

**RADA NAUKOWA DYSCYPLINY
INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI WARSZAWSKIEJ**

zaprasza na

OBRONĘ ROZPRAWY DOKTORSKIEJ

mgr. inż. Tomasza Marka Lehmana

która odbędzie się w dniu **29 września 2026 roku**, o godzinie **13:15** w trybie hybrydowym

Temat rozprawy:

***“Zastosowanie lekkich architektur głębokich w cyfrowej percepcyjnej
analizie i korekcji degradacji obrazu”***

Promotor: prof. dr hab. inż. Przemysław Rokita – Politechnika Warszawska

Recenzenci: dr hab. inż. Leszek Chmielewski, prof. uczelni – Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

dr hab. inż. Radosław Mantiuk, prof. uczelni – Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

prof. dr hab. inż. Adam Wojciechowski – Politechnika Łódzka

Obrona odbędzie się online oraz stacjonarnie **w Sali nr 116** Wydziału Elektroniki i Technik Informatycznych Politechniki Warszawskiej w Warszawie.

Z rozprawą doktorską i recenzjami można zapoznać się w Czytelni Biblioteki Głównej Politechniki Warszawskiej, Warszawa, Plac Politechniki 1.

Streszczenie rozprawy doktorskiej i recenzje są zamieszczone na stronie internetowej: <https://www.bip.pw.edu.pl/Postepowania-w-sprawie-nadania-stopnia-naukowego/Doktoraty/Wszczete-po-30-kwietnia-2019-r/Rada-Naukowa-Dyscypliny-Informatyka-Techniczna-i-Telekomunikacja/mgr-inz.-Tomasz-Lehmann>

Przewodniczący Rady Naukowej Dyscypliny
Informatyka Techniczna i Telekomunikacja
Politechniki Warszawskiej
prof. dr hab. inż. Jarosław Arabas

Zastosowanie lekkich architektur głębokich w cyfrowej percepcyjnej analizie i korekcji degradacji obrazu

Celem pracy było rozwiązanie wybranych problemów rekonstrukcji obrazu poprzez opracowanie oraz ewaluację algorytmów sztucznej inteligencji. Zaproponowane metody, przy zachowaniu złożoności obliczeniowej porównywalnej z wiodącymi rozwiązaniami (ang. *state-of-the-art*), zapewniają jakość percepcyjną na poziomie najlepszych podejść opisywanych w literaturze przedmiotu.

Wyodrębniono trzy główne obszary rekonstrukcji obrazu: usuwanie efektu mory, poprawę jakości w słabym oświetleniu oraz redukcję rozmycia. Równoległym celem prowadzonych prac było opracowanie i praktyczne wykorzystanie bezwzorcowych algorytmów obiektywnej estymacji jakości percepcyjnej.

Nadrzędnym założeniem było wykorzystanie zaawansowanego zarządzania danymi oraz specjalizowanych funkcji straty, zgodnie z ideą zrównoważonego rozwoju sieci neuronowych. Podejście to umożliwiło podstawowym architekturom osiągnięcie wyników możliwie zbliżonych do tych uzyskiwanych przez rozwiązania o znacznie większej liczbie parametrów lub wymagające wieloetapowej inferencji, a w wybranych przypadkach także ich przewyższanie.

Zrealizowano cele badawcze obejmujące opracowanie metod redukcji degradacji obrazu oraz zaprojektowanie modelu oceny jakości. Opracowano model bezwzorcowej oceny jakości obrazu (ang. *Blind Image Quality Assessment*, BIQA) do estymacji miary wrażeniowej (ang. *Mean Opinion Score*, MOS) wraz z nowatorską procedurą równoważenia etykiet. Zaprojektowano wydajne algorytmy redukcji efektu mory oraz rekonstrukcji obrazów w słabym oświetleniu. Wykazano również skuteczność zastosowania kompozytowych funkcji straty zawierających składniki percepcyjne — zarówno w powyższych zadaniach, jak i w problemie redukcji rozmycia.

Dodatkowym wkładem w rozwój dziedziny jest udostępnienie ustandaryzowanego zbioru obrazów typu *text-to-image*. Został on wykorzystany do weryfikacji modelu predykcji jakości obrazu w kontekście generatywnej sztucznej inteligencji, gdzie metody bezwzorcowe znajdują szczególne zastosowanie.

Wyniki badań były sukcesywnie publikowane w czasopismach ujętych w wykazie ministerialnym oraz prezentowane na konferencjach o zasięgu międzynarodowym.

dr hab. inż. Leszek J. Chmielewski, prof. SGGW
Instytut Informatyki Technicznej
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
ul. Nowoursynowska 159, 02-776 Warszawa

Recenzja rozprawy doktorskiej

autor rozprawy:

mgr inż. Tomasz Lehmann

Instytut Informatyki, Politechnika Warszawska

tytuł rozprawy:

Zastosowanie lekkich architektur głębokich w cyfrowej percepcyjnej analizie i korekcji degradacji obrazu

Recenzja została przygotowana w odpowiedzi na powołanie, uchwałą Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej z dnia 24 marca 2026 r., na recenzenta rozprawy doktorskiej Pana mgr inż. Tomasz Lehmana. Promotorem rozprawy jest prof. dr hab. inż. Przemysław Rokita. Przewód jest prowadzony w dyscyplinie *informatyka techniczna i telekomunikacja*.

1 Zawartość rozprawy

1.1 Charakterystyka ogólna

Rozprawa doktorska dotyczy zastosowania lekkich architektur głębokich w percepcyjnej analizie i korekcji degradacji obrazu cyfrowego. Jej zasadniczym celem jest opracowanie metod poprawy jakości obrazów, które łączą wysoką skuteczność rekonstrukcji z umiarkowanym, czy też ograniczonym kosztem obliczeniowym. Autor koncentruje się na trzech istotnych problemach przetwarzania obrazów: redukcji efektu mory, wzmacnianiu obrazów wykonanych w warunkach niedoświetlenia oraz usuwaniu rozmycia. Równolegle rozwijany jest wątek bezwzorcowej oceny jakości obrazu, w szczególności predykcji miary MOS, która jest własnym osiągnięciem Doktoranta, a następnie wykorzystania tej wiedzy percepcyjnej w zadaniach rekonstrukcyjnych i w ocenie obrazów generowanych przez modele *text-to-image*.

Rozprawa składa się z ośmiu rozdziałów i bibliografii.

Rozdział pierwszy pełni funkcję wprowadzenia do tematyki rozprawy. Przedstawiono w nim kontekst naukowy pracy, wskazując na znaczenie percepcyjnej jakości obrazu we współczesnym widzeniu komputerowym oraz na ograniczenia, jakimi charakteryzują się klasyczne miary jakości. Jak zauważa Doktorant, są one słabo skorelowane z subiektywnym odbiorem człowieka. Omówiono główne typy degradacji obrazu rozważane w rozprawie, uzasadnia potrzebę stosowania metod efektywnych obliczeniowo oraz formułuje cele szczegółowe pracy. W rozdziale tym wprowadzono również operacyjne definicje takich pojęć, jak efektywność obliczeniowa, ograniczona złożoność, lekka architektura oraz wydajny algorytm, dla ujednolicenia sensu tych definicji i umożliwienia nadania im cechy mierzalności.

Rozdział drugi zawiera przegląd literatury dotyczącej badanych zagadnień. Omówiono w nim metody obiektywnej oceny jakości obrazów, przy czym szczególnie zwrócono uwagę na metody bezwzorcowe (ang. *Blind Image Quality Assessment*, BIQA), a także aktualny stan badań nad rekonstrukcją obrazów zdegradowanych przez mórę, niedoświetlenie i rozmycie. Rozdział ten porządkuje tło badawcze rozprawy oraz wskazuje luki, które motywują zaproponowane przez autora rozwiązania.

Rozdział trzeci poświęcono badaniom nad algorytmami percepcyjnej, bezwzorcowej oceny jakości obrazów. Autor wykorzystuje zbiór KonIQ-10k i analizuje zaproponowane architektury predykcji MOS, w tym modele dual-Xception oraz ConvNeXt+MLP. Szczególną uwagę zwrócono na wpływ strategii uczenia, rozdzielczości wejściowej i sposobu doboru danych na jakość predykcji oraz koszt obliczeniowy.

Rozdział czwarty dotyczy redukcji efektu mory w obrazach cyfrowych. Zaproponowano w nim metodę opartą na architekturze Attention U-Net, a następnie przeanalizowano wpływ różnych funkcji straty i strategii doboru próbek na skuteczność rekonstrukcji. Badania prowadzone są z wykorzystaniem zbiorów danych przeznaczonych do oceny algorytmów usuwania mory, a ich celem jest wykazanie, że konkurencyjna jakość percepcyjna może być osiągnięta przy niższej złożoności obliczeniowej.

Rozdział piąty przedstawia badania nad wzmacnianiem obrazów wykonanych w warunkach niedoświetlenia. Autor analizuje architekturę Swin-UNet, strategię integracji danych oraz złożone funkcje straty łączące składniki pikselowe i percepcyjne. Rozdział ten rozwija główną ideę rozprawy, zgodnie z którą poprawa jakości obrazu powinna być oceniana nie tylko za pomocą miar klasycznych, lecz również z uwzględnieniem jakości postrzeganej przez człowieka.

Rozdział szósty rozszerza wykorzystanie opracowanego predyktora MOS na problem redukcji rozmycia. Przedstawiono w nim hybrydową funkcję straty, w której składnik percepcyjny oparty na estymowanej wartości MOS uzupełnia kryteria pikselowe i strukturalne. Celem tego rozdziału jest sprawdzenie, czy wiedza o percepcyjnej jakości obrazu może korzystnie wpływać na proces uczenia modelu rekonstrukcyjnego.

Rozdział siódmy wprowadza zestaw testowy DIQA – zrównoważony zbiór obrazów – oraz protokół oceny obrazów generowanych z opisów tekstowych. Autor wykorzystuje opracowany wcześniej predyktor MOS do porównania jakości obrazów uzyskanych z różnych modeli dyfuzyjnych. Ten fragment rozprawy rozszerza zakres zastosowania metod BIQA poza klasyczną fotografię, obejmując także treści syntetyczne generowane przez współczesne modele sztucznej inteligencji.

Rozdział ósmy zawiera podsumowanie uzyskanych wyników, wnioski końcowe oraz rekomendacje dotyczące dalszych badań. Autor syntetyzuje w nim najważniejsze osiągnięcia rozprawy i podkreśla znaczenie lekkich architektur, funkcji straty zorientowanych na percepcję człowieka, oraz bezwzorcowych metod oceny jakości obrazu dla efektywnego rozwoju algorytmów rekonstrukcji obrazu.

1.2 Cele pracy

Kontekst naukowy i problematyka badań zostały bardzo czytelnie przedstawione w rozdziale 1.1. Na tej podstawie, w rozdziale 1.2, sformułowano pięć celów badań, których nazwy tu zacytuje – z opisem skróconym:

1. Opracowanie bezwzorcowego predyktora MOS o ograniczonym koszcie zasobowym – realizacja modelu lekkiego, w tym zaprojektowanie modeli BIQA.
2. Opracowanie metody redukcji efektu mory przy umiarkowanej złożoności – realizacja wydajnego algorytmu.
3. Zaprojektowanie metody rekonstrukcji obrazów niedoświetlonych z uwzględnieniem metryk percepcyjnych – realizacja poprawy jakości względem metryk zarówno klasycznych, jak i percepcyjnych
4. Zdefiniowanie funkcji straty ze składnikiem opartym na MOS dla usuwania rozmycia – weryfikacja wpływu użycia funkcji na jakość rekonstrukcji w porównaniu do funkcji klasycznych – umożliwienie powtarzalnych porównań architektur i raportowanie wyników w mierzalnej postaci.

Autor zdefiniował pojęcia takie, jak efektywność, wydajność, ograniczona złożoność, lekka architektura, aby w każdym przypadku móc uzyskać mierzalne wyniki.

2 Dyskusja zawartości i wyników rozprawy

Tematyka rozprawy jest istotna zarówno z punktu widzenia rozwoju informatyki technicznej, jak i z perspektywy praktycznych zastosowań metod sztucznej inteligencji. Cyfrowe obrazy są obecnie podstawowym nośnikiem informacji w diagnostyce, systemach bezpieczeństwa, multimedialach, robotyce, dokumentacji technicznej i licznych usługach konsumenckich. Ich degradacja, wynikająca między innymi z niedoświetlenia, rozmycia, ograniczeń sensorów lub efektu

mory, może obniżać nie tylko estetyczną jakość obrazu, lecz także jego użyteczność interpretacyjną. Dlatego badania nad automatyczną oceną i poprawą jakości obrazów należą do ważnych kierunków widzenia komputerowego.

Szczególnie wartościowe jest przyjęcie w rozprawie ukierunkowanie na jakość percepcyjną, a więc taką, która jest bliższa rzeczywistemu odbiorowi obrazu przez człowieka niż klasyczne miary błędu wynikające z dziedziny analizy sygnałów. Ma to duże znaczenie, ponieważ w wielu zastosowaniach obraz nie jest oceniany wyłącznie jako macierz piksli, ale jako informacja wizualna przeznaczona do ludzkiej interpretacji albo do dalszej analizy przez systemy automatyczne. Często, albo nawet najczęściej, w praktyce nie dysponujemy wzorcowym obrazem niezakłóconym. Dlatego opracowanie bezwzorcowych metod oceny jakości odpowiada na realny problem praktyczny.

Pomimo wielkiego wzrostu możliwości obliczeniowych, wynikających tak z przyspieszania zegarów, jak i ze zrównoleglenia obliczeń, efektywność obliczeniowa jest niezmiernie ważna. Dostępność wielkich zasobów jako usług chmurowych sprawia, że zbliżamy się już do granic możliwości, jakie można uzyskać w praktyce, jednocześnie doświadczając poważnego wzrostu kosztów takich obliczeń. Stąd, wielu badań, praktycznie biorąc, nie można powtórzyć, co ogranicza wiarygodność wyników. Dlatego poświęcenie znaczącego wysiłku Doktoranta ukierunkowanego na uzyskanie metod efektywnych obliczeniowo jest cenną cechą tej dysertacji. Poszukiwanie lekkich architektur głębokich, które zachowują wysoką jakość wyników przy umiarkowanej złożoności, jest zatem ważne nie tylko naukowo, lecz również ekonomicznie i zdroworozsądkowo. W tym sensie rozprawa wpisuje się w aktualny nurt badań nad skutecznymi, ale zarazem racjonalnie uzasadnionymi metodami sztucznej inteligencji.

2.1 Uwagi pozytywne

2.1.1 Ważność tematyki rozprawy

Jakość obrazu cyfrowego ma znaczenie w wielu obszarach współczesnej informatyki i techniki, od multimediów po zastosowania diagnostyczne i przemysłowe. Praktyka pokazuje, że często nie można uniknąć degradacji obrazu, zatem automatyczna korekcja stanowi aktualny i praktycznie istotny problem badawczy.

Istotnym walorem pracy jest ukierunkowanie na jakość percepcyjną, bliższą rzeczywistemu odbiorowi obrazu przez człowieka niż klasyczne miary sygnałowe. Szczególne znaczenie ma tu rozwijanie metod bezwzorcowej oceny jakości, które mogą być stosowane także w tych częstych przypadkach, gdy obraz niezakłócony nie jest dostępny.

Dążenie do wykorzystania lekkich architektur głębokich, nie zaś uzyskiwanie lepszych efektów dzięki zwiększaniu zapotrzebowania na moce obliczeniowe, jest istotne tak z naukowego, jak i praktycznego punktu widzenia.

2.1.2 Opis stanu wiedzy

Doktorant prawidłowo opisał stan wiedzy w zakresie problematyki rozprawy, przyczym umieścił Można to można ująć w następujących punktach:

1. Co do obiektywnej oceny jakości obrazów rozwijane są metody pełnowzorcowe, ze wzorcem zredukowanym oraz bezwzorcowe. Szczególne znaczenie praktyczne mają metody bezwzorcowe, ze względu na częsty brak obrazu referencyjnego. Kierunek jest trudny, między innymi ze względu na ograniczoną dostępność oznaczonych zbiorów danych oraz złożoność modelowania subiektywnej percepcji jakości.
2. W zakresie redukcji efektu mory stosowane są obecnie metody oparte na głębokich sieciach neuronowych, w tym architektury splotowe, rozwiązania wieloskalowe, sieci typu U-Net, mechanizmy uwagi oraz modele wyspecjalizowane dla obrazów wysokiej rozdzielczości. Istniejące rozwiązania osiągają dobre wyniki, lecz często wiążą się z wysokim kosztem obliczeniowym. Uzasadnia to poszukiwanie metod prostszych i lżejszych, przy porównywalnej jakości rekonstrukcji.
3. Rekonstrukcja obrazów niedoświetlonych przeszła ewolucję od klasycznych metod mapowania tonalnego i podejść typu Retinex do złożonych modeli uczenia głębokiego. Współ-

czesne metody muszą godzić kilka sprzeczne wymagania: poprawę jasności, ograniczenie szumu, zachowanie barw i tekstur oraz akceptowalną złożoność obliczeniową. Obiecującym kierunkiem są architektury łączące własności sieci spłotowych, struktur enkoder-dekoder oraz mechanizmów uwagi.

4. W zadaniach rekonstrukcji obrazu coraz większe znaczenie zyskują funkcje straty lepiej powiązane z percepcją człowieka, niż klasyczne straty pikslowe. Obok miar takich jak SSIM i strat percepcyjnych opartych na cechach głębokich sieci, rozwijane są podejścia wykorzystujące modele oceny jakości obrazu. Włączenie estymatora MOS (autorskiego estymatora typu BIQA Doktoranta) do funkcji straty stanowi próbę zmniejszenia rozbieżności między optymalizacją numeryczną a subiektywną oceną jakości.
5. Ocena jakości obrazów generowanych z tekstu jest nowym i szybko rozwijającym się problemem, związanym z postępem modeli dyfuzyjnych i systemów *text-to-image*. Pomimo dużego zainteresowania tą tematyką brakuje powszechnie przyjętego standardu oceny jakości takich obrazów. Z tego względu wydaje się potrzebne opracowanie protokołów testowych i miar pozwalających porównywać modele generatywne z uwzględnieniem jakości percepcyjnej.

2.1.3 Oryginalne elementy rozprawy

W tym zakresie Doktorant, przeprowadzając autoanalizę swojej pracy w rozdziałach, wymienił elementy oryginalne i opisał je w sposób, z którym się zgadzam. Tym samym, sam dostarczył recenzentowi odpowiedni opis, który tu zrekapituluje.

1. Opracowanie bezwzorcowych metod estymacji jakości percepcyjnej obrazu, opartych na predykcji wartości MOS. Wkład ten obejmuje zaproponowanie architektur dual-Xception oraz ConvNeXt+MLP z dodatkowym torem przetwarzania parametrów globalnych obrazu, a także zastosowanie procedury równoważenia rozkładu etykiet i kryterium selekcji modelu ukierunkowanego na poprawę jakości estymacji w różnych zakresach skali MOS.
2. Opracowanie metody redukcji efektu mory z wykorzystaniem architektury Attention U-Net. Istotnym elementem tego wkładu jest analiza funkcji straty łączących składniki pikslowe i strukturalne oraz zaproponowanie strategii rotacyjnego doboru próbek, pozwalającej ograniczyć koszt uczenia przy zachowaniu konkurencyjnej jakości rekonstrukcji.
3. Zaprojektowanie algorytmu rekonstrukcji obrazów wykonanych w warunkach niedoświetlenia, opartego na architekturze Swin-UNet i złożonej funkcji straty zawierającej składniki percepcyjne. Na uwagę zasługuje także opracowanie jednolitego protokołu treningowego dla kilku zbiorów danych oraz analiza wpływu strategii doboru obrazów ze zbioru SID na jakość i stabilność uczenia.
4. Zweryfikowanie użyteczności estymatora MOS jako składnika percepcyjnego w hybrydowej funkcji straty dla zadania redukcji rozmycia. Praca pokazuje, że włączenie takiego komponentu może poprawiać jakość rekonstrukcji względem wariantów wykorzystujących wyłącznie kryteria pikslowe lub strukturalne.
5. Zdefiniowanie i udostępnienie zestawu testowego DIQA dla obrazów generowanych przez modele *text-to-image* oraz zaproponowanie protokołu ich oceny opartego na estymowanej wartości MOS. Wkład ten rozszerza zastosowanie metod bezwzorcowej oceny jakości poza klasyczne obrazy fotograficzne, obejmując również treści syntetyczne generowane przez modele dyfuzyjne.

Wspólną cechą wymienionych osiągnięć jest dążenie do poprawy jakości percepcyjnej obrazów bez eskalowania złożoności modeli, a więc połączenie skuteczności rekonstrukcji, oceny zgodnej z ludzką percepcją, oraz racjonalnego wykorzystania zasobów obliczeniowych.

2.1.4 Weryfikacja metod

Autor przywiązuje wielką wagę do weryfikowalności i powtarzalności wyników badanych metod. Wszystkie metody są przetestowane na powszechnie dostępnych, dużych bazach obrazów, w tym baz obrazów z opisami wykonanymi przez ludzi.

Każdy rozdział jest zakończony wnioskami, gdzie podsumowuje się treść na podstawie eksperymentów. Dla przykładu, wnioski w rozdziale 3.4, w których są podsumowane cechy autorskiej

metody Doktoranta ConvNeXt+MLP, wskazują, że ta metoda jest konkurencyjna dzięki temu, że uzyskuje się za jej pomocą wskaźniki bliskie najwyższych dla metod konkurencyjnych, przy użyciu zasobów znacznie mniejszym. Jest to oparte na wynikach pokazanych w tabelach i na wykresach w tym rozdziale. W szczególności, oszczędność czasu treningu pokazano w Tab. 3.3.4.

W pozostałych rozdziałach jest podobnie.

2.1.5 Propozycje dalszych badań

Każdy rozdział jest uzupełniony przez propozycję dalszych badań. Propozycje te są mocno oparte na doświadczeniach zdobytych podczas prac nad metodami i są solidnie uzasadnione. Propozycje te mogą stać się zaczątkiem dalszych badań Autora i kontynuatorów jego koncepcji.

2.1.6 Publikacje

Z zakresu tematycznego rozprawy doktorskiej Doktorant ma pięć publikacji: cztery we współautorstwie (wraz z promotorem) jedną samodzielną. Są to trzy artykuły w *MGEV* i dwie publikacje konferencyjne na *ICAISC*, wydane w serii *LNCS*. Ich punktacja ministerialna to 20 punktów oraz jedna praca w *MGEV* za 140 punktów. Znam publikacje Doktoranta wydane w *MG&V* i oceniam je bardzo wysoko. Poza tematyką rozprawy Doktorant ma dwa artykuły w czasopiśmie, za 12 i 40 punktów, oraz jedną pracę konferencyjną za 70 punktów. Najwyższa punktacja to 140, sumaryczna punktacja to 342 punkty MNiSW, dwa cytowania w Scopus (autocytowania). Jest to dorobek zupełnie wystarczający na etapie rozprawy doktorskiej.

2.1.7 Podsumowane

Praca jest bardzo obszerna, obejmuje co najmniej cztery tematy, w których doktorant zaproponował i zweryfikował nowe metody, oraz piąty, w którym Doktorant zaproponował nowy zbiór obrazów testowych wraz z protokołem ewaluacji na podstawie tego zbioru. Prawdopodobnie można by wydzielić z tego zbioru kilka dobrych rozpraw doktorskich.

2.2 Uwagi dyskusyjne i krytyczne

2.3 Uwagi merytoryczne

Nie mam istotnych merytorycznych uwag krytycznych. Zawsze można twierdzić, że dana metoda mogłaby być przetestowana jeszcze na innych zbiorach danych, lub że liczba metod porównawczych jest zbyt mała. Jednak takie uwagi są mało konstruktywne, gdyż prowadziłyby do niekończących się badań, co w obecnym świecie nadmiaru danych jest raczej niebezpieczeństwem, niż szansą.

2.4 Uwagi redakcyjne i techniczne

Praca została zredagowana bardzo starannie. Potknięcia są drobne i nieliczne. Wskazuję je w celu edukacyjnym i nie spodziewam się omawiania ich w odpowiedzi na recenzję.

Cytowanie źródeł innych, niż podstawowe Lepiej jest cytować publikacje recenzowane, niż preprinty. Np. praca cytowana jako [29] jest preprintem w serwisie arXiv, podczas gdy ta sama praca została opublikowana w *ACM Transactions on Intelligent Systems and Technology* w 2025 roku i jest dostępna pod adresem <https://doi.org/10.1145/3744746>. Być może w czasie pisania rozprawy artykuł mógł jeszcze nie być opublikowany.

Bibliografia – edycja Program BibTeX sprowadza litery do małych w tytułach artykułów. Aby tego uniknąć, należy wziąć takie litery w dodatkową klamrę. Na przykład, w pozycji [33] należałoby poprawić: ~~konie~~KonIQ, zaś w [34] ~~u-net~~U-net.

Mieszanie języków Nieco utrudnia czytanie pracy pomieszanie języków. Dobrze to widać na rys. 20, gdzie teksty na grafikach są po angielsku, w podpisie dwa języki są pomieszane. Prawdopodobnie grafiki zostały przygotowane dla artykułów po angielsku i dla rozprawy doktorskiej nie zostały przetłumaczone.

Anakolut Doktorantowi należą się gratulacje. Niemal w każdej pracy, od inżynierskiej do habilitacyjnej, jest przynajmniej jeden *anakolut*. Doktorant ustrzegł się od tej plagi języka polskiego. W takich zdaniach z imiesłowem przysłówkowym współczesnym, jak np. „Na podstawie zbalansowanego zbioru zapytań (...) przeprowadzono analizę porównawczą (...), wykorzystując opracowany wcześniej predyktor MOS.” podmiot dla orzeczenia (przeprowadzający analizę) i w okoliczniku sposobu (wykorzystujący predyktor) jest ten sam.

Drobne błędy typograficzne

Str. 5 i dalej: ~~Blind Image Quality Assessment~~*Blind Image Quality Assessment*, ~~Mean Opinion Score~~*Mean Opinion Score*, ~~text-to-image~~*text-to-image* – wyrazy obcojęzyczne z zasady piszemy *czcionką wyróżniającą*. Jest jasne, że obecnie wiele wyrazów angielskich wchodzi do języka polskiego i ulega całkowitemu spolszczeniu, ale w tych przypadkach nie mamy do czynienia z tym zjawiskiem. Raz fraza *text-to-image* została napisana właściwie, str. 68 dół.

Str. 36 i inne: nieprawidłowe użycie cudzysłowu do wyróżnienia terminów; powinna być użyta czcionka wyróżniająca, np.: „*klasy*”*klasy*.

Str. 39: w podpisie Rys. 7 występują nieprecyzyjne i błędne określenie: ~~czarna linia~~*czarna linia*, ~~żółta linia~~*żółta linia*, ~~czarna linia~~*czarna linia*.

Str. 40 i inne: literówki: ~~w wdrożeniu~~*w wdrożeniu*.

3 Podsumowanie

W rozprawie można wyróżnić istotne elementy nowatorskie oraz inne pozytywne aspekty omówione w rozdziale 2.1, zaś uwagi krytyczne praktycznie nie mają miejsca.

Cele pracy zostały osiągnięte.

W rozprawie zaproponowano i przebadano cztery nowatorskie metody dotyczące poprawiania jakości obrazów, uwzględniające, obok klasycznych miar pikselowych i strukturalnych, również wartość percepcyjną. Przedstawiono testowy zbiór obrazów, który powinien ułatwić dalsze badania.

Wniosek ogólny Powyższy opis, uwzględniający uwagi pozytywne jak również uwagi dyskusyjne i krytyczne, uzasadnia moje ostateczne wnioski o następującej treści.

Recenzowana rozprawa w postaci opracowania pisemnego **wykazuje i prezentuje ogólną wiedzę teoretyczną kandydata w dyscyplinie naukowej informatyka techniczna i telekomunikacja oraz umiejętność samodzielnego prowadzenia pracy naukowej**. Rozprawa stanowi **oryginalne rozwiązanie problemu naukowego** w tej dyscyplinie. Opisane metody mogą także zostać wykorzystane do tworzenia **oryginalnych rozwiązań w zakresie zastosowania wyników badań naukowych w sferze gospodarczej**. Zagadnienie badawcze zostało prawidłowo postawione i skutecznie rozwiązane, a rozwiązanie zostało rzetelnie zweryfikowane. Tym samym rozprawa spełnia wymagania określone w art. 187.1 i 2 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce. **Rozprawę oceniam pozytywnie** i stawiam wniosek o skierowanie jej do dalszych etapów przewodu doktorskiego.

Wniosek o wyróżnienie rozprawy W nawiązaniu do podsumowania w rozdz. 2.1.7, w którym stwierdzam, że dorobek przedstawiony w rozprawie jest bardzo znaczny, stawiam wniosek o uznanie rozprawy za wyróżniającą się.



dr hab. inż. Radosław Mantiuk, prof. ZUT
Wydział Informatyki
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
ul. Żołnierska 52, 71-210, Szczecin
e-mail: radoslaw.mantiuk@zut.edu.pl

Recenzja rozprawy doktorskiej wykonana dla Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja

Tytuł recenzowanej rozprawy:

Zastosowanie lekkich architektur głębokich w cyfrowej percepcyjnej analizie i korekcji degradacji obrazu.

Autor rozprawy:

Mgr inż. Tomasz Lehmann

Recenzja wykonana została na podstawie uchwały Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej z dnia 24 marca 2026 r.

- 1 **Jakie zagadnienie naukowe jest rozpatrzone w pracy (teza rozprawy) i czy zostało ono dostatecznie jasno sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?**

Rozprawa doktorska omawia zagadnienia związane z zastosowaniem sztucznej inteligencji, w szczególności lekkich architektur głębokich sieci neuronowych, do korekcji degradacji obrazów cyfrowych oraz obiektywnej estymacji ich jakości percepcyjnej. Praca wpisuje się w nurt rozwoju technologii uczenia maszynowego nakierowanego na efektywność obliczeniową (ang. Green AI), dążąc do wykazania, że możliwe jest osiągnięcie jakości wyników porównywalnych z rozwiązaniami klasy state-of-the-art przy niższym zapotrzebowaniu na zasoby obliczeniowe.

Autor nie formułuje w tekście pracy pojedynczej tezy w tradycyjnym ujęciu. Zamiast tego, we Wprowadzeniu (Rozdz. 1.2) określono cele rozprawy, do których należą:

- opracowanie bezwzorcowego predyktora MOS o ograniczonym koszcie zasobowym,
- opracowanie metody redukcji efektu mory przy umiarkowanej złożoności,
- zaprojektowanie metody rekonstrukcji obrazów niedoświetlonych z uwzględnieniem metryk percepcyjnych,
- zdefiniowanie funkcji straty ze składnikiem opartym na MOS dla usuwania rozmycia,
- przygotowanie zasobów testowych oraz protokołu oceny generowanych obrazów.

Powyższe cele są realizowane w kolejnych rozdziałach na przykładach redukcji efektu mory, poprawy czytelności obrazów w warunkach niedoświetlenia, usuwania rozmycia oraz oceny treści generowanych (text-to-image). Zagadnienia te zostały sformułowane jasno, a wprowadzone definicje operacyjne (np. "model lekki", "wydajny algorytm") precyzują ramy badawcze.

Rozprawa ma charakter doświadczalny i aplikacyjny (inżynierski). Autor opiera swoje wnioski na wynikach eksperymentów numerycznych, projektując, trenując i poddając ewaluacji (abstrakcyjnej i porównawczej) modele głębokie (np. dual-Xception, Attention U-Net, Swin-UNet) w odniesieniu do referencyjnych zbiorów danych.

2 Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł / w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle / świadczący o dostatecznej wiedzy autora. Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

Autor przeprowadził ustrukturyzowany przegląd literatury (Rozdział 2), odpowiadający tematyce rozprawy. Analiza źródeł została podzielona na sekcje odpowiadające wyzwaniom badawczym: obiektywnej ocenie jakości obrazów (w tym modelom bezwzorcowym - BIQA), technikom redukcji efektu mory, rekonstrukcji obrazów niedoświetlonych oraz usuwaniu rozmycia. Przegląd wskazuje na znajomość trendów w dziedzinie widzenia komputerowego, w tym koncepcji takich jak sieci splotowe (CNN), mechanizmy uwagi, architektury typu Transformer czy modele dyfuzyjne.

Warto odnotować, że autor analizuje omawiane podejścia, wskazując na ich ograniczenia w kontekście złożoności obliczeniowej, co stanowi punkt wyjścia dla własnych propozycji inżynierskich.

Należy wskazać na zwięzłość przeglądu literatury. Choć Rozdział 2 zajmuje relatywnie niewiele miejsca, to obejmuje on elementy istotne z punktu widzenia treści i celów badawczych rozprawy. Autor wyselekcjonował pozycje istotne dla tematu, unikając dygresji. Dobór cytowanej literatury jest odpowiedni. Wnioski sformułowane na podstawie analizy źródeł są jasne i stanowią zaplecze do dalszych działań eksperymentalnych.

3 Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

Głównymi osiągnięciami zaprezentowanymi w rozprawie są:

- Opracowanie i zbadanie dwóch bezwzorcowych predyktorów oceny MOS (architektury dual-Xception oraz ConvNeXt+MLP).
- Wprowadzenie metody równoważenia rozkładu klas podczas treningu (ERCU).
- Zaprojektowanie systemów redukcji degradacji obrazu (Attention U-Net dla mory oraz Swin-UNet dla obrazów niedoświetlonych).
- Opracowanie uniwersalnego zbioru zapytań i wygenerowanie otwartego zbioru danych (DIQA).

W zakresie badań nad algorytmem do percepcyjnej oceny jakości obrazów (Rozdział 3), Autor zaproponował dwa modele bezwzorcowej oceny jakości obrazów (BIQA), opierające się na dwustrumieniowej architekturze, która równolegle analizuje piksele obrazu (tor wizualny) oraz cztery skojarzone z nim metadane: ostrość, jasność, kontrast i poziom kompresji (tor parametryczny). Pierwszy z modeli, dual-Xception, przetwarza oba strumienie wykorzystując sieć splotową Xception, rozszerzając parametry skalarne na obszar przestrzenny. Drugi model, ConvNeXt+MLP, wykorzystuje sieć ConvNeXt w torze wizualnym, a analizę metadanych upraszcza poprzez zastosowanie perceptronu wielowarstwowego (MLP), co pozwala na zmniejszenie złożoności obliczeniowej rozwiązania. W celu weryfikacji przeprowadzono ablację działania modeli. Analiza wykazała, że zastosowanie autorskiej strategii podpróbkiowania klas ERCU pozwala przyspieszyć proces uczenia (skracając czas trwania epoki z 357 do 85 sekund), a jednocześnie zapewnia spójność wyników dla całego zakresu MOS z danych referencyjnych. Wyrównuje to błąd predykcji i poprawia trafność ocen dla niedoreprezentowanych w zbiorze obrazów o najniższej i najwyższej jakości. Autor wykazał, że mniejszy model ConvNeXt+MLP (ok. 88 mln parametrów) osiąga korelację oceny

zbliżone do danych referencyjnych ($SROCC = 0.921$, $PLCC = 0.932$) w zestawieniu z architekturami posiadającymi większą liczbę parametrów (jak nowsze modele typu Q-Align), wymagając mniejszych nakładów obliczeniowych. *Należy jednak zauważyć, że w treści rozprawy (w tym w opisie wyników oraz w tabeli 3.3.7) nie podano informacji dotyczących złożoności obliczeniowej, takich jak liczba parametrów, dla analizowanych modeli referencyjnych, co utrudnia pełne porównanie korzyści wydajnościowych. Wybór wielkości wejścia obrazu na podstawie testu zaledwie trzech rozdzielczości należy traktować jako osiągnięcie praktycznego kompromisu, a nie wyłonienie bezwzględnego optimum.*

W kontekście redukcji efektu mory (Rozdział 4), Autor skoncentrował się na wykorzystaniu architektury splotowej Attention U-Net do usuwania prążków interferencyjnych. Architektura ta została uzupełniona o mechanizm bramek uwagi (attention gates) w połączeniach skrótowych, co ułatwia sieci selektywne wzmacnianie cech w obszarach występowania artefaktów przy jednoczesnym tłumieniu informacji z tła. Opracowano strategię rotacyjnego krzyżowego próbkowania o stałym budżecie (RFBCS), umożliwiającą łączenie i równoważenie danych uczących pochodzących ze zbiorów o różnej wielkości (UHDM oraz TIP2018). Wykazano, że zastosowanie hybrydowej funkcji straty (kombinacja błędu MSE i indeksu SSIM) pozwala uzyskać wyższe wartości wskaźników jakości obrazu niż optymalizacja oparta na błędzie pikselowym. Autor wskazał, prezentując liczbę parametrów w tabeli, że model charakteryzuje się mniejszą złożonością (6,4 mln parametrów w zestawieniu z np. 58,6 mln w modelu MopNet), osiągając przy tym zbliżone wyniki wierności strukturalnej. *Z uwagi na ograniczenia sprzętowe eksperymenty oraz uczenie modelu prowadzono na obrazach o zredukowanej rozdzielczości 512x512 pikseli, co stanowi różnicę względem natywnej rozdzielczości 4K wykorzystywanego zbioru danych UHDM. W przypadku analizy efektu mory skalowanie obrazów wpływa na fizyczną strukturę tego artefaktu, który jest zależny od częstotliwości przestrzennej obrazu, co może rzutować na poprawność porównań z metodami z literatury. Ponadto, zdolność proponowanego modelu do generalizacji nie została zweryfikowana z wykorzystaniem zewnętrznych zbiorów testowych.*

W przypadku rekonstrukcji obrazów niedoświetlonych (Rozdział 5), osiągnięcie polega na wdrożeniu architektury Swin-UNet, wykorzystującej mechanizmy transformera (okienkowej uwagi). Autor zbadał strategię łączenia danych treningowych z różnych zbiorów (LOL i SID), wskazując, że trening uwzględniający wyłącznie najciemniejsze ekspozycje wpływa pozytywnie na zdolność modelu do generalizacji. Zaprojektowano także wieloskładnikową funkcję straty integrującą metryki percepcyjne (LPIPS) z sygnałowymi. *Należy jednak zauważyć zjawisko "Perception-Distortion Tradeoff", gdzie liniowe łączenie kryteriów pikselowych (MSE) z percepcyjnymi (LPIPS) wpływa na spadek wierności sygnałowej, co sprawia, że jednoczesna maksymalizacja obu tych wskaźników jest utrudniona metodologicznie. Ponadto nie przeprowadzono testów generalizacji na zbiorach zewnętrznych (np. MEF), a dobór wag łączących elementy w funkcji celu wymaga szerszego uzasadnienia.*

Kolejne osiągnięcie (Rozdział 6) to zintegrowanie zbudowanego wcześniej estymatora MOS jako różniczkowalnego składnika percepcyjnego w funkcji straty dla zadania usuwania rozmycia (deblurring). W tym celu Autor połączył zamrożony model predykcji jakości (BIQA) z trenowanym regulatorem U-Net odpowiedzialnym za rekonstrukcję. Wyjście sieci usuwającej rozmycie było bezpośrednio przekazywane na wejście estymatora jakości, co umożliwiło propagację gradientów błędu percepcyjnego przez całą architekturę hybrydową. Ostateczna funkcja straty stanowiła liniową kombinację tradycyjnych miar odtworzenia sygnału (MSE oraz SSIM) oraz składnika maksymalizującego prognozowaną ocenę MOS. Doświadczenia wykazały, że gradient propagowany przez predyktor percepcji kierunkuje sieć U-Net na wyniki zbliżone do preferencji obserwatorów. *Wskazane byłoby przeprowadzenie oceny podatności sieci odszumiającej na przykłady kontradykcyjne, co mogłoby prowadzić do modyfikacji pikseli maksymalizujących metrykę bez realnej poprawy jakości. Ponadto weryfikacja oparta na zbiorze HIDE jest ograniczona, podczas gdy standardem metodologicznym w dziedzinie deblurringu jest ewaluacja na dodatkowych, powszechnie uznanych*

bazach testowych, takich jak zbiory GoPro oraz RealBlur (zarówno w podzbiorze RealBlur-J, jak i RealBlur-R), w celu jednoznacznego potwierdzenia ogólnej użyteczności estymatora.

Ostatnim elementem dorobku (Rozdział 7) jest ewaluacja jakości treści generowanych (text-to-image) i stworzenie zasobu testowego DIQA, obejmującego 1000 obrazów wygenerowanych za pomocą 10 modeli dyfuzyjnych. Autor podjął próbę systematyzacji oceny jakości grupy modeli generatywnych. W tym celu zdefiniował zbiór zróżnicowanych promptów tekstowych, reprezentujących pięć kategorii tematycznych, takich jak portrety, krajobrazy, fantastyka, martwa natura oraz architektura i abstrakcja. Na ich podstawie, przy użyciu dziesięciu modeli generatywnych, wygenerowano zbiór obrazów o nazwie DIQA. Głównym celem badawczym było sprawdzenie, jak opracowane wcześniej predyktory bezwzorcowe (BIQA) radzą sobie z oceną specyficznych zniekształceń generatywnych w porównaniu ze zniekształcenia fotograficznymi. Autor zaproponował protokół ewaluacyjny, który pozwala na automatyczne sortowanie, filtrowanie i ocenę estetyczno-techniczną wygenerowanych treści bez udziału ludzkich obserwatorów. Zaprezentowane podejście stanowi próbę systematyzacji kontroli jakości w systemach opartych na generatywnej sztucznej inteligencji. *Ewaluacja DIQA wykazuje jednak ograniczenia w postaci zastosowania miary typu Out-of-Distribution (OOD). Modele dyfuzyjne generują artefakty odmienne od zniekształceń fotograficznych, na jakich trenowany był wyjściowy predyktor MOS, co może wpływać na ocenę. Ponadto zbiór został oceniony metodami automatycznymi; brak weryfikacji manualnej ("Ground Truth") z udziałem ludzi ogranicza możliwość pełnej oceny wiarygodności protokołu dla obrazów syntetycznych, a stosowane miary nie uwzględniają zgodności obrazu z tekstem (text-alignment).*

Podsumowując, stwierdzam, że w badaniach użyto adekwatnych metod inżynierii sztucznej inteligencji, a przyjęte założenia o możliwości kompresji złożoności architektur przy użyciu odpowiednich technik treningowych zostały potwierdzone. Choć zasygnalizowany przeze mnie szereg potencjalnych ulepszeń odnosi się do konkretnych aspektów metodologicznych, należy w pełni docenić szeroki zakres podjętej tematyki badawczej, przy którym naturalnym zjawiskiem jest brak możliwości wyczerpującego i wszechstronnego wyeksplorowania każdego z analizowanych elementów. Proszę o ustosunkowanie się do moich uwag zawartych w tym podpunkcie w czasie obrony pracy.

4 Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Oryginalność rozprawy wynika z połączenia inżynierii głębokich architektur i psychofizyki widzenia (IQA). Zamiast podążać trendem powiększania modeli i używania wielkoskalowych transformerów czy modeli dyfuzyjnych, Autor skupia się na modyfikacji funkcji straty i selekcji danych uczących.

Za oryginalny, samodzielny dorobek Autora można uznać:

- Zaprojektowanie dwustrumieniowej architektury predyktora MOS (dual-Xception oraz ConvNeXt+MLP), operującej na hybrydyzacji tensorów obrazowych i zreplikowanych meta-danych.
- Sformułowanie procedur treningowych równoważących rozkłady danych (ERCu, RFBCS), co stanowi wskazówkę metodyczną dla konstruowania algorytmów uczenia maszynowego w scenariuszach z nie zrównoważonymi zbiorami.
- Opracowanie i udostępnienie zbioru DIQA wraz z protokołem wykorzystującym estymatory BIQA do oceny algorytmów text-to-image. Odnosi się to do problemu obiektywizacji oceny generatywnej sztucznej inteligencji.

Wykorzystanie komponentu estymującego MOS jako części różniczkowalnej funkcji straty przy odsumowaniu wskazuje na systemowe podejście Doktoranta do rozpatrywanego zagadnienia. Publikacje w materiałach konferencyjnych i czasopismach potwierdzają obecność tych prac w dyskursie naukowym.

5 Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników / zwięzłość, jasność, poprawność redakcyjna rozprawy/?

Praca została napisana językiem precyzyjnym i poprawnym pod kątem akademickim. Konstrukcja rozprawy jest spójna i logiczna. Wyniki badań są zilustrowane adekwatnymi wykresami, tabelami podsumowującymi parametry i wyniki, oraz materiałami wizualnymi porównującymi jakość obrazów przed i po przetwarzaniu.

W pracy podano definicje kluczowych pojęć (efektywność, lekka architektura) na początku rozprawy, opisy procedur ablacyjnych oraz zraportowano ograniczenia zbadanych architektur. Jest to odpowiednia praktyka badawcza. Zwięzłość, szczególnie we wspomnianym wcześniej przeglądzie literatury, miejscami wymusza stosowanie skrótów myślowych, co nie wpływa na spójność dokumentu.

W pracy można zidentyfikować drobne uchybienia edytorskie i nieścisłości zapisu matematycznego. Nie utrudniają one jednak zrozumienia treści pracy.

6 Jakie są słabe strony rozprawy i jej główne wady?

Słabszym merytorycznie punktem rozprawy jest brak szerszej dyskusji nad istotnością statystyczną osiąganych różnic, szczególnie tam, gdzie zaproponowane warianty (jak w przypadku modyfikacji funkcji straty dla problemu niedoświetlenia) dają poprawę metryk na poziomie setnych części lub pojedynczych dziesiątych decybel. *Wskazane byłoby zastosowanie rozszerzonego aparatu statystycznego. Prezentowanie samych średnich może być niewystarczające dla pracy doktorskiej; pożądane byłoby wdrożenie testów istotności statystycznej (np. test t-Studenta lub test Wilcozona), które pozwoliłyby rozróżnić rzeczywistą przewagę metody od przypadkowego "szumu" treningowego. Ponadto, wskazane byłoby wprowadzenie estymacji niepewności (np. przedziały ufności dla średniej), co pozwoliłoby ocenić stabilność modelu i zweryfikować, czy wysoka średnia wydajność nie wiąże się z dużą wariancją wyników. Choć przeprowadzenie tych analiz nie było konieczne dla wykazania ogólnej poprawności zaproponowanych rozwiązań i obrony celów pracy, to rozszerzona ocena statystyczna uwiarygodniłaby uzyskane wyniki.*

Drugą kwestią jest struktura i stopień autorskiego wkładu w badaniach z Rozdziału 6 (usuwanie rozmycia). Autor deklaruje we wstępie do tego rozdziału, że jego wkład "obejmował nadzór merytoryczny nad realizacją i przebiegiem eksperymentów". Choć rozumiem, że jest to kontynuacja (włączająca autorski estymator MOS), tak zaprezentowany podział rodzi wątpliwość co do kwalifikacji tego rozdziału jako w pełni samodzielnego dorobku w ramach tekstu rozprawy.

Na marginesie uwag technicznych, niektóre przykłady wizualne (np. Rys. 12 i 13) mogłyby zostać lepiej skadrowane (tzw. "zoom-in") w publikacji, by pokazać konkretny detal redukcji artefaktów; w formacie PDF drobne różnice na powiększeniach często zacierają się przez skalowanie dokumentu.

7 Jaka jest przydatność rozprawy do nauk technicznych?

Przydatność rozprawy dla dyscypliny informatyka techniczna i telekomunikacja wiąże się z jej potencjałem aplikacyjnym. Poszukiwanie rozwiązań obliczeniowo "lekkich", efektywnych energetycznie i pamięciowo jest dziś, w dobie brzegowego przetwarzania danych (ang. Edge AI) oraz ro-

snących kosztów utrzymywania modeli serwerowych, zagadnieniem istotnym. Wypracowane heurystyki uczenia i zarządzania niebalansowanymi danymi obrazowymi mogą zostać zaadaptowane w potokach przetwarzania wideo (transkodowanie, streaming), systemach fotografii mobilnej czy aplikacjach monitoringu środowiskowego (kamery w trudnych warunkach oświetleniowych).

Opracowany zbiór zapytań i system referencyjnej oceny modeli generatywnych (DIQA) stanowi narzędzie do prowadzenia benchmarków modeli udostępnianych jako usługi (MaaS).

Do której z następujących kategorii Recenzent zalicza rozprawę:

- nie spełniająca wymagań stawianych rozprawom doktorskim przez obowiązujące przepisy
- wymagająca wprowadzenia poprawek i ponownego recenzowania
- spełniająca wymagania
- spełniająca wymagania z wyraźnym nadmiarem
- **bardzo dobra, zasługująca na wyróżnienie**

Stwierdzam, że Autor wniósł wkład do dyscypliny nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja, wykazując się odpowiednimi umiejętnościami teoretycznymi i inżynieryjno-badawczymi. Uważam, że przedłożona do recenzji rozprawa spełnia wymagania ustawowe i wnioskuję o dopuszczenie mgr inż. Tomasza Lehmana do publicznej obrotu.

Radosław Stankiewicz



prof. dr hab. inż. Adam Wojciechowski
Instytut Informatyki
Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej
Politechnika Łódzka
Al. Politechniki 8, 93-590 Łódź

RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy: **Zastosowanie lekkich architektur głębokich w cyfrowej percepcyjnej analizie i korekcji degradacji obrazu**

Autor rozprawy: **mgr inż. Tomasz Lehmann**

Promotor rozprawy: **prof. dr hab. inż. Przemysław Rokita**

Jakie zagadnienie naukowe jest rozpatrywane w pracy (teza rozprawy) i czy zostało ono dostatecznie jasno sformułowane przez Autora? Czy tematyka rozprawy jest aktualna lub dostatecznie ważna? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?

Pan mgr inż. Tomasz Lehmann podjął w swojej pracy doktorskiej wybrane aspekty metod oceny obrazów według kryteriów jakościowych skorelowanych z ludzką percepcją oraz zagadnienia, warunkowanej percepcyjnie, poprawy jakości obrazów cyfrowych, przy ograniczonych zasobach obliczeniowych. W ramach zagadnienia badawczego sformułowano różnorodne cele pracy. Obejmują one opracowanie bezwzorcowego predyktora średniej oceny ważonej *MOS* (ang. *Mean Opinion Score*), opracowanie metody redukcji efektu mory, zaprojektowanie metody rekonstrukcji obrazów niedoświetlonych, zdefiniowanie funkcji straty ze składnikiem opartym na *MOS* dla usuwania rozmycia obrazów oraz przygotowanie zasobów testowych i protokołu oceny obrazów generowanych na podstawie tekstu. Zagadnienia badawcze zostały zatem precyzyjnie sformułowane i wyjaśnione.

Tematyka rozprawy jest aktualna i ważna, nie tylko w kontekście funkcjonalnej korekty obrazów cyfrowych, celem zapewnienia im wysokiej jakości, ale przede wszystkim w kontekście analizy czynników opisujących percepcyjną jakość obrazu. Stanowi to podstawę do konstruowania obiektywnych, łączących aspekty techniczne i subiektywne, bezwzorcowych miar oceny jakości obrazów, tak istotnych przy dynamicznie rozwijających się modelach i technikach przetwarzania i generowania obrazów cyfrowych. Istotne są też badania nad metodami tworzenia efektywnych i lekkich modeli, które wymagają znacznie niższych zasobów i krótszego czasu uczenia.

Rozprawa ma charakter eksperymentalny i nosi w sobie istotny potencjał projektowo-wdrożeniowy. Doktorant w ramach każdego z postawionych celów, przygotował lub dobrał odpowiednie zbiory danych, zaprojektował i zaimplementował autorskie architektury lekkich modeli głębokich oraz przeprowadził szereg ciekawych eksperymentów. Praca z racji podjętej problematyki, zakresu oraz metodologii badawczej jest pełnoprawnym osiągnięciem o charakterze naukowo-badawczym w dziedzinie nauk inżynieryjno-technicznych, w dyscyplinie Informatyka Techniczna i Telekomunikacja.

Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł (w tym literatury światowej i stanu zagadnień w przemyśle) świadczącą o dostatecznej wiedzy Autora? Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

W przedstawionej do oceny rozprawie doktorskiej, przeglądowi literatury poświęcono nominalnie rozdział drugi, przy czym każda z sekcji rozdziału drugiego dokonuje zwięzłego przeglądu literatury w ramach poszczególnych celów badawczych podjętych w pracy: oceny jakości obrazów, redukcji efektu mory, rekonstrukcji obrazów niedoświetlonych, wykorzystania estymatora MOS w funkcji straty modelu redukującego rozmycie obrazów oraz oceny jakości obrazów generowanych z tekstu. Poszczególne sekcje koncentrują się na powierzchownym przywołaniu kluczowych prac z każdego podjętego podobszaru. Opis ma charakter narracyjny, selektywny i mało krytyczny, ale wnioski wyciągnięte z przeglądu są trafne i naświetlają tło do innowacji zaproponowanych przez Doktoranta. Wnioski z przeglądu lokują zainteresowania Doktoranta w eksperymentach weryfikujących przydatność, znanych z literatury, lekkich modeli rekonstrukcyjnych typu U-Net, w zadaniu korekcji obrazów cyfrowych. Autorskie modyfikacje obejmują adaptację wybranych architektur do przedmiotowych zadań, w szczególności poprzez wzbogacenie funkcji kosztu o kryteria odzwierciedlające percepcję ludzką oraz weryfikację technik uczenia modeli w kontekście zmniejszenia złożoności procesu. Przegląd jest też okazją do przywołania referencyjnych zbiorów danych, które stanowią podstawę do uczenia i testowania opracowanych modeli. Analiza wskazuje jednocześnie na potrzebę rozwijania metodologii w obszarze bezwzorcowej oceny obrazu.

Bibliografia jest obszerna i liczy 158 pozycji, w tym 5 prac z udziałem Doktoranta. Przegląd należy uznać za rzeczowy, choć umiarkowanie aktualny, gdyż odczuwalna jest niedoreprezentacja publikacji z 3 ostatnich lat. W dorobku Autora znajdują się trzy prace opublikowane w czasopiśmie *Machine Graphics and Vision* oraz dwie publikacje zaprezentowane na międzynarodowej konferencji *ICAISC*. Całokształt dokonań publikacyjnych Doktoranta jest w zupełności wystarczający w świetle obowiązujących wymagań.

Czy Autor rozwiązał postawione zagadnienie, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

W obszarze podjętej problematyki, Doktorant rzetelnie przeprowadził analizę zagadnienia. Określił kluczowe etapy prac badawczych, przygotował zbiory danych i zaproponował autorskie adaptacje lekkich architektur sieci głębokich. W szczególności, zaprezentował udoskonalenia funkcji kosztu i modyfikacje procesu uczenia modeli. Przeprowadził szeroko zakrojone eksperymenty weryfikujące przydatność zaproponowanych modyfikacji w szacowaniu jakości obrazów oraz niwelowaniu ich degradacji, co systematycznie wspierało kontrybucje Doktoranta.

Dobór referencyjnych architektur głębokich, do przedmiotowych problemów, był prawdopodobnie intuicyjny, gdyż nie wszystkie rozwiązania referencyjne, pojawiające się w końcowych zestawieniach porównawczych (tab. 3.3.7, 5.3.5), zostały omówione w przeglądzie literatury. Doktorant nie doprecyzował kryteriów doboru modeli referencyjnych, ale ich zakres nie budzi zastrzeżeń.

Rozdział 3 poświęcony jest badaniom nad algorytmem percepcyjnej oceny jakości obrazów, który stanowiłby predyktor wartości *MOS*. Opis charakterystyk, jednego z największych, w obszarze badań nad *IQA*, zbioru danych *KoniQ-10k*, uzasadnia motywację stojącą za dwustrumieniowymi architekturami *dual-Xception*, czy *ConvNeXt+MLP*, analizującymi, poza cechami głębokimi, również metadane obrazów tj.: jasność, kontrast, ostrość, czy gęstość bitowa. Ciekawe wydaje się przetestowanie różnych strategii łączenia danych, wielokryterialnej funkcji straty integrującej *MSE*, *RMSE*, *PLCC* i *SROCC* oraz strategii epokowego podpróbkowania z limitem proporcji (*ERCU*). Drobiazgowa analiza poszczególnych strumieni, wariantów łączenia danych, czy podpróbkowania

z limitem proporcji w architekturze *dual-Xception* jest wartościowa. Analogiczne, choć skromniejsze są eksperymenty dotyczące architektur *ConvNeXt* i *ConvNeXt+MLP*. Analiza pozwala zrozumieć wpływ każdego z komponentów i istotność poszczególnych modyfikacji finalnej architektury. Prezentacja wyników korelacji *PLCC* i *SROCC*, na tle rozwiązań referencyjnych, jest inspirująca, ale dopełnieniem obrazu byłaby prezentacja parametrów złożoności pamięciowej i czasowej - być może na wzór tabel 3.3.4 lub 4.3.3. Doktorant bowiem kontestuje wielkość modeli, komentuje to w przeglądzie literatury, ale wyniki prezentuje głównie w obszarze wskaźników jakościowych.

Rozdział 4 podejmuje temat redukcji efektu mory w obrazach cyfrowych. Badania odbywają się w oparciu o zbiory danych posiadających referencje (*UHDM*, *TIP2018*) oraz zaadaptowaną splotową architekturę *Attention U-Net*. W szczególności, weryfikowano skuteczność hybrydowej funkcji straty łączącej błąd średniokwadratowy (*MSE*) z podobieństwem strukturalnym (*SSIM*) i stosunkiem sygnału do szumu (*PSNR*). Eksperymenty przeprowadzone na różnych schematach łączenia tych miar wykazały względną przewagę hybrydowego podejścia. Ciekawe jest też odniesienie do wcześniej wprowadzonej predykcji wartości *MOS*, która zachowywała się spójnie z oczekiwaniami, pomimo relatywnie niskiej wartości samego współczynnika (*MOS*). Ciekawa jest również strategia rotacyjnego doboru próbek, która zwiększa efektywność nauki modelu. Zestawienie wyników z rozwiązaniami literaturowymi, w sposób przekonujący, demonstruje przewagi opracowanych rozwiązań, choć, jak słusznie zauważa Doktorant, kontrybucje można wzmocnić prowadząc eksperymenty także na innych zbiorach danych.

Rozdział 5 poświęcony jest badaniom nad wzmacnianiem obrazów niedoświetlonych. Badania zostały przeprowadzone w kontekście dwóch popularnych, posiadających wzorce właściwej ekspozycji, zbiorów danych (*LOL*, *SID*) oraz lekkiej architektury *Swin-UNet*, adaptującej mechanizm transformera do architektury typu *U-Net*. Badania koncentrowały się na wieloskładnikowej funkcji straty, uwzględniającej między innymi: zgodność percepcyjną (np. bazującą na pre-trenowanej sieci *VGG19*, na metryce *LPIPS*, czy percepcyjnej przestrzeni barw), zgodność strukturalną, zgodność histogramową, czy jakościową (np. bazującą na mierze *PSNR* oraz *MSE*). Zaproponowaną architekturę przetestowano dla kilku początkowych szerokości kanałów, z różnymi schematami głębokości sieci oraz dla różnych funkcji straty (*MSE*, *FN-loss*, *LYT*, *LYT+LPIPS*). Doktorant przeprowadził wartościową próbę łączenia zbiorów danych w procesie nauki, zarówno naturalnych jak i syntetycznych, ale bardzo ciekawe byłyby eksperymenty weryfikujące zdolności generalizacyjne modeli, przykładowo poprzez uczenie modelu wyłącznie na jednym zbiorze (np. *SID*), a testowania na drugim (np. *LOL*). O ile optymalizacja hiperparametrów modelu wydaje się częścią rzetelnego warsztatu badacza to analiza porównawcza testowanych funkcji straty dostarcza ciekawych wniosków zarówno w kwestiach jakości (*PSNR*, *SSIM*) skorygowanego (rozjaśnionego) obrazu, jak i w kontekście bezwzorcowej oceny jakości obrazów (*NIQE*, *MOS*, *BRISQUE*). Analiza zbieżności procesu uczenia, w funkcji poszczególnych funkcji straty, jest inspirująca, ale nie wyjaśniono do końca wszystkich wzajemnych relacji. Czy fakt, że funkcja *FN-loss* lepiej przekłada się na wartość *MOS*, zaś funkcje będące adaptacją *LYT* skutkują lepszą oceną *NIQE* i *BRISQUE*, zależy od zbiorów danych, czy niespójności bezwzorcowych miar jakości obrazu? W pracy zaprezentowano obszerne porównanie z literaturowymi metodami referencyjnymi (tab. 5.3.5), ale w świetle zaprezentowanych wyników i braku jawnego odniesienia do złożoności obliczeniowej, trudno znaleźć jawne dowody, że, jak twierdzi Doktorant, „*Swin-UNet* ... oferuje korzystny kompromis między złożonością, kosztem obliczeniowym, a jakością rekonstrukcji”.

Rozdział 6 implementuje opracowany wcześniej model predykcji wartości *MOS* w funkcji straty architektury *Attention U-Net*, redukującej rozmycie obrazu. Badania były prowadzone z użyciem zbioru *HIDE*, zawierającym systemowo rozmyte oraz referencyjne obrazy. Zaproponowany model uczenia wykorzystywał autorską, wieloskładnikową funkcję straty, łączącą składnik fotometryczny (*MSE*), strukturalny (*SSIM*) i percepcyjny (*MOS*). Wyniki prezentują względną przewagę uwzględnienia

komponentu percepcyjnego, ale z racji braku odniesienia do wyników literaturowych oraz innych zbiorów danych, trudno wnioskować o uniwersalności sformułowanych stwierdzeń.

Rozdział 7 wprowadza natomiast autorski zbiór *DIQA* (1000 obrazów), stworzony z obrazów generowanych komputerowo, przez wybrane dyfuzyjne modele generatywne, według pięciu przykładowych schematów w formule *text-to-image*. Rozdział wprowadza również referencyjny protokół oceny jakościowej *MOS*, który może służyć środowisku naukowemu jako punkt odniesienia w tworzeniu bezwzorcowych miar jakości obrazu. Badania są na pewno inspirujące. Uwagę zwraca relatywnie niska wartość współczynnika *MOS* dla obrazów generowanych komputerowo, co może sugerować potencjał badań w zakresie metod bezreferencyjnej oceny *IQA* w generowaniu obrazów o wyższej jakości percepcyjnej.

Dysertację wieńczy sekcja podsumowująca kontrybucje Doktoranta, wkład w dyscyplinę naukową oraz perspektywy dalszych badań. Doktorant, ze względu na reporterski styl narracji, koncentruje się mocniej na sprawozdawczym referowaniu nowości architektonicznych (architektury, funkcje straty), usprawnień proceduralnych i metodologicznych oraz nowościach praktycznych, opracowanych rozwiązań. Ogólny zakres osiągnięć jest okazały i wartościowy, ale w uzupełnieniu perspektywy dalszych badań można byłoby przeprowadzić pogłębioną dyskusję nad własnościami generalizacyjnymi architektur i istotnością statystyczną wyników. Niekiedy różnice pomiędzy wartościami liczbowymi były nieznaczne i mogły wynikać z konkretnego splitu stosowanych zbiorów danych. Z reguły, warto taką dyskusję zamieścić w ramach sekcji poświęconej ograniczeniom przeprowadzonych badań. Niemniej, cele badawcze, które sformułowano na początku pracy, należy uznać za osiągnięte.

Podsumowując, przedstawione rozwiązania, eksperymenty oraz wyniki są ciekawe i nowatorskie. Szczególnie wartościowy jest zbiór różnych koncepcji wplatania czynników percepcyjnych do funkcji straty oraz seria przeprowadzonych na nich eksperymentów, usprawniających nowoczesne i lekkie modele głębokich sieci neuronowych w zadaniach korekty obrazów cyfrowych i bezwzorcowej ich oceny.

Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek Autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy, czy poziomowi techniki reprezentowanych przez literaturę światową?

Oryginalność rozprawy doktorskiej leży przede wszystkim w szeroko pojętym zakresie eksperymentów przeprowadzonych z użyciem nowoczesnych i lekkich architektur głębokich, których zadaniem była bezwzorcowo ocena jakości obrazów oraz ich wykorzystanie w korekcji rozmycia, mory czy niedoświetlenia obrazów. W szczególności najważniejszymi, moim zdaniem, osiągnięciami są:

- opracowanie dwustrumieniowych modeli *BIIQA* do estymacji wskaźnika *MOS* wraz z procedurami równoważenia rozkładu etykiet;
- opracowanie metody redukcji efektu mory opartej na architekturze *Attention U-NET* wraz z analizą funkcji straty oraz strategią rotacyjnego doboru próbek;
- zaprojektowanie algorytmu rekonstrukcji obrazów w słabym oświetleniu, opartego na architekturze *Swin-UNet* i złożonej funkcji straty, uwzględniającej składnik percepcyjny;
- zdefiniowanie hybrydowej funkcji straty z różniczkowalnym komponentem *MOS*, w zadaniu usuwania rozmyć obrazu;
- opracowanie i udostępnienie środowisku naukowemu zbioru testowego *DIQA* oraz protokołu ewaluacji modeli typu *text-to-image* w oparciu o wskaźnik *MOS*;

Wymienione osiągnięcia stanowią bezsprzecznie oryginalny wkład Doktoranta w rozwój dyscypliny Informatyka Techniczna i Telekomunikacja.

Czy Autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników (zwięźłość, jasność, poprawność redakcyjna rozprawy)?

Praca stanowi bardzo szerokie, wieloaspektowe studium oceny percepcyjnej obrazów cyfrowych oraz wykorzystania kryteriów i markerów jakości percepcyjnej w uczeniu lekkich architektur głębokich, korygujących wybrane degradacje obrazów cyfrowych. Autor podzielił pracę na niezależne sekcje, które realizowane były prawdopodobnie w innej kolejności niż zostało to zaprezentowane w pracy. Może to rodzić pytania o następstwo badań i przyczynowo-skutkowe implikacje poszczególnych rozważań, ale nie obniżają one wartości prezentowanych osiągnięć.

Docenienia wymaga spójny i przyjemny w lekturze układ pracy, który względnie kompleksowo ujmuje przedmiotowy obszar badań. Pracę czyta się bardzo dobrze. Pod względem edycyjnym praca została zredagowana starannie. Praca zawiera wszystkie niezbędne spisy oraz bibliografię. Drobną kwestią pozostają rysunki zamieszczone w pracy, które z racji swojej szczegółowości i małego rozmiaru wydruku, są czasami słabo czytelne, np. rys. 4, 5, 20.

Jakie są słabe strony rozprawy i jej główne wady?

Rozprawa doktorska nie zawiera istotnych uchybień, a poniższe uwagi mają przede wszystkim charakter dyskusyjny. Poniżej przedstawiłem przykładowe kwestie, które mogłyby zostać omówione, lecz nie ujmują pozytywnej oceny rozprawy.

Pewnej refleksji wymaga dokładność prezentowania wyników liczbowych, istotność statystyczna ich różnic oraz natura zagregowanych miar i wskaźników. W niektórych miejscach (np. *SROCC* w tab. 3.3.1) precyzja przedstawianych wyników, do dwóch miejsc po przecinku, zdawała się niewystarczająca, aby wskazać rozwiązanie lepsze. Stąd prawdopodobnie, w kolejnych tabelach, wyniki podawane były z dokładnością do 3 lub nawet 4 miejsc po przecinku. Nasuwa się jednak pytanie, na ile deklarowane zmiany w wynikach, rzędu 0.01 czy 0.001, implikują realną zmianę i mają przełożenie na percepcję różnic jakościowych? Czy dla obrazów o stosunkowo niewielkiej liczbie pikseli (~250k), mocno ograniczonej liczbie zbiorów danych i zmianach wskaźników na trzecim lub czwartym miejscu po przecinku, mamy do czynienia z istotnymi statystycznie różnicami? Jakie różnice dla poszczególnych miar i wskaźników można uznać za nieznaczące, tzn. „jakości wyników na poziomie porównywalnym w przyjętych metrykach” (str. 15). Na ile dobór zbioru danych ma wpływ na wyniki i jak charakter/zawartość obrazów wpływa na zagregowane miary (np. *MSE*, *SSIM*, *PSNR*, *RMSE*, in.)?

W kontekście deklarowanej we wstępie efektywności obliczeniowej, autorsko zdefiniowanej jako funkcja złożoności obliczeniowej (pamięć, czas) i liczby parametrów, zaskakująco rzadko pojawiały się w pracy analizy kluczowych parametrów implikujących efektywność. W sekcji 3 znajduje się analiza liczby parametrów i czas treningu (tab. 3.3.4), w sekcji 4 znajduje się analiza liczby parametrów (tab. 4.3.3), ale w innych sekcjach nie zamieszczono odpowiednich zestawień. Jak zatem wygląda deklarowana efektywność obliczeniowa opracowanych rozwiązań na tle referencyjnych?

W kontekście ograniczonej liczby zbiorów danych, w każdym z podjętych problemów, oraz wynikających z tego prawdopodobnych ograniczeń metod, warto byłoby podjąć dyskusję nad zdolnościami generalizacyjnymi zaproponowanych architektur oraz uniwersalnością opracowanych funkcji straty. Jakie mogą być metody zwiększania ich uniwersalności i niwelowania biasu wynikającego z naturalnego przeuczania na konkretnych zbiorach danych?

Konkluzja

Uważam, że cele rozprawy doktorskiej zostały w pełni zrealizowane. Doktorant przedstawił w rozprawie nowe badania z zakresu analizy percepcyjnej obrazów oraz jej wykorzystania w efektywnym uczeniu lekkich architektur głębokich, znajdujących zastosowanie w korygowaniu obrazów cyfrowych. Uzyskane przez Doktoranta wyniki uważam za oryginalne, wartościowe i ciekawe poznawczo. Zakres prowadzonych badań był szeroki, analiza opracowanych rozwiązań wszechstronna, a warsztat badawczy Doktoranta był właściwy. Doktorant może poszczycić się zauważalną aktywnością publikacyjną.

Tym samym rozprawa prezentuje wartościowe osiągnięcia naukowe w obszarze dyscypliny Informatyka Techniczna i Telekomunikacja oraz potwierdza umiejętność samodzielnego prowadzenia przez Doktoranta pracy naukowej.

Bez najmniejszej wątpliwości stwierdzam, że recenzowana rozprawa doktorska Pana mgr inż. Tomasza Lehmana spełnia wymagania stawiane rozprawom doktorskim, przez obowiązującą ustawę. Wnoszę o jej przyjęcie i dopuszczenie rozprawy do publicznej obrony.

