyjne dane HR. W przeciwieństwie do zbiorów danych rzeczywistych, nie wymaga się w tym wypadku rzeczywistych obrazów referencyjnych HR, co zmniejsza koszt przygotowania danych treningowych. We wspomnianych badaniach przystosowano sieć HighRes-net [68] do rekonstrukcji obrazów Sentinel-2 generowanych z wykorzystaniem pełnego procesu degradacji, jak i techniki częściowej symulacji danych. Tak wytrenowane modele zostały ocenione na danych rzeczywistych (PROBA-V). Wyniki przeprowadzonych eksperymentów wskazują, że modele szkolone na tak pozyskanych danych dostarczają obrazów wyższej jakości z mniejszą liczbą wizualnych artefaktów niż modele trenowane tylko na danych degradowanych w sposób klasyczny.

Satelity Sentinel-2 posiadają instrumenty optyczne zdolne do rejestrowania jednego obrazu w 13 różnych pasmach spektralnych i trzech różnych rozdzielczościach przestrzennych. Aby uprościć eksperyment, wykorzystano tylko dane o GSD (gęstość próbkowania, ang. *Ground-Sampling Distance*) 10 m odpowiadające pasmu 8 (bliska podczerwień, ang. *near infra-red*, NIR). Dla każdego obrazu zostało zebranych minimum dziesięć obserwacji minimalizując różnice w czasie ich rejestrowania i maksymalizując ilości określonych (bez chmur i innych zaburzeń) pikseli. Dane te, początkowo o wymiarach 10980×10980 pikseli, zostały podzielone na fragmenty o wymiarach 288×288 pikseli.

W końcowym zbiorze danych [120] wykorzystano dane przekształcone do poziomu L2A¹², czyli po korekcji geometrycznej oraz usunięciu wpływu atmosfery. Zastosowano dwie różne techniki generowania wielu obrazów LR. Do każdego obrazu HR zastosowano wiele globalnych operacji, aby zasymulować zmienność informacji, która naturalnie występuje między rzeczywistymi obserwacjami tej samej sceny. Te operacje obejmują zmiany jasności i kontrastu, dodawanie losowego szumu, stosowanie filtrów rozmycia Gaussa oraz translację o przesunięcia subpikselowe pobrane z rozkładu jednorodnego, podobnie jak w pracy [24]. Obrazy HR zostały przeskalowane w dół do rozmiaru LR (stosunek rozmiaru HR do LR wynosi 3:1) z użyciem interpolacji dwusześcienniej, aby usunąć informacje wysokiej częstotliwości i dopasować się do wymaganej rozdzielczości przestrzennej obrazów LR - ten krok jest wykonywany tuż po zastosowaniu losowych translacji i jest poprzedzony wcześniej wymienionymi operacjami. Korzystając z tej techniki, stworzono zbiór danych w pełni symulowany, składający się z 5000, 1000 i 1000 przykładów dla podzbiorów, odpowiednio: treningowego, walidacyjnego i testowego. Każdy zestaw składa się z dziewięciu symulowanych obrazów LR i odpowiadają-

¹²Poziomy obróbki obrazów opisane są na stronie internetowej misji ESA Sentinel-2: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/processing-levels>

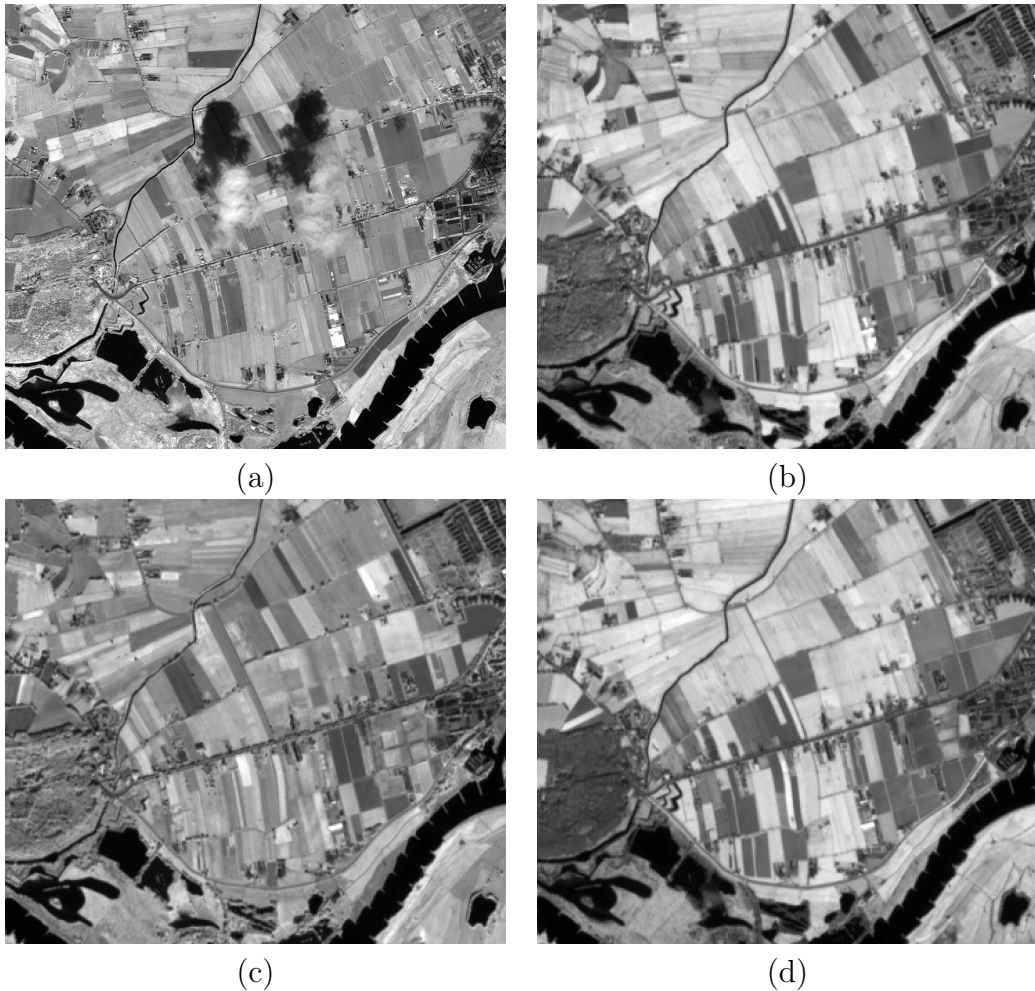


Rysunek 2.9: Przykładowe obrazy ze zbioru danych symulowanych Sentinel-2: a) - obraz HR, b-d) - przykładowe obrazy LR przeskalowane do rozmiaru HR. Widoczne są większe niż w danych PROBA-V zmiany kolorystyczne obrazów.

cego im obrazu referencyjnego HR. Przykłady obrazów ze wspomnianego zbioru są na rys. 2.9.

2.3.3 Zbiór MuS2

Problem niewystarczającej rozdzielczości przestrzennej w teledetekcji – obrazowaniu satelitarnym został podjęty w pracy [27]. Jedną z przeszkód w MISR jest brak rzeczywistych danych odniesienia, ponieważ dane symulowane mogą nie odzwierciedlać w pełni warunków operacyjnych. Również dane pochodzące z jednego instrumentu takiego jak PROBA-V mają jednorodną charakterystykę kolorymetryczną [27].



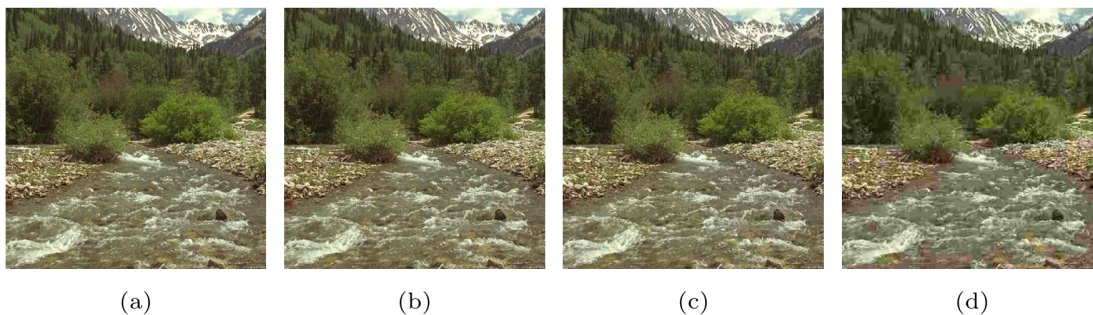
Rysunek 2.10: Przykładowe obrazy ze zbioru danych MuS2: a) - obraz HR, b-d) - przykładowe obrazy LR przeskalowane do rozmiaru HR. Widoczne są większe niż w danych PROBA-V zmiany kolorystyczne obrazów oraz znaczne różnice lokalnego wyglądu obszarów zarejestrowanych przez satelitę, nieobecne w poprzednio wymienionych zbiorach.

Aby rozwiązać ten problem, autorzy przedstawiają nowy benchmark o nazwie Multi-image Sentinel-2 Super-Resolution (MuS2), który ma na celu zapewnienie danych do testowania i porównywania różnych metod MISR. Zbiór danych składa się z 91 zróżnicowanych scen obejmujących około 2500 km², z których każda składa się z co najmniej 14 obrazów wielospektralnych Sentinel-2 i jednego obrazu referencyjnego WorldView-2 o wysokiej rozdzielczości. Obrazy LR są zarejestrowane w rozdzielczości 16-bitowej, natomiast HR - w 8 bitach.

Zaproponowano również protokół oceny wyników SR dla wszystkich pasm Sentinel-2 10 m na podstawie odpowiadających im pasm WorldView-2. Wykazano, że metryka LPIPS (ang. *learned perceptual image patch similarity*) jest odpowiednia do oceny rezultatu SR, nawet jeśli referencyjne obrazy wysokiej rozdzielczości zostały pozyskane z innego satelity. Aby to zweryfikować, został przeprowadzony sondaż opinii. Wreszcie, wprowadzono w MuS2 maski istotności, które wskazują regiony obrazu odpowiednie do oceny na podstawie metryki podobieństwa obrazów. Przykłady obrazów ze zbioru MuS2 pokazane są na rys. 2.10.

2.3.4 Zbiór LIVE

Ocena jakości powinna być zgodna z subiektywnymi odczuciami ludzkich obserwatorów. Aby skalibrować algorytmy i przetestować ich wydajność, wymagany jest zestaw obrazów których jakość została oceniona przez ludzi. W LIVE (Laboratory for Image & Video Engineering, University of Texas) oraz na wydziale psychologii tego samego uniwersytetu przeprowadzono szeroko zakrojony eksperyment w celu uzyskania wyników od osób badanych dla wielu obrazów zdegradowanych różnymi typami zniekształceń. Tak zebrane dane zostały opublikowane w zbiorze LIVE [110], [121].



Rysunek 2.11: Przykłady obrazów ze zbioru LIVE. a) obraz oryginalny, b)-d) różne wersje obrazu zdegradowanego

Zbiór ten zawiera obrazy zdegradowane w kontrolowany sposób, przy użyciu

różnych metod takich jak m.in. kompresja JPEG (169 obrazów), JPEG200 (175 obrazów) rozmycie Gaussa (145 obrazów), biały szum (145) oraz JPEG200 z błędami (145 obrazów). Badania nad oceną jakości w dużym stopniu zależą od subiektywnych danych, które mogą posłużyć do kalibracji metryk, a także mechanizmu testowania. Wymagany jest zatem zestaw danych obrazów, których jakość została oceniona przez ludzi. Algorytm oceny jakości obrazu może być trenowany na takim zestawie danych. Dlatego w omawianym zbiorze dla każdego zdegradowanego obrazu jest dostępna ocena dokonana przez ludzkich obserwatorów, agregowana do postaci różnicowego średniego wyniku opinii (ang. *differential mean opinion score*, DMOS). Przykłady ze zbioru danych LIVE pokazano na rys. 2.11.

2.3.5 Zbiór B4MultiSR

W ramach prac w projekcie zleconym przez Europejską Agencję Kosmiczną – Satellite Image SPAtial Resolution Enhancement (SISPARE), w którym autor rozprawy brał udział, został opracowany zbiór danych do rekonstrukcji nadrozdzielczej z wielu obrazów [122]. Zbiór ten zawiera:

1. Obrazy syntetyczne o różnej rozdzielczości przestrzennej. Kategoria ta obejmuje obrazy przedstawiające ten sam wzór, wygenerowane w różnych rozdzielczościach, a także obrazy wygenerowane w wysokiej rozdzielczości, sztucznie przeskalowane w dół i zdegradowane do mniejszych rozdzielczości.
2. Rzeczywiste obrazy zdegradowane przy użyciu różnych modeli obrazowania dla każdego obrazu o wysokiej rozdzielczości – obrazy o niskiej rozdzielczości są sztucznie uzyskiwane. Pozwala to na weryfikację teoretycznych założeń algorytmów SR. Dane należące do tej kategorii są zwykle wykorzystywane do oceny algorytmów SR opisanych w literaturze.
3. Rzeczywiste obrazy przedstawiające tę samą scenę (lub region Ziemi w przypadku obrazów satelitarnych) wykonane w różnych rozdzielczościach natywnych. Może istnieć pojedynczy lub wiele obrazów o wysokiej rozdzielczości w połączeniu z wieloma obrazami o niższej rozdzielczości. W zestawie danych zestawiono ze sobą obrazy pochodzące z różnych satelitów. Dane te pozwalają na walidację algorytmów SR w rzeczywistych scenariuszach. Ważną kwestią jest tutaj właściwa rejestracja obrazów o niskiej i wysokiej rozdzielczości.

2.3.6 Zbiór WorldStrat

Jednym z ostatnio opublikowanych zbiorów jest *World Stratified* [123]¹³, skrótowo nazywany WorldStrat. Według autorów jest to obecnie największy i najbardziej zróżnicowany publicznie dostępny zbiór danych zawierający obrazy LR oraz HR tych samych obszarów. Obrazy wysokiej rozdzielczości pochodzą z satelitów Airbus SPOT 6/7 o GSD 1,5 m – ten zbiór został stworzony w ramach projektu "Query-Planet", finansowanego przez ESA w ramach programu Phi-Lab. Każdy obraz HR jest zestawiony z wieloma obrazami LR pochodzącymi z satelitów Sentinel-2 o GSD 10 m. W zbiorze znajdują się obrazy ok. 10000 km² powierzchni Ziemi, przedstawiające różne rodzaje użytkowania ziemi na całym świecie: od rolnictwa po czapy lodowe, od lasów po różne rodzaje przestrzeni zurbanizowanej. Dane zawierają również lokalizacje takie jak miejsca klęsk humanitarnych czy nielegalne miejsca wydobycia surowców oraz osiedla. Wraz ze zbiorem danych udostępniono pakiet Python¹⁴, który pozwala na użycie lub rozszerzenie zbioru danych.

¹³<https://zenodo.org/record/6810792>

¹⁴<https://github.com/worldstrat/worldstrat>

Rozdział 3

Opracowane metody

Rozdział zawiera opis wypracowanych w ramach badań podejść do rozwoju metod rekonstrukcji nadrozdzielczej oraz sposobów porównywania obrazów. Zwrócono uwagę na zastosowanie metod opartych na głębokim uczeniu w MISR (rozdział 3.1)¹. W ramach sekcji dotyczącej miar podobieństwa pokazane są możliwości użycia detektorów i deskryptorów punktów charakterystycznych (rozdział 3.2) oraz wykorzystania architektury głębokich sieci konwolucyjnych (rozdział 3.2.2). Następnie omówiono jak te miary mogą być użyte w rekonstrukcji. W ostatniej sekcji został podjęty temat zastosowania miar podobieństwa opartych na uczeniu głębokim w treningu sieci MISR (rozdział 3.3).

3.1 Połączenie klasycznych metod rekonstrukcji nadrozdzielczej oraz sieci głębokich

3.1.1 Ewolucyjna optymalizacja hiperparametrów metod wieloobrazowych – EvoIM

W rozdziale 2.1 opisana została metoda FRSSR [20], która zakłada z góry postać funkcji rozmycia punktu (ang. *point spread function*, PSF), zdefiniowaną jako rozmycie dwuwymiarową funkcją Gaussa, której odchylenie standardowe jest hiperparametrem metody. Oprócz tego w algorytmie istnieje szereg innych parametrów, opisanych w dalszej części rozdziału, będących w istocie hiperparametrami metody optymalizacji. Artykuł [20] podaje przykładowe wartości tych parametrów dla kilku konkretnych przykładów obrazów, nie podając sposobu ich uzyskania, zaszła zatem potrzeba by wypracować metodę pozwalającą na automatyczne do-

¹Ta część, mimo że nie jest obecnie najnowszym stanem wiedzy została tutaj zamieszczona ze względu na fakt, że było to jedno z pierwszych podejść do MISR z użyciem sieci głębokich.

branie tych parametrów.

Model obrazowania w metodzie FRSR zakłada rozmycie Gaussa (opisane macierzą \mathbf{B}), po którym wykonywane jest zmniejszenie rozmiaru obrazu tak by otrzymać próbkę LR. W celu dokonania rekonstrukcji proces ten jest odwracany, tak by otrzymać obraz wysokiej rozdzielczości jako rozwiązanie problemu optymalizacyjnego (\mathcal{X}). Używana jest tutaj metoda gradientu prostego, z krokiem zdefiniowanym następująco:

$$\Delta\mathcal{X} = -\beta \left[\mathbf{B}'\mathbf{A}^T \text{signum}(\mathbf{A}\mathbf{B}\mathcal{X}_n - \mathbf{A}\mathcal{X}_0) + \lambda \frac{\delta U(\mathcal{X})}{\delta \mathcal{X}}(\mathcal{X}_n) \right], \quad (3.1)$$

$$U(\mathcal{X}) = \lambda \sum_{l=-P}^P \sum_{m=0}^P \alpha^{|m|+|l|} [\mathbf{I} - \mathbf{S}_y^{-m} \mathbf{S}_X^{-l}] \text{signum}(\hat{X}_n - \mathbf{S}_X^l \mathbf{S}_y^m \hat{X}_n) \quad (3.2)$$

gdzie: β jest parametrem kontrolującym krok metody, \mathbf{A} jest macierzą zawierającą informację o wkładzie poszczególnych obrazów LR do \mathcal{X}_0 , a $\mathbf{B}' = \mathbf{B}^T$ jest macierzą używaną w formie rozplotu (ang. *deconvolution*) rozmycia Gaussa. S jest oszacowaniem przesunięcia obrazu LR, $U(\mathcal{X})$ jest składnikiem odpowiedzialnym za regularyzację i jest kontrolowane hiperparametrami: λ – współczynnikiem wpływu, α – czynnikiem określającym zasięg operacji oraz $P \in 1, 2, 3$ – liczbą pełnopikselowych przesunięć wykonywanych w trakcie regularyzacji.

W metodzie EvoIM, oprócz optymalizacji hiperparametrów, używana jest zamiast jąder splotu Gaussa jako \mathbf{B} oraz \mathbf{B}' , swobodnie optymalizowana postać tych macierzy, tak by było możliwe zamodelowanie dowolnych funkcji rozmycia punktu (PSF) zamiast stosowanej zwykle funkcji Gaussa – jest to autorskie rozwiązanie. Zakłada się, że macierze te są ograniczone do postaci 5×5 i są symetryczne w ośmiu kierunkach (rys. 3.1), co redukuje liczbę optymalizowanych parametrów do sześciu b_0, \dots, b_5 . Oprócz wartości budujących macierze \mathbf{B} oraz \mathbf{B}' pozostałymi optymalizowanymi parametrami są wartości: α , β , λ , P , co daje w sumie przestrzeń poszukiwań złożonych z 16 zmiennych.

Optymalizacja realizowana jest za pomocą algorytmu ewolucyjnego. Wybór tej strategii był podyktowany głównie wymaganiem by metoda optymalizacji nie wymagała różniczkowalności funkcji celu. Oczywiście możliwe są inne strategie, które mają tę cechę [124]. Algorytmy ewolucyjne efektywnie przeszukują przestrzeń poszukiwań [125] i mają małą skłonność do utykania w lokalnych minimach w porównaniu np. do optymalizacji metodą roju cząstek [124], [126]. Również w literaturze opisywano próby podejścia do realizacji zadania SR za pomocą tego typu metod ([127]). Szczegóły jego działania są opisane w pseudokodzie w alg. 1. Hiperparametry wymienione wcześniej są zakodowane w chromosomie osobników jako liczby rzeczywiste, z wyjątkiem P , które jest liczbą naturalną. Proces obliczania funkcji przystosowania η osobnika p_i wymaga zbioru treningowego \mathbf{T} o rozmiarze

b5	b4	b2	b4	b5
b4	b3	b1	b3	b4
b2	b1	b0	b1	b2
b4	b3	b1	b3	b4
b5	b4	b2	b4	b5

Rysunek 3.1: Sposób kodowania jąder splotu w metodzie EvoIM. Maska jest symetryczna względem środka. Takie podejście pozwoliło ograniczyć liczbę parametrów optymalizowanych w sposób ewolucyjny. Pokazano na przykładzie maski 5×5 .

\mathbf{M} , gdzie każdy przykład HR jest związany ze zbiorem obrazów LR $\mathbf{I}^{(L)}$ przedstawiających tą samą scenę, ale w niższej rozdzielczości. Aby ocenić wartość funkcji przystosowania η osobnika p_i , hiperparametry zdefiniowane przez jego chromosom są wykorzystywane do konfiguracji algorytmu SR, który jest uruchamiany w celu przetworzenia każdego $I^{(L)}$ i rekonstruuje obraz o wysokiej rozdzielczości $\mathcal{I}^{(h)}$. Jakość rekonstrukcji jest oceniana na podstawie podobieństwa między $\mathcal{I}^{(h)}$ i $\mathcal{I}^{(h)}$ mierzonego za pomocą wybranej miary. Ostateczną wartość $\eta(p_i)$ uzyskuje się jako średnią z M wyników podobieństwa. Zasadniczo, im wyższa jest końcowa wartość, tym bardziej podobny jest zrekonstruowany obraz do obrazu referencyjnego, a to podobieństwo powinno wzrosnąć podczas optymalizacji ewolucyjnej. Sposób w jaki obliczana jest wartość funkcji przystosowania dla osobnika jest opisany na rys. 3.2. W trakcie eksperymentów w tej roli były użyte różnorodne miary, w szczególności KFS opisane w rozdziale 3.2.

3.1.2 Metoda EvoNet

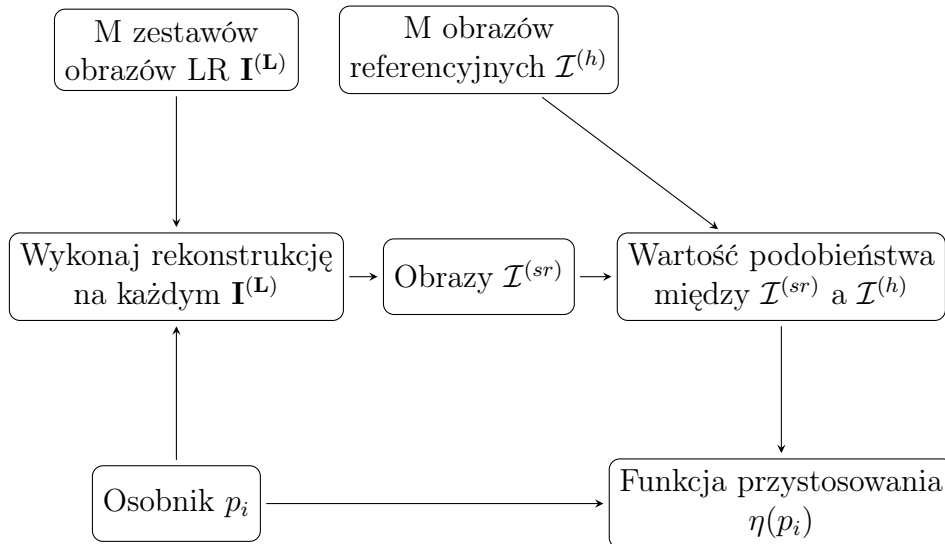
Sieci głębokie były przed 2019 rokiem z powodzeniem stosowane do realizacji zadania rekonstrukcji jednoobrazowej. Nie było jednak znane zastosowanie takich modeli do MISR. Przedstawiona w dalszej części metoda EvoNet, w której powstaniu autor rozprawy miał znaczący udział, była jednym z pierwszych podejść

Algorytm 1 Algorytm optymalizacji parametrów w metodzie EvoIM

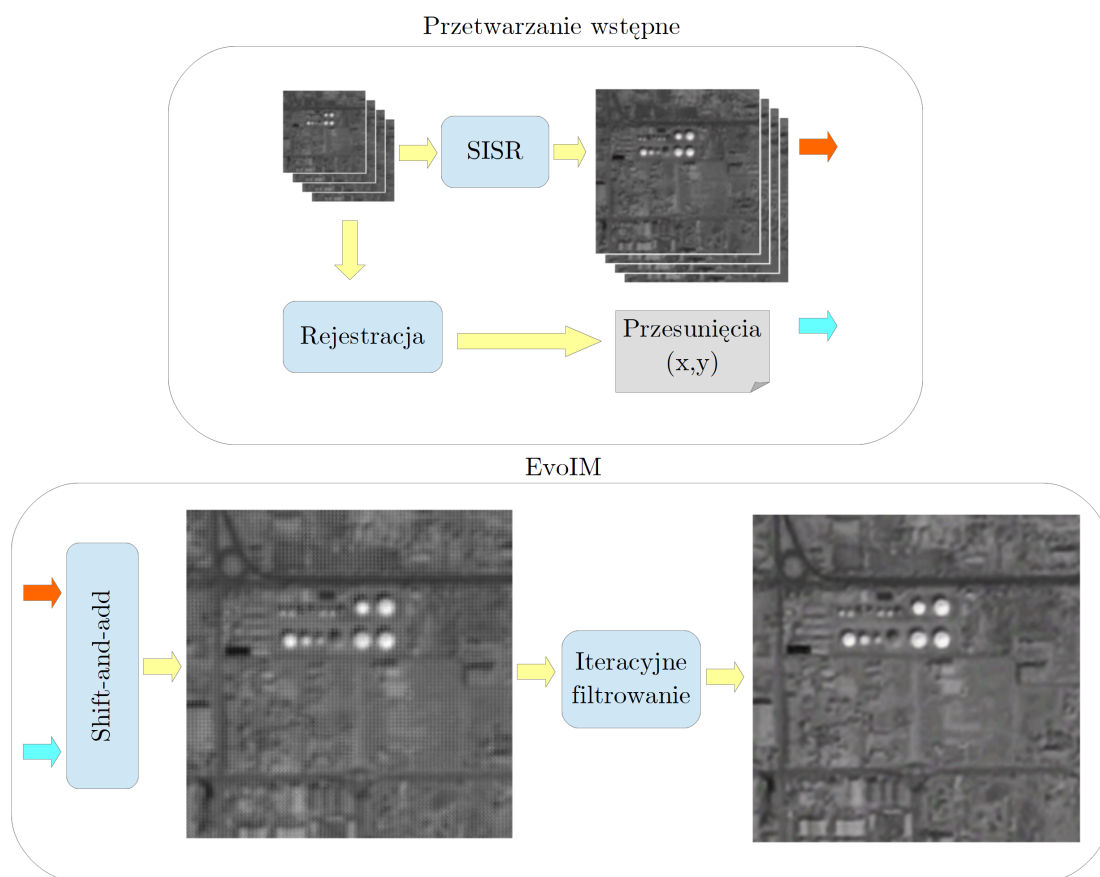
```

1: Inicjalizuj populację  $\mathcal{P} = \{p_i\}$  rozmiaru  $N_P$ 
2: repeat
3:    $\mathcal{P}_M \leftarrow \text{WybierzOsobnikiDoMutacji}(\mathcal{P}, \mathcal{P}_M)$ 
4:    $\mathcal{P}_C \leftarrow \mathcal{P} \setminus \mathcal{P}_M$ 
5:    $\mathcal{P}'_C \leftarrow \text{Krzyzowanie}(\mathcal{P}_C)$ 
6:    $\mathcal{P}'_M \leftarrow \text{Mutacja}(\mathcal{P}_M)$ 
7:    $\mathcal{P}' \leftarrow \mathcal{P} \cup \mathcal{P}'_M \cup \mathcal{P}'_C$ 
8:   for each  $p_i \in \mathcal{P}'$  do
9:      $\eta(p_i) \leftarrow \text{FunkcjaPrzystosowania}(p_i)$ 
10:  end for
11:   $\mathcal{P} \leftarrow \text{Selekcja}(\mathcal{P}', N_P)$ 
12:   $p_B \leftarrow \arg \max_{p_i \in \mathcal{P}} \{\eta(p_i)\}$ 
13: until WarunekZakonczenia
14: return  $p_B$ 

```



Rysunek 3.2: Obliczanie wartości funkcji przystosowania osobnika w algorytmie EvoIM. Osobnik p_i w genomie ma zakodowane wartości hiperparametrów metody FRSSR [20] oraz elementy jąder splotu stosowanych w tej metodzie jako funkcja rozmycia punktu. Każdy z zestawów obrazów LR jest przetwarzany za pomocą tych parametrów dając obraz zrekonstruowany $\mathcal{I}^{(sr)}$. Jakość rekonstrukcji jest oceniana za pomocą wybranej metryki i uśredniana dla wszystkich M zestawów. Wynik tej operacji stanowi wartość funkcji przystosowania $\eta(p_i)$ (ang. *fitness*).



Rysunek 3.3: Schemat działania metody EvoNet. Zbiór obrazów wejściowych LR jest przetwarzany wybraną metoda SISR. Równolegle obliczane są przesunięcia między tymi obrazami. Tak przetworzone dane zasilają metodę EvoIM, która w pierwszym kroku wykonuje operację shift-and-add, a następnie dokonuje kroków optymalizacji gradientowej, dając w efekcie wynikowy obraz SR.

do zastosowania uczenia głębokiego w rekonstrukcji wieloobrazowej [128]. Jak już wspomniano – do tych badań odnosiły się pionierskie prace dotyczące zadania MISR realizowanego za pomocą głębokich sieci konwolucyjnych [66]–[68]. Metoda EvoNet [128] polega na wstępnym przetworzeniu obrazów wejściowych $\mathbf{I}^{(L)}$ za pomocą jednoobrazowej rekonstrukcji nadrozdzielczej, a następnie użyciu wcześniej opisanej metodą ewolucyjnej optymalizacji SR EvoIM [129]. Metoda ta bywa wciąż używana w trakcie najnowszych badań nad MISR [130]. We wspomnianym artykule została wykonana analiza możliwych sposobów połączeń przetwarzania SISR oraz MISR. Metoda EvoNet została użyta w badaniach jako reprezentant podejścia, gdzie użycie metod SISR poprzedza MISR. Wyniki pokazują, że w najbardziej

złożonym przebadanym scenariuszu (dla danych zaszumionych przy trzykrotnym zwiększaniu rozdzielczości) EvoNet zapewnia najlepszą jakość rekonstrukcji spośród badanych metod.

Zasada działania EvoNet jest następująca:

1. Każdy z obrazów wejściowych $\mathcal{I}_i^{(l)}$ jest poddawany operacji zwiększania rozdzielczości za pomocą wybranej metody SISR dając w wyniku N obrazów $\mathbf{I}^{(l+)} = \{\mathcal{I}_i^{(l+)}\}$. Rozdzielczość na tym etapie jest K razy większa od $\mathcal{I}_i^{(l)}$. Może zostać użytych wiele różnych metod SISR, co jest przeanalizowane w rozdziale 4.1.
2. Równoległe do kroku (1), obrazy wejściowe $\mathcal{I}_i^{(l)}$ są rejestrowane w odniesieniu do pierwszego z nich, dając w wyniku przesunięcia \mathbf{S} . Przesunięcia te są wyznaczone na poziomie subpikselowym.
3. Wartości przesunięć \mathbf{S} oraz wstępnie przetworzone w kroku (1) obrazy $\mathbf{I}^{(l+)}$ są używane w kroku fuzji, realizowanej za pomocą metody shift-and-add (opisanej w rozdziale 2.1.2, a zobrazowanej na rys. 2.2. Daje to w wyniku obraz \mathcal{X}_0 , który jest L razy większy od $\mathbf{I}^{(l+)}$, a zarazem $K \cdot L$ większy od obrazów wejściowych LR $\mathcal{I}_i^{(l)}$.
4. \mathcal{X}_0 jest następnie poddawany iteracyjnej filtracji za pomocą metody EvoIM, w efekcie czego otrzymywany jest obraz $\mathcal{I}^{(sr)}$.

3.2 Miary podobieństwa obrazów

Jak wspomniano w rozdziale 2.2.3, miary operujące bezpośrednio w przestrzeni pikseli takie jak L1 i L2 mają znane ograniczenia i wady, które redukują możliwość ich zastosowania do oceny jakości rekonstrukcji nadrozdzielczej.

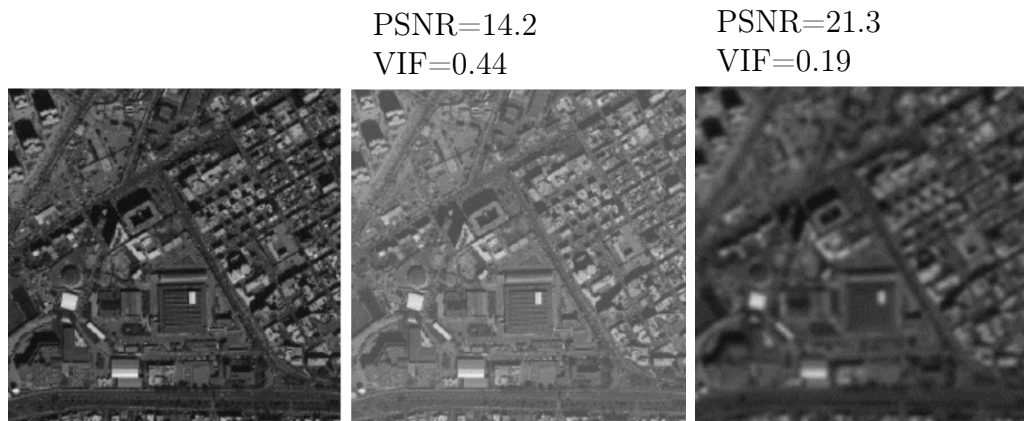
Zadanie jest jeszcze trudniejsze w przypadku, gdy obrazy pochodzą z różnych źródeł, a nie są rezultatem różnego rodzaju degradacji obrazu oryginalnego (kompresja stratna, rozmywanie, szum). Przykładowe obrazy tego samego obszaru pochodzące z dwóch różnych instrumentów, uchwyconych przy różnych natywnych rozdzielczościach i skalowanych do tej samej rozdzielczości, są pokazane na rys. 3.4). Można zaobserwować, że oprócz innej rozdzielczości przestrzennej (GSD dla Sentinel-2 – 10 m, dla WorldView-2 – 1,84 m) fragmenty obrazów różnią się pod względem jasności. Użycie tego typu danych wejściowych w SR powoduje, że miary takie jak PSNR nie są najlepszą metodą oceny jakości rekonstrukcji. Jest to spowodowane tym, że wynik SR posiada charakterystykę kolorystyczną zbliżoną do obrazów LR, a różną od obrazu HR. Ta różnica kolorystyczna dominuje nad porównywaniem samych szczegółów obrazu, które są kluczowe dla oceny jakości rekonstrukcji [24].



Rysunek 3.4: Porównanie wyglądu obrazów z różnych źródeł. Po lewej - obraz niskiej rozdzielczości z satelity Sentinel-2, po prawej - ten sam obszar widziany przez instrumenty misji WorldView-2 [24]

Dane tego typu można łatwo zasymulować, co jest pokazane na rys. 3.5. Oryginalny obraz został poddany zmianie histogramu w postaci rozjaśnienia (na środku) oraz rozmyciu przy zastosowaniu filtra Gaussa (po prawej). Pierwsza operacja nie zmienia widoczności szczegółów widocznych na obrazie, podczas gdy druga redukuje ją drastycznie. Od miary oceny jakości zgodnej z ludzką percepcją oraz poprawnie oceniającej widoczność szczegółów należy oczekiwać, że oceni podobieństwo po operacji zmiany histogramu jako większe niż obrazu ze znacznie niższą ilością detali. Jak jednak widać po wartościach, miara PSNR działa niezgodnie z tymi oczekiwaniami: $PSNR(a, b) < PSNR(a, c)$, gdzie relacja w zdefiniowanym wcześniej kontekście powinna być odwrotna. Miara VIF zachowuje się inaczej, uwzględniając szczegóły obrazu bardziej niż zmiany jasności.

Zasadne zatem wydaje się używanie miar, które nie są zbyt czułe na różnice charakterystyk instrumentów dokonujących akwizycji obrazu. Ponadto, dobra miara powinna koncentrować się na ekstrakcji i porównywaniu szczegółów widocznych w obrazach. Można domniemywać, że zaawansowane metody analizy cech lokalnych takie jak detektory punktów charakterystycznych, powinny zwracać podobne wyniki dla obrazu HR i odtworzonego obrazu. Większe podobieństwo wartości cech w punktach charakterystycznych znalezionych w HR do cech obliczonych na SR w tych samych punktach, będzie wskazywało na lepszą jakość rezultatu SR [24].



Rysunek 3.5: Działanie klasycznych miar - symulacja zmian histogramu obrazów. Obraz z satelity SPOT (po lewej) poddany zmianie histogramu (środek) oraz rozmyciu (po prawej). Widoczna jest większa czułość miary PSNR na zmiany kolorystyki niż utratę szczegółów. [24]. Podane tutaj dla porównania wartości miary VIF wskazują, że ta działa inaczej, uwzględniając informację o szczegółach bardziej niż zmiany kolorystyki.

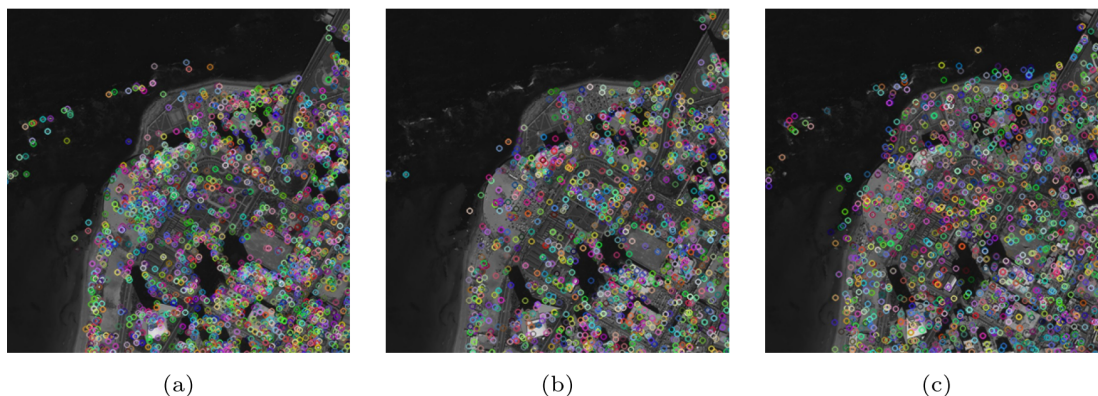
W zasadzie można uznać, że celem SR jest odtworzenie szczegółów z obrazów LR, tak aby lokalne cechy w wynikowym obrazie przypominały te wyodrębnione z obrazu HR.

3.2.1 Miary oparte na detektorach punktów kluczowych

Na podstawie przedstawionych wcześniej obserwacji i analiz zaproponowana została miara podobieństwa obrazów oparta na analizie punktów charakterystycznych. W takim podejściu podobieństwo obrazów rozważane jest w przestrzeni cech, która ignoruje zmiany histogramu obrazu, tak globalne jak i lokalne.

W dziedzinie przetwarzania obrazów często wykorzystywane są najczęściej detektory punktów charakterystycznych takie jak: SIFT (ang. *scale-invariant feature transform*) [131], BRISK (ang. *binary robust invariant scalable keypoints*) [132] oraz SURF (ang. *speeded-up robust features*) [133]. Wszystkie te metody analizują lokalne cechy obrazu w celu zidentyfikowania punktów charakterystycznych. Przykłady wykrytych punktów charakterystycznych przy użyciu BRISK, SIFT i SURF są przedstawione na rys. 3.6.

Wykorzystywane do zlokalizowania punktów charakterystycznych lokalne cechy obrazu dostarczają istotnych informacji na temat otoczenia punktów charakterystycznych. Ta informacja jest kodowana w celu dalszej analizy jako tzw. deskryptory punktów charakterystycznych. Są to wektory liczb, które charakteryzują oto-



Rysunek 3.6: Działanie detektorów punktów kluczowych: a) SIFT, b) BRISK, c) SURF. [24]

czenie punktu kluczowego. W przypadku SIFT i SURF są to informacje oparte na histogramie gradientu obrazu wokół punktu i są reprezentowane liczbami rzeczywistymi – dla SIFT jest to wektor 128 liczb, dla SURF – 64 lub 128 liczb [131], [133]. W metodzie BRISK deskryptorem jest natomiast wektor konstruowany na podstawie lokalnych wzorców binarnych (ang. *local binary patterns*, LBP) i jest zapisany w formie binarnej zawierającej 512 bitów. Takie cechy charakteryzują otoczenie punktu i mogą być ze sobą porównywane. Podobieństwo w punktach kluczowych jest obliczane na podstawie odległości między ich deskryptorami w dziedzinie lokalnych cech (dla SIFT i SURF stosuje się odległość euklidesową, podczas gdy odległość Hamminga jest odpowiednia dla BRISK). Istotne jest (co jest widoczne na rys. 3.6), że praktycznie nie wykrywa się punktów charakterystycznych w jednolitych obszarach nie zawierających żadnych szczegółów (w przykładzie jest to morze), dlatego takie obszary nie są brane pod uwagę przy mierzeniu podobieństwa. Kontrastuje to z innymi metodami oceny podobieństwa obrazów, które używają wszystkich pikseli obrazu do określenia podobieństwa, co w takich wypadkach wpływa zwykle na zafałszowanie wyniku.

Sposób obliczania odległości między obrazami w opisany wcześniej sposób jest zapisany w pseudokodzie w alg. 3.2.1. Proces przebiega w ten sposób, że w obrazie referencyjnym $\mathcal{I}^{(h)}$ wykrywane są punkty charakterystyczne (\mathcal{P}^h), które są następnie opisywane za pomocą deskryptorów ($\mathcal{D}^{(h)}$). W następnym kroku dla obrazu porównywanego są obliczane deskryptory w miejscach o współrzędnych takich samych jak uprzednio wykryte punkty kluczowe ($\mathcal{D}^{(sr)}$). Następnie obliczana jest średnia odległość ρ (euklidesowa dla SIFT i SURF, Hamminga dla BRISK) między parami deskryptorów z obrazów: referencyjnego $\mathcal{I}^{(h)}$ oraz porównywanego $\mathcal{I}^{(sr)}$. Ta wartość stanowi odległość między tymi obrazami w przestrzeni cech – de-

Algorytm 2 Obliczanie odległości między obrazami SR i HR przy użyciu detektorów punktów kluczowych

Require: *detektor*: detektor punktów kluczowych (SIFT, SURF, BRISK)

Require: *odleglosc*: metoda obliczania odległości między deskryptorami (euklidesowa dla SIFT i SURF, Hamminga dla BRISK)

Require: ϵ : stała o niewielkiej wartości wymagana dla uniknięcia dzielenia przez zero w przypadku gdy porównywane obrazy są identyczne

```

1:
2: function REJESTRACJA( $\mathcal{I}_1, \mathcal{I}_2$ )
3:   Rejestracja obrazu  $\mathcal{I}_1$  do  $\mathcal{I}_2$ 
4:   return  $\mathcal{I}_1$ 
5: end function
6:
7: function PUNKTYKLUCZOWE( $\mathcal{I}^{(h)}$ )
8:    $\mathcal{P} \leftarrow \text{detektor.wykryjPunktyKluczowe}(\mathcal{I}^{(h)})$ 
9:   return  $\mathcal{P}$ 
10: end function
11:
12: function DESKRYPTORY( $\mathcal{P}, \mathcal{I}$ )
13:    $\mathcal{D} \leftarrow \text{detektor.obliczDeskryptory}(\mathcal{P}, \mathcal{I})$ 
14:   return  $\mathcal{D}$ 
15: end function
16:
17: function ODLEGŁOŚĆ( $\mathcal{D}^{(h)}, \mathcal{D}^{(sr)}$ )
18:   odleglosci  $\leftarrow 0$ 
19:   for ( $\mathbf{d}^h, \mathbf{d}^{sr}$ ) in zip( $\mathcal{D}^{(h)}, \mathcal{D}^{(sr)}$ ) do
20:      $o \leftarrow o + \text{odleglosc}(\mathbf{d}^h, \mathbf{d}^{sr})$ 
21:   end for
22:   return odleglosci/len( $\mathcal{D}^{(h)}$ )
23: end function
24:
25: function KFS( $\mathcal{I}^{(sr)}, \mathcal{I}^{(h)}$ )
26:    $\mathcal{I}^{(sr)} \leftarrow \text{REJESTRACJA}(\mathcal{I}^{(sr)}, \mathcal{I}^{(h)})$ 
27:    $\mathcal{P}^h \leftarrow \text{PUNKTYKLUCZOWE}(\mathcal{I}^{(h)})$ 
28:    $\mathcal{D}^{(h)} \leftarrow \text{DESKRYPTORY}(\mathcal{P}^h, \mathcal{I}^{(h)})$ 
29:    $\mathcal{D}^{(sr)} \leftarrow \text{DESKRYPTORY}(\mathcal{P}^h, \mathcal{I}^{(sr)})$ 
30:    $d \leftarrow \text{ODLEGŁOŚĆ}(\mathcal{D}^{(h)}, \mathcal{D}^{(sr)})$ 
31:   return  $1/(d + \epsilon)$ 
32: end function

```

skryptorów. Aby otrzymać wynik interpretowalny w sposób najczęściej stosowany w ocenie podobieństwa (większa wartość wskazująca na większe podobieństwo), otrzymana odległość jest odwracana oraz skalowana za pomocą współczynnika C ($C = 1000$ dla SIFT i SURF oraz $C = 1$ dla BRISK). Sposób obliczania wartości miar KFS można zatem zapisać za pomocą wyrażenia:

$$\text{KFS}(\mathcal{I}^{(sr)}, \mathcal{I}^{(h)}) = \frac{C \cdot N}{\sum_{i=1}^N \rho(\mathcal{D}^{(h)}, \mathcal{D}^{(sr)})}, \quad (3.3)$$

gdzie N jest liczbą punktów kluczowych wykrytych na obrazie $\mathcal{I}^{(h)}$. W przypadku gdy nie zostaną wykryte żadne punkty kluczowe wartość podobieństwa zostaje określona na 1000, co wykracza poza spotykany normalnie zakres wartości miar KFS.

Istotne jest, że tego typu metoda jest asymetryczna – ważne jest, który obraz zostanie podany jako referencyjny, w przeciwieństwie do algorytmów takich jak PSNR czy SSIM. Autorskie metody oceny podobieństwa obrazów w ten sposób zdefiniowane [24] zostały nazwane KFS (ang. *keypoint feature similarity*). Zaproponowano zatem trzy miary: KFS_{SIFT} , KFS_{SURF} , $\text{KFS}_{\text{BRISK}}$.

3.2.2 Miary oparte na uczeniu maszynowym

Jak wspomniano w rozdziale 2.2.3, obecnie nacisk w rozwoju miar podobieństwa obrazu jest położony na modele utworzone za pomocą metod sztucznej inteligencji, w szczególności głębokie konwolucyjne sieci neuronowe. Jedną z szeroko stosowanych tego typu miar jest LPIPS. Dzięki publicznie dostępnej implementacji, w szczególności sposobowi treningu, istnieje możliwość jej modyfikowania – co zostanie opisane w dalszej części rozdziału.

W trakcie oryginalnego treningu LPIPS w każdej epoce przetwarzany jest zbiór treningowy, który składa się z zestawu par obrazów porównywanych: \mathcal{I}_0 , \mathcal{I}_1 oraz etykiety, którą jest liczba z zakresu $\langle 0, 1 \rangle$. Liczba ta jest interpretowana w następująco: 0 oznacza lepszą jakość obrazu \mathcal{I}_0 , 1 zaś - obrazu \mathcal{I}_1 . Etykiety zostały uzyskane za pomocą badania ankietowego. Dla każdej pary obrazów uzyskano dwa głosy obserwatorów $\{\mathcal{H}_1, \mathcal{H}_2\}$ – po ich uśrednieniu otrzymano możliwe wartości wartości: $\{0; 0,5; 1\}$.

Widać zatem, że w trakcie treningu można zmieniać dane etykiet podając inne wartości z zakresu $\langle 0, 1 \rangle$, modyfikując liczbę i źródło głosów – oczywiście w przypadku gdy źródło stanowią wyłącznie głosy inne niż ludzkie, trudniej mówić o perceptualności miary w ścisłym sensie. Istnieje możliwość wytrenowania architektury sieci wykorzystywanej do wyznaczania wartości metryki LPIPS używając wyłącz-

nie lub dodatkowo głosu \mathcal{V} pochodzącego z innej metryki \mathcal{M} w następujący sposób:

$$\mathcal{V} = \begin{cases} 0 & \text{jeżeli } \mathcal{M}(\mathcal{I}_0) < \mathcal{M}(\mathcal{I}_1), \\ 0,5 & \text{jeżeli } \mathcal{M}(\mathcal{I}_0) = \mathcal{M}(\mathcal{I}_1), \\ 1 & \text{jeżeli } \mathcal{M}(\mathcal{I}_0) > \mathcal{M}(\mathcal{I}_1), \end{cases} \quad (3.4)$$

przy założeniu, że miara \mathcal{M} jest interpretowana jako odległość między obrazami, czyli mniejsza wartość oznacza większe podobieństwo.

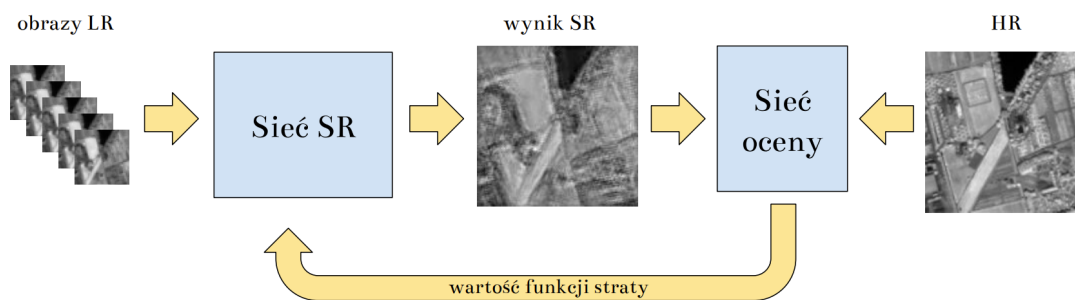
W ten sposób zdefiniowanie zostały dodatkowe miary:

1. DKFS (ang. *Deep Keypoint Feature Similarity*), gdzie jedyny głos pochodzi z opisaną wzorem 3.4 interpretacji wartości miary KFS: $\{\mathcal{V}\}$.
2. DKFS_{MIX}, gdzie do dwóch oryginalnych głosów obserwatorów ludzkich zostaje dodany głos pochodzący KFS: $\{\mathcal{H}_1, \mathcal{H}_2, \mathcal{V}\}$.

Końcową wartością miar DKFS jest średnia ze zbioru głosów. Jako KFS może być zastosowana każda z miar KFS_{BRISK}, KFS_{SIFT}, KFS_{SURF}. Wybór używanej w walidacji eksperymentalnej miary KFS_{SIFT} został podyktowany wynikami badań opisanych w rozdziale 4.2. W odróżnieniu od samych miar KFS, w ten sposób zdefiniowana i wytrenowana metoda jest symetryczna – nie ma znaczenia, w jakiej kolejności obrazy są do niej podawane.

3.3 Użycie miary opartej na cechach głębokich w rekonstrukcji wieloobrazowej

Jak wspomniano w rozdziale 2.1.3, w treningu obecnie występujących w literaturze metod SR bazujących na sieciach głębokich dominującymi funkcjami straty są metryki L1 oraz L2 (w szczególności PSNR lub cPSNR) i z powodzeniem są one używane w zastosowaniach takich jak rekonstrukcja nadrozdzielcza na danych PROBA-V. W rozdziałach 2.2 oraz 3.2 omówiono zaś wady typowych miar jakości obrazów i wskazano możliwe inne rozwiązania, takie jak KFS. Mimo ich zalet nad wspomnianymi najprostszyimi metrykami, te miary nie są stosowane ze względu na brak różniczkowalności, który uniemożliwia ich bezpośrednie zastosowanie. Celem jest zatem znalezienie sposobu jak tego typu miary zastosować w treningu sieci. Możliwe jest użycie w miejscu klasycznej funkcji straty modułu obliczającego podobieństwo, który sam w sobie stanowi osobną całość. Należy oczywiście w takim podejściu zablokować możliwość trenowania tego fragmentu sieci, czyli potraktować zmienne (wagi) tej części jako stałe. Sposób rozwiązania treningu z użyciem miary opartej na sieci głębokiej powoduje, że sama sieć nie jest modyfikowana, natomiast zmieniany jest sposób treningu – pokazane to jest na rys. 3.7.



Rysunek 3.7: Sposób treningu sieci głębokiej z użyciem miar opartych na sieciach głębokich: wstępnie przetworzone obrazy wejściowe są podawane do sieci SR. Wynik rekonstrukcji jest porównywany z obrazem referencyjnym za pomocą funkcji straty opartej również na sieci konwolucyjnej

W chwili pisania tej rozprawy, jedną z najlepszych metod MISR był model PIUnet [69], opisany w rozdziale 2.1. W publicznym rankingu metod MISR [72] dla obrazów satelitarnych pochodzących z satelity PROBA-V algorytm ten wciąż uzyskuje jedne z najwyższych wyników. Stąd decyzja autora niniejszej rozprawy by eksperymenty oprzeć na tej metodzie. Jednak zaprezentowane tutaj podejście jest ogólne i nie wymaga użycia specyficznie tego modelu. Celowe wydaje się użycie w treningu sieci modeli bardziej zbliżonych do ludzkiego postrzegania podobieństwa obrazów, takich jak LPIPS [93] lub DKFS (rozdział 3.2), które przełamują ograniczenia tradycyjnych metod, w szczególności nadmierną wrażliwość na zmiany histogramu, a małą – na degradację szczegółów, które są kluczowe w zadaniu rekonstrukcji nadrozdzielczej.

Istnieje ryzyko, że takie podejście może prowadzić do powstawania artefaktów w wynikowych obrazach. Wynika to z faktu, że miary oparte na cechach głębokich mogą kierować trening tak, że pewne cechy obrazu są preferowane nad innymi, co może wprowadzać niepożądane artefakty. Przykładowo, model trenowany z użyciem miary opartej na cechach głębokich może preferować odwzorowanie tekstury kosztem dokładności odwzorowania krawędzi lub kolorystyki obrazu. Aby uniknąć powstawania tego typu efektów w wynikowych obrazach, wprowadzono dwie strategie. Pierwszą jest łączenie w postaci funkcji straty różnych miar oceny jakości obrazów – co pozwala na uwzględnienie w treningu różnych aspektów jakości, takich jak krawędzie, tekstura i kolory. W praktyce będzie to sumowanie lub uśrednianie wartości funkcji straty opartej na cechach głębokich z funkcją straty L1 / L2.

$$\mathcal{L} = w_1\mathcal{L}_1 + w_2\mathcal{L}_2, \quad (3.5)$$

gdzie \mathcal{L} jest wynikową funkcją straty, \mathcal{L}_1 oraz \mathcal{L}_2 są funkcjami straty składowymi,

a w_1 oraz w_2 są wagami, które pozwalają na sterowanie istotnością poszczególnych elementów tak skonstruowanej funkcji straty. Oczywiście, to podejście może zostać uogólnione na więcej niż dwa składniki, natomiast w niniejszej rozprawie zostały przetestowane jako \mathcal{L}_1 miary LIPS, DKFS i DKFS_{MIX}, a jako \mathcal{L}_2 zostało użyte L1. Kolejną strategią jest dotrenowanie modelu wytrenowanego z funkcją straty opartą na cechach głębokich za pomocą innej funkcji straty L1 / L2 oraz scenariusz odwrotny.

Rozdział 4

Walidacja eksperymentalna

Opis eksperymentów składa się z kilku podrozdziałów, w których jest przedstawione jak zastosowanie odpowiednich metod porównywania obrazów oraz sposobów optymalizacji może wpłynąć na polepszenie jakości rekonstrukcji za pomocą metod MISR. W pierwszym kroku została zweryfikowana jest teza T1, zgodnie z którą zastosowanie sieci neuronowych w zadaniu MISR w połączeniu z metodami klasycznymi powoduje poprawienie jakości rekonstrukcji (rozdział 4.1). Uzyskane wyniki wskazały na fakt, że tradycyjnie używane miary podobieństwa nie prowadzą do dobrych jakościowo wyników rekonstrukcji na danych satelitarnych. Stąd też w następnym kroku miary zostały przebadane pod kątem pożądanych w SR cech (rozdział 4.2). Dalej – wybrane uprzednio miary zostały zastosowane w trakcie treningu jednej z uznanych metod MISR tak by zweryfikować tezę T2, mówiącą że dla rzeczywistych obrazów satelitarnych możliwa jest poprawa względem tradycyjnie stosowanych miar L1 i L2 (rozdział 4.3).

4.1 Połączenie metod klasycznych z sieciami głębokimi

Algorytmy opisane w rozdziale 3.1 służą optymalizacji hiperparametrów metody FRSSR [20] oraz samych jąder splotu używanych w autorskiej modyfikacji (rozdział 3.1) – razem stanowi to metodę EvoIM. Zastosowane w tym eksperymencie wartości parametrów tej metody zostały uzyskane w ramach badań opisanych w artykule [129]. Celem badań i analiz w niniejszym rozdziale, tak jak w artykule [128], jest potwierdzenie, że przetwarzanie wstępne obrazów wybranymi metodami SISR przed zastosowaniem algorytmu MISR daje efekty lepsze niż składowe elementy stosowane samodzielnie. Było to jedno z pierwszych podejść, w którym do rekonstrukcji wieloobrazowej zostały wykorzystane głębokie sieci splotowe, co zostało między innymi odnotowane w artykułach opisujących pierwsze sieci głębokie im-

plementujące cały proces rekonstrukcji wieloobrazowej: DeepSUM [67], RAMS [66] oraz HighRes-net [68].

W celu przeprowadzenia walidacji eksperymentalnej metody EvoNet zostały użyte następujące dane wejściowe:

1. AD (ang. *artificially degraded images*, obrazy sztucznie degradowane) – 10 obrazów z satelity Sentinel-2 (część zbioru danych B4MultiSR) w rozmiarze 500×500 pikseli; degradacja polegała na zastosowaniu przesunięć subpikselowych (po 0,5 piksela w każdym kierunku – co daje 4 obrazy zdegradowane) przed zastosowaniem zniekształceń: rozmycia Gaussa, interpolacji zmniejszającej rozdzielczość przestrzenną oraz dodania szumu.
2. RS (ang. *real satellite images*, rzeczywiste obrazy satelitarne), dla których dla każdego z obrazów HR (dwóch scen z satelity SPOT oraz jednej z Digital Globe WorldView-4) zostało dobranych 10 obrazów LR zarejestrowanych w ramach misji Sentinel-2.

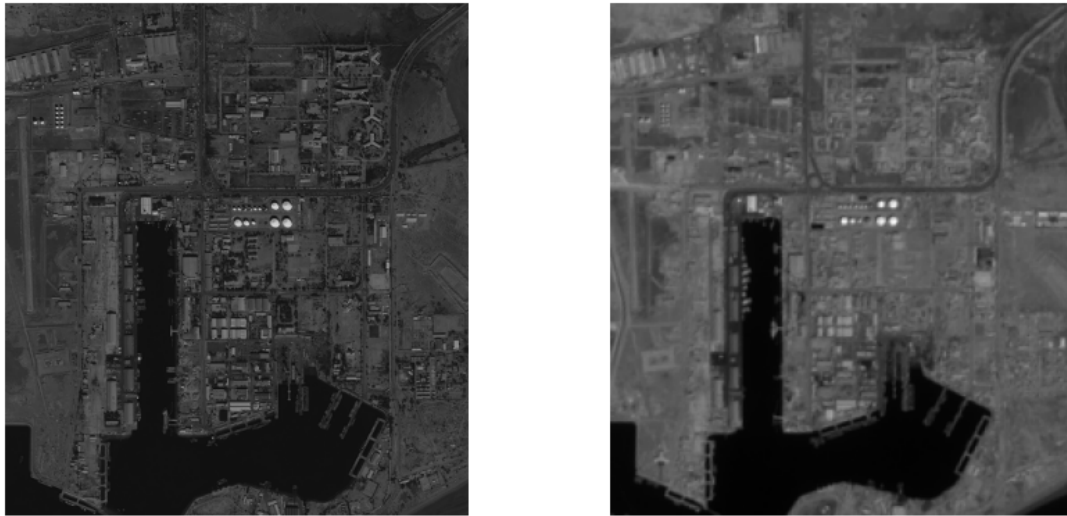
Wyniki EvoNet zostały porównywane z:

- 1) interpolacją dwusześcienną,
- 2) metodami SISR: FSRCNN [47], SRResNet [48] i LapSRN [56] wytrenowanymi w ten sam sposób na obrazach ze zbioru DIV2K (algorytmy te są używane w kroku przetwarzania wstępnego badanych wersji EvoNet),
- 3) uśrednionymi obrazami przetworzonymi wspomnianymi metodami SISR,
- 4) metodami MISR: GPA [59] i EvoIM [129].

Walidacji poddane zostały cztery różne wersje metody EvoNet: ze wstępnym przetwarzaniem obrazów wejściowych za pomocą interpolacji dwusześciennej oraz za pomocą wspomnianych wcześniej trzech sieci SISR: FSRCNN, SRResNet i LapSRN.

Metoda EvoIM, której walidacja opisana jest w tym rozdziale, została wytrenowana osobno na obrazach z grup AD i RS [129]. Jako funkcji oceny jakości w tych badaniach używano PSNR z przetwarzaniem wstępnym w postaci filtru górnoprzepustowego. Algorytmy zostały zaimplementowane w języku C++ i tak przygotowane moduły były uruchamiane w środowisku Java (optymalizacja ewolucyjna) oraz Python / Keras, tak by otrzymać wyniki z metody EvoNet. Eksperymenty przeprowadzono na komputerze z procesorem graficznym RTX 2080 z 8 GB pamięci VRAM.

Metody EvoIM oraz wymienione SISR zostały uruchomione w trybie dwukrotnego powiększania obrazu. Ponieważ schemat EvoNet w każdym kroku dwukrotnie

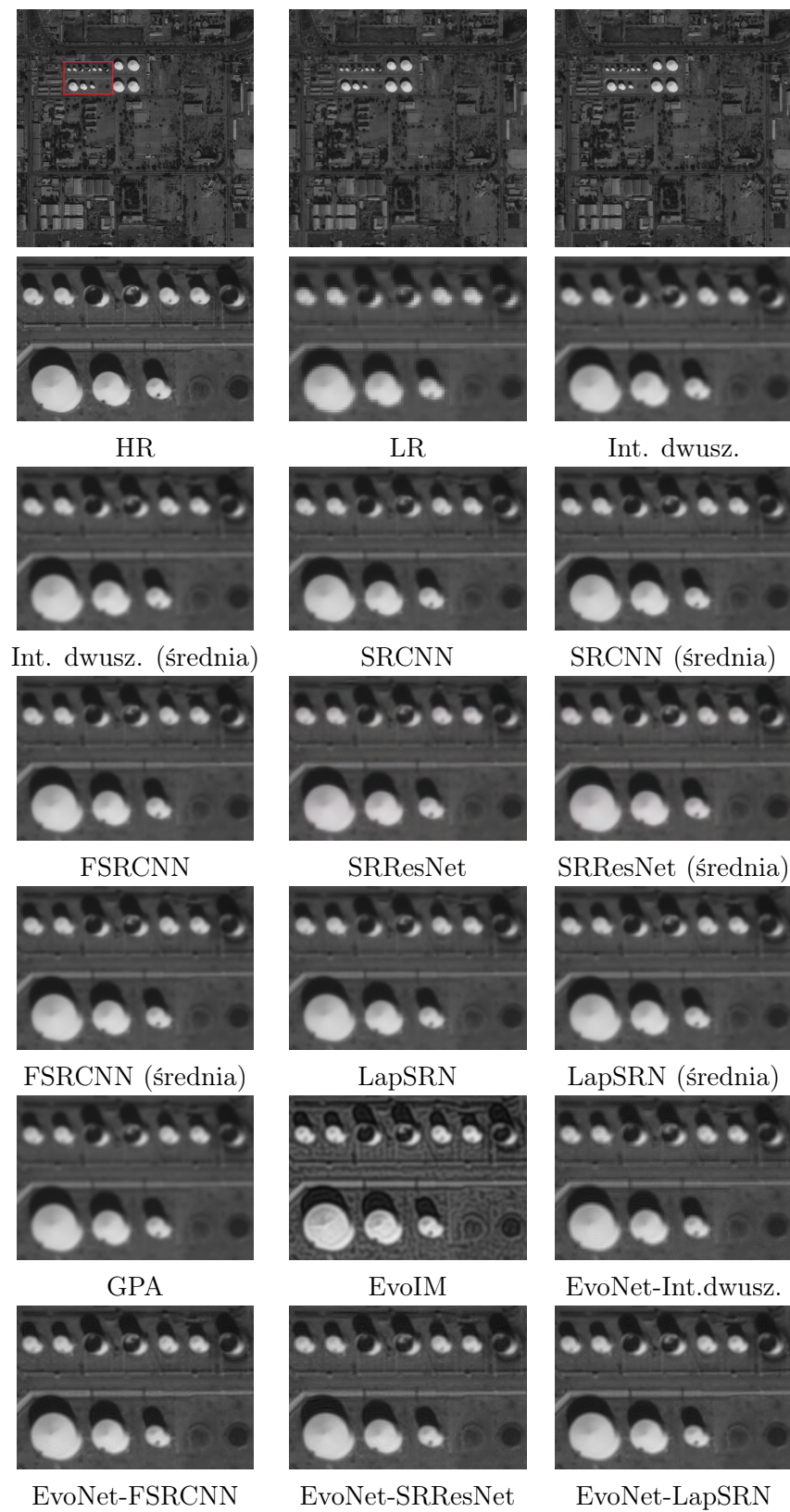


Rysunek 4.1: Przykład lokalnych różnic jasności i kontrastu między obrazami HR oraz LR pochodzącymi z różnych satelitów w zbiorze 2 (RS). Po lewej – obraz HR z satelity SPOT, po prawej – jeden z obrazów LR z satelity Sentinel-2.

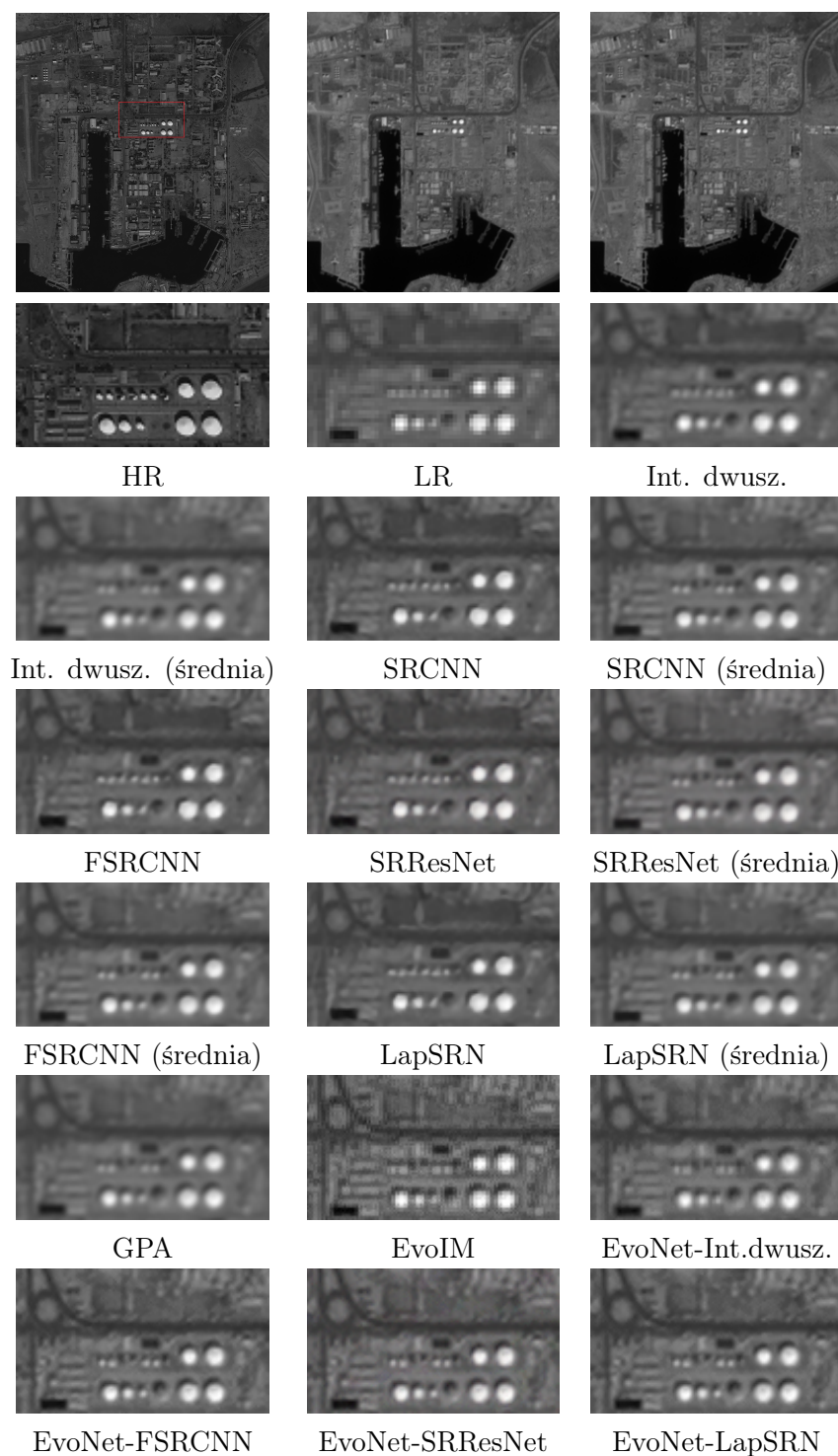
zwiększa rozdzielczość, wynikowe wymiary obrazu SR są czterokrotnie większe od LR. Aby wyniki działania metod były porównywalne, należy sprowadzić porównywane obrazy do tego samego rozmiaru, zatem możliwe jest dodatkowe dwukrotne powiększenie obrazów EvoIM/SISR, lub dwukrotne zmniejszenie rozdzielczości wyniku EvoNet. Wadą pierwszego wskazanego podejścia jest fakt, że obraz otrzymany w wyniku działania metody EvoNet zostanie porównany z obrazem dodatkowo przetworzonym, co może faworyzować EvoNet. Dlatego też zdecydowano się otrzymane obrazy EvoNet poddać zmniejszaniu za pomocą interpolacji dwusześcienną, by otrzymać w wyniku obraz dwa razy większy od LR – czyli o rozdzielczości obrazów wynikowych SISR/EvoIM.

Analiza wyników

Wyniki zostały przedstawione w tabeli 4.1 w postaci wartości miar SSIM, PSNR oraz VIF. Wizualnie wyniki dla przykładowych obrazów zostały zaprezentowane na rysunku 4.2 (obraz ze zbioru AD) oraz na rysunku 4.3 (obraz ze zbioru RS). Dla zbioru AD wyraźnie widać, że najlepsze wyniki daje zastosowanie EvoNet, na co wskazują wartości wszystkich trzech wspomnianych miar. Sytuacja jest nieco inna dla rzeczywistych danych satelitarnych – zbioru RS, gdzie LR i HR pochodzą z różnych instrumentów (odpowiednio Sentinel-2 oraz SPOT). Tutaj, co prawda



Rysunek 4.2: Przykłady wyników EvoNet i pozostałych metod badanych w niniejszym rozdziale na zbiorze obrazów sztucznie degradowanych (AD). Poza pierwszym wierszem pokazano powiększenia fragmentu zaznaczonego czerwoną ramką na obrazie HR.



Rysunek 4.3: Przykłady wyników EvoNet i pozostałych metod badanych w niniejszym rozdziale na zbiorze obrazów rzeczywistych (RS). Poza pierwszym wierszem pokazano powiększenia fragmentu zaznaczonego czerwoną ramką na obrazie HR.

Tabela 4.1: Wyniki działania metody EvoNet [128]. W sekcji SISR powiększenie stosowane na pojedynczym obrazie LR. W sekcji MISR dla metod innych niż GPA i EvoIM – uśrednienie powiększeń na obrazach LR. Trzy najlepsze wartości miar w każdej z kolumn są pogrubione.

		AD			RS		
		PSNR	SSIM	VIF	PSNR	SSIM	VIF
SISR	Interp. dwusz.	28,12	0,791	0,422	16,09	0,445	0,113
	SRCNN	26,79	0,755	0,374	15,48	0,378	0,099
	FSRCNN	29,04	0,826	0,458	16,13	0,439	0,115
	SRResNet	28,77	0,823	0,453	16,47	0,426	0,111
	LapSRN	29,00	0,826	0,457	16,17	0,446	0,116
MISR	Interp. dwusz.	28,13	0,791	0,423	16,15	0,494	0,129
	SRCNN	27,41	0,763	0,388	15,88	0,455	0,121
	FSRCNN	29,05	0,827	0,460	16,21	0,499	0,134
	SRResNet	28,72	0,818	0,449	16,17	0,506	0,134
	LapSRN	29,02	0,827	0,459	16,23	0,500	0,134
	GPA	28,05	0,792	0,422	15,95	0,474	0,125
	EvoIM	31,19	0,863	0,561	15,64	0,402	0,114
EvoNet	EvoNet-Interp. dwusz.	33,56	0,916	0,587	15,89	0,454	0,125
	EvoNet-FSRCNN	36,27	0,948	0,670	15,91	0,459	0,128
	EvoNet-SRResNet	35,06	0,948	0,661	15,83	0,482	0,134
	EvoNet-LapSRN	36,79	0,951	0,686	15,93	0,459	0,128

miara VIF daje najlepszy wynik dla EvoNet-SRRResNet, ale podobne wartości przyjmuje również dla trzech innych metod (pogrubienie w tabeli 4.1). Pozostałe dwie miary (PSNR oraz SSIM) dla metod EvoNet dają wartości niższe od innych metod. W zbiorze RS dane LR oraz HR mają różniącą się charakterystykę spektralną (rys. 4.1). Dodatkowo należy zwrócić uwagę, że względu na wady metryk PSNR oraz SSIM opisane w rozdziale 2.2 uzyskiwane wartości wspomnianych metryk są niższe niż powszechnie spotykane (zwykle więcej niż 20 dB dla PSNR i więcej niż 0,7 dla SSIM [81], [111]).

Ocena wyników przez obserwatorów

W poprzednim punkcie opisano niekorzystne dla metody EvoNet oceny uzyskane na zbiorze RS z użyciem metryk literaturowych. Z uwagi na ten fakt postanowiono przeprowadzić badanie ankietowe. Zestaw pytań utworzono na bazie 10 przykładów ze zbioru AD oraz pięciu – z RS. Prezentowane wersje obrazów po-

Tabela 4.2: Wyniki działania metody EvoNet – opinie ankietowanych. Wartości oznaczają liczbę oddanych na badane metody głosów dla każdego z pytań [128]. Prezentowane są również rangi metod (kolumny #) w obrębie każdego z pytań. Pytania oznaczone gwiazdką zostały zadane w interpretacji odwrotnej – np. *Wskaż najmniej ostry obraz*, co skutkuje odwrotnym sortowaniem rang.

Zbiór	Pyt.	Interp. Dwusześć	#	EvoNet- -LapSRN	#	LapSRN	#	EvoIM	#
RS	3	0	3	241	1	0	3	66	2
	5	0	4	272	1	13	3	22	2
	9	0	4	238	1	7	3	62	2
	10	1	4	257	1	22	3	27	2
	15	0	4	263	1	6	3	38	2
AD	1	2	4	288	1	3	3	14	2
	2*	224	4	10	1	55	3	18	2
	4	4	4	260	1	5	3	38	2
	6	0	4	240	1	11	3	56	2
	7	2	4	88	2	11	3	206	1
	8*	115	3	8	1	17	2	167	4
	11	2	4	182	1	3	3	120	2
	12	0	4	264	1	10	3	33	2
	13	0	4	260	1	14	3	33	2
	14*	283	4	4	1	9	2	11	3

chodziły z czterech metod, które zostały wybrane w następujący sposób: metoda z najlepszym numerycznym wynikiem dla zbioru AD (algorytm EvoNet-LapSRN), metody składowe tego podejścia (EvoIM oraz LapSRN) oraz interpolacja dwusześcienna jako bazowe odniesienie. Każdy ankietowany otrzymał zestaw pytań jednokrotnego wyboru skonstruowanych w większości na bazie zadań dla obserwatora [134], w postaci podobnej do następujących przykładów: *Na którym obrazie drzewa i drogi wyglądają najbardziej naturalnie? Który obraz wygląda najbardziej naturalnie i ostro? Na którym obrazie najłatwiej policzyć drzewa? Który obraz jest najmniej ostry?* Przykłady pytań z prezentowanymi obrazami są przedstawione na rys. 4.4¹. Otrzymano 307 odpowiedzi, z których został wyznaczony wskaźnik opinii MOS (ang. *Mean Opinion Score*) pokazujący jaka część ankietowanych wskazała daną metodę jako najlepszą – osobno dla zbiorów AD oraz RS. Sposób opracowania wyników był następujący:

¹Szczegóły wszystkich pytań znajdują się pod linkiem: <https://sun.aei.polsl.pl/mkawulok/srr/evonet.pdf>

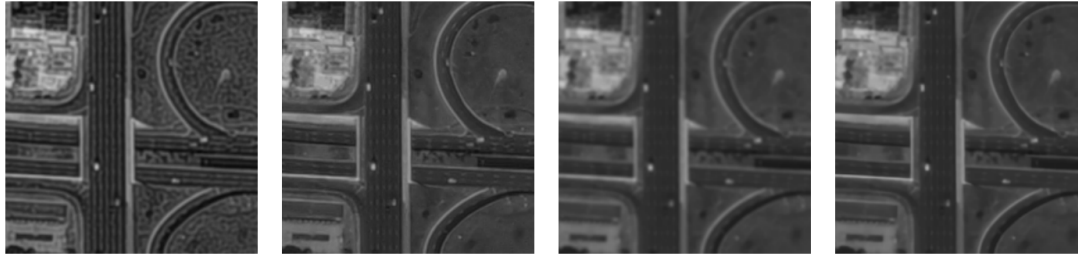
Tabela 4.3: Wyniki działania metody EvoNet – podsumowanie opinii ankietowanych. Wartości procentowe oznaczają, jaka część ankietowanych wskazała daną metodę w danym zbiorze jako dającą najlepsze wyniki [128].

Metoda	AD	RS
Bicubic	6,29%	0,07%
LapSRN	18,26%	3,13%
EvoIM	24,94%	14,01%
EvoNet-LapSRN	50,51%	82,80%

1. Zebranie odpowiedzi w postaci czwórek (Q, S, M, A) , gdzie Q to numer pytania, S to informacja o zbiorze (AD / RS), którego dotyczy pytanie, M to wskazana jako odpowiedź metoda. A to sposób interpretacji odpowiedzi: czy w pytaniu chodzi o wskazanie najlepszego ($A = 1$), czy najgorszego ($A = -1$) jakościowo obrazu spośród prezentowanych. Liczba wszystkich ankietowanych osób oznaczana jest jako N .
2. Dla każdego pytania obliczane jest: (Q, S, M, Σ_Q) , gdzie Σ_Q jest liczbą wskazującą w ilu wypadkach ankietowani w danym pytaniu Q wskazali jako najlepszą metodę M . Dla pytań zadanych odwrotnie ($A = -1$) wartość Σ_Q jest modyfikowana następująco: $\Sigma_Q = N - \Sigma_Q$.
3. W każdym ze zbiorów S obliczana jest suma odpowiedzi dla każdej z metod: (S, M, Σ_S) .
4. Obliczana jest suma wszystkich odpowiedzi dla zbioru: (S, Σ) .
5. Dla każdej z metod w każdym zbiorze obliczany jest procent odpowiedzi dla metody M : $(S, M, \Sigma_S / \Sigma \cdot 100\%)$.

Wyniki są zaprezentowane w tabelach 4.2 (szczegóły dla pytań) oraz 4.3 (podsumowanie). Widać, że tylko w wypadku pojedynczego pytania ze zbioru AD (drugie pytanie na rysunku 4.2) za najlepszą jakościowo została uznana metoda inna niż EvoNet-LapSRN – i była to metoda EvoIM. Wyniki są spójne zarówno dla pytań zadanych wprost (*Gdzie najłatwiej policzyć drzewa?*) oraz odwrotnie (*Wskazaj najmniej ostry obraz*). Wyniki te potwierdzają, że według obserwatorów najlepiej rekonstrukcję przeprowadza metoda EvoNet – zarówno w zbiorze AD, jak i w znacznie wyraźniejszy sposób w zbiorze RS.

Na którym z obrazów przerywane linie na drodze są najlepiej widoczne?



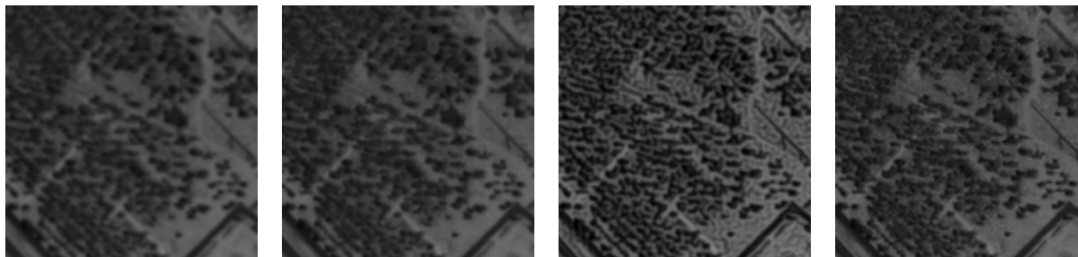
EvoIM
(56)

EvoNet-LapSRN
(240)

Interp. dwusz.
(0)

LapSRN
(11)

Na którym z obrazów najłatwiej policzyć drzewa?



Interp. dwusz.
(2)

LapSRN
(11)

EvoIM
(206)

EvoNet-LapSRN
(88)

Rysunek 4.4: Przykładowe pytania z ankiety dotyczącej wyników metody EvoNet. Poniżej nazwy metody podano liczbę oddanych głosów.

Podsumowanie

W świetle zaprezentowanych wyników uzyskanych dla metody EvoNet można uznać, że zarówno w wypadku danych symulowanych AD (gdzie zarówno ocena przez obserwatorów oraz wartości miar literaturowych są zgodne), jak i RS (gdzie ocena ankietowanych w przeważającej większości to pokazuje) złożenie metod klasycznych z algorytmami opartymi na sztucznych sieciach neuronowych daje wyniki lepsze jakościowo niż każde z tych podejść z osobna. Stanowi to potwierdzenie tezy T1. Zarazem niespójność między ocenami ilościowymi za pomocą miar PSNR i SSIM a ocenami obserwatorów dla zbioru składającego się z rzeczywistych obrazów satelitarnych poskutkowało zaplanowaniem i wykonaniem badań opisanych w dalszej części rozprawy.

4.2 Badanie cech metryk pod kątem rekonstrukcji nadrozdzielczej

Jak podkreślono w rozdziale 2.1, istniejące metryki definiowane analitycznie, w większości są czułe na zmiany kolorystyki, często w stopniu większym niż zmiany w szczegółach obrazu. Również eksperymenty opisane w rozdziale 4.1 pokazały znaczącą niespójność między wartościami miar oceny jakości takimi jak PSNR lub SSIM a ocenami ankietowanych obserwatorów. Stąd też celowe okazało się zbadanie jak poszczególne metody oceny jakości obrazów zachowują się w zadaniu rekonstrukcji nadrozdzielczej. W niniejszej pracy przeanalizowano różne metryki podobieństwa obrazów, w tym wprowadzone wcześniej algorytmy KFS (rozdział 3.2.1). Ilościowo i jakościowo zbadano ich wrażliwości na różne operacje degradacji obrazu. Do walidacji wykorzystano dane ze zbioru danych B4MultiSR (2.3.5) oraz zbioru LIVE (2.3.4).

Zastosowano następujące podejścia:

- A Sposób oceny wskaźników podobieństwa obrazów z LIVE, który zawiera kilka serii obrazów poddanych różnym procedurom degradacji. W badaniach wykorzystano podzbiór 175 obrazów, zawierający zniekształcenia wynikające z kompresji JPEG. Chociaż ten test porównawczy nie koncentruje się na kwestiach specyficznych dla SR, zamiarem było wykorzystanie zestawu wzorcowego jako punktu odniesienia podczas oceny nowych metryk.
- B Badanie odporności miar podobieństwa na różne czynniki degradacji. Jest to wykonywane na sztucznie zdegradowanych obrazach satelitarnych ze zbioru danych B4MultiSR.
- C Używając dwóch iteracyjnych algorytmów SR: IBP [58] oraz GPA [20], obserwowane są wartości podobieństwa w kolejnych iteracjach dla rzeczywistych obrazów satelitarnych ze zbioru danych B4MultiSR. Celem tego eksperymentu było wykorzystanie technik, które stopniowo poprawiają wynik rekonstrukcji w kolejnych iteracjach. Chciano dokonać obserwacji, czy miary wskazują istotnie trend polepszania jakości obrazu. Co ważne, priorytetem było poleganie na metodach, których działanie – iteracyjna poprawa jakości – zostało potwierdzone w wielu przypadkach.
- D Kilka wybranych metryk zostało użytych jako funkcje celu w algorytmie EvoIM w celu dostrojenia hiperparametrów i jąder splotu metody FRSSR [20]. W tym przypadku jakościowo badane jest, w jaki sposób zastosowana metryka podobieństwa wpływa na wynik rekonstrukcji.

Badane miary podobieństwa

W badaniu uwzględniono następujące istniejące referencyjne miary podobieństwa: IFC [135], PSNR [78], SSIM [81] oraz VIF [82]. Ponadto zbadano bezreferencyjną miarę jakości obrazu w postaci entropii [136]. Zweryfikowano również zachowanie uprzednio wprowadzonych miar KFS, w których punkty kluczowe są wykrywane przy użyciu detektorów BRISK [132], SIFT [131] i SURF [133].

Badanie korelacji miar z oceną obserwatorów ludzkich

Zbiór danych LIVE [110], opisany w rozdziale 2.3.4, zawiera obrazy zdegradowane w określony sposób, dla każdego z których podane są wartości DMOS (ang. *differential mean opinion score*). Są to liczby wynikające z oceny ludzkich obserwatorów, opracowane w ten sposób, by wyższa wartość oznaczała lepszą jakość obrazu. Oficjalny protokół dla zbioru danych LIVE, zastosowany w dalszej części pracy, polega na sprawdzeniu, w jaki sposób poszczególne miary podobieństwa między oryginalnym obrazem a zdegradowanymi obrazami korelują z DMOS. W tabeli 4.4 pokazane są współczynniki korelacji Spearmana między DMOS a badanymi miarami, uzyskane dla podzbioru zniekształceń JPEG 175 zdegradowanych obrazów.

Badanie wrażliwości miar na degradację obrazów

Aby sprawdzić wrażliwość miar podobieństwa na różne operacje degradujące oryginalny obraz, zestaw 13 scen (każda o wymiarach około 1000×1000 pikseli) ze zbioru danych B4MultiSR, obejmujący dwa obrazy naturalne i 11 obrazów satelitarnych. Dla każdego obrazu z zestawu oryginalny obraz został zdegradowany przez zastosowanie pięciu różnych scenariuszy degradacji:

- a) obrazy poddane modyfikacji jasności – 256-poziomowe wartości skali szarości są przesunięte o stałą wartość w zakresie $\langle -50; 50 \rangle$ (rozważanych jest 12 poziomów w tym zakresie, aby uzyskać 12 zdegradowanych obrazów z każdego obrazu wejściowego); jeśli zaktualizowana wartość piksela spadnie poniżej 0 lub przekroczy 255, zostanie ona obcięta odpowiednio do 0 lub 255,
- b) obrazy poddane modyfikacji kontrastu – modyfikowane są wartości pikseli, przez mnożenie ich przez współczynnik z zakresu $\langle 0, 5; 1, 5 \rangle$ (rozważa się 10 poziomów degradacji z tego zakresu),
- c) obrazy rozmyte przy użyciu filtra gaussowskiego – przy użyciu splotu z maską Gaussa o odchyleniu standardowym $\sigma \in \langle 0, 3; 10 \rangle$ (rozważa się 13 wartości σ z tego zakresu),

- d) obrazy zanieczyszczone szumem gaussowskim $\mu = 0$ i $\sigma \in \langle 1; 50 \rangle$ (rozważanych jest 9 wartości σ),
- e) obrazy zdegradowane przez następującą operację: zmniejszanie rozdzielczości o współczynnik $s \in \langle 0, 3; 10 \rangle$ i ponowne powiększenie o ten sam współczynnik do oryginalnych wymiarów przy użyciu interpolacji Lanczosa (rozważa się 16 wartości s z tego zakresu).

Przykłady obrazów zdegradowanych na opisane wcześniej sposoby są pokazane na rys. 4.5. Zasadniczo miara stosowana do oceny wyniku SR powinna być odporna na przekształcenia (a) i (b), które nie wpływają na szczegóły obrazu, będąc jednocześnie wrażliwą na operacje, które wpływają na szczegóły (c-e).

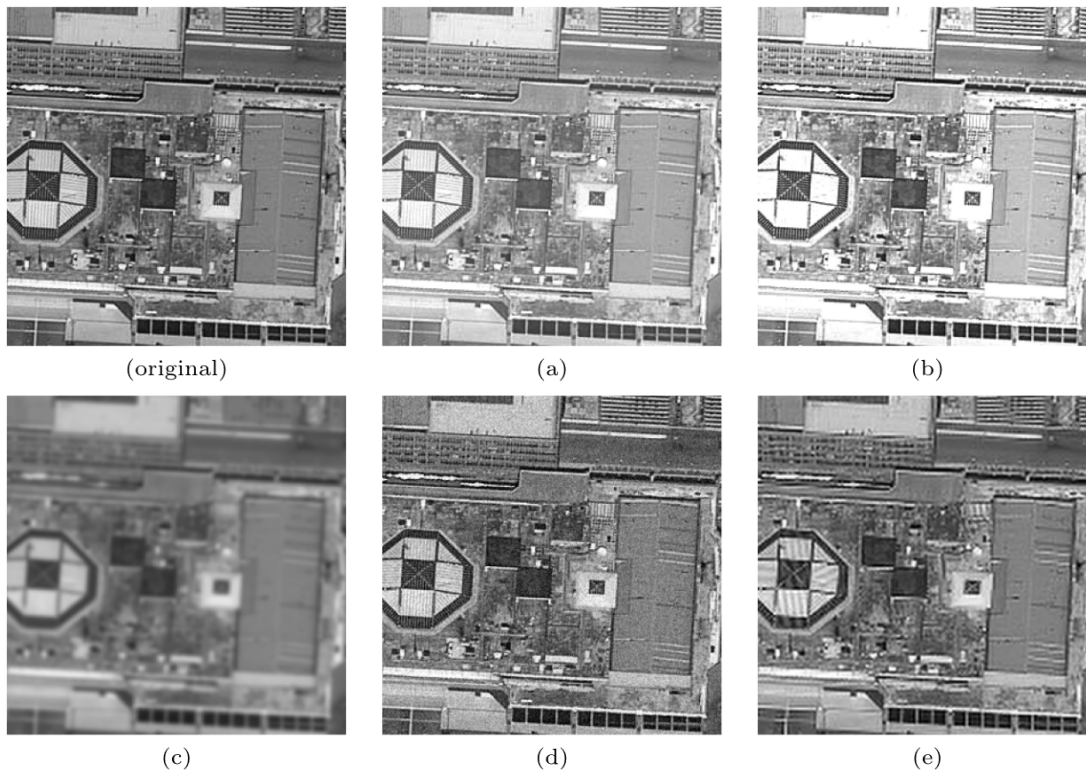
W celu ilościowego określenia wrażliwości metryki podobieństwa na określony scenariusz degradacji, tworzony jest zestaw wyników podobieństwa obliczonych między oryginalnym obrazem a jego zdegradowanymi odpowiednikami, przetwarzanymi z różnymi poziomami degradacji. Dla takiego zestawu miarą zmienności jest względne odchylenie standardowe (ang. *relative standard deviation*, $RSD = \frac{\sigma}{\mu}$, gdzie σ jest odchyleniem standardowym wartości podobieństwa, a μ jest ich średnią wartością). Na podstawie wszystkich serii stopni degradacji w zbiorze danych, obliczana jest średnia RSD (ang. *average relative standard deviation*, $ARSD$), która wskazuje wrażliwość miary na zastosowaną procedurę degradacji. Interpretacja tych wartości jest następująca: jeśli miara nie jest w ogóle wrażliwa w danej kategorii, to wszystkie wartości w zbiorze są takie same, stąd $ARSD = 0$). W przypadku miary jakości obrazu, która nie porównuje wyniku z obrazem referencyjnym (entropia), wartości metryki zostały wzięte bezpośrednio w celu otrzymania $ARSD$.

Pożądane jest, żeby miara używana do oceny rezultatów algorytmów SR była:

- 1) odporna na zmiany związane z histogramem, czyli jasność i kontrast (a) oraz (b), cechy niskiej częstotliwości; stąd wartość $ARSD$ powinna być jak najmniejsza;
- 2) wrażliwa na operacje zmieniające cechy wysokiej częstotliwości - z grup (c), (d) i (e), które wpływają na poziom szczegółowości, dlatego wartości $ARSD$ powinny być stosunkowo duże.

Walidacja za pomocą iteracyjnych metod rekonstrukcji nadrozdzielczej

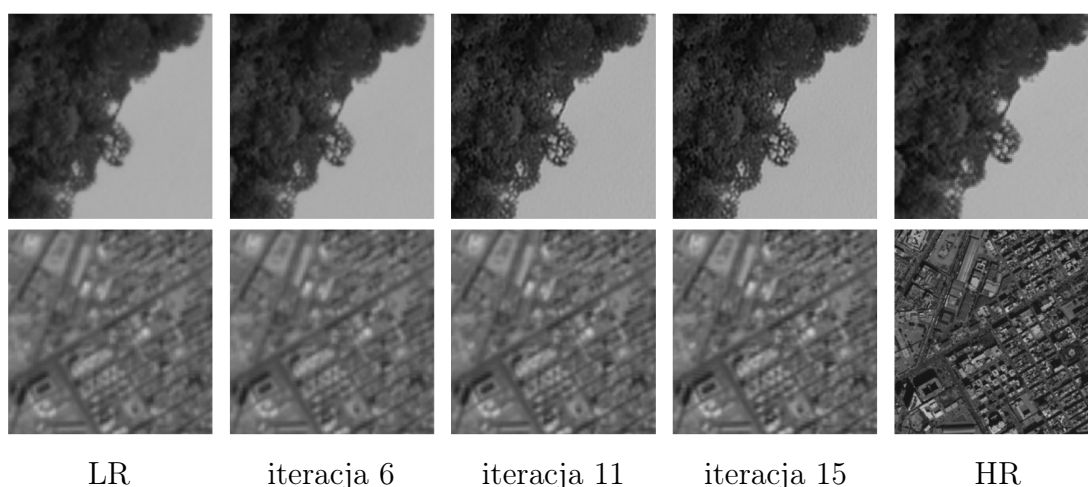
W tej części badań sprawdzane było jak miary zachowują się dla uznanych metod rekonstrukcji nadrozdzielczej GPA [20] oraz IBP [58]. Są to algorytmy, które wykorzystują iteracyjne poprawianie wyglądu obrazów, gdzie z każdą iteracją jakość rekonstrukcji rośnie. Dokonano zatem rekonstrukcji trzech scen ze zbioru danych



Rysunek 4.5: Przykłady degradacji obrazów w badaniu czułości metryk: obraz oryginalny oraz a) zmiana jasności, b) zmiana kontrastu, c) rozmycie, d) szum, e) degradacja wynikająca z interpolacji

B4MultiSR [122]. W przypadku obu algorytmów SR wynik rekonstrukcji jest poprawiany w kolejnych iteracjach, dając wizualnie lepsze wyniki, co można zobaczyć na rys. 4.6. Celem było sprawdzenie, czy ta obserwacja jest poprawnie odzwierciedlana przez wartości miar podobieństwa – powinny one dawać wyniki wskazujące na lepszą jakość rekonstrukcji w kolejnych iteracjach. Wzięto pod uwagę typowe wartości liczby iteracji: pierwszych 15 w przypadku GPA i 6 dla IBP. Dla każdej z iteracji obliczono podobieństwo między wynikiem rekonstrukcji a obrazem referencyjnym. Wszystkie testowane miary poprawnie wykazały poprawę jakości w kolejnych iteracjach, jednak ich zmienność była różna. Obliczono wyniki ARSD, które przedstawiono w tabeli 4.4. Można zauważyć, że miara KFS_{SIFT} w sposób najbardziej wyraźny pokazuje poprawę jakości. Z kolei IFC i VIF wykazują tutaj mniejszą zmienność wartości.

W tabeli 4.4 zaprezentowano wyniki opisanych eksperymentów. Dla oceny przydatności miar w rekonstrukcji nadrozdzielczej w ramach każdej z kategorii de-



Rysunek 4.6: Wyniki rekonstrukcji nadrozdzielczej w badaniu czułości metryk – przykładowe obrazy z wybranych iteracji algorytmu GPA [20]. Obrazy LR przeskalowane do wymiarów HR za pomocą interpolacji metodą najbliższego sąsiada. Widoczne powinno być stopniowe wyostanie szczegółów na obrazach wraz z postępem iteracji.

gradacji utworzono ranking. Utworzono również zbiorczy ranking – podsumowano rangi w każdej z kategorii, tak by ocenić które miary otrzymały najlepsze wyniki. Miara KFS_{SIFT} okazała się konkurencyjna względem ugruntowanej w literaturze miary VIF, szczególnie uzyskując znacznie większą czułość w zakresie degradacji dotyczących wysokich częstotliwości obrazu. Pokazuje to jej potencjał i przydatność w zastosowaniu do oceny rezultatów metod SR, w których główny nacisk powinien być położony na porównywanie wysokoczęstotliwościowych cech obrazu. Miary o takiej charakterystyce zostały wybrane w trakcie walidacji eksperymentalnej jako funkcja straty w treningu metod rekonstrukcji nadrozdzielczej opartych na głębokich sieciach konwolucyjnych. Przedstawione w bieżącym rozdziale wyniki stanowią zatem przyczynek do zaprezentowanych dalej badań skupiających się na użyciu miary opartej na wykrywaniu punktów charakterystycznych w rekonstrukcji nadrozdzielczej – co jest **elementem prac nad potwierdzeniem tezy T2** niniejszej rozprawy.

Tabela 4.4: Wyniki analizy rozpatrywanych miar: korelacja z DMOS z bazy LIVE, średnie względne odchylenie standardowe (ARSD) – miara zmienności wartości metryk dla różnych stopni degradacji. Pożądane są niewielkie wartości w przypadku: jasności i kontrastu (odporność na zmiany w charakterystyce niskich częstotliwości) oraz jak największe w pozostałych wypadkach (czułość na zmiany w charakterystyce częstotliwości wysokich). Pokazane są wyniki rankingu miar w ramach kategorii oraz rankingu całościowego dla sumy rang w poszczególnych kategoriach. Trzy najlepsze wyniki są pogrubione, a najlepszy wynik w każdej kolumnie jest podkreślony.

Miara	DMOS		Niskie Częst. (ARSD)			Wysokie Częst. (ARSD)			SR (ARSD)			Ranking					
	Korelacja	#	Jasność	#	Kontrast	#	Roznycie	#	Szum	#	Interp.		#	GPA	#	IBP	#
Entropia	0,675	8	0,084	<u>1</u>	0,065	<u>1</u>	0,036	8	0,023	8	0,001	8	0,006	8	0,003	6	7
IFC	0,883	6	0,174	3	0,074	2	0,815	5	0,434	6	0,584	4	0,029	5	0,008	5	5
PSNR	0,839	7	0,32	5	0,241	5	0,452	6	0,364	7	0,204	6	0,022	6	0,002	8	8
SSIM	0,901	3	0,194	4	0,099	3	0,321	7	0,467	5	0,167	7	0,018	7	0,017	4	6
VIF	0,905	<u>1</u>	0,138	2	0,154	4	0,823	4	0,661	4	0,464	5	0,059	3	0,020	3	2
KFS _{BRISK}	0,895	5	1,787	7	0,503	8	2,953	2	1,118	2	0,912	2	0,059	3	0,003	6	4
KFS _{SIFT}	0,905	<u>1</u>	2,047	8	0,358	7	3,038	<u>1</u>	1,253	<u>1</u>	1,095	<u>1</u>	0,074	<u>1</u>	0,022	<u>1</u>	<u>1</u>
KFS _{SURF}	0,899	4	1,618	6	0,316	6	2,915	3	0,895	3	0,739	3	0,069	2	0,022	<u>1</u>	3

4.3 Sieci głębokie z funkcją straty opartą na wykrywaniu punktów charakterystycznych

4.3.1 Dobór funkcji straty

Bazując na wynikach przedstawionych w rozdziale 4.2, gdzie pokazano potencjalną przydatność metryk opartych na wykrywaniu punktów charakterystycznych w rekonstrukcji nadrozdzielczej, podjęto próbę zastosowania opracowanych miar podobieństwa do treningu sieci głębokiej realizującej zadanie MISR. Badanie takie było ukierunkowane na weryfikację tezy T2 niniejszej rozprawy. Jako sieć MISR została użyta metoda PIUnet [69], która była jednym z wiodących rozwiązań w czasie pisania niniejszej rozprawy. Trening PIUnet oryginalnie używa metryki L1. W niniejszym rozdziale zostało opisane jak zastosowanie miar opartych na wykrywaniu punktów charakterystycznych wpływa na jakość rekonstrukcji tej metody. Zostało sprawdzone, czy dla rzeczywistych danych, gdzie obrazy LR i HR pochodzą z różnych instrumentów satelitarnych, przy zastosowaniu funkcji straty opartej na sieci głębokiej, da się uzyskać wyniki ilościowo i jakościowo lepsze od tych generowanych przy użyciu oryginalnego schematu treningu.

Zostało przebadane zastosowanie następujących funkcji straty:

1. Oryginalna funkcja straty PIUnet oparta na metryce L1.
2. cPSNR – miara oparta na metryce L2, niwelująca przesunięcia jasności między obrazami [26].
3. LPIPS w wersji oryginalnej [93], nieużywającej informacji o punktach charakterystycznych obrazu.
4. DKFS oraz $DKFS_{MIX}$ – w celu sprawdzenia czy pośrednie użycie miar opartych na punktach charakterystycznych polepsza jakość rekonstrukcji w stosunku do innych miar, w szczególności LPIPS.
5. Sekwencje trenowania sieci w pierwszym etapie za pomocą DKFS albo $DKFS_{MIX}$ a następnie L1, bądź też w pierwszym etapie za pomocą L1 a następnie DKFS albo $DKFS_{MIX}$ – opisane w rozdziale 3.7. Dotrenowywanie następuje dla całkowicie wytrenowanych modeli przez liczbę epok taką samą jak w treningu od podstaw.
6. Złożone funkcje straty wykorzystujące zarówno L1 jak i LPIPS, DKFS lub $DKFS_{MIX}$ w postaci wspomnianej w równaniu 3.5, z użyciem wag o wartości 1.

Dla porównania zbadano również wyniki uzyskane za pomocą uśrednienia obrazów LR powiększonych interpolacją dwusześcienną. Do tego dołączone są wyniki otrzymane za pomocą oryginalnych modeli PIUnet – wytrenowanych przez autorów na danych PROBA-V. Eksperymenty zostały zrealizowane przy pomocy środowiska Python / PyTorch na komputerze z procesorem graficznym RTX 1070 Ti z 11 GB VRAM.

Trening metod oceny jakości opartych na cechach obrazu

Jak opisano w rozdziale 3.2.2, metody DKFS oraz $DKFS_{MIX}$ są oparte na architekturze LPIPS [93]. Wykorzystano też oryginalną implementację treningu LPIPS, zmieniając jednak dane używane jako etykiety w opisany dalej sposób. Tak wytrenowane modele zostały użyte następnie w dalszych badaniach.

1. Dla miary DKFS jako etykieta zostaje podana wyłącznie wartość \mathcal{V} otrzymana za pomocą miary KFS_{SIFT} według wyrażenia 3.4.
2. Dla miary $DKFS_{MIX}$ do dwóch głosów obserwatorów $\{\mathcal{H}_1, \mathcal{H}_2\}$ pochodzących z oryginalnych danych LPIPS zostaje dodany głos \mathcal{V} otrzymany za pomocą miary KFS_{SIFT} według wyrażenia 3.4. Podawana do treningu etykieta zostaje wyliczona według wzoru: $(\mathcal{H}_1 + \mathcal{H}_2 + \mathcal{V})/3$.

Przygotowanie danych

Dane, które są podawane do modelu PIUnet są wstępnie przetwarzane w sposób opisany w pseudokodzie jako alg. 3, tak by zapewnić na wejściu sieci dane o wartościach z zakresu wymaganego w implementacji treningu PIUnet. Sposób przygotowywania danych jest następujący:

1. Obrazy są rejestrowane w dokładności pełnopikselowej – do pierwszego obrazu w serii obrazów LR.
2. Dokonywana jest transformacja obrazów LR za pomocą standaryzacji: $\mathcal{I}^{(l)} = (\mathcal{I}^{(l)} - \mu_{LR}) / \sigma_{LR}$, gdzie μ_{LR} to średnia wartość pikseli dla wszystkich obrazów LR, σ_{LR} oznacza zaś odchylenie standardowe wartości pikseli we wszystkich obrazach z tego zbioru.
3. Podobnie przekształcane są obrazy HR używane jako wartości referencyjne w treningu: $\mathcal{I}^{(h)} = (\mathcal{I}^{(h)} - \mu_{HR}) / \sigma_{HR}$.

Trening sieci PIUnet

Do treningu sieci użyto zbioru MuS2 [27]. Dane HR oraz LR pochodzące z dwóch instrumentów o różnych charakterystykach powodują większą trudność w trenowaniu modeli SR [128]. MuS2 zawiera takie właśnie dane, co odróżnia go od danych sztucznie degradowanych, ale także od zbioru PROBA-V. Ma to duże znaczenie w sposobie działania metryk [24], [27]. Do testów użyto pasma bliskiej podczerwieni (ang. *near infrared*, NIR), cechującego się wysoką zmiennością czasową z powodu różnic w wyglądzie roślinności na powierzchni Ziemi, również między obserwacjami LR, co dodatkowo wpływa na trudność w adaptacji modelu SR, uwydatniając słabości klasycznych metryk. Dla satelity Sentinel-2 jest to pasmo B08 (NIR), dla WorldView-2 brane jest pod uwagę pasmo NIR1. Rozkład pasm jest pokazany na rys. 4.7. Jak widać, pasma NIR mają najszersze zakresy, co skutkuje dużą ilością informacji zawartą w obrazach, są też nieco przesunięte względem siebie w sensie spektralnym.

Podział na podzbiory: treningowy, walidacyjny i testowy dla MuS2 wygląda następująco:

1. Podział obrazów na nienachodzące na siebie fragmenty o rozmiarze 40×40 pikseli w skali LR i 120×120 pikseli w skali HR.
2. 50% fragmentów losowo wybieranych do podzbioru treningowego.
3. 10% wszystkich fragmentów tworzy podzbiór walidacyjny.

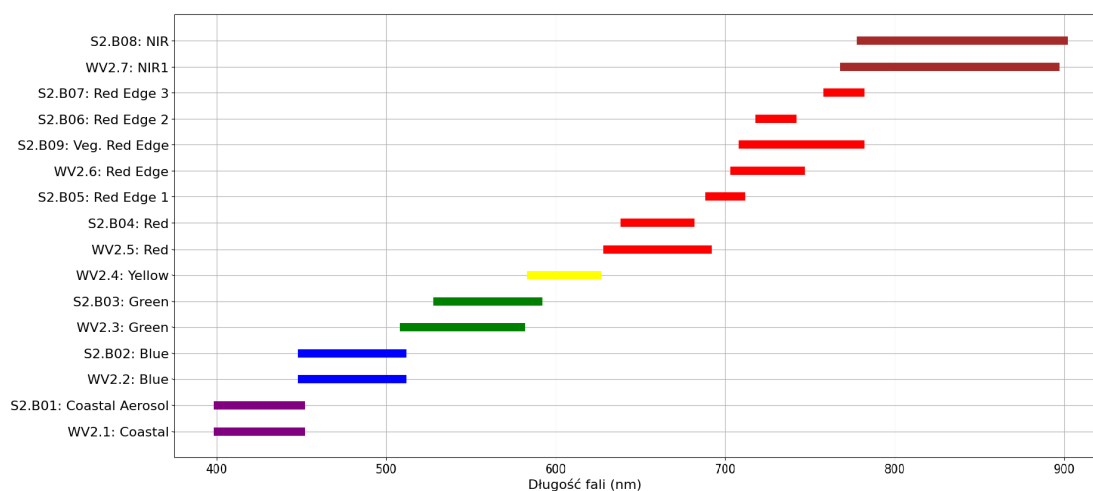
Algorytm 3 Wstępne przygotowanie danych

Require: *katalog*: bazowy katalog z danymi**Require:** *pasmo*: pasmo spektralne

```

1:
2: function ODCZYTAJZBIÓR(katalog, pasmo)
3:   imgsets ← sorted list of image directories in base_dir
4:   X ← []; y ← []
5:   for scena in katalog do
6:     LR ← []
7:     for img in scena do
8:       LR.append(odczytaj(img))
9:     end for
10:    X.append(LR)
11:    y.append(odczytaj(HR))
12:  end for
13:  return X, y
14: end function
15:
16: function ZAREJESTRUJSCENĘ(scena)
17:   ref ← scena[0]
18:   imgset_reg ← []
19:   for i in (0..len(X)) do
20:     x ← i-ty element of scena
21:     s, PHASECROSSCORRELATION(ref, x)
22:     x ← TRANSLACJA(x, s)
23:     imgset_reg[..., i] ← x
24:   end for
25:   return imgset_reg
26: end function
27:
28: function ZAREJESTRUJZBIÓR(X)
29:   X_reg ← []
30:   for i in range(length(X)) do
31:     img_reg ← ZAREJESTRUJSCENĘ(X[i])
32:     X_reg.append(img_reg)
33:   end for
34:   return X_reg
35: end function
36:
37: x, y ← ODCZYTAJZBIÓR("katalog", "b8")
38: x ← ZAREJESTRUJZBIÓR(x)
39: zapisz(x)
40: zapisz(y)

```



Rysunek 4.7: Pasma spektralne satelitów Sentinel-2 (S2) oraz WorldView-2 (WV2). W eksperymentach zostały użyte pasma S2.B08 oraz WV2.NIR

4. Pozostałe fragmenty (40%) stanowią podzbiór testowy².

Daje to w wyniku 18956 fragmentów, z których 9478 stanowią podzbiór treningowy, 1895 - walidacyjny, a pozostałe 7583 - testowy. Każdy fragment LR jest w postaci 14 obserwacji pochodzących z obrazów pozyskanych z Sentinel-2 w różnym czasie.

Wybór stosunkowo niewielkiego rozmiaru fragmentu został podyktowany możliwościami sprzętu, na którym były realizowane eksperymenty (RTX 1070 Ti z 11 GB VRAM) i był kompromisem między wykonalnością a maksymalizacją rozmiaru fragmentu. Należy wspomnieć, że w oryginalnym treningu PIUnet używane są fragmenty o rozmiarze 32×32 pikseli. Rozmiar fragmentu jest o tyle ważny, że zwykle na brzegach powstają artefakty. W trakcie oceny jakości piksele z krawędzi obrazów są zatem ignorowane [26], jednak lepiej jest maksymalizować stosunek obszaru bez artefaktów do tego, który musi zostać odrzucony.

Każdy trening trwał 100 epok, z możliwością zakończenia w przypadku 10 następujących po sobie epok bez poprawy wartości funkcji straty na zbiorze walidacyjnym. Sposób treningu opisany jest w pseudokodzie jako alg. 4, a można go opisać następująco: na wejściu wymagane są: liczba epok, model, funkcja straty, dane treningowe i walidacyjne oraz warunek zakończenia. W trakcie treningu w każdej epoce przetwarzane są całe dane treningowe. Model jest wtedy ustawiony w tryb treningu, czyli aktualizowane są wagi sieci. Po przetworzeniu każdego wsadu (ang. *batch*) obliczana jest funkcja straty i wykonywana jest

²stosunkowo duża proporcja jest podyktowana niewielkim rozmiarem całego zbioru MuS2, tak by dane testowe były reprezentatywne

Algorytm 4 Trening sieci głębokiej MISR

Require: *epoch_count*: maksymalna liczba epok
Require: *loss*: funkcja straty
Require: *model*: model sieci głębokiej
Require: *train_loader*: dane treningowe
Require: *validation_loader*: dane walidacyjne
Require: *ending_condition*: warunek zakończenia

```

1:
2: for epoka = 0; epoka < epoch_count; epoka+ = 1 do
3:   model.train()                                ▷ ustaw model w tryb treningu
4:   for batch in train_loader do                ▷ pętla treningu:
5:     train_result ← model(batch)
6:     train_loss ← loss(batch)
7:     loss.backward()                            ▷ propagacja wsteczna funkcji straty
8:   end for
9:
10:  model.eval()                                  ▷ ustaw model w tryb ewaluacji
11:  for batch in validation_loader do           ▷ pętla walidacji:
12:    val_result ← model(batch)
13:    val_loss ← loss(batch)
14:    if val_loss < best_val_loss then save_best(model)
15:    if ending_condition(val_loss) then return
16:  end for
17:  save(model)
18: end for

```

wsteczna propagacja, która aktualizuje wagi sieci. Po zakończeniu przetwarzania zbioru treningowego sprawdzana jest wartość funkcji straty na zbiorze walidacyjnym. Podobnie – przetwarzany jest on wsadowo, a dla wyniku działania modelu na każdym wsadzie obliczana jest wartość funkcji straty. Jeśli jest ona lepsza niż poprzednia najlepsza, taki model jest zapisywany. Dodatkowo sprawdzane jest czy nie wystąpił warunek wcześniejszego zakończenia. W takim wypadku trening jest kończony przed osiągnięciem maksymalnej liczby epok.

Pojedyncze wykonanie dla pojedynczej ewaluowanej funkcji straty trwało przeciętnie około 48 godzin.

Ocena rekonstrukcji na zbiorze testowym MuS2

Po wytrenowaniu modeli wykonana została ocena ich jakości na zbiorze testowym, zawierającym fragmenty nieużywane w trakcie treningu. Dla każdego zostały ob-

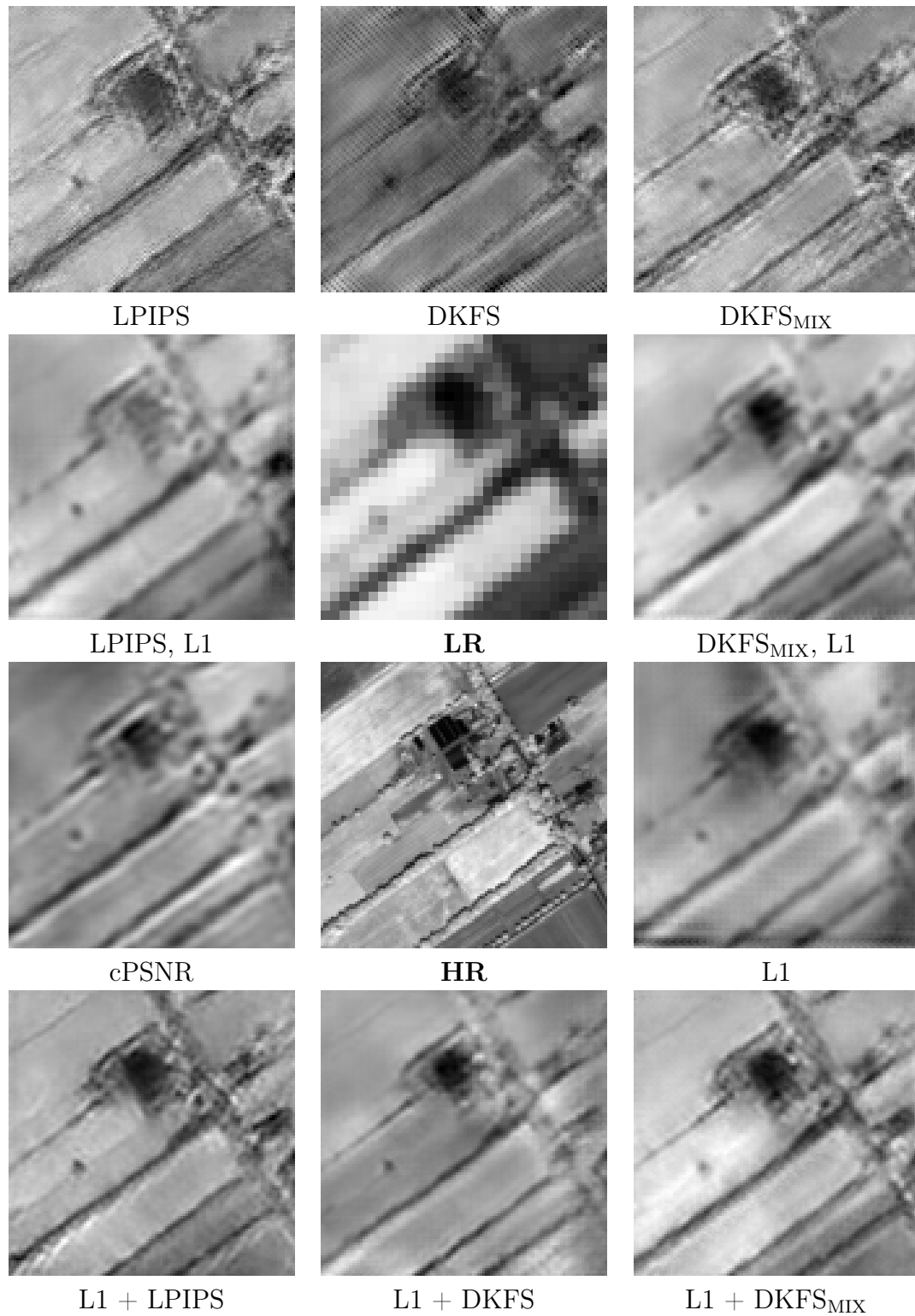
liczone miary: cPSNR, cSSIM, LPIPS, DKFS oraz $DKFS_{MIX}$. Trzy pierwsze są miarami najczęściej wykorzystywanymi w literaturze. Wartości DKFS nie są wykorzystywane podczas dalszych analiz – założono, że wszelkie znaczące pomiary przedstawione w tej sekcji będą realizowane przy użyciu metryk występujących w literaturze.

4.3.2 Wyniki dla zbioru MuS2

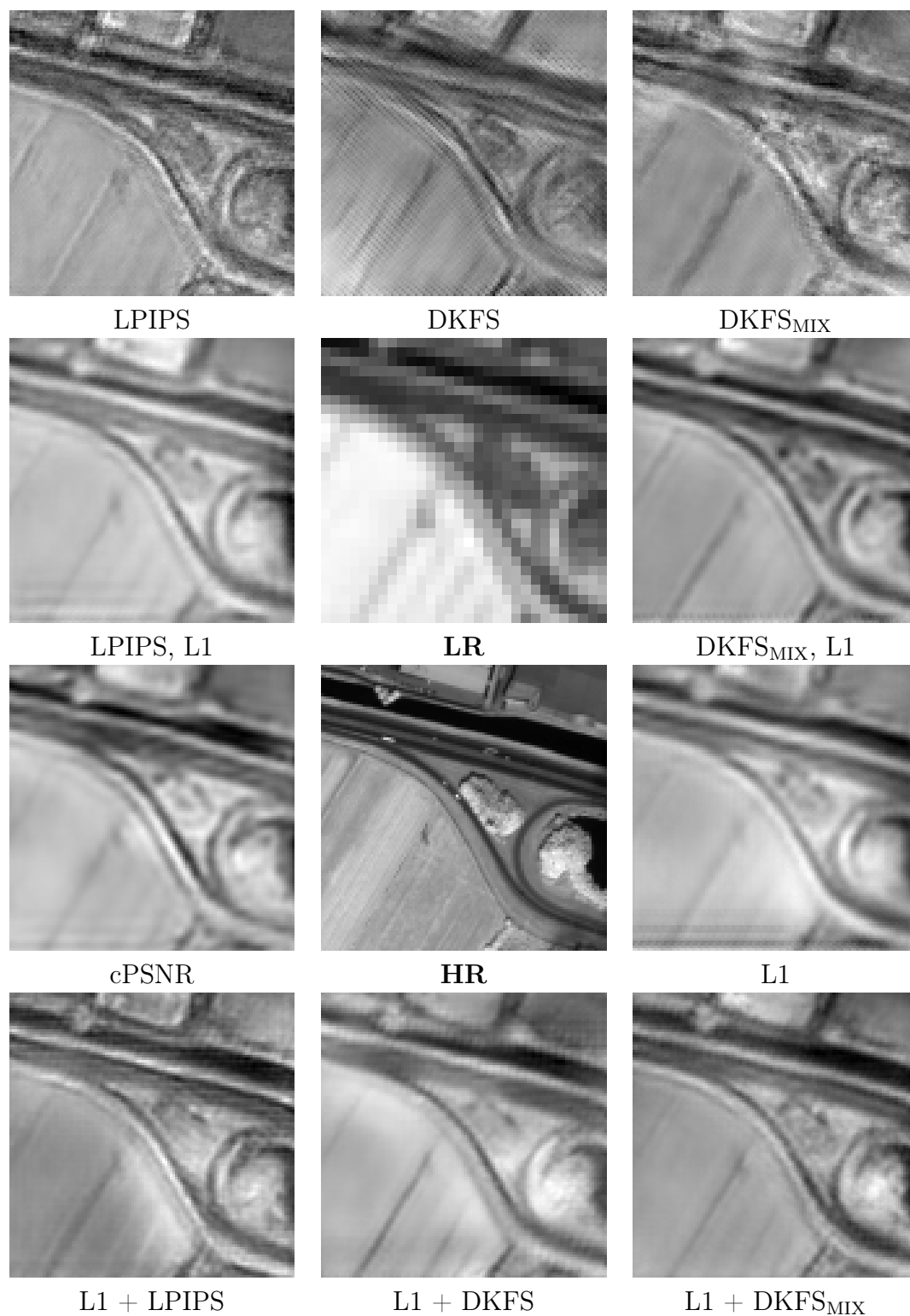
Najważniejszy pod względem rozpatrywanej tezy rozprawy jest eksperyment na zbiorze MuS2. Przykładowe wizualizacje dla modeli wytrenowanych z użyciem różnych funkcji straty pokazane są na rysunkach 4.8- 4.11. Można zauważyć, że użycie jako funkcji straty LPIPS, DKFS oraz $DKFS_{MIX}$, oprócz wyostrenienia widocznych szczegółów, generuje w obrazie spore natężenie artefaktów. Dotrenowanie takich modeli z użyciem funkcji straty L1 nieco ogranicza zawartość artefaktów. Z kolei użycie funkcji straty L1 oraz cPSNR powoduje brak ostrości rekonstrukcji. Najbardziej zbalansowane między natężeniem artefaktów a widocznością rozróżnialnych szczegółów wydają się obrazy uzyskane za pomocą modeli wytrenowanych z użyciem funkcji straty L1+LPIPSi L1+ $DKFS_{MIX}$.

W tabeli 4.5 zebrano wyniki dla tych danych. W ramach każdej miary wyróżniono trzy najlepsze modele. Na podstawie tych wyników można zauważyć, że wartości wszystkich metryk cPSNR, cSSIM i LPIPS są lepsze dla modeli wytrenowanych za pomocą funkcji straty opartych na LPIPS i DKFS, a dla modeli wytrenowanych w oparciu o miary DKFS wyniki dla metryk cPSNR oraz cSSIM są najlepsze. Ciekawy jest fakt, że modele te uzyskują wartości cPSNR wyższe niż model wytrenowany bezpośrednio z funkcją straty cPSNR. Równocześnie przeprowadzono testy statystyczne t-Studenta na wszystkich możliwych kombinacjach (*model, metryka oceniająca jakość rekonstrukcji*). Testy zostały wykonane dla wartości miar cPSNR, cSSIM oraz LPIPS, a ich wyniki są zaprezentowane w tabeli 4.6, 4.7 oraz 4.8. Pokazują one, że w wielu wypadkach funkcji straty innych niż L1 polepszenie jakości wyników w stosunku do modelu wytrenowanego za pomocą L1 jest statystycznie istotnie. Zaś w relacji do modelu uzyskanego za pomocą treningu z metryką L2 (cPSNR) istotnie lepsze w sensie statystycznym są wyniki uzyskane przy użyciu modeli opartych na LPIPS i DKFS, ale tylko w testach statystycznych dla miary LPIPS.

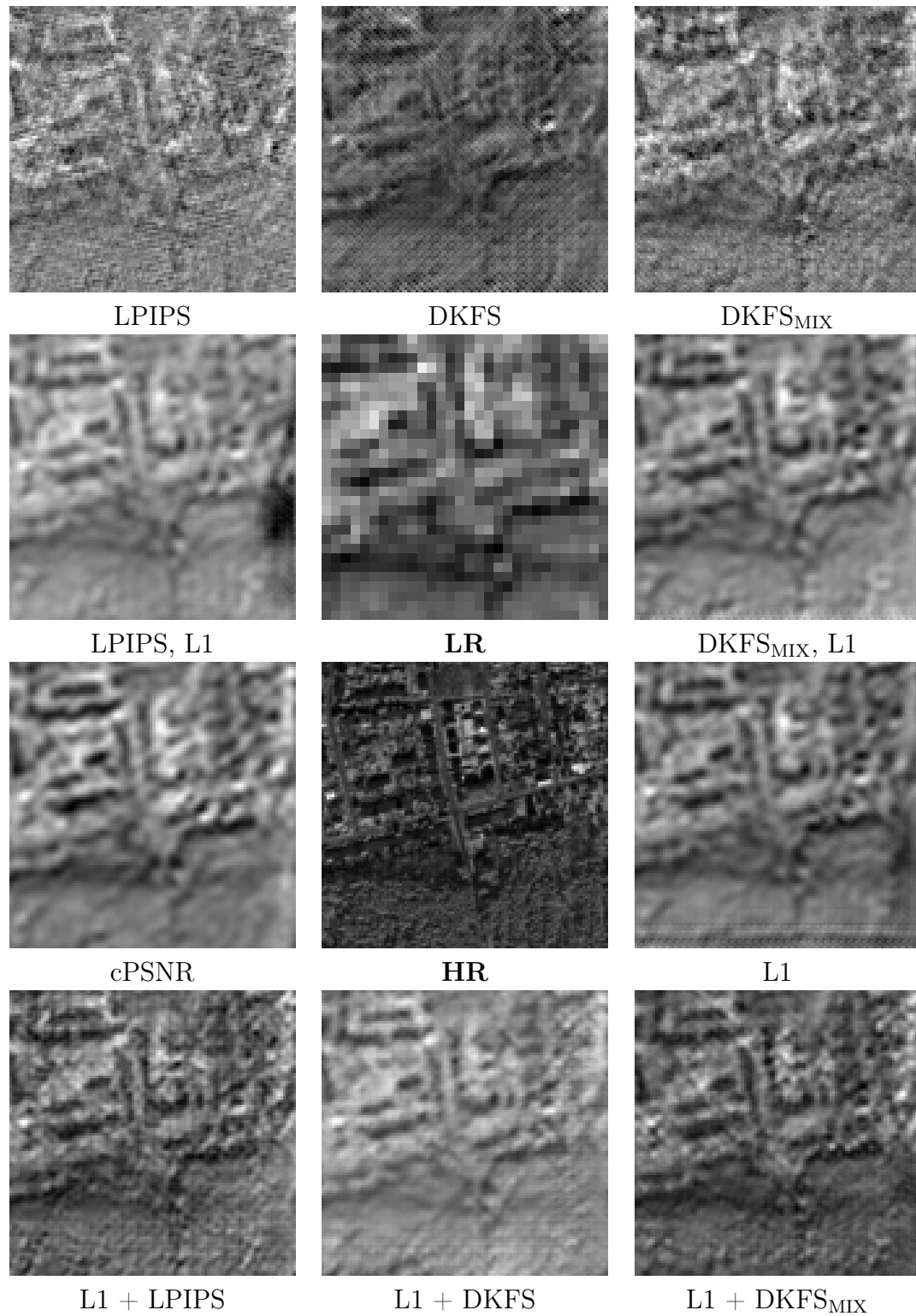
W celu wskazania modeli, które dają ogółem najlepsze wyniki dla różnych metryk zastosowano podejście oparte na rankingu badanych metod w ramach miar. Wzięto pod uwagę miary cPSNR, cSSIM, które są klasycznymi metrykami oraz dobrze już ugruntowaną w literaturze miarę LPIPS. Wyniki tego rankingu są pokazane w tabeli 4.9. Okazało się, że najlepsze wyniki uzyskują modele używające funkcji straty zdefiniowanej jako suma metryki opartej na L1 (tak jak w oryginalnym treningu PIUnet) oraz LPIPS albo dwóch autorskich miar DKFS bazujących



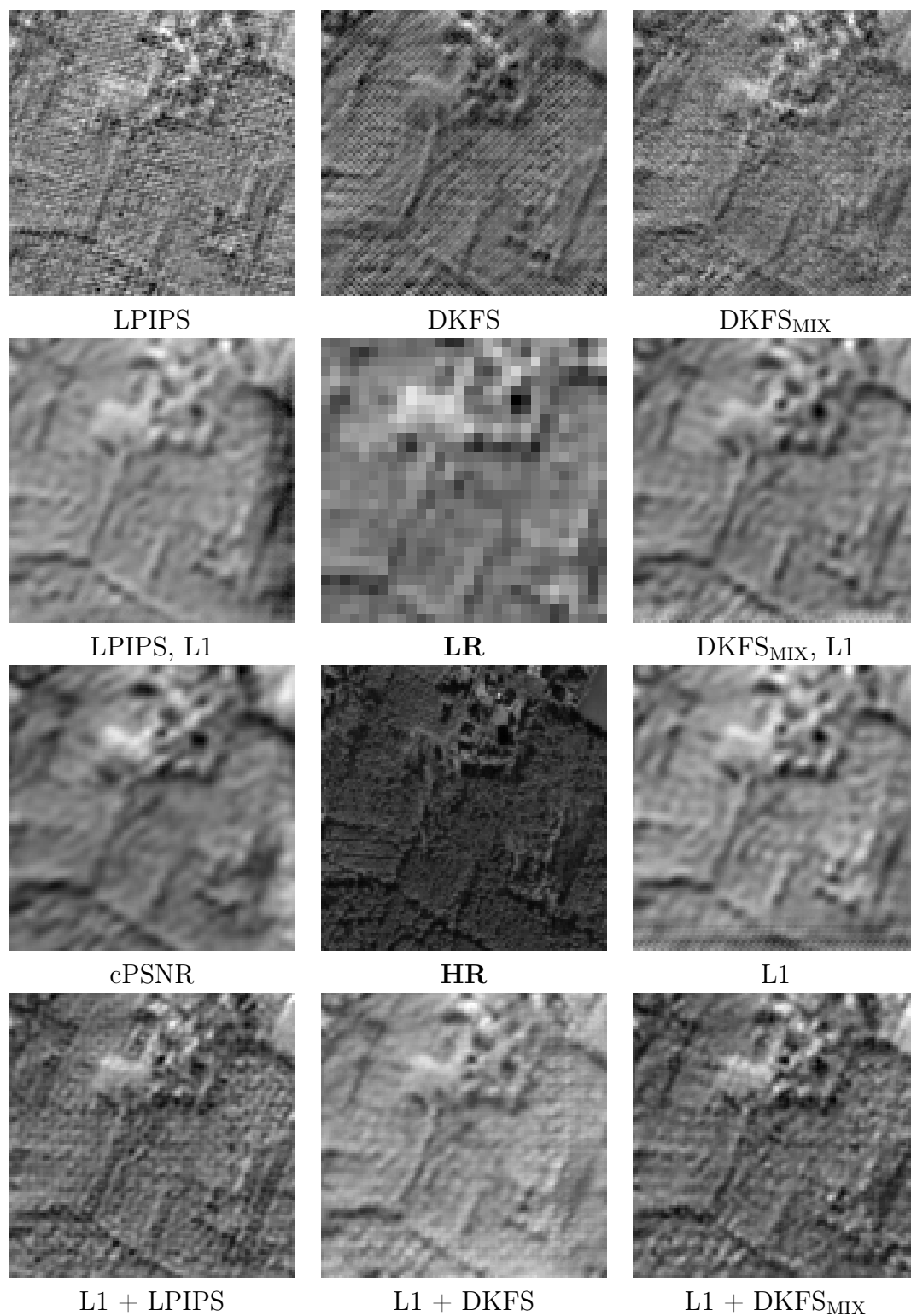
Rysunek 4.8: Przykładowe wyniki dla wybranych funkcji straty na zbiorze testowym MuS2. Dla porównania pokazano również obrazy HR i LR (przeskalowany do rozmiaru HR).



Rysunek 4.9: Przykładowe wyniki dla wybranych funkcji straty na zbiorze testowym MuS2. Dla porównania pokazano również obrazy HR i LR (przeskalowany do rozmiaru HR).



Rysunek 4.10: Przykładowe wyniki dla wybranych funkcji straty na zbiorze testowym MuS2. Dla porównania pokazano również obrazy HR i LR (przeskalowany do rozmiaru HR).



Rysunek 4.11: Przykładowe wyniki dla wybranych funkcji straty na zbiorze testowym MuS2. Dla porównania pokazano również obrazy HR i LR (przeskalowany do rozmiaru HR).

Tabela 4.5: Wyniki działania modeli na zbiorze MuS2. Kolumny oznaczone znakiem # oznaczają wyniki rankingu w ramach metryki. Pierwsze trzy miejsca są pogrubione. Najlepszy wynik został podkreślony.

		cPSNR	#	cSSIM	#	LPIPS	#	DKFS	#	DKFS _{MIX}	#
Interp.	NN	26,0619	10	0,5572	13	0,4928	16	0,3097	17	0,3924	17
	Interp	26,1199	8	0,5654	11	0,5705	18	0,3136	19	0,4197	18
	bilinear	26,1485	7	0,5593	12	0,589	19	0,3115	18	0,4303	19
PIUnet - funkcja straty	cPSNR	26,6476	3	0,6053	4	0,5016	17	0,282	11	0,373	15
	L1	25,9267	11	0,598	7	0,4635	14	0,2813	10	0,3595	13
	LPIPS	24,3555	17	0,4667	17	0,2102	3	0,3056	15	0,2532	6
	DKFS	22,7129	18	0,4618	18	0,3923	7	0,2279	1	0,2802	7
	DKFS _{MIX}	25,1566	15	0,5281	15	0,2137	4	0,2731	7	0,2386	3
	L1,LPIPS	24,9298	16	0,5238	16	0,2030	1	0,277	9	0,2394	4
	L1,DKFS	22,4422	19	0,3777	19	0,4355	9	0,2295	2	0,2962	8
	L1,DKFS _{MIX}	25,1594	14	0,5419	14	0,2071	2	0,2602	5	0,2304	1
	DKFS,L1	26,3023	6	0,601	5	0,4409	10	0,2739	8	0,3428	10
	DKFS _{MIX} ,L1	26,4656	5	0,6129	1	0,4621	13	0,2836	13	0,3486	11
	LPIPS,L1	25,3964	13	0,5941	8	0,4542	11	0,2834	12	0,3559	12
	L1+LPIPS	26,7336	2	0,5988	6	0,2316	5	0,2383	4	0,2347	2
	L1+DKFS	26,5136	4	0,6116	2	0,4025	8	0,2648	6	0,3118	9
	L1+DKFS _{MIX}	26,7566	1	0,6087	3	0,2604	6	0,2378	3	0,2453	5
	L1 (PROBA-V NIR)	26,0774	9	0,5685	9	0,4606	12	0,3006	14	0,367	14
L1 (PROBA-V NIR)	25,7553	12	0,5661	10	0,4682	15	0,3059	16	0,3735	16	

na architekturze tej metryki. Modele uczone za pomocą funkcji straty L2 (cPSNR) lub L1 osiągają gorsze wyniki. Wspomniany ranking wskazuje na zasadność użycia jako funkcji straty w treningu metody LPIPS oraz autorskiej miary DKFS_{MIX}, która okazuje się lepsza od samego LPIPS. L1+DKFS_{MIX} okazuje się z kolei lepszy od L1+LPIPS.

Wyniki eksperymentów na zbiorze symulowanym Sentinel-2 oraz zestawie PROBA-V

W celu sprawdzenia zachowania otrzymanych modeli PIUnet na innych źródłach danych przetworzono dane pochodzące ze zbiorów PROBA-V (rozdział 2.3.1) oraz dane symulowane z Sentinel-2 (rozdział 2.3.2). Są to zbiory o innej charakterystyce: PROBA-V zawiera dane pochodzące z tego samego satelity, ale wykonywane przy użyciu instrumentów różniących się wyłącznie rozdzielczością (przy czym same dane w zestawie zawierają również różnice temporalne), zaś drugi ze zbiorów składa się z danych symulowanych. W tej części eksperymentu zadano pytanie jak modele wytrenowane na danych bardziej złożonych działają dla danych innego typu – jak kształtują się wyniki dla poszczególnych funkcji straty względem siebie. Wyniki dla zbioru PROBA-V są pokazane w tabeli 4.11 (PROBA-V RED) oraz tabeli 4.12 (PROBA-V NIR), dla zbioru symulowanego Sentinel-2 są zaś w tabeli 4.10.

Tabela 4.6: Testy statystyczne na zbiorze MuS2 dla wartości cPSNR uzyskanych za pomocą poszczególnych modeli. Wartości w komórkach tabeli prezentują: 1) relację między wartością średnią cPSNR dla obrazów otrzymanych za pomocą modelu z wiersza do wartości średniej dla obrazów przetworzonych modelem z kolumny; 2) wynik testu t-Studenta na wartościach cPSNR dla obrazów otrzymanych z tych modeli. Plus na drugiej pozycji oznacza, że różnica jest statystycznie istotna (p-wartość $<0,05$), minus oznacza zaś brak istotnej statystycznie różnicy. Zielonym kolorem oznaczone przypadki, gdy test wskazuje istotną poprawę.

test dla cPSNR	cPSNR	L1	DKFS _{MIX}	LPIPS	DKFS	L1,LPIPS	L1,DKFS	L1,DKFS _{MIX}	LPIPS,L1	DKFS,L1	DKFS _{MIX} ,L1	L1+LPIPS	L1+DKFS	L1+DKFS _{MIX}	L1 (PROBA-V NIR)	L1 (PROBA-V RED)
cPSNR	.	>+	>+	>+	>+	>+	>+	>+	>+	>+	>	<	>	<	>+	>+
L1	<+	.	>+	>+	>+	>+	>+	>+	>+	>+	>	<+	<+	<+	<	>
DKFS _{MIX}	<+	<+	.	>+	>+	>	>+	<	<	<+	<+	<+	<+	<+	<+	<+
LPIPS	<+	<+	<+	.	>+	<+	>+	<+	<+	<+	<+	<+	<+	<+	<+	<+
DKFS	<+	<+	<+	<+	.	<+	>	<+	<+	<+	<+	<+	<+	<+	<+	<+
L1,LPIPS	<+	<+	<	>+	.	.	>+	<	<+	<+	<+	<+	<+	<+	<+	<+
L1,DKFS	<+	<+	<+	<+	<	<+	.	<+	<+	<+	<+	<+	<+	<+	<+	<+
L1,DKFS _{MIX}	<+	<+	>	>+	>+	>	>+	.	<	<+	<+	<+	<+	<+	<+	<+
LPIPS,L1	<+	<+	>	>+	>+	>+	>+	>	.	<+	<+	<+	<+	<+	<+	<+
DKFS,L1	<+	>+	>+	>+	>+	>+	>+	>+	>+	.	<	<+	<+	<+	>	>+
DKFS _{MIX} ,L1	<	>+	>+	>+	>+	>+	>+	>+	>+	>+	>	.	<	<	<	>+
L1+LPIPS	>	>+	>+	>+	>+	>+	>+	>+	>+	>+	>	.	>	<	>+	>+
L1+DKFS	<	>+	>+	>+	>+	>+	>+	>+	>+	>	>	<	.	<	>+	>+
L1+DKFS _{MIX}	>	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>	>	.	>+	>+
L1 (PROBA-V NIR)	<+	>	>+	>+	>+	>+	>+	>+	>+	>+	<	<+	<+	<+	.	>
L1 (PROBA-V RED)	<+	<	>+	>+	>+	>+	>+	>+	>+	>+	<+	<+	<+	<+	<	.

Najlepsze okazują się modele udostępnione przez twórców PIUnet – dla zbioru PROBA-V jest to zachowanie oczekiwane, ponieważ zostały one wytrenowane na tym zbiorze. Zaskakująca okazała się wysoka pozycja interpolacji metodą najbliższego sąsiada. Przyczyną może być charakterystyka zbiorów – dane z PROBA-V są spójne między wersjami LR i HR w sensie histogramu obrazów. Również w wypadku danych symulowanych Sentinel-2 jest podobnie wysoka spójność tego typu. Zbiór MuS2 jest, jak wspomniano w rozdziale 2.3.3 [27] pod tym względem odmienny. Należy wspomnieć również, że dla danych PROBA-V NIR na trzecim miejscu znajduje się jeden z modeli opracowanych w ramach rozprawy (L1+DKFS_{MIX}), ten sam, który daje najlepsze numerycznie wyniki dla zbioru MuS2 (z którego, jak wspomniano wcześniej używane jest właśnie pasmo bliskiej podczerwieni - NIR).

Tabela 4.7: Testy statystyczne na zbiorze MuS2 dla wartości cSSIM. Zawartość tabeli zaraportowana w sposób opisany w tabeli 4.6

test dla cSSIM	cPSNR	L1	DKFS _{MIX}	LPIPS	DKFS	L1,LPIPS	L1,DKFS	L1,DKFS _{MIX}	LPIPS,L1	DKFS,L1	DKFS _{MIX} ,L1	L1+LPIPS	L1+DKFS	L1+DKFS _{MIX}	L1 (PROBA-V NIR)	L1 (PROBA-V RED)
cPSNR	.	>	>>	>>	>>	>>	>>	>>	>	>	>	>	>	>	>>	>>
L1	<	.	>>	>>	>>	>>	>>	>>	>	<	<	<	<	<	>>	>>
DKFS _{MIX}	<>	<>	.	>>	>>	>	>>	<>	<>	<>	<>	<>	<>	<>	<>	<>
LPIPS	<>	<>	<>	.	>	<>	>>	<>	<>	<>	<>	<>	<>	<>	<>	<>
DKFS	<>	<>	<>	.	>	<>	>>	<>	<>	<>	<>	<>	<>	<>	<>	<>
L1,LPIPS	<>	<>	<	>>	>>	.	>>	<>	<>	<>	<>	<>	<>	<>	<>	<>
L1,DKFS	<>	<>	<>	<>	<>	.	>>	<>	<>	<>	<>	<>	<>	<>	<>	<>
L1,DKFS _{MIX}	<>	<>	>>	>>	>>	>>	>>	.	<>	<>	<>	<>	<>	<>	<>	<>
LPIPS,L1	<	<	>>	>>	>>	>>	>>	>>	.	<	<	<	<	<	>>	>>
DKFS,L1	<	>	>>	>>	>>	>>	>>	>>	>	.	<	>	>	>	>>	>>
DKFS _{MIX} ,L1	>	>>	>>	>>	>>	>>	>>	>>	>>	>	.	>>	>	>	>>	>>
L1+LPIPS	<	>	>>	>>	>>	>>	>>	>>	>	<	<	.	<	<	>>	>>
L1+DKFS	>	>>	>>	>>	>>	>>	>>	>>	>>	>	<	>	.	>	>>	>>
L1+DKFS _{MIX}	>	>	>>	>>	>>	>>	>>	>>	>>	>	<	>	>	.	>>	>>
L1 (PROBA-V NIR)	<>	<>	>>	>>	>>	>>	>>	>>	<>	<>	<>	<>	<>	<>	.	>
L1 (PROBA-V RED)	<>	<>	>>	>>	>>	>>	>>	>>	<>	<>	<>	<>	<>	<>	<	.

4.3.3 Ocena wyników przez obserwatorów

W niniejszym rozdziale przedstawiono wyniki badania, które miało na celu dodatkową niezależną od metryk ocenę jakości przetworzonych obrazów ze zbioru MuS2. W ramach wcześniejszych eksperymentów otrzymano wyniki dla różnie wytrenowanych modeli w postaci zestawu obrazów, które poddano ocenie ilościowej za pomocą trzech popularnych metryk: cPSNR, cSSIM i LPIPS. Jednakże, zważywszy na to, że wymienione metryki uwzględniają różne aspekty jakości obrazów, celowe wydawało się sprawdzenie jak wspomniane wyniki zostaną ocenione przez obserwatorów. W tym celu zdecydowano się przeprowadzić ankietę, której uczestnicy oceniali jakość przetworzonych obrazów.

Sposób ankietowania

Dobierając zestaw metod, których wyniki miały zostać przedstawione w ankiecie, uwzględniono trzy modele dające najlepsze wyniki zgodnie z rankingiem z tabeli 4.9. Dla porównania dodano modele otrzymane za pomocą funkcji straty typowo stosowanych w literaturze (L1 i cPSNR). Ankieta została zatem przeprowadzona dla wyników pięciu modeli wytrenowanych za pomocą funkcji straty: L1+LPIPS, L1+DKFS, L1+DKFS_{MIX}, cPSNR, L1.

W celu skonstruowania miarodajnej ankiety, zastosowano metodę wymuszo-

Tabela 4.8: Testy statystyczne na zbiorze MuS2 dla wartości LPIPS. Zawartość tabeli zaraportowana w sposób opisany w tabeli 4.6 - z wyjątkiem odwrotnej relacji poprawy jakości obrazów względem wartości metryki: LPIPS wskazuje większe podobieństwo obrazów wraz ze spadkiem wartości, w przeciwieństwie do miar cPSNR, cSSIM.

test dla LPIPS	cPSNR	L1	DKFS _{MIX}	LPIPS	DKFS	L1,LPIPS	L1,DKFS	L1,DKFS _{MIX}	LPIPS,L1	DKFS,L1	DKFS _{MIX} ,L1	L1+LPIPS	L1+DKFS	L1+DKFS _{MIX}	L1 (PROBA-V NIR)	L1 (PROBA-V RED)
cPSNR	.	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+
L1	<+	.	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+
DKFS _{MIX}	<+	<+	.	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+
LPIPS	<+	<+	<+	.	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+
DKFS	<+	<+	>+	>+	.	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+
L1,LPIPS	<+	<+	<+	<+	<+	.	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+
L1,DKFS	<+	<+	>+	>+	>+	>+	.	>+	>+	>+	>+	>+	>+	>+	>+	>+
L1,DKFS _{MIX}	<+	<+	<+	<+	<+	>+	>+	.	>+	>+	>+	>+	>+	>+	>+	>+
LPIPS,L1	<+	<+	>+	>+	>+	>+	>+	>+	.	>+	>+	>+	>+	>+	>+	>+
DKFS,L1	<+	<+	>+	>+	>+	>+	>+	>+	>+	.	>+	>+	>+	>+	>+	>+
DKFS _{MIX} ,L1	<+	<+	>+	>+	>+	>+	>+	>+	>+	>+	.	>+	>+	>+	>+	>+
L1+LPIPS	<+	<+	>+	>+	>+	>+	>+	>+	>+	>+	>+	.	>+	>+	>+	>+
L1+DKFS	<+	<+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	.	>+	>+	>+
L1+DKFS _{MIX}	<+	<+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	.	>+	>+
L1 (PROBA-V NIR)	<+	<+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	.	>+
L1 (PROBA-V RED)	<+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	>+	.

nego wyboru (ang. *two-alternative forced choice*, 2AFC) [101]. W podejściu tym uczestnicy ankiety byli proszeni o dokonanie wyboru między dwoma alternatywnymi wariantami obrazów bez możliwości wstrzymania się od głosu. Ankietowani porównywali zatem obraz referencyjny HR z obrazami przetworzonymi za pomocą różnych metod SR. W każdym pytaniu ankietowemu uczestnikowi prezentowano trójkę obrazów: obraz referencyjny oraz obrazy przetworzone przez dwie z pięciu metod. Ankietowany musiał wybrać preferowany obraz, który jego zdaniem, był bardziej zbliżony do oryginału lub też postrzegany jako lepszy w kontekście jakości wizualnej. Podejście to jest analogiczne do zastosowanego przez autorów LPIPS w trakcie zbierania danych do treningu tej metryki [93]. Przez zadawanie serii takich pytań, uwzględniając wszystkie możliwe kombinacje obrazów referencyjnych i przetworzonych, otrzymuje się zestaw wyników preferencji. Analizując te wyniki, można wyciągnąć wnioski dotyczące jakości poszczególnych metod przetwarzania. Kilka przykładowych pytań z ankiety pokazano na rys. 4.12.

Do ankiety wybrano zestaw 12 obrazów, będący podzbiorem zbioru testowego MuS2 (rys. 4.13). Liczba możliwych kombinacji dwuelementowych dla pięciu badanych metod wynosi 10, co sumarycznie daje 120 przykładów do oceny w poje-

Tabela 4.9: Ranking modeli względem różnych miar na zbiorze MuS2. Do utworzenia rankingu omawianych modeli użyto jakości wyników modeli w ujęciu dla klasycznych metryk cPSNR i cSSIM oraz już ugruntowanej w literaturze miary LPIPS. Sumaryczna pozycja w rankingu jest wyznaczona na podstawie sumy cząstkowych pozycji w ramach miar. Modele uszeregowane według całościowego rankingu.

Model	#cPSNR	#cSSIM	#LPIPS	#
L1+DKFS _{MIX}	1	3	6	1
L1+LPIPS	2	6	5	2
L1+DKFS	4	2	8	3
DKFS _{MIX} ,L1	5	1	13	4
DKFS,L1	6	5	10	5
cPSNR	3	4	17	6
L1,DKFS _{MIX}	14	14	2	7
L1 (PROBA-V RED)	9	9	12	7
L1	11	7	14	9
LPIPS,L1	13	8	11	9
L1,LPIPS	16	16	1	11
DKFS _{MIX}	15	15	4	12
bicubic	8	11	18	13
LPIPS	17	17	3	13
L1 (PROBA-V NIR)	12	10	15	13
bilinear	7	12	19	16
NN	10	13	16	17
DKFS	18	18	7	18
L1,DKFS	19	19	9	19

dynczej ankiecie, takich jak pokazanych na rys. 4.12. Aby uniknąć prezentowania ankietowanym dużej liczby bardzo podobnie wyglądających obrazów, dla każdego z pytań losowo zastosowano jedno z niezmiennających szczegółów przekształceń obrazów: symetryczne odbicie w pionie, poziomie oraz obroty o wielokrotności kąta prostego.

Metoda analizy wyników ankiety była następująca:

1. Zebranie odpowiedzi w postaci czwórek $(X, Y, \mathcal{I}^{(h)}, V)$, gdzie X oraz Y są metodami, za pomocą których zostały otrzymane badane obrazy, $\mathcal{I}^{(h)}$ to obraz referencyjny, a V to wynik postaci 0, 1, wskazujący czy w danej kombinacji (X, Y) bliższy obrazowi referencyjnemu jest X ($V = 0$) czy Y ($V = 1$).

Tabela 4.10: Wyniki działania modeli na zbiorze symulowanym Sentinel-2. Oprócz wartości metryk wskazano ranking w ramach miar oraz ranking całkowity (kolumna ##) oparty na miarach cPSNR, cSSIM i LPIPS (podobnie jak w tabeli 4.9)

		cPSNR	#	cSSIM	#	LPIPS	#	DKFS	#	DKFS _{MIX}	#	##
Interp.	NN	42,5371	4	0,9501	4	0,2233	14	0,0553	6	0,1500	14	4
	bicubic	42,6107	3	0,9512	3	0,2874	17	0,0626	11	0,1760	17	5
	bilinear	42,1067	5	0,9452	5	0,3498	19	0,0730	16	0,2124	19	9
PIUnet - funkcja straty	cPSNR	37,7839	14	0,8824	14	0,2927	18	0,0963	19	0,1897	18	18
	L1	40,7958	7	0,9439	6	0,2415	15	0,0567	8	0,1507	15	8
	LPIPS	37,1859	16	0,8551	18	0,1843	7	0,0673	14	0,1405	11	15
	DKFS	37,0552	17	0,8703	15	0,1979	9	0,0590	10	0,1402	9	15
	DKFS _{MIX}	37,3992	15	0,8699	16	0,1836	6	0,0669	13	0,1403	10	14
	L1,LPIPS	38,7305	13	0,9039	13	0,1595	3	0,0580	9	0,1155	3	9
	L1,DKFS	35,3371	19	0,8061	19	0,2436	16	0,0765	18	0,1614	16	19
	L1,DKFS _{MIX}	36,9945	18	0,8652	17	0,1997	10	0,0739	17	0,1489	13	17
	LPIPS,L1	40,6663	8	0,9404	8	0,2103	13	0,0537	5	0,1320	7	9
	DKFS _{MIX} ,L1	40,9051	6	0,9432	7	0,1835	5	0,0560	7	0,1177	4	3
	DKFS,L1	40,0700	10	0,9343	10	0,2073	12	0,0629	12	0,1338	8	12
	L1+LPIPS	39,4764	11	0,9253	11	0,1767	4	0,0509	4	0,1202	5	6
	L1+DKFS	39,2067	12	0,9250	12	0,2015	11	0,0716	15	0,1414	12	13
	L1+DKFS _{MIX}	40,6148	9	0,9345	9	0,1860	8	0,0497	3	0,1231	6	6
	L1 (PROBA-V NIR)	43,8980	1	0,9645	1	0,1125	2	0,0317	1	0,0745	2	1
L1 (PROBA-V RED)	43,6338	2	0,9622	2	0,1073	1	0,0324	2	0,0719	1	2	

2. Obliczenie wskaźników preferencji: dla każdej z badanych metod M_i obliczana jest liczba P_i – oznaczająca w ilu wypadkach wynik metody M_i był wybrany jako bardziej zbliżony do $\mathcal{I}^{(h)}$. Zliczane zatem są przypadki gdy ($X = M_i$ oraz $V = 0$) lub ($Y = M_i$ oraz $V = 1$).
3. Normalizacja wskaźników: każde P_i dzielone jest przez liczbę wszystkich odpowiedzi w ankiecie N .
4. Tworzony jest ranking metod na podstawie wartości P_i posortowanych w porządku malejącym.

Zebrano 38 ankiet, co daje razem 4560 odpowiedzi na pojedyncze pytania, a tym samym 456 przykładów dla każdej kombinacji porównywanych metod. Taka ilość danych pozwala na przeprowadzenie analizy preferencji obserwatorów. Zbiorcze wyniki przedstawiono w tabeli 4.13, zaś w tabeli 4.14 zebrano wyniki w postaci podzielonej na wszystkie kombinacje porównywanych metod.

Ankieta potwierdziła, że zgodnie z oceną z rankingu w ramach metryk cPSNR, cSSIM i LPIPS (tabela 4.9), modele wytrenowane za pomocą złożonych funkcji straty: L1+LPIPS oraz L1+DKFS_{MIX} dają wyniki znacznie lepsze wizualnie od klasycznych funkcji straty L1+LPIPS i tym bardziej cPSNR. Jednak w zakresie porównania wyników modeli wytrenowanych za pomocą funkcji straty: L1+LPIPS

Tabela 4.11: Wyniki działania modeli na zbiorze PROBA-V RED. Oprócz wartości metryk wskazano ranking w ramach miar oraz ranking całkowity (kolumna ##) oparty na miarach cPSNR, cSSIM i LPIPS (podobnie jak w tabeli 4.9)

		cPSNR	#	cSSIM	#	LPIPS	#	DKFS	#	DKFS _{MIX}	#	##
Interp.	NN	34,3636	4	0,8510	4	0,3405	10	0,1810	10	0,2685	11	3
	bicubic	34,6701	3	0,8607	3	0,4017	13	0,1717	6	0,2662	10	4
	bilinear	34,2881	5	0,8441	5	0,4170	17	0,1760	8	0,2792	15	7
PU _{net} - funkcja straty	cPSNR	31,0944	14	0,8247	10	0,4098	16	0,2043	15	0,2897	17	15
	L1	31,2678	12	0,8187	11	0,4087	15	0,1797	9	0,2889	16	14
	LPIPS	30,2981	16	0,6798	18	0,2682	7	0,2209	19	0,2548	9	17
	DKFS	31,6448	10	0,7898	14	0,3050	9	0,1668	5	0,2355	7	12
	DKFS _{MIX}	30,4571	15	0,7122	17	0,2504	5	0,2065	16	0,2402	8	13
	L1,LPIPS	31,8246	9	0,7857	15	0,2423	4	0,1897	13	0,2247	5	8
	L1,DKFS	23,3267	19	0,4735	19	0,6477	19	0,2078	17	0,3626	19	19
	L1,DKFS _{MIX}	31,1620	13	0,7621	16	0,2289	3	0,1910	14	0,2204	4	11
	LPIPS,L1	30,2587	17	0,8062	12	0,3823	11	0,1832	11	0,2732	12	15
	DKFS _{MIX} ,L1	31,3286	11	0,8437	6	0,4019	14	0,1832	12	0,2756	14	10
	DKFS,L1	28,4586	18	0,8054	13	0,4600	18	0,2107	18	0,3198	18	18
	L1+LPIPS	32,4616	7	0,8301	7	0,2709	8	0,1658	4	0,2265	6	6
	L1+DKFS	32,2540	8	0,8263	9	0,3917	12	0,1747	7	0,2744	13	9
	L1+DKFS _{MIX}	32,7375	6	0,8299	8	0,2568	6	0,1552	3	0,2101	3	5
	L1 (PROBA-V NIR)	38,2321	2	0,9347	2	0,2103	2	0,1433	2	0,1593	2	2
L1 (PROBA-V RED)	38,4413	1	0,9370	1	0,1976	1	0,1420	1	0,1534	1	1	

oraz L1+DKFS_{MIX} otrzymano inną ich kolejność w stosunku do wyników liczbowych uzyskanych z rankingu. Postanowiono zatem przeanalizować te wyniki wizualnie. Na rysunkach 4.14-4.16 pokazano obrazy poddawane ocenie wraz z liczbą oddanych głosów (posortowane względem rosnącej liczby głosów oddanych na L1+LPIPS). Da się zauważyć, że dla obrazów po lewej stronie natężenie zauważalnych artefaktów jest większe (przy czym tutaj ocena jest subiektywna, oparta na doświadczeniu autora rozprawy, jaką postać mają zwykle artefakty w rekonstrukcji) – widoczne jest to szczególnie na powiększeniu pokazanym na rysunku 4.17. Te artefakty jednak współgrają z ogólną strukturą obrazu, co może dawać wrażenie lepszej jakości. Dla kontekstu należy dodać, że w ankiecie brały w zdecydowanej większości osoby nie zajmujące się na co dzień tematyką przetwarzania obrazów³.

Podsumowanie

W niniejszym rozdziale opisano eksperymenty polegające na użyciu w treningu sieci głębokiej realizującej zadanie MISR różnych funkcji straty. Wykonano badania polegające na rekonstrukcji za pomocą tak wytrenowanych modeli przy użyciu danych ze zbioru MuS2. Otrzymane wyniki wskazują, że zarówno w sensie miar literaturowych, jak i w ocenie ankietowanych obserwatorów użycie funkcje straty opartych na cechach głębokich obrazów oraz wykrywaniu punktów charak-

³Wersje cyfrowe obrazów są opublikowane pod linkiem: <http://tiny.cc/w5q8vz>

Tabela 4.12: Wyniki działania modeli na zbiorze PROBA-V NIR. Oprócz wartości metryk wskazano ranking w ramach miar oraz ranking całkowity (kolumna ##) oparty na miarach cPSNR, cSSIM i LPIPS (podobnie jak w tabeli 4.9)

		cPSNR	#	cSSIM	#	LPIPS	#	DKFS	#	DKFS _{MIX}	#	##
Interp.	NN	32,4810	4	0,8169	4	0,3611	10	0,2091	9	0,2953	16	3
	bicubic	32,7760	3	0,8285	3	0,3880	15	0,1912	4	0,2745	10	5
	bilinear	32,4171	5	0,8077	6	0,3954	16	0,1927	5	0,2809	13	8
PIUnet - funkcja straty	cPSNR	27,7624	16	0,7628	13	0,4414	18	0,2447	17	0,3225	19	18
	L1	28,6430	13	0,7932	10	0,3640	12	0,2043	7	0,2756	11	13
	LPIPS	27,2406	18	0,5684	18	0,2912	8	0,2723	19	0,2906	14	17
	DKFS	28,0435	15	0,7059	15	0,3641	13	0,2222	13	0,2913	15	16
	DKFS _{MIX}	27,6753	17	0,6226	17	0,2625	7	0,2463	18	0,2643	7	15
	L1,LPIPS	29,3568	10	0,7310	14	0,2215	4	0,2299	14	0,2300	4	10
	L1,DKFS	25,4731	19	0,5130	19	0,4850	19	0,2198	12	0,3146	17	19
	L1,DKFS _{MIX}	28,7579	12	0,7015	16	0,2251	5	0,2338	15	0,2381	6	12
	LPIPS,L1	29,6995	9	0,7979	7	0,3638	11	0,2070	8	0,2701	9	8
	DKFS _{MIX} ,L1	29,8106	7	0,8119	5	0,3855	14	0,2178	11	0,2793	12	6
	DKFS,L1	28,4372	14	0,7943	8	0,4359	17	0,2430	16	0,3175	18	14
	L1+LPIPS	29,7029	8	0,7818	12	0,2507	6	0,2015	6	0,2331	5	6
	L1+DKFS	29,3498	11	0,7868	11	0,3592	9	0,2096	10	0,2697	8	11
	L1+DKFS _{MIX}	30,3115	6	0,7935	9	0,2159	3	0,1858	3	0,2116	3	3
	L1 (PROBA-V NIR)	36,3904	1	0,9211	1	0,2140	2	0,1601	1	0,1717	1	1
L1 (PROBA-V RED)	35,7811	2	0,9165	2	0,2089	1	0,1628	2	0,1732	2	2	

terystycznych daje polepszenie jakości rekonstrukcji. Stanowi to przyczynek do potwierdzenia tezy T2.

4.4 Dyskusja wyników i wnioski

W rozdziale 4.1 wskazano jaki może być schemat przetwarzania obrazów (EvoNet), który efektywnie polepsza możliwości klasycznych metod rekonstrukcji nadrozdzielczej w połączeniu z metodami stosującymi modele w postaci głębokich sieci

Tabela 4.13: Wyniki ankiety dla metod rekonstrukcji

Metoda	Liczba głosów	Procent głosów
L1+LPIPS	1568	34.4
L1+DKFS _{MIX}	1483	32.5
L1+DKFS	674	14.8
L1	467	10.2
cPSNR	368	8.1

Tabela 4.14: Wyniki ankiety dla poszczególnych kombinacji metod. Dane należy odczytywać w następujący sposób: komórka wskazuje w ilu wypadkach metoda z wiersza była częściej preferowana od metody z kolumny.

	L1	cPSNR	L1+LPIPS	L1+DKFS	L1+DKFS_{MIX}
L1	-	258 (57%)	14 (3%)	172 (38%)	23 (5%)
cPSNR	198 (43%)	-	17 (4%)	136 (30%)	17 (4%)
L1+LPIPS	442 (97%)	439 (96%)	-	422 (93%)	265 (58%)
L1+DKFS	284 (62%)	320 (70%)	34 (7%)	-	38 (8%)
L1+DKFS _{MIX}	433 (95%)	439 (96%)	191 (42%)	420 (92%)	-

konwolucyjnych. Przez optymalizację ewolucyjną hiperparametrów metody EvoIM oraz wstępne przetwarzanie obrazów wejściowych za pomocą sieci głębokiej rozwiązującej problem SISR uzyskano dla obrazów degradowanych wyniki liczbowo lepsze niż dla każdej ze składowych metod oddzielnie dla tego samego powiększenia. Dla obrazów rzeczywistych sytuacja nie była jednoznaczna, ale wyniki ankiety wskazały, że ludzie obserwatorzy postrzegają te wyniki jako lepsze jakościowo. Biorąc to pod uwagę, **można uznać za potwierdzoną tezę T1** niniejszej rozprawy:

Połączenie klasycznych metod wieloobrazowej rekonstrukcji nadrozdzielczej z technikami jednoobrazowymi wykorzystującymi głębokie spłotowe sieci neuronowe, pozwala na uzyskanie lepszych jakościowo wyników niż zastosowanie którejkolwiek z metod bazowych z osobna

Wspomniane wyniki są szczególnie ważne ze względu na znaczenie prac, które były cytowane w artykułach wprowadzających pierwsze sieci głębokie realizujące zadanie MISR: DeepSUM [67] i RAMS [66].

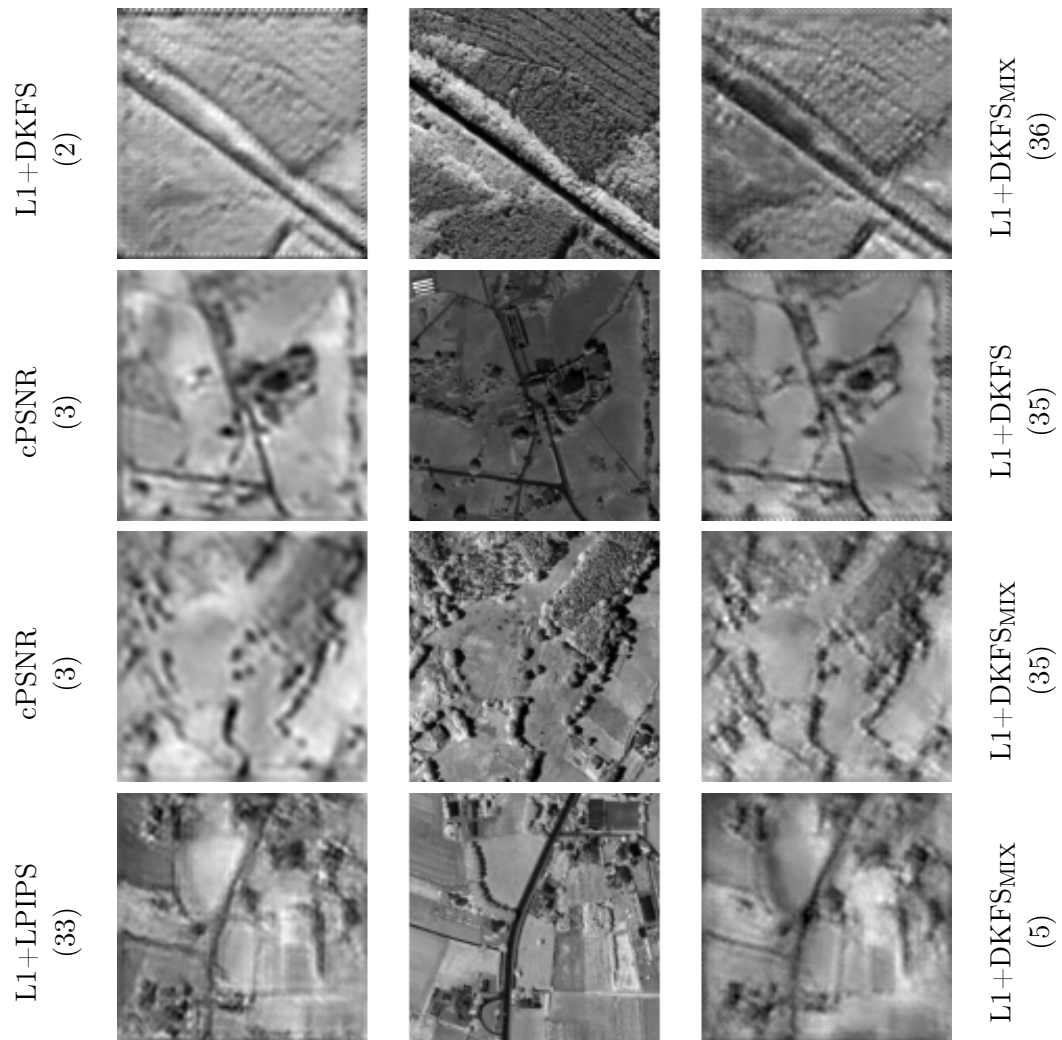
Wyniki uzyskane w rozdziale 4.2 potwierdziły przydatność miar innych niż PSNR i SSIM (typowo stosowanych w literaturze) do zastosowań w problemach spotykanych w rekonstrukcji nadrozdzielczej, przy uwzględnieniu rekonstrukcji z użyciem obrazów o różnej charakterystyce pochodzących z innego źródła w stosunku do obrazów HR. Obiecujące okazały się wyniki uzyskane przez miary oparte na detektorach i deskryptorach punktów kluczowych – w szczególności KFS_{SIFT}, który stosuje ugruntowaną w wizji komputerowej metodę wykrywania punktów kluczowych [131]. Te wyniki posłużyły do wybrania miar stosowanych w eksperymentach dotyczących możliwości treningu sieci MISR z zastosowaniem miar innych niż L1 i L2.

W trakcie prac nad zastosowaniem metryk innych niż tradycyjne w treningu MISR (rozdział 4.3) zostały wybrane modele: LPIPS jako metryka podobieństwa obrazów oraz PIUnet jako jedna z wiodących metod MISR. Podjęto szerokie badania w ramach możliwych do użycia metryk, ich sekwencji w treningu lub też

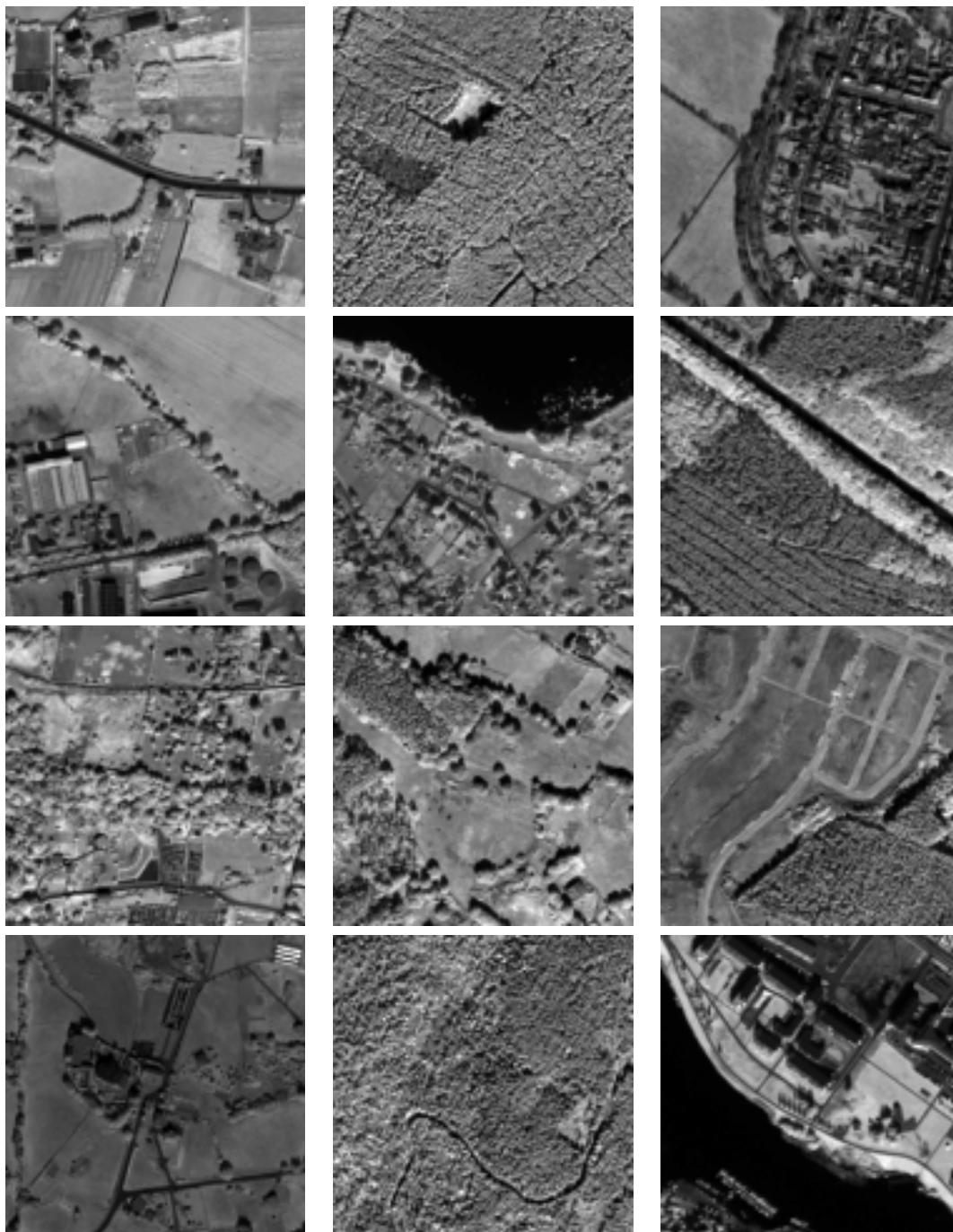
kombinacji. Przez odpowiednio skonstruowany trening miary DKFS opartej na architekturze LPIPS – została użyta miara KFS. Podejście to okazało się skuteczne. W wypadku użycia miar LPIPS i DKFS, opartych na tej samej architekturze, a szczególnie przy użyciu w treningu sum miar L1 oraz LPIPS uzyskano wyniki lepsze numerycznie. Są one istotnie statystycznie lepsze od wyników generowanych przez modele wytrenowane z użyciem zarówno L1 jak L2. To **potwierdza słusność tezy T2**:

W wieloobrazowej rekonstrukcji nadrozdzielczej obrazów satelitarnych, realizowanej za pomocą głębokich splotowych sieci neuronowych, możliwe jest uzyskanie poprawy wyników poprzez zastosowanie przy obliczaniu funkcji straty metryki opartej na cechach głębokich służących do wykrywania punktów charakterystycznych w obrazach

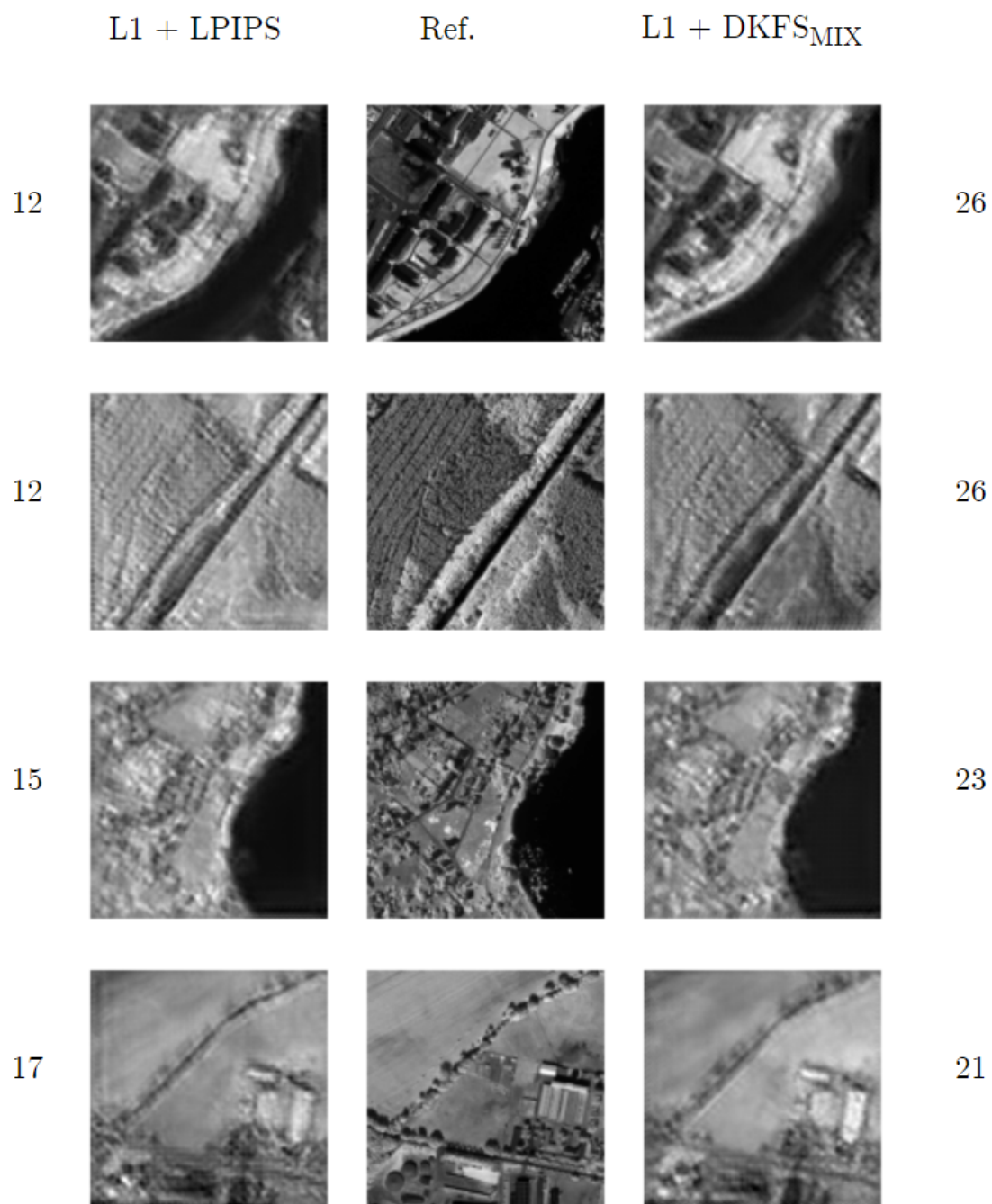
Dodatkowo pośrednie zastosowanie miary KFS – przez użycie w treningu $DKFS_{MIX}$, która to metoda jest następnie użyta jako funkcja straty w treningu PIUnet, poprawia wyniki w sensie tradycyjnych miar (cPSNR i cSSIM) w stosunku do użycia jako funkcji straty LPIPS. Ankieta przeprowadzona wśród obserwatorów nie dostarczyła jednoznacznego potwierdzenia tego polepszenia, co jest na pewno wskazówką do dalszych badań. Pokazała jednak zdecydowaną przewagę wszystkich modeli opartych na treningu z miarami opartymi na sieciach głębokich nad modelami wytrenowanymi z funkcjami straty L1 i L2, **co również jest przyczynkiem do potwierdzenia tezy T2**.



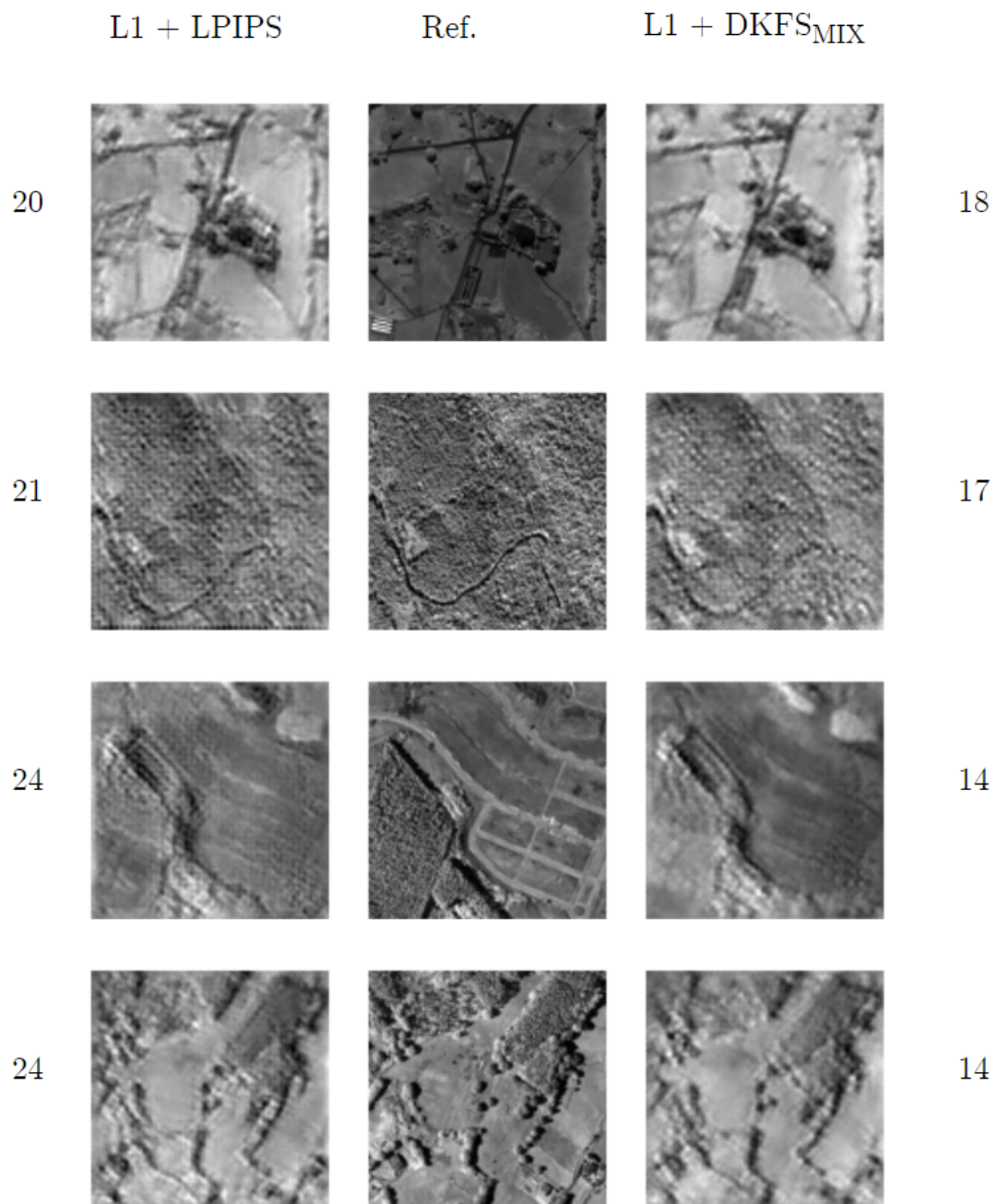
Rysunek 4.12: Przykładowe pytania z ankiety 2AFC. Pytanie dla każdego z przykładów brzmiało „Który obraz (po prawej, czy po lewej stronie) jest bardziej podobny do środkowego jeśli chodzi o widoczne szczegóły (drzewa, drogi itp.)”. W środku - obraz referencyjny. Podano liczbę oddanych głosów.



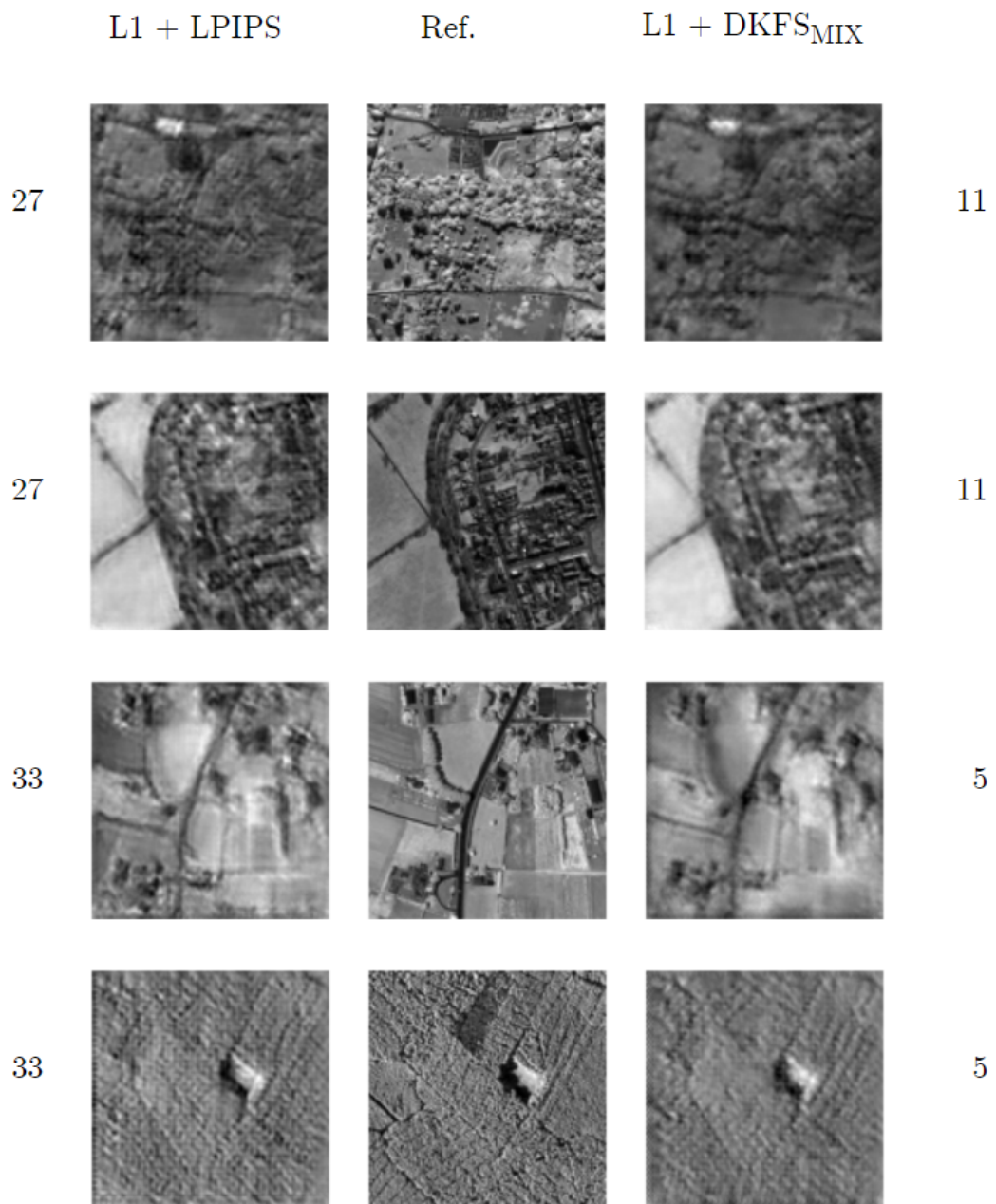
Rysunek 4.13: Obrazy wybrane do ankiety. Sposób wyboru: początkowo większa liczba obrazów została wylosowana z podzbioru testowego zbioru MuS2, a następnie spośród nich dokonano selekcji obrazów prezentujących: lasy, obszary wiejskie, miejskie, zawierające drogi i rozróżnialne drzewa oraz wody powierzchniowe. Zostały również odrzucone fragmenty zawierające chmury i gdzie ponad połowę obszaru stanowiły wody.



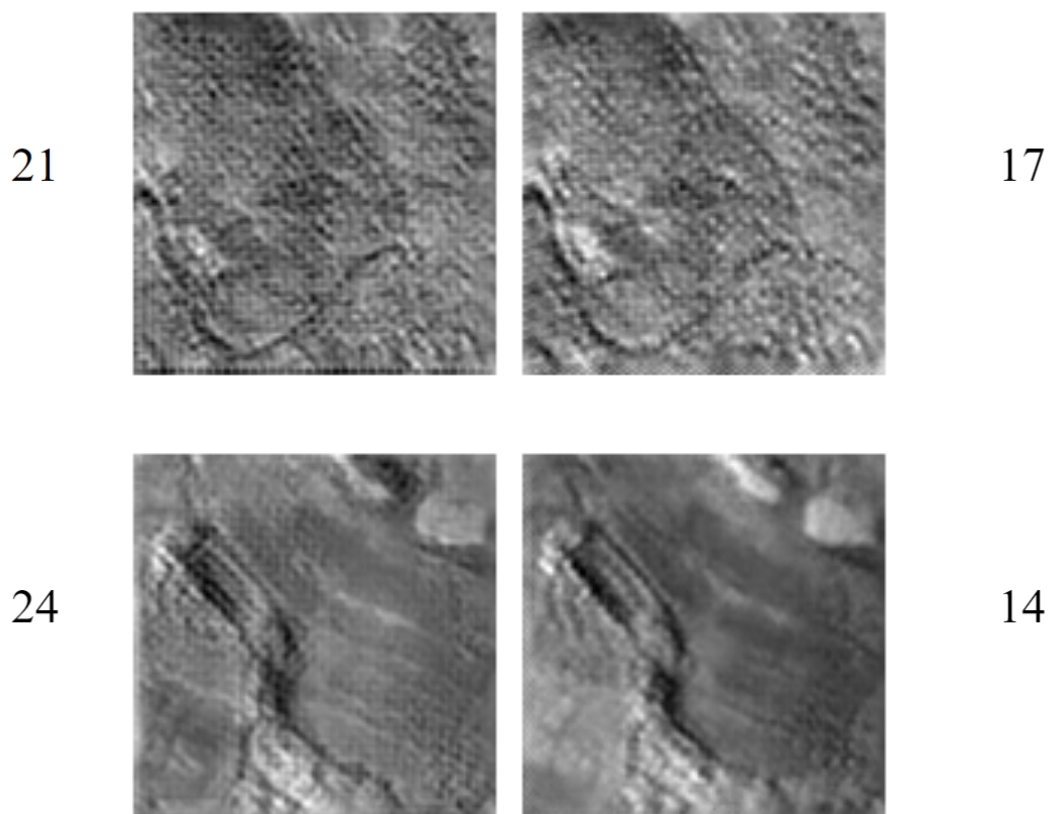
Rysunek 4.14: Porównanie obrazów SR L1+LPIPS z L1+DKFS_{MIX} (obrazy 1 – 4). Po lewej – rekonstrukcja z użyciem modelu wytrenowanego z funkcją straty L1+LPIPS, w środku – obraz referencyjny HR, po prawej – rekonstrukcja z użyciem modelu wytrenowanego z funkcją straty L1+DKFS_{MIX}. Po bokach liczba głosów oddana na rekonstrukcje dla danego przykładu.



Rysunek 4.15: Porównanie obrazów SR L1+LPIPS z L1+DKFS_{MIX} (obrazy 5 – 8).



Rysunek 4.16: Porównanie obrazów SR L1+LPIPS z L1+DKFS_{MIX} (obrazy 9 – 12).



Rysunek 4.17: Porównanie obrazów SR L1+LPIPS z L1+DKFS_{MIX}, przykład artefaktów. Po lewej – rekonstrukcja z użyciem modelu wytrenowanego z funkcją straty L1+LPIPS, po prawej – rekonstrukcja z użyciem modelu wytrenowanego z funkcją straty L1+DKFS_{MIX}. Po bokach liczba głosów oddana na rekonstrukcje dla danego przykładu. Dla obrazu po lewej widoczne większe natężenie artefaktów, które jednak współgrają z ogólną strukturą obrazu.

Rozdział 5

Podsumowanie i możliwe kierunki dalszych badań

5.1 Podsumowanie

W niniejszej rozprawie zostały zaprezentowane badania służące rozwojowi metod rekonstrukcji nadrozdzielczej ze szczególnym uwzględnieniem danych pochodzących z obrazowania satelitarnego. Problem rekonstrukcji nadrozdzielczej został wprowadzony wraz ze zdefiniowaniem typowych wyzwań. Wskazano cele pracy i postawiono tezy, w celu potwierdzenia których zaprojektowano stosowne metody i zaplanowano serię badań.

Dokonano przeglądu literatury związanej z tematem rekonstrukcji nadrozdzielczej (przedstawiony w rozdziale 2). Omówiono metody SR, ze szczególnym uwzględnieniem tych, które stanowią podstawę do opracowywania nowych podejść do rekonstrukcji wieloobrazowej. Skupiono się także na przedstawieniu metryk stosowanych do oceny jakości obrazów, omawiając ich zalety i ograniczenia w kontekście MISR (przedstawione w rozdziale 2.2). Istotnym aspektem badań nad SR są dane służące do treningu oraz walidacji tworzonych technik, dlatego też opisano zbiory danych (rozdział 2.3), które znajdują zastosowanie w eksperymentach opisanych w rozdziale 4. Przedstawiono krytyczną analizę problemów występujących w procesie rekonstrukcji, w tym brak adekwatnych narzędzi oceny jakości obrazów SR, co stanowiło istotne wyzwanie.

Rozdział 3 przedstawia proponowane rozwiązania wskazanych problemów. Opiszano jak sieci głębokie mogą być użyte w połączeniu z klasycznymi metodami rekonstrukcji nadrozdzielczej (**cel 1 rozprawy**: *Zaprojektować i zaimplementować sposób połączenia metod rekonstrukcji nadrozdzielczej opartych na sieciach głębokich z metodami klasycznymi*). Zaprojektowano miary oparte o wykrywanie punktów charakterystycznych (**cel 3 rozprawy**: *Zaprojektować i zaimplemento-*

wać miary podobieństwa obrazu, które posiadają cechy pożądane w SR) oraz sposób w jaki mogą być one zaimplementowane w architekturze sieci głębokiej. Pokazano również jak takie modele mogą być użyte jako funkcja straty w procesie trenowania modeli rozwiązujących problem SR (**cel 5 rozprawy**: *Zaprojektować sposób, w jaki zaprojektowane miary podobieństwa mogą zostać użyte w treningu algorytmu MISR opartego na sieci głębokiej*).

Opis eksperymentów (rozdział 4) składa się z kilku sekcji, w których analizowane jest, jak zastosowanie odpowiednich narzędzi do porównywania obrazów i technik optymalizacji może wpłynąć na poprawę jakości rekonstrukcji przy użyciu metod MISR. W pierwszym etapie przeprowadzono weryfikację tezy T1: że zastosowanie sieci neuronowych w zadaniu MISR w połączeniu z podejściami klasycznymi prowadzi do poprawy jakości rekonstrukcji (**cel 2 rozprawy**: *Zweryfikować eksperymentalnie poprawę jakości dla danych symulowanych oraz rzeczywistych obrazów satelitarnych z użyciem zdefiniowanego zaproponowanego podejścia*). Następnie, w świetle uzyskanych wyników poszukiwano metryk, które charakteryzują się pożądanymi cechami w zadaniu SR, przeprowadzono również walidację zaprojektowanych miar podobieństwa (**cel 4 rozprawy**: *Zweryfikować eksperymentalnie przydatność wspomnianych miar podobieństwa obrazu w rekonstrukcji nadrozdzielczej w porównaniu do metryk spotykanych w literaturze*). Kolejnym krokiem było zastosowanie wybranych wcześniej metryk podczas treningu jednej z uznanych metod MISR, aby potwierdzić tezę T2: że w przypadku rzeczywistych obrazów satelitarnych istnieje możliwość uzyskania lepszych rezultatów w porównaniu do tradycyjnych miar L1 i L2 (**cel 6 rozprawy**: *Zweryfikować eksperymentalnie poprawę jakości rekonstrukcji z wykorzystaniem zaproponowanej metody na rzeczywistych danych satelitarnych*). Wskazane w rozprawie rozwiązania skutkowały poprawą jakości rekonstrukcji zarówno w sensie ilościowym (obiektywnym) – z perspektywy tradycyjnych metryk literaturowych, jak i przy uwzględnieniu ocen dokonanych przez obserwatorów ludzkich. Wykazano, że modele wytrenowane z użyciem funkcji straty wykorzystujących głębokie sieci konwolucyjne (tak autorskie miary DKFS, jak i znana z literatury metoda LPIPS) osiągnęły lepsze rezultaty od modeli trenowanych z użyciem miar L1 lub L2. W ocenie ogółu obserwatorów najlepsze okazały się wyniki modeli wytrenowanych z użyciem metryk LPIPS i DKFS.

5.2 Wnioski

Przeprowadzone w ramach prac nad rozprawą badania pokazały lub potwierdziły następujące fakty:

1. Rekonstrukcja nadrozdzielcza wieloobrazowa działająca na danych rzeczywistych (w szczególności obrazach satelitarnych) wymaga innego podejścia do

treningu i oceny wyników niż dla często stosowanych w literaturze danych sztucznie zdegradowanych.

2. Zbiory danych zawierające rzeczywiste dane obrazowe z instrumentów satelitarnych, które powstały w ostatnich latach są coraz obszerniejsze, co daje nadzieję na rozwój lepszych metod MISR.
3. Miary stosowane typowo jako funkcje straty w treningu modeli opartych na sieciach głębokich (bazujące na metrykach $L1 / L2$) nie kierują treningu w stronę najlepszego odwzorowania szczegółów.
4. Celowe wydaje się używanie jako funkcji straty metod złożonych, które korzystają zarówno z ograniczających artefakty funkcji bazujących na metrykach $L1 / L2$ oraz metod faworyzujących dodawanie szczegółów do obrazu.

5.3 Możliwe kierunki dalszych badań

Prezentowane badania otwierają nowe możliwości w dziedzinie rekonstrukcji nadrozdzielczej. Wskazują one na potrzebę dalszej eksploracji i zrozumienia subiektywnych aspektów oceny jakości obrazów oraz na konieczność dostosowania narzędzi badawczych do realnych potrzeb dziedziny. Niniejsza rozprawa nie tylko poszerza wiedzę w tej dziedzinie, ale także pozwala na identyfikację kierunków dalszych prac badawczych.

Przede wszystkim celowe wydawałoby się sprawdzenie działania metryk dla danych bardziej obszernych niż zbiór MuS2, takich jak WorldStrat (rozdział 2.3.6). Również w ramach samego zbioru MuS2 należałoby sprawdzić działanie opisanych metod na zakresach widmowych innych niż bliska podczerwień, która jak pokazują badania w pracy [27] jest najbardziej zmiennym zakresem – widać to po wartościach metryk niższych w stosunku do innych pasm.

Kolejnym aspektem, który należałoby przebadać byłoby zastosowanie w ramach metryk złożonych (takich jak $L1+DKFS_{MIX}$) metryki VIF, która została pozytywnie oceniona w części dotyczącej badania czułości metryk w zastosowaniach SR. Opracowany i sprawdzony w ramach badań schemat umożliwiający efektywne użycie miar opartych na architekturze LPIPS, tak by uwzględniała dodatkowe dane wejściowe, wykonanie takich eksperymentów nie powinno nastęrczać trudności.

Zastosowane w rozprawie miary oparte na punktach charakterystycznych używają detektorów i deskryptorów zdefiniowanych w sposób analityczny. W literaturze pojawiły się od czasu zaprojektowania tych miar podejścia do wykrywania punktów kluczowych oparte na uczeniu głębokim, takie jak [137]. Wartościowe byłoby sprawdzenie czy tego typu metody mogą znaleźć zastosowanie w ocenie

jakości rekonstrukcji – miara podobieństwa oparta na tego typu rozwiązaniu mogłaby zostać zweryfikowana w treningu zarówno metod SISR jak MISR.

Podziękowania

Mojej żonie Justynie i córce Anastazji dziękuję za cierpliwość, wyrozumiałość i ogromną miłość. Jesteście moim największym źródłem sił i motywacji, szczególnie w tych najtrudniejszych momentach. Wasza gotowość do poświęcenia swojego czasu na rzecz moich badań była nieoceniona i sprawiła, że byłem w stanie skoncentrować się na pracy.

Moim promotorom: Michałowi Kawulokowi i Danielowi Kostrzewie dziękuję za wprowadzenie mnie w świat badań naukowych i kierowanie moją pracą. Dziękuję za poświęcenie niezliczonych godzin, dostępność, anielską cierpliwość i pomoc w trudnych momentach, a było ich sporo. Wasze wsparcie było bezcenne, a obserwowanie Waszej codziennej tytanicznej pracy jest dla mnie wielką inspiracją.

Dziękuję również współpracownikom z Future Processing, KP Labs, Katedry Informatyki Stosowanej i innym osobom, których życzliwość, rady i pomoc pomogły mi w przeprowadzeniu badań i napisaniu rozprawy.

Bibliografia

- [1] R. G. Driggers, *Encyclopedia of Optical Engineering: Las-Pho, pages 1025-2048*. CRC press, 2003, t. 2.
- [2] R. Klette i A. Rosenfeld, *Digital geometry: Geometric methods for digital picture analysis*. Morgan Kaufmann, 2004.
- [3] M. Sozzi, F. Marinello, A. Pezzuolo i L. Sartori, “Benchmark of satellites image services for precision agricultural use,” w *Proceedings of the AgEng Conference, Wageningen, The Netherlands*, 2018, s. 8–11.
- [4] S.-E. Qian, “Hyperspectral satellites, evolution, and development history,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, t. 14, s. 7032–7056, 2021.
- [5] K. Nasrollahi i T. Moeslund, “Super-resolution: A comprehensive survey,” *Machine Vision and Applications*, t. 25, s. 1423–1468, sierp. 2014. DOI: 10.1007/s00138-014-0623-4.
- [6] S. Anwar, S. Khan i N. Barnes, “A deep journey into super-resolution: A survey,” *ACM Computing Surveys (CSUR)*, t. 53, nr. 3, s. 1–34, 2020.
- [7] H. Greenspan, “Super-resolution in medical imaging,” *The Computer Journal*, t. 52, nr. 1, s. 43–63, 2009.
- [8] N. N. A. N. Ghazali, N. A. Zamani, S. N. H. S. Abdullah i J. Jameson, “Super resolution combination methods for CCTV forensic interpretation,” w *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)*, IEEE, 2012, s. 853–858.
- [9] M. R. Arefin, V. Michalski, P.-L. St-Charles, A. Kalaitzis, S. Kim, S. E. Kahou i Y. Bengio, “Multi-image super-resolution for remote sensing using deep recurrent networks,” w *Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) workshops*, 2020, s. 206–207.
- [10] R. Fernandez-Beltran, P. Latorre-Carmona i F. Pla, “Single-frame super-resolution in remote sensing: A practical overview,” *International Journal of Remote Sensing*, t. 38, nr. 1, s. 314–354, 2017.

- [11] S. Farsiu, D. Robinson, M. Elad i P. Milanfar, “Advances and challenges in super-resolution,” *International Journal of Imaging Systems and Technology*, t. 14, nr. 2, s. 47–57, 2004.
- [12] Z. Wang, J. Chen i S. C. Hoi, “Deep learning for image super-resolution: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, t. 43, nr. 10, s. 3365–3387, 2020.
- [13] X. Wang, J. Yi, J. Guo, Y. Song, J. Lyu, J. Xu, W. Yan, J. Zhao, Q. Cai i H. Min, “A review of image super-resolution approaches based on deep learning and applications in remote sensing,” *Remote Sensing*, t. 14, nr. 21, s. 5423, 2022.
- [14] J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza, A. Plaza i F. Pla, “A new deep generative network for unsupervised remote sensing single-image super-resolution,” *IEEE Transactions on Geoscience and Remote Sensing*, t. 56, nr. 11, s. 6792–6810, 2018.
- [15] L. Liebel i M. Körner, “Single-image super resolution for multispectral remote sensing data using convolutional neural networks,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, t. 41, s. 883–890, 2016.
- [16] J. Zhang, T. Xu, J. Li, S. Jiang i Y. Zhang, “Single-image super resolution of remote sensing images with real-world degradation modeling,” *Remote Sensing*, t. 14, nr. 12, s. 2895, 2022.
- [17] J. Tian i K.-K. Ma, “A survey on super-resolution imaging,” *Signal, Image and Video Processing*, t. 5, s. 329–342, 2011.
- [18] X. Zhang, H. Zeng i L. Zhang, “Edge-oriented convolution block for real-time super resolution on mobile devices,” w *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, s. 4034–4043.
- [19] E. Zamfir, M. V. Conde i R. Timofte, “Towards real-time 4k image super-resolution,” w *Proceedings of the 2023 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, s. 1522–1532.
- [20] S. Farsiu, M. D. Robinson, M. Elad i P. Milanfar, “Fast and robust multiframe super resolution,” *IEEE Transactions on Image Processing*, t. 13, nr. 10, s. 1327–1344, 2004.
- [21] D. Yang, Z. Li, Y. Xia i Z. Chen, “Remote sensing image super-resolution: Challenges and approaches,” w *2015 IEEE International Conference on Digital Signal Processing (DSP)*, IEEE, 2015, s. 196–200.
- [22] J. Yang i T. Huang, “Image super-resolution: Historical overview and future challenges,” w *Super-resolution imaging*, CRC Press, 2017, s. 1–34.

- [23] D. C. Lepcha, B. Goyal, A. Dogra i V. Goyal, “Image super-resolution: A comprehensive review, recent trends, challenges and applications,” *Information Fusion*, 2022.
- [24] P. Benecki, M. Kawulok, D. Kostrzewa i L. Skonieczny, “Evaluating super-resolution reconstruction of satellite images,” *Acta Astronautica*, t. 153, s. 15–25, 2018.
- [25] J. Adler, J. Kawulok i M. Kawulok, “Toward Understanding the Impact of Input Data for Multi-Image Super-Resolution,” w *Asian Conference on Intelligent Information and Database Systems*, Springer, 2022, s. 329–342.
- [26] M. Märtens, D. Izzo, A. Krzic i D. Cox, “Super-resolution of PROBA-V images using convolutional neural networks,” *Astrodynamics*, t. 3, s. 387–402, 2019.
- [27] P. Kowaleczko, T. Tarasiewicz, M. Ziaja, D. Kostrzewa, J. Nalepa, P. Rokita i M. Kawulok, “MuS2: A Benchmark for Sentinel-2 Multi-Image Super-Resolution,” *arXiv preprint arXiv:2210.02745*, 2022.
- [28] S. Dai, M. Han, Y. Wu i Y. Gong, “Bilateral Back-Projection for Single Image Super Resolution,” w *Proceedings of 2007 IEEE International Conference on Multimedia and Expo*, 2007, s. 1039–1042. DOI: 10.1109/ICME.2007.4284831.
- [29] E. Van Reeth, I. W. Tham, C. H. Tan i C. L. Poh, “Super-resolution in magnetic resonance imaging: a review,” *Concepts in Magnetic Resonance Part A*, t. 40, nr. 6, s. 306–325, 2012.
- [30] X. He, Y. Lei, Y. Fu, H. Mao, W. J. Curran, T. Liu i X. Yang, “Super-resolution magnetic resonance imaging reconstruction using deep attention networks,” w *Medical Imaging 2020: Image Processing*, SPIE, t. 11313, 2020, s. 642–647.
- [31] J. Kamenicky, M. Bartos, J. Flusser, B. Mahdian, J. Kotera, A. Novozamsky, S. Saic, F. Sroubek, M. Sorel, A. Zita i in., “PIZZARO: Forensic analysis and restoration of image and video data,” *Forensic Science International*, t. 264, s. 153–166, 2016.
- [32] S. Villena, M. Vega, J. Mateos, D. Rosenberg, F. Murtagh, R. Molina i A. K. Katsaggelos, “Image super-resolution for outdoor digital forensics. Usability and legal aspects,” *Computers in Industry*, t. 98, s. 34–47, 2018.
- [33] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang i L. Zhang, “Image super-resolution: The techniques, applications, and future,” *Signal Processing*, t. 128, s. 389–408, 2016.

- [34] R. Zhou i S. Susstrunk, “Kernel modeling super-resolution on real low-resolution images,” w *Proceedings of the 2019 IEEE International Conference on Computer Vision (ICCV)*, 2019, s. 2433–2443.
- [35] B. Wronski, I. Garcia-Dorado, M. Ernst, D. Kelly, M. Krainin, C.-K. Liang, M. Levoy i P. Milanfar, “Handheld multi-frame super-resolution,” *ACM Transactions on Graphics (ToG)*, t. 38, nr. 4, s. 1–18, 2019.
- [36] H. Liu, Z. Ruan, P. Zhao, C. Dong, F. Shang, Y. Liu, L. Yang i R. Timofte, “Video super-resolution based on deep learning: a comprehensive survey,” *Artificial Intelligence Review*, t. 55, nr. 8, s. 5981–6035, 2022.
- [37] A. Kappeler, S. Yoo, Q. Dai i A. K. Katsaggelos, “Video super-resolution with convolutional neural networks,” *IEEE Transactions on Computational Imaging*, t. 2, nr. 2, s. 109–122, 2016.
- [38] S. Son, S. Lee, S. Nah, R. Timofte i K. M. Lee, “NTIRE 2021 challenge on video super-resolution,” w *Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, s. 166–181.
- [39] K. Egiazarian i V. Katkovnik, “Single image super-resolution via BM3D sparse coding,” w *2015 23rd European signal processing conference (EUSIPCO)*, IEEE, 2015, s. 2849–2853.
- [40] S. Mandal, A. Bhavsar i A. K. Sao, “Noise adaptive super-resolution from single image via non-local mean and sparse representation,” *Signal Processing*, t. 132, s. 134–149, 2017.
- [41] H. Chavez-Roman i V. Ponomaryov, “Super resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation,” *IEEE Geoscience and Remote Sensing Letters*, t. 11, nr. 10, s. 1777–1781, 2014.
- [42] J. Yang, J. Wright i Y. Ma, “Image super-resolution as sparse representation of raw image patches,” czer. 2008. DOI: 10.1109/CVPR.2008.4587647.
- [43] M. A. Herman i T. Strohmer, “High-resolution radar via compressed sensing,” *IEEE Transactions on Signal Processing*, t. 57, nr. 6, s. 2275–2284, 2009.
- [44] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng i L. Zhang, “Convolutional Sparse Coding for Image Super-Resolution,” w *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, grud. 2015.
- [45] M. Zhang, Z. Liu i L. Yu, *CRNet: Image Super-Resolution Using A Convolutional Sparse Coding Inspired Network*, 2019. arXiv: 1908.01166 [eess.IV].

- [46] C. Dong, C. C. Loy, K. He i X. Tang, “Image Super-Resolution Using Deep Convolutional Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, t. 38, nr. 2, s. 295–307, 2016. DOI: 10.1109/TPAMI.2015.2439281.
- [47] C. Dong, C. C. Loy i X. Tang, “Accelerating the Super-Resolution Convolutional Neural Network,” w *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe i M. Welling, red., Cham: Springer International Publishing, 2016, s. 391–407, ISBN: 978-3-319-46475-6.
- [48] C. Ledig, Z. Wang, W. Shi i in., “Photo-realistic single image super-resolution using a generative adversarial network,” w *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, s. 4681–4690.
- [49] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville i Y. Bengio, “Generative adversarial nets,” *Advances in Neural Information Processing Systems*, t. 27, 2014.
- [50] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta i A. A. Bharath, “Generative Adversarial Networks: An Overview,” *IEEE Signal Processing Magazine*, t. 35, nr. 1, s. 53–65, 2018. DOI: 10.1109/MSP.2017.2765202.
- [51] J. Feng, “Generative Adversarial Networks for Remote Sensing,” w *2022 2nd International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR)*, 2022, s. 108–112. DOI: 10.1109/ICBAR58199.2022.00028.
- [52] B. Lim, S. Son, H. Kim, S. Nah i K. M. Lee, “Enhanced deep residual networks for single image super-resolution,” w *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) workshops*, 2017, s. 136–144.
- [53] Y. Zhang, Y. Tian, Y. Kong, B. Zhong i Y. Fu, “Content-conditioned GAN for super-resolution,” w *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, s. 1827–1836.
- [54] T. Karras, T. Aila, S. Laine i J. Lehtinen, “High-resolution image synthesis with progressive growing of GANs,” w *Proceedings of the International Conference on Learning Representations*, 2018.
- [55] Y. Tao i J.-P. Muller, “Super-resolution restoration of spaceborne ultra-high-resolution images using the UCL OpTiGAN system,” *Remote Sensing*, t. 13, nr. 12, s. 2269, 2021.

- [56] W.-S. Lai, J.-B. Huang, N. Ahuja i M.-H. Yang, “Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, t. 41, nr. 11, s. 2596–2608, 2019. DOI: 10.1109/TPAMI.2018.2873593.
- [57] A. Panagiotopoulou i V. Anastassopoulos, “Super-resolution image reconstruction techniques: trade-offs between the data-fidelity and regularization terms,” *Information Fusion*, t. 13, nr. 3, s. 185–195, 2012.
- [58] M. Irani i S. Peleg, “Improving resolution by image registration,” *CVGIP: Graphical models and image processing*, t. 53, nr. 3, s. 231–239, 1991.
- [59] R. R. Schultz i R. L. Stevenson, “Extraction of high-resolution frames from video sequences,” *IEEE Transactions on Image Processing*, t. 5, nr. 6, s. 996–1011, 1996.
- [60] Y. Wang, R. Fevig i R. R. Schultz, “Super-resolution mosaicking of UAV surveillance video,” w *Proceedings of 2008 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2008, s. 345–348.
- [61] T. Akgun, Y. Altunbasak i R. M. Mersereau, “Super-resolution reconstruction of hyperspectral images,” *IEEE Transactions on Image Processing*, t. 14, nr. 11, s. 1860–1875, 2005.
- [62] R. Hardie, “A fast image super-resolution algorithm using an adaptive Wiener filter,” *IEEE Transactions on Image Processing*, t. 16, nr. 12, s. 2953–2964, 2007.
- [63] F. Li, X. Jia i D. Fraser, “Universal HMT based super resolution for remote sensing images,” w *Proceedings of 2008 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2008, s. 333–336.
- [64] M. Haris, G. Shakhnarovich i N. Ukita, “Deep Back-Projection Networks for Super-Resolution,” w *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, s. 1664–1673.
- [65] Z. Zhang, K. Zhang, Z. Li i Y. Qiao, “Attention-Guided Back-Projection Networks for Image Super-Resolution,” w *Proceedings of the AAAI Conference on Artificial Intelligence*, t. 33, 2019, s. 8990–8997.
- [66] F. Salvetti, V. Mazzia, A. Khaliq i M. Chiaberge, “Multi-image super resolution of remotely sensed images using residual attention deep neural networks,” *Remote Sensing*, t. 12, nr. 14, s. 2207, 2020.
- [67] A. Bordone Molini, D. Valsesia, G. Fracastoro i E. Magli, “DeepSUM: Deep Neural Network for Super-Resolution of Unregistered Multitemporal Images,” *IEEE Transactions on Geoscience and Remote Sensing*, t. 58, nr. 5, s. 3644–3656, 2020. DOI: 10.1109/TGRS.2019.2959248.

- [68] M. Deudon, A. Kalaitzis, I. Goytom, M. R. Arefin, Z. Lin, K. Sankaran, V. Michalski, S. E. Kahou, J. Cornebise i Y. Bengio, “Highres-net: Recursive fusion for multi-frame super-resolution of satellite imagery,” *arXiv preprint arXiv:2002.06460*, 2020.
- [69] D. Valsesia i E. Magli, “Permutation Invariance and Uncertainty in Multi-temporal Image Super-Resolution,” *IEEE Transactions on Geoscience and Remote Sensing*, t. 60, s. 1–12, 2022. DOI: 10.1109/TGRS.2021.3130673.
- [70] J. Johnson, A. Alahi i L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” w *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, Springer, 2016, s. 694–711.
- [71] M. S. Rad, B. Bozorgtabar, U.-V. Marti, M. Basler, H. K. Ekenel i J.-P. Thiran, “Srobb: Targeted perceptual loss for single image super-resolution,” w *Proceedings of the 2019 IEEE International Conference on Computer Vision (ICCV)*, 2019, s. 2710–2719.
- [72] M. A. Research, *Papers With Code: state-of-the-art in MISR*, <https://paperswithcode.com/sota/multi-frame-super-resolution-on-proba-v>, Accessed on: May 30, 2023, 2023.
- [73] Z. Wang, A. C. Bovik i L. Lu, “Why is image quality assessment so difficult?” W *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, t. 4, 2002, s. IV–3313.
- [74] Z. Wang i A. C. Bovik, “Modern image quality assessment,” *prac. dokt.*, Springer, 2006.
- [75] G. Zhai i X. Min, “Perceptual image quality assessment: a survey,” *Science China Information Sciences*, t. 63, s. 1–52, 2020.
- [76] X. Li, “Blind image quality assessment,” w *Proceedings of 2002 IEEE International Conference on Image Processing (ICIP)*, IEEE, t. 1, 2002, s. I–I.
- [77] P. J. Durka, *Wstęp do współczesnej statystyki*. Wydaw. Adamantan, 2003.
- [78] J. Korhonen i J. You, “Peak signal-to-noise ratio revisited: Is simple beautiful?” W *2012 Fourth International Workshop on Quality of Multimedia Experience*, IEEE, 2012, s. 37–38.
- [79] A. Amer i E. Dubois, “Fast and reliable structure-oriented video noise estimation,” *IEEE Transactions on Circuits and Systems for Video Technology*, t. 15, nr. 1, s. 113–118, 2005.

- [80] S. Winkler i P. Mohandas, “The evolution of video quality measurement: From PSNR to hybrid metrics,” *IEEE Transactions on Broadcasting*, t. 54, nr. 3, s. 660–668, 2008.
- [81] Z. Wang, A. Bovik, H. Sheikh i E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, t. 13, nr. 4, s. 600–612, 2004. DOI: 10.1109/TIP.2003.819861.
- [82] H. R. Sheikh i A. C. Bovik, “A visual information fidelity approach to video quality assessment,” w *The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, sn, t. 7, 2005, s. 2117–2128.
- [83] H. Sheikh i A. Bovik, “Image information and visual quality,” *IEEE Transactions on Image Processing*, t. 15, nr. 2, s. 430–444, 2006. DOI: 10.1109/TIP.2005.859378.
- [84] E. Simoncelli i W. Freeman, “The steerable pyramid: a flexible architecture for multi-scale derivative computation,” w *Proceedings of 1995 IEEE International Conference on Image Processing (ICIP)*, t. 3, 1995, 444–447 vol.3. DOI: 10.1109/ICIP.1995.537667.
- [85] R. Reisenhofer, S. Bosse, G. Kutyniok i T. Wiegand, “A Haar wavelet-based perceptual similarity index for image quality assessment,” *Signal Processing: Image Communication*, t. 61, s. 33–43, 2018.
- [86] S. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, t. 11, nr. 7, s. 674–693, 1989.
- [87] I. Daubechies, *Ten Lectures on Wavelets*. Society for Industrial i Applied Mathematics, 1992.
- [88] R. Eckhorn, R. Bauer, W. Jordan, M. Brosch, W. Kruse, M. Munk i H. J. Reitboeck, “Coherent oscillations: A mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat,” *Biological Cybernetics*, t. 60, nr. 2, s. 121–130, 1988.
- [89] E. P. Simoncelli i E. H. Adelson, “Noise removal via Bayesian wavelet coring,” w *Proceedings of 1996 IEEE International Conference on Image Processing (ICIP)*, t. 3, 1996, s. 379–382.
- [90] X. Yu, Z. Wang i D. Chen, “Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index,” *IEEE Transactions on Image Processing*, t. 25, nr. 10, s. 4869–4881, 2016.
- [91] J. M. Prewitt i in., “Object enhancement and extraction,” *Picture Processing and Psychopictorics*, t. 10, nr. 1, s. 15–19, 1970.

- [92] S. S. Aghaian, K. Panetta i A. M. Grigoryan, “A new measure of image enhancement,” w *IASTED International Conference on Signal Processing and Communication*, Citeseer, 2000, s. 19–22.
- [93] R. Zhang, P. Isola, A. A. Efros, E. Shechtman i O. Wang, *The Unreasonable Effectiveness of Deep Features as a Perceptual Metric*, 2018. arXiv: 1801.03924 [cs.CV].
- [94] H. Talebi i P. Milanfar, “NIMA: Neural image assessment,” *IEEE Transactions on Image Processing*, t. 27, nr. 8, s. 3998–4011, 2018.
- [95] E. Prashnani, H. Cai, Y. Mostofi i P. Sen, “PieAPP: Perceptual Image-Error Assessment through Pairwise Preference,” w *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, s. 2437–2445.
- [96] Y. Jo, S. Yang i S. J. Kim, “Investigating loss functions for extreme super-resolution,” w *Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) workshops*, 2020, s. 424–425.
- [97] T. Miyata, “Training of Perceptual Image Denoising Network with Weighted Sum of IQA Methods,” w *2023 IEEE 20th Consumer Communications & Networking Conference (CCNC)*, IEEE, 2023, s. 658–659.
- [98] K. Simonyan i A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv preprint arXiv:1409.1556*, 2014. adr.: <https://arxiv.org/abs/1409.1556>.
- [99] A. Krizhevsky, I. Sutskever i G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” w *Advances in Neural Information Processing Systems (NIPS)*, 2012, s. 1097–1105.
- [100] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally i K. Keutzer, “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size,” w *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, s. 6861–6869. DOI: 10.1109/CVPR.2017.732.
- [101] R. Bogacz, E. Brown, J. Moehlis, P. Holmes i J. D. Cohen, “The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks.,” *Psychological Review*, t. 113, nr. 4, s. 700, 2006.
- [102] R. D. Luce i W. Edwards, “The derivation of subjective scales from just noticeable differences.,” *Psychological Review*, t. 65, nr. 4, s. 222, 1958.
- [103] I. L. Vér i L. L. Beranek, *Noise and vibration control engineering: principles and applications*. John Wiley & Sons, 2005.

- [104] R. Ulichney, "Dithering with blue noise," *Proceedings of the IEEE*, t. 76, nr. 1, s. 56–79, 1988. DOI: 10.1109/5.3288.
- [105] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Transactions on Consumer Electronics*, t. 38, nr. 1, s. 18–34, 1992.
- [106] I. Goodfellow, Y. Bengio i A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [107] L. Gondara, "Medical Image Denoising Using Convolutional Denoising Autoencoders," w *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, 2016, s. 241–246. DOI: 10.1109/ICDMW.2016.0041.
- [108] A. Majumdar, "Blind Denoising Autoencoder," *IEEE Transactions on Neural Networks and Learning Systems*, t. 30, nr. 1, s. 312–317, 2019. DOI: 10.1109/TNNLS.2018.2838679.
- [109] E. Agustsson i R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," w *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) workshops*, 2017, s. 126–135.
- [110] H. R. Sheikh, M. F. Sabir i A. C. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *IEEE Transactions on Image Processing*, t. 15, nr. 11, s. 3440–3451, 2006. DOI: 10.1109/TIP.2006.881959.
- [111] E. C. Larson i D. M. Chandler, "Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy," *Journal of Electronic Imaging*, t. 19, nr. 1, s. 011 006, 2010. DOI: 10.1117/1.3281673.
- [112] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli i F. Battisti, "TID2008 – A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics," *Advances of Modern Radioelectronics*, t. 10, nr. 4, s. 30–45, 2009. adr.: <http://www.computing.dcu.ie/~kieran/TID2008.pdf>.
- [113] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti i in., "Image database TID2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, t. 30, s. 57–77, 2015. DOI: 10.1016/j.image.2014.10.021.
- [114] N. Murray, L. Marchesotti i F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," w *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2012, s. 2408–2415.

- [115] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto i H. Adam, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.
- [116] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke i A. Rabinovich, “Going deeper with convolutions,” w *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, s. 1–9.
- [117] c. Moscow State University: Video processing i quality research group, *PSNR and SSIM: application areas and criticism*, <https://videoprocessing.ai/metrics/ways-of-cheating-on-popular-objective-metrics.html>, Accessed on: May 30, 2023, 2023.
- [118] M. Sharif, L. Bauer i M. K. Reiter, “On the suitability of lp-norms for creating and preventing adversarial examples,” w *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) workshops*, 2018, s. 1605–1613.
- [119] H. Zhao, O. Gallo, I. Frosio i J. Kautz, “Loss functions for image restoration with neural networks,” *IEEE Transactions on Computational Imaging*, t. 3, nr. 1, s. 47–57, 2016.
- [120] T. Tarasiewicz, J. Nalepa i M. Kawulok, “Semi-Simulated Training Data for Multi-Image Super-Resolution,” w *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, s. 481–484. DOI: 10.1109/IGARSS46834.2022.9884565.
- [121] H. Sheikh, “LIVE image quality assessment database release 2,” <http://live.ece.utexas.edu/research/quality>, 2005.
- [122] D. Kostrzewa, Ł. Skonieczny, P. Benecki i M. Kawulok, “B4MultiSR: a benchmark for multiple-image super-resolution reconstruction,” w *Beyond Databases, Architectures and Structures. Facing the Challenges of Data Proliferation and Growing Variety: 14th International Conference, BDAS 2018, Held at the 24th IFIP World Computer Congress, WCC 2018, Poznan, Poland, September 18-20, 2018, Proceedings 14*, Springer, 2018, s. 361–375.
- [123] J. Cornebise, I. Orsolice i F. Kalaitzis, “Open High-Resolution Satellite Imagery: The WorldStrat Dataset – With Application to Super-Resolution,” w *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022. adr.: <https://openreview.net/forum?id=DEigo9L8xZA>.
- [124] X.-S. Yang, *Nature-inspired optimization algorithms*. Academic Press, 2020.
- [125] G. Venter, “Review of optimization techniques,” 2010.

- [126] J. Kennedy i R. Eberhart, "Particle swarm optimization," w *Proceedings of ICNN'95-international conference on neural networks*, IEEE, t. 4, 1995, s. 1942–1948.
- [127] B. Ahrens, "Genetic algorithm optimization of superresolution parameters," w *Proceedings of the 7th annual Conference on Genetic and Evolutionary Computation*, 2005, s. 2083–2088.
- [128] M. Kawulok, P. Benecki, S. Piechaczek, K. Hrynczenko, D. Kostrzewa i J. Nalepa, "Deep Learning for Multiple-Image Super-Resolution," *IEEE Geoscience and Remote Sensing Letters*, t. 17, nr. 6, s. 1062–1066, 2020. DOI: 10.1109/LGRS.2019.2940483.
- [129] M. Kawulok, P. Benecki, D. Kostrzewa i L. Skonieczny, "Evolving imaging model for super-resolution reconstruction," w *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 2018, s. 284–285.
- [130] M. M. Afrasiabi, R. Hosseini i A. Abbasfar, "Combination of Single and Multi-frame Image Super-resolution: An Analytical Perspective," *arXiv preprint arXiv:2303.03212*, 2023.
- [131] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, t. 60, s. 91–110, 2004.
- [132] S. Leutenegger, M. Chli i R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," w *Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV)*, Ieee, 2011, s. 2548–2555.
- [133] H. Bay, A. Ess, T. Tuytelaars i L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, t. 110, nr. 3, s. 346–359, 2008.
- [134] M. Leszczuk, "Assessing Task-Based Video Quality — A Journey from Subjective Psycho-Physical Experiments to Objective Quality Models," w *Multimedia Communications, Services and Security*, A. Dziech i A. Czyżewski, red., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, s. 91–99, ISBN: 978-3-642-21512-4.
- [135] H. R. Sheikh, A. C. Bovik i G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, t. 14, nr. 12, s. 2117–2128, 2005.
- [136] S. F. Gull i J. Skilling, "Maximum entropy method in image processing," w *IET Proceedings on Communications, Radar and Signal Processing*, IET, t. 131, 1984, s. 646–659.

- [137] A. Barroso-Laguna, E. Riba, D. Ponsa i K. Mikolajczyk, “Key. net: Keypoint detection by handcrafted and learned cnn filters,” w *Proceedings of the 2019 IEEE International Conference on Computer Vision (ICCV)*, 2019, s. 5836–5844.

Lista skrótów i akronimów

2AFC – ang. *two-alternative forced choice*, metoda zbierania danych ankietowych

AD – ang. *artificially degraded*, obrazy sztucznie degradowane

ARSD – ang. *average relative standard deviation*, średnie względne odchylenie standardowe

BRISK – ang. *binary robust invariant scalable keypoints*, jedna z metod wykrywania i opisywania punktów charakterystycznych obrazu

CNN – ang. *convolutional neural network*, spłotowa sieć neuronowa

cPSNR – ang. *corrected peak signal-to-noise ratio*, skorygowany stosunek sygnału szczytowego do szumu, jedna z metryk oceny podobieństwa obrazów

CSBA – ang. *compressed-sensing based approach*, grupa metod rekonstrukcji nadrozdzielczej

CUDA – ang. *compute unified device architecture*, interfejs programistyczny procesorów graficznych firmy NVIDIA

DBPN – ang. *deep back-projection network*, jedna z metod rekonstrukcji nadrozdzielczej

DMOS – ang. *differential mean opinion score*, sposób opisu danych zebranych z ankiet

EDSR – ang. *enhanced deep residual networks*, jedna z metod rekonstrukcji nadrozdzielczej

EME – ang. *enhancement by entropy*, bezreferencyjna miara oceny jakości obrazu

ESA – ang. *European Space Agency*, Europejska Agencja Kosmiczna

EvoIM – metoda rekonstrukcji nadrozdzielczej działająca w oparciu na algorytmie ewolucyjnym

EvoNet – metoda rekonstrukcji nadrozdzielczej działająca w oparciu na algorytmie ewolucyjnym i sieci konwolucyjnej

FSRCNN – ang. *fast super resolution convolutional network*, jedna z metod rekonstrukcji nadrozdzielczej

FSRS – ang. *fast and robust super resolution*, jedna z metod rekonstrukcji nadrozdzielczej

GA – ang. *genetic algorithm*, algorytm genetyczny

GAN – ang. *generative adversarial network*, sieć generatywna

GMSD – ang. *gradient magnitude similarity deviation*, jedna z metryk oceny podobieństwa obrazów

GPA – ang. *gradient projection algorithm*, jedna z metod rekonstrukcji nadrozdzielczej

GPU – ang. *graphics processing unit*, procesor graficzny

GSD – ang. *ground sampling distance*, miara rozdzielczości przestrzennej obrazów lotniczych / satelitarnych

HR – ang. *high resolution*, wysoka rozdzielczość

- IBP** – ang. *iterative back projection*, projekcja wsteczna
- IFC** – ang. *information fidelity criterion*, jedna z metryk podobieństwa obrazów
- IM** – ang. *imaging model*, model obrazowania
- JND** – ang. *just noticeable differences*, metoda zbierania danych ankietowych
- KFS** – ang. *keypoint feature similarity*, jedna z metryk podobieństwa obrazów
- LIVE** – Laboratory for Image and Video Engineering, University of Texas
- LPIPS** – ang. *learned perceptual image patch similarity*, jedna z metryk podobieństwa obrazów
- LR** – ang. *low resolution*, niska rozdzielczość
- MAE** – ang. *mean absolute error*, średni błąd bezwzględny
- MISR** – ang. *multiple image super resolution*, rekonstrukcja nadrozdzielcza wieloobrazowa
- MOS** – ang. *mean opinion score*, sposób opisu danych zebranych z ankiet
- MSA** – ang. *median shift and add*, metoda fuzji obrazów przesuniętych na poziomie subpikselowym
- MSE** – ang. *mean squared error*, błąd średniokwadratowy
- NIMA** – ang. *neural image assessment*, jedna z metryk podobieństwa obrazów
- NIR** – ang. *near infrared*, bliska podczerwień
- NN** – ang. *nearest-neighbor interpolation*, interpolacja metodą najbliższego sąsiada
- POCS** – ang. *projection onto convex sets*, projekcja na zbiory wypukłe
- PSF** – ang. *point spread function*, funkcja opisująca mapowanie punktu między rozdzielczościami
- PSNR** – ang. *peak signal-to-noise ratio*, stosunek sygnału szczytowego do szumu, jedna z metryk oceny podobieństwa obrazów
- RAMS** – ang. *residual attention multi-image super-resolution*, jedna z metod rekonstrukcji nadrozdzielczej
- RS** – ang. *real satellite images*, zbiór obrazów satelitarnych niesymulowanych
- RSD** – ang. *relative standard deviation*, względne odchylenie standardowe
- SIFT** – ang. *scale invariant feature transform*, jedna z metod wykrywania i opisu punktów charakterystycznych obrazu
- SISPARE** – ang. *Satellite Image Spatial Resolution Enhancement*, projekt dotyczący rozwoju metod rekonstrukcji nadrozdzielczej
- SISR** – ang. *single image super resolution*, rekonstrukcja nadrozdzielcza jednoobrazowa
- SR** – ang. *super resolution*, rekonstrukcja nadrozdzielcza
- SRBA** – ang. *sparse representation-based approach*, grupa podejść do rekonstrukcji nadrozdzielczej oparta na reprezentacji rzadkiej
- SRCNN** – ang. *super resolution convolutional neural network*, jedna z metod rekonstrukcji nadrozdzielczej

SRGAN – ang. *super-resolution generative adversarial network*, jedna z metod rekonstrukcji nadrozdzielczej

SSIM – ang. *structure similarity index metric*, jedna z metryk podobieństwa obrazów

SURF – ang. *speeded up robust features*, jedna z metod wykrywania i opisu punktów charakterystycznych obrazu

TEFA – ang. *temporally equivariant feature attention*, element modeli rekonstrukcji nadrozdzielczej

TERN – ang. *temporally equivariant registration network*, element modeli rekonstrukcji nadrozdzielczej

VGG – ang. *visual geometry group*, standardowa sieć głęboka ekstrahująca wielopoziomowo cechy obrazu

VIF – ang. *visual information fidelity*, jedna z metryk podobieństwa obrazów

Spis symboli

Symbole ogólne

L1 – metryka L1

L2 – metryka L2

μ – średnia

σ – odchylenie standardowe

σ^2 – wariancja

\mathcal{I} – obraz

$\mathcal{I}^{(h)}$ – obraz wysokiej rozdzielczości

$\mathcal{I}^{(l)}$ – obraz niskiej rozdzielczości

$\mathcal{I}^{(sr)}$ – obraz zrekonstruowany

$\mathbf{I}^{(L)}$ – zbiór obrazów niskiej rozdzielczości

Model obrazowania

\mathbf{D} – macierz opisująca zmniejszanie rozdzielczości

\mathbf{B} – macierz opisująca operację rozmycia

\mathbf{W} – macierz opisująca przekształcenie geometryczne homograficzne: skalowanie, przesunięcie i obrót

n – szum dodany do obrazu

Metody SR z regularyzacją

ρ – miara niepodobieństwa między obrazami

U – regularyzacja

λ – współczynnik wpływu regularyzacji

Metoda FRSR

\hat{Z} – wynik operacji shift-and-add

\mathbf{H} – macierz opisująca rozmycie punktu (ang. *point spread function*)

\mathbf{S} – macierz przesunięcia

\mathbf{A} – macierz opisująca liczbę obserwacji LR dla danego piksela we współrzędnych HR w wyniku operacji shift-and-add

α, β, λ – hiperparametry metody FRSR

Miary podobieństwa obrazu

\mathcal{M} – maksymalna możliwa wartość przyjmowana przez piksele obrazów w stosowanej rozdzielczości bitowej w metodzie SSIM

σ_{ij} – kowariancja między wartościami obrazu w oknach i oraz j

$c, c1, c2$ – stałe stabilizujące numerycznie wynik w miarach SSIM oraz GMSD

b – przesunięcie jasności między obrazami porównywanymi w cPSNR i cSSIM

$m(i)$ – magnituda gradientu obrazu w punkcie i (metoda GMSD)

H – wartość entropii

$p(i)$ – prawdopodobieństwo wystąpienia określonej intensywności piksela

$w_{k,l}$ – okna, na które jest dzielony wejściowy obraz (miara EME)

$I_{max;k;l}^w$ oraz $I_{min;k;l}^w$ – minimum i maksimum wartości pikseli w obrazie X wewnątrz bloku $w_{k,l}$ (miara EME)

EvoIM / EvoNet

\mathbf{B} – macierz opisująca gaussowskie rozmycie punktu

\mathcal{X}_0 – pierwsze oszacowanie obrazu wynikowego

$\Delta\mathcal{X}$ – krok gradientu w optymalizacji

β – hiperparametr kontrolujący krok metody

\mathbf{A} – macierz zawierająca informację o wkładzie poszczególnych obrazów LR do \mathcal{X}_0

$U(\mathcal{X})$ – składnik odpowiedzialny za regularyzację

\mathbf{S} – macierz przesunięcia

λ – hiperparametr określający wpływ operacji regularyzacji

α – hiperparametr określający zasięg operacji regularyzacji

P – hiperparametr określający liczbę pełnopikselowych przesunięć wykonywanych w trakcie regularyzacji

\mathcal{P} – populacja w algorytmie ewolucyjnym

p – osobnik w algorytmie ewolucyjnym

η – wartość funkcji przystosowania osobnika

Metody KFS / DKFS

C – współczynnik skalujący wynik

\mathcal{P} – punkty kluczowe obrazu

\mathcal{D} – deskryptory punktów kluczowych

ρ – sposób obliczania odległości między deskryptorami

$\mathcal{H}_1, \mathcal{H}_2$ – głosy uzyskane od obserwatorów LPIPS

\mathcal{M} – metryka używana w DKFS

\mathcal{V} – wartość głosu uzyskanego za pomocą metryki \mathcal{M} w DKFS

Miara oparta na cechach głębokich w MISR

\mathcal{L} – funkcja straty

$\mathcal{L}_1, \mathcal{L}_2$ – składowe funkcje straty

w_1, w_2 – wagi składowych funkcji straty

Połączenie SISR i MISR – ankieta

Q – numer pytania

S – informacja o zbiorze (AD / RS), którego dotyczy pytanie

M – wskazana jako odpowiedź metoda

A – sposób interpretacji odpowiedzi: czy w pytaniu chodzi o wskazanie najlepszego ($A = 1$), czy najgorszego ($A = -1$) jakościowo obrazu spośród prezentowanych

Miara oparta na cechach głębokich w MISR – ankieta

M, X, Y - metody, za pomocą których zostały otrzymane badane obrazy

$\mathcal{I}^{(h)}$ – obraz referencyjny

V – wynik postaci 0, 1, wskazujący czy w danej kombinacji (X, Y) bliższy obrazowi referencyjnemu jest X ($V = 0$) czy Y ($V = 1$)

P_i — liczba oznaczająca w ilu wypadkach wynik metody M był wybrany jako bardziej zbliżony do $\mathcal{I}^{(h)}$