

Poznań, 24.05.2026

dr hab. inż. Rafał Długosz, prof. PBS
Wydział Telekomunikacji, Informatyki i Elektrotechniki
Politechnika Bydgoska im. Jana i Jędrzeja Śniadeckich
ul. Al. Prof. S. Kaliskiego 7, 85-796 Bydgoszcz
e-mail: rafal.dlugosz@pbs.edu.pl

WYDZIAŁ ELEKTRONIKI
TELEKOMUNIKACJI I INFORMATYKI
Politechniki Gdańskiej
RPU/25680/2026
Wpłynęło dnia 18.06.2026/
25.06.2026

1 Wprowadzenie

Recenzja została przygotowana na podstawie uchwały z dnia 10.02.2026 (212.511.1.2026.3)
Rady Dyscypliny Naukowej Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne,
Politechniki Gdańskiej.

Tytuł rozprawy:

“Efficient Deep Neural Networks for Video Denoising”

„Wydajne głębokie sieci neuronowe do usuwania szumów w sygnale wideo”

Autor rozprawy:

mgr inż. Piotr Kopa Ostrowski

Promotor rozprawy:

dr hab. inż. Tomasz Stefański

2 Wprowadzenie

2.1 Ogólna charakterystyka pracy

Przedstawiona rozprawa doktorska podejmuje aktualny i istotny problem z obszaru przetwarzania obrazów a dokładniej sekwencji wideo. Problem ten to odszumianie sekwencji wideo z wykorzystaniem metod głębokiego uczenia. Autor zwrócił uwagę na kompromis pomiędzy jakością rekonstrukcji a efektywnością obliczeniową. Przedstawione wyniki badań mają duże znaczenie w zastosowaniach praktycznych, m.in. w tzw. “edge computing”, w urządzeniach mobilnych czy też w systemach czasu rzeczywistego. Problematyka ta dobrze wpisuje się we współczesny nurt badań nad efektywnymi metodami “video restoration” (cyfrowa rekonstrukcja wideo) oraz architekturami sieci neuronowych przeznaczonymi do pracy przy ograniczonych zasobach obliczeniowych i energetycznych.

Główna teza pracy zakłada, że istniejące modele niewystarczająco wykorzystują kontekst czasowy, nadmiernie polegając na klatce referencyjnej, szczególnie w przypadku lekkich modeli sieciowych. Problem ten stanowi aktualne wyzwanie badawcze w obszarze współczesnych metod odszumiania sekwencji wideo. Autor zaproponował dwie uzupełniające się strategie, które mają na celu poprawę wykorzystania informacji temporalnej przy zachowaniu wysokiej efektywności obliczeniowej.

Pierwsza z tych strategii polega na wymuszeniu silniejszego wykorzystania kontekstu czasowego poprzez odpowiednie modyfikacje procesu uczenia (FIP – Frame Interpolation Pre-training). Zaproponowane podejście obejmuje m.in. ograniczanie jakości lub dostępności informacji z klatki odszumianej podczas treningu, co prowadzi do większego wykorzystania informacji z sąsiednich klatek bez zwiększania kosztów obliczeniowych na etapie inferencji. Strategię tę można uznać za ciekawą i jednocześnie oryginalny wkład metodologiczny w rozwój efektywnych metod odszumiania sekwencji wideo.

Rafał Długosz

Druga strategia opiera się na uproszczonym modelowaniu ruchu poprzez wykorzystanie tzw. przepływu optycznego (ang. optical flow) estymowanego w oparciu o dane ze zredukowaną rozdzielczością. Podejście to pozwala ograniczyć koszt obliczeniowy estymacji ruchu przy zachowaniu użytecznej informacji temporalnej wspomagającej proces rekonstrukcji. Choć wykorzystanie przepływu optycznego w zadaniach video restoration jest znane w literaturze, proponowane rozwiązanie zostało właściwie ukierunkowane na potrzeby lekkich modeli przeznaczonych do pracy w środowiskach o ograniczonych zasobach sprzętowych.

Istotnym atutem rozprawy jest również połączenie aspektów teoretycznych z praktyczną weryfikacją proponowanych metod, obejmującą zarówno szerokie badania eksperymentalne, jak i implementację modeli na urządzeniach "brzegowych". Takie podejście odpowiada aktualnym trendom badań w dziedzinie wizji komputerowej, gdzie coraz większy nacisk kładzie się nie tylko na jakość wyników, ale również na możliwość praktycznego wdrożenia proponowanych rozwiązań.

Podsumowując, tematyka rozprawy jest dobrze osadzona w aktualnym stanie badań nad efektywnym przetwarzaniem wideo i odpowiada na istotne wyzwania współczesnych metod opartych na głębokim uczeniu. Przedstawione kierunki badawcze należy uznać za wartościowe zarówno z perspektywy poznawczej, jak i aplikacyjnej.

2.2 Oryginalność i wkład naukowy Autora rozprawy

Za jedno z najważniejszych osiągnięć pracy można uznać zaproponowanie przez Autora nowej strategii pre-treningu (FIP). Jest to interesująca i dobrze umotywowana koncepcja, oparta na uczeniu poprzez „zadanie zastępcze”, jakim jest interpolacja klatek, przy jednoczesnym ograniczeniu dostępu do klatki centralnej oraz stopniowym przejściu do właściwego zadania w ramach etapu Post-FIP. Należy podkreślić, że stanowi to istotny wkład naukowy, ponieważ dotyczy procesu uczenia, a nie jedynie architektury modelu, a także wykazuje niezależność od konkretnej struktury sieci, co zostało potwierdzone eksperymentalnie.

Kolejnym ważnym elementem pracy jest wprowadzenie nowych metryk analizy wykorzystania kontekstu czasowego, takich jak SFD (Single Frame Difference) oraz PCIF (Percentage of Correct Intrinsic Flow). Szczególnie metryka PCIF zasługuje na uwagę, gdyż łączy informacje o gradientach sieci z interpretacją ruchu, stanowiąc oryginalne narzędzie do analizy zachowania modeli.

2.3 Zastosowana metodologia badawcza

Metodologia pracy ma kilka mocnych stron. Autor przeprowadził szereg eksperymentów obejmujących różne metody, takie jak Unet, NAFNet, Restormer czy CGNet, a także różne poziomy szumu oraz dane zarówno syntetyczne, jak i rzeczywiste (CRVD, ReCRVD). Przeprowadzone analizy są solidne i pozwalają na rzetelną ocenę wpływu poszczególnych komponentów. Uwzględniono zarówno metryki jakościowe (PSNR, SSIM, LPIPS), jak i analizę spójności czasowej oraz wykorzystania kontekstu.

Jednocześnie można wskazać pewne słabsze elementy metodologiczne. Przede wszystkim brakuje głębszego uzasadnienia teoretycznego proponowanych metod. Ponadto analiza przypadków niepowodzeń ("failure cases") jest ograniczona, a wyniki wykazują pewną zależność od doboru hiperparametrów.

2.4 Ocena Rozdziału pierwszego

Rozdział ten dostarcza obszerne i dobrze przygotowane wprowadzenie do problematyki rozprawy. Autor przedstawił szeroki kontekst badań nad odsumowaniem obrazów i sekwencji wideo, poczynwszy od konwencjonalnych metod przetwarzania sygnałów, aż po współczesne podejścia oparte na głębokim uczeniu. W przekonujący sposób uzasadnił znaczenie problemu zarówno z perspektywy jakości wizualnej, jak i zastosowań praktycznych związanych z systemami wizji komputerowej,

“edge computing” oraz przetwarzaniem w czasie rzeczywistym. Istotnym wkładem tego rozdziału jest też czytelne sformułowanie celu rozprawy oraz dwóch głównych hipotez badawczych. Autor logicznie uzasadnił potrzebę lepszego wykorzystania informacji temporalnej w lekkich modelach oraz przedstawił zaproponowane przez siebie strategie badawcze.

2.5 Ocena Rozdziału drugiego

W tym rozdziale Autor zaprezentował dobrze uporządkowany przegląd literatury związanej ze współczesnymi metodami odszumiania obrazów oraz sekwencji wideo, jak również efektywnych architektur głębokiego uczenia. Autor omówił zarówno klasyczne podejścia, jak i nowoczesne rozwiązania oparte na głębokim uczeniu, wykazując dobrą orientację w aktualnym stanie wiedzy. Omówił m.in. problematykę modelowania szumu, danych treningowych oraz ograniczenia współczesnych zbiorów danych, takich jak CRVD czy ReCRVD.

Autor trafnie zidentyfikował główne kierunki rozwoju, obejmujące podejścia wykorzystujące “optical flow”, mechanizmy “attention” oraz architektury transformatorowe. Poprawnie wskazał też najważniejsze problemy istniejących metod, takie jak ograniczone wykorzystanie informacji temporalnej przez lekkie modele, wysoki koszt estymacji ruchu czy trudności związane z pozyskiwaniem danych referencyjnych.

Pewnym mankamentem pozostaje bardzo szeroki zakres przeglądu, przez co niektóre fragmenty mają charakter bardziej encyklopedyczny niż analityczny. Porównanie metod pod kątem kompromisu “jakość – złożoność obliczeniowa – efektywność” mogłoby zostać przedstawione bardziej dogłębnie. Mimo tych uwag rozdział stanowi solidne opracowanie aktualnego stanu wiedzy i tworzy dobre uzasadnienie dla dalszych części rozprawy.

2.6 Ocena Rozdziału trzeciego

Rozdział trzeci można uznać za jeden z kluczowych elementów rozprawy. Prezentuje autorską metodę FIP (Frame Interpolation for Pre-training), której celem jest zwiększenie wykorzystania informacji temporalnej w lekkich modelach odszumiania sekwencji wideo. Autor trafnie zidentyfikował problem nadmiernego uzależnienia modeli od klatki referencyjnej i zaproponował interesującą strategię treningową polegającą na wstępnym uczeniu zadania interpolacji klatek z maskowaniem klatki centralnej. Dzięki temu model jest zmuszony do lepszego wykorzystania informacji z klatek sąsiednich.

Rozdział został przygotowany bardzo starannie pod względem metodologicznym i zawiera szeroką analizę eksperymentalną obejmującą różne architektury, poziomy szumu oraz dane syntetyczne i rzeczywiste. Na uwagę zasługuje również wprowadzenie autorskich metryk SFD oraz PCIF służących do analizy “implicit motion estimation”.

Wyniki eksperymentów pokazują, że metoda FIP poprawia wykorzystanie kontekstu temporalnego i pozwala uzyskać korzystny kompromis pomiędzy jakością odszumiania, a złożonością obliczeniową. Szczególnie dobre rezultaty uzyskano dla lekkich modeli przeznaczonych do zastosowań czasu rzeczywistego.

Pewnym ograniczeniem pozostaje stosunkowo niewielka analiza kosztu treningowego oraz ograniczona weryfikacja metody dla architektur wykorzystujących długoterminowy kontekst czasowy. Mimo tych uwag rozdział stanowi wartościowy i oryginalny wkład badawczy do omawianej problematyki.

2.7 Ocena Rozdziału czwartego

W rozdziale tym Autor rozwija swoje wcześniejsze badania nad efektywnymi metodami odszumiania strumieni wideo, koncentrując się na dwóch głównych elementach: wykorzystaniu niskorozdzielczego przepływu optycznego (LROF) oraz uproszczonej strategii treningowej Noise Imbalancing (NI). Autor pokazał, że nawet silnie zredukowany przepływ optyczny może skutecznie wspierać wyrównywanie klatek i poprawiać wykorzystanie informacji temporalnej przy

niewielkim koszcie obliczeniowym. Szczególnie wartościowa jest analiza kompromisu pomiędzy jakością odszumiania a wydajnością, jak również szeroki zakres eksperymentów obejmujących różne strategie, poziomy szumu oraz platformy sprzętowe.

Istotnym osiągnięciem rozdziału jest wykazanie, że uproszczona metoda NI daje efekty zbliżone do bardziej złożonego FIP, co sugeruje, że kluczowe znaczenie ma ograniczenie dominacji klatki referencyjnej podczas treningu. Autor bardzo dobrze analizuje również praktyczne aspekty wdrażania rozwiązań, pokazując rzeczywiste ograniczenia związane z operacjami odkształcania (warpingu) i przetwarzaniem na urządzeniach mobilnych oraz systemach wbudowanych.

Do mocnych stron należą wysoka jakość metodologiczna, szeroko analiza eksperymentalna, świadome podejście do ograniczeń zbiorów danych oraz praktyczny charakter proponowanych rozwiązań.

Jako uwagi krytyczne można wskazać nieco ograniczoną analizę wpływu degradacji przepływu optycznego na jakość rekonstrukcji oraz brak klasycznych metryk jakości przepływu, takich jak EPE (Endpoint error), co utrudnia ilościową ocenę dokładności estymacji ruchu. Niedostatecznie przeanalizowano również przypadki, w których LROF pogarsza wyniki, szczególnie dla zbiorów danych z ograniczonym ruchem, takich jak CRVD. Pewnym problemem pozostaje także silna zależność skuteczności metody od heurystycznego doboru parametrów, takich jak skala s czy próg ρ , przy braku bardziej systematycznej analizy ich wpływu na stabilność i generalizację rozwiązania.

Pewnym ograniczeniem jest też stosunkowo niewielka nowość samej idei "low-resolution flow", która częściowo wpisuje się w istniejące wcześniej podejścia "coarse-to-fine". Dodatkowo większość eksperymentów przeprowadzono na ograniczonej liczbie architektur, co utrudnia ocenę uniwersalności proponowanych metod w kontekście nowoczesnych modeli transformerowych i bardziej złożonych architektur video restoration.

Mimo tych uwag rozdział stanowi wartościowy wkład w rozwój efektywnych metod wykorzystania informacji czasowej w odszumianiu wideo.

2.8 Ocena wniosków końcowych

Rozdział piąty stanowi spójne podsumowanie rozprawy. Autor w uporządkowany sposób odniósł się do najważniejszych wyników do wcześniej postawionych hipotez badawczych. W syntetyczny sposób przedstawił osiągnięcia związane z metodami pre-treningu, niskorozdzielczego przepływu optycznego (Low-Resolution Optical Flow) oraz efektywnego wykorzystania informacji czasowej w odszumianiu sekwencji wideo. Wskazał też na różnice pomiędzy metodami FIP i NI, podkreślając większą prostotę oraz elastyczność podejścia NI.

3 Odpowiedzi na pytania recenzji

3.1 Jaki jest problem naukowy (teza) rozprawy i czy został on trafnie i jasno sformułowany?

Rozprawa dotyczy aktualnego i istotnego problemu z obszaru przetwarzania obrazów, jakim jest efektywne odszumianie sekwencji wideo z wykorzystaniem metod głębokiego uczenia. Autor koncentruje się na kompromisie pomiędzy jakością rekonstrukcji a wydajnością obliczeniową modeli, co ma szczególne znaczenie w zastosowaniach praktycznych, takich jak "edge computing", urządzenia mobilne czy systemy czasu rzeczywistego.

Autor trafnie zidentyfikował dwa główne ograniczenia współczesnych metod odszumiania obrazów, stawiając dwie tezy badawcze. Są to nadmierne uzależnienie modeli od klatki referencyjnej oraz wysoki koszt obliczeniowy klasycznych metod kompensacji ruchu. Główna teza rozprawy zakłada, że istniejące modele niewystarczająco wykorzystują informacje temporalne, szczególnie w przypadku lekkich architektur sieciowych, oraz że odpowiednio zaprojektowane strategie pre-

treningu i uproszczona estymacja ruchu mogą poprawić wykorzystanie kontekstu czasowego bez istotnego zwiększania kosztów inferencji.

Problem badawczy został sformułowany jasno, logicznie i w sposób dobrze osadzony w aktualnym stanie wiedzy. Rozprawa ma charakter przede wszystkim eksperymentalny i aplikacyjny, jednak zawiera również wartościowe elementy metodologiczne oraz analityczne.

3.2 Czy Autor rozwiązał postawiony problem i czy użył do tego właściwych metod dowodząc, że posiadał umiejętności związane z metodyką i metodologią prowadzenia badań naukowych?

Autor w sposób przekonujący rozwiązał postawiony problem badawczy, proponując dwie uzupełniające się strategie poprawy wykorzystania informacji temporalnej w odszumianiu sekwencji wizyjnych: metodę FIP (Frame Interpolation Pre-training) oraz metodę NI (Noise Imbalancing). Obie techniki zostały szczegółowo opisane, dobrze uzasadnione oraz szeroko zweryfikowane eksperymentalnie. Dodatkowo zaproponowano koncepcję Low-Resolution Optical Flow (LROF), polegającą na estymacji przepływu optycznego na silnie zredukowanej rozdzielczości w celu ograniczenia kosztów obliczeniowych przy zachowaniu użyteczności informacji o ruchu.

Zastosowana metodyka badań została przygotowana starannie i obejmuje szeroki zakres analiz eksperymentalnych, w tym badania różnych architektur sieciowych, danych syntetycznych i rzeczywistych, zróżnicowanych poziomów szumu, a także rozbudowane analizy ablacyjne. Istotnym elementem jest również połączenie ewaluacji ilościowej i jakościowej oraz testy wdrożeniowe na platformach edge AI, co znacząco wzmacnia praktyczny wymiar pracy.

Na podkreślenie zasługuje wprowadzenie autorskich metryk SFD oraz PCIF, umożliwiających analizę wykorzystania informacji temporalnej oraz "implicit motion estimation". Rozwiązania te świadczą o pogłębionym podejściu do oceny modeli, wykraczającym poza standardowe metryki jakości obrazu.

Przeprowadzone badania potwierdzają, że Autor posiada bardzo dobre przygotowanie metodologiczne oraz umiejętność samodzielnego projektowania i realizacji badań naukowych w obszarze głębokiego uczenia. Całość eksperymentów została zaplanowana w sposób spójny, systematyczny i adekwatny do postawionych hipotez badawczych.

3.3 Czy tematyka rozprawy jest aktualna lub dostatecznie ważna?

Tematyka rozprawy jest bez wątpienia aktualna oraz istotna zarówno z naukowego, jak i praktycznego punktu widzenia. Problem odszumiania sekwencji wideo z wykorzystaniem metod głębokiego uczenia należy do jednego z rozwijanych kierunków badań w obszarze komputerowego przetwarzania obrazu i widzenia maszynowego. Szczególnego znaczenia nabiera przy tym zagadnienie kompromisu pomiędzy jakością rekonstrukcji a efektywnością obliczeniową modeli, zwłaszcza w kontekście zastosowań czasu rzeczywistego oraz wdrożeń na urządzeniach brzegowych i mobilnych.

Współczesne systemy obrazowania, wykorzystywane m.in. w urządzeniach mobilnych, monitoringu, robotyce, pojazdach autonomicznych czy systemach medycznych, wymagają metod zdolnych do skutecznego usuwania szumu przy ograniczonych zasobach sprzętowych i energetycznych. W tym kontekście problematyka podjęta w rozprawie ma wyraźny wymiar aplikacyjny. Można by dodać, że tego rodzaju rozwiązania mają duże znaczenie w systemach aktywnego bezpieczeństwa w systemach w motoryzacji, w algorytmach z obszaru ADAS.

Autor nie ogranicza się jedynie do poprawy jakości rekonstrukcji, lecz analizuje także aspekty praktycznej implementacji i wdrożenia proponowanych metod na rzeczywistych platformach sprzętowych. Z tego względu temat rozprawy należy uznać za w pełni aktualny, naukowo wartościowy oraz istotny z perspektywy współczesnych zastosowań systemów przetwarzania wideo.

3.4 Na czym polega oryginalny dorobek Autora i jakie jest jego znaczenie poznawcze lub przydatność praktyczna dla nauki bądź techniki?

Najistotniejszym osiągnięciem rozprawy jest zaproponowanie nowego podejścia do zwiększania wykorzystania kontekstu czasowego w modelach do odszumiania sekwencji wideo bez istotnego wzrostu kosztów obliczeniowych. Autor pokazał, że problem niewystarczającego wykorzystania informacji z klatek sąsiednich nie wynika wyłącznie z ograniczeń architektury sieci, lecz również ze sposobu jej trenowania.

Za oryginalny wkład należy uznać opracowanie metod pre-treningu wymuszających silniejsze wykorzystanie informacji czasowej. W pierwszym podejściu (FIP – Frame Interpolation Pre-training) ograniczany jest dostęp modelu do odszumianej klatki centralnej poprzez jej maskowanie, co zmusza sieć do rekonstrukcji informacji na podstawie klatek sąsiednich i prowadzi do lepszego “implicit motion learning”. Drugie podejście (Noise Imbalancing – NI) polega na czasowym zwiększeniu poziomu szumu w klatce centralnej przy jednoczesnym zmniejszeniu szumu w klatkach sąsiednich. Rozwiązanie to stanowi interesującą i praktycznie bardzo użyteczną alternatywę dla FIP, gdyż nie wymaga dodatkowego etapu przejściowego podczas treningu.

Istotnym elementem nowatorskim rozprawy jest również wykazanie, że estymacja przepływu optycznego może być wykonywana na obrazach o radykalnie zmniejszonej rozdzielczości bez istotnej utraty jakości odszumiania. Zaproponowane wykorzystanie LROF (Low-Resolution Optical Flow) pozwala znacząco ograniczyć koszt obliczeniowy kompensacji ruchu, dzięki czemu mechanizmy przepływu optycznego mogą zostać wykorzystane także w modelach przeznaczonych do zastosowań w systemach czasu rzeczywistego oraz urządzeniach brzegowych. Wyniki eksperymentalne pokazują, że nawet bardzo silna redukcja rozdzielczości zachowuje wystarczającą informację o ruchu, aby poprawić jakość rekonstrukcji.

Na uwagę zasługuje również analityczna część pracy dotycząca “implicit motion learning”. Autor proponuje własne metryki oceny wykorzystania kontekstu czasowego, w szczególności PCIF, która stanowi interesujące narzędzie do badania stopnia zgodności działania modelu z rzeczywistym ruchem pomiędzy klatkami.

Dodatkową wartością pracy jest jej silny aspekt praktyczny. Autor nie ograniczył się do analizy jakości rekonstrukcji, lecz przeprowadził również kompleksową ocenę efektywności obliczeniowej proponowanych rozwiązań, obejmującą analizę FPS, GMACs, opóźnień czasowych oraz wdrożeń na platformach brzegowych oraz mobilnych. Szczególnie wartościowe jest pokazanie, że proponowane rozwiązania mogą działać w czasie rzeczywistym na urządzeniach takich jak np. Jetson AGX Orin czy Samsung Galaxy S24.

Całość rozprawy stanowi spójne i dobrze uzasadnione połączenie elementów metodologicznych, analitycznych i praktycznych. W mojej ocenie najważniejszą nowością pracy jest wykazanie, że odpowiednio zaprojektowana strategia uczenia może znacząco poprawić wykorzystanie informacji czasowej przez lekkie modele odszumiania strumieni wideo, bez konieczności stosowania kosztownych architektur lub złożonych mechanizmów estymacji ruchu.

3.5 Czy rozprawa świadczy o dostatecznej wiedzy Autora, wiedzy na zaawansowanym poziomie, o charakterze podstawowym dla dziedziny nauk technicznych oraz o charakterze szczegółowym, odpowiadającej obszarowi prowadzonych badań naukowych

Rozprawa świadczy o bardzo dobrej wiedzy Autora zarówno w zakresie metod głębokiego uczenia, przetwarzania obrazu i sekwencji wideo, jak również praktycznych aspektów implementacji modeli sztucznej inteligencji na platformach edge AI. Autor wykazuje bardzo dobrą orientację w aktualnym stanie wiedzy oraz swobodnie porusza się w zagadnieniach związanych z architekturami sieci neuronowych, metodami odszumiania strumieni wideo, “implicit motion learning”, “optical flow” oraz optymalizacją modeli pod kątem efektywności obliczeniowej.

Na szczególne podkreślenie zasługuje bardzo dobra znajomość współczesnej literatury przed-

miotu. Rozprawa odwołuje się do licznych aktualnych prac badawczych i właściwie osadza prezentowane wyniki w kontekście istniejących metod state-of-the-art. Autor wykazuje umiejętność krytycznej analizy istniejących rozwiązań, identyfikacji ich ograniczeń oraz formułowania własnych hipotez badawczych.

Praca świadczy również o wysokich kompetencjach eksperymentalnych Autora. Przeprowadzono dużą liczbę złożonych eksperymentów obejmujących różne architektury sieciowe, poziomy szum, zbiory danych oraz platformy sprzętowe. Szczególnie wartościowe są obszerne badania ablacyjne oraz analiza kompromisu pomiędzy jakością rekonstrukcji a efektywnością obliczeniową modeli. Autor potrafi poprawnie interpretować uzyskane wyniki i formułować na ich podstawie uzasadnione wnioski.

Rozprawa wskazuje także na dużą świadomość praktycznych aspektów implementacyjnych. Autor analizuje zagadnienia związane z wydajnością obliczeniową, opóźnieniami przetwarzania, wykorzystaniem pamięci oraz wdrożeniem modeli na rzeczywistych platformach sprzętowych, takich jak Jetson AGX Orin, Qualcomm Snapdragon czy Samsung Galaxy S24. Świadczy to o dojrzałym podejściu do problemu badawczego oraz bardzo dobrym przygotowaniu Autora zarówno w warstwie teoretycznej, jak i praktycznej.

3.6 Czy rozprawa obejmuje najnowsze osiągnięcia nauki i świadczy o znajomości współczesnej literatury z dyscypliny naukowej, której dotyczy?

Tak. Rozprawa obejmuje bardzo szeroki i aktualny zakres literatury dotyczącej współczesnych metod przetwarzania obrazu i wideo z wykorzystaniem głębokiego uczenia. Autor przedstawił obszerny przegląd badań związanych z takimi zagadnieniami jak: "video denoising", "optical flow", "lightweight deep learning", "transformer architectures", "implicit motion learning" oraz "edge AI deployment", a także wielu zagadnień pokrewnych.

Na szczególne podkreślenie zasługuje fakt, że przegląd literatury nie ogranicza się wyłącznie do klasycznych metod odsumowania, lecz obejmuje również najnowsze kierunki badań związane z efektywnymi architekturami głębokich sieci neuronowych, kompresją modeli, quantization-aware training, pruningiem, knowledge distillation, metodami self-supervised learning oraz współczesnymi strategiami poprawy wykorzystania kontekstu czasowego w przetwarzaniu sekwencji wideo.

W rozprawie szczegółowo omówiono również aktualny stan badań w zakresie restauracji wideo, w tym ograniczenia podejścia "frame-by-frame" oraz znaczenie wykorzystania informacji temporalnej. Uwzględniono strategie modelowania zależności czasowych, takie jak podejścia rekurencyjne, przetwarzanie wieloklatkowe, mechanizmy propagacji dwukierunkowej. Autor odnosi się także do problemów związanych z opóźnieniami i ograniczeniami modeli offline, które wymagają dostępu do całych sekwencji wideo.

Istotnym elementem przeglądu jest również uwzględnienie współczesnych benchmarków i zbiorów uczących wykorzystywanych w odsumowaniu sekwencji wideo, takich jak DAVIS, Set8, CRVD oraz ReCRVD, wraz z omówieniem trudności związanych z pozyskiwaniem danych rzeczywistych. Rozprawa pokazuje znajomość aktualnych metod generowania danych treningowych, oraz problemów związanych z realistycznym odwzorowaniem degradacji, szczególnie w kontekście ruchu, niedopasowań między klatkami oraz ograniczeń procedur uśredniania.

Rozprawa świadczy tym samym o bardzo dobrej orientacji Autora we współczesnym stanie badań w dyscyplinie oraz o umiejętności osadzenia własnych wyników badawczych w aktualnym kontekście światowej literatury naukowej.

3.7 Jakie są wady i słabe strony rozprawy?

Do słabszych stron rozprawy należy zaliczyć:

- Nieco ograniczoną analizę teoretyczną,
- Dominację podejścia eksperymentalnego,
- Niewystarczającą analizę "failure cases",
- Ograniczoną analizę niezawodności w kontekście ewentualnych zastosowań przemysłowych,
- Częściowo heurystyczny dobór parametrów,
- Ograniczoną analizę odporności na dane spoza rozkładu treningowego,
- Miejscami uproszczony sposób interpretacji wyników eksperymentalnych.

Moim zdaniem wskazane ograniczenia wynikają w znacznej mierze z charakteru rozprawy, która koncentruje się na projektowaniu i empirycznej weryfikacji złożonych schematów uczenia dla odszumiania sekwencji wideo, obejmujących m.in. FIP, NI oraz LROF. W żaden sposób nie stanowią więc zarzutów dyskwalifikujących.

Pewien niedosyt pozostawia analiza sytuacji, w których estymacja ruchu ulega degradacji, w szczególności w scenach o niskiej teksturze, przy wysokim poziomie szumu lub w obecności błędów estymacji przepływu optycznego. W systemach o znaczeniu komercyjnym jest to istotna kwestia, w kontekście ich koniecznej niezawodności w różnych scenariuszach warunków zewnętrznych, w których te systemy będą pracować (np. w zastosowaniach w obszarze automotive).

Podsumowując, przedstawione ograniczenia mają charakter typowy dla prac o silnie eksperymentalnym profilu badawczym, w których główny nacisk położony jest na projektowanie i empiryczną weryfikację nowych mechanizmów uczenia. Nie wpływają one w sposób istotny na ogólną wartość naukową rozprawy, która wnosi istotny wkład w problematykę wykorzystania informacji temporalnej w odszumianiu strumieni wideo.

3.8 8. Do której z następujących kategorii Recenzent zalicza rozprawę:

- a/ nie spełniająca wymagań,
- b/ wymagająca wprowadzenia poprawek i ponownego recenzowania
- c/ zadowalająco spełniająca wymagania
- d/ wyraźnie wykraczająca poza poziom przeciętny (spełniająca wymagania z nadmiarem),
- e/ wybitna.

Rozprawę doktorską należy ocenić jako pracę wyraźnie wykraczającą poza poziom przeciętny, jednak nieposiadającą cech typowych dla prac wybitnych.

Z jednej strony, praca wyróżnia się wysokim poziomem dojrzałości badawczej oraz inżynierskiej, co znajduje odzwierciedlenie zarówno w zaproponowanej architekturze rozwiązania, jak i w zakresie przeprowadzonych eksperymentów. Kluczowym osiągnięciem jest spójne połączenie estymacji ruchu w postaci Low-Resolution Optical Flow (LROF) z mechanizmem kontrolowanego zarządzania rozkładem szumu (NI), co prowadzi do poprawy jakości procesu odszumiania przy jednoczesnym zachowaniu wysokiej efektywności obliczeniowej. Istotnym atutem pracy jest również konsekwentne podejście systemowe, w którym poszczególne komponenty tworzą dobrze zintegrowany "pipeline" przetwarzania wideo.

Na podkreślenie zasługuje także szeroka i wieloaspektowa ewaluacja eksperymentalna, obejmująca różne poziomy degradacji, zróżnicowane typy ruchu oraz kilka odmiennych zbiorów danych. Analiza została dodatkowo uzupełniona o testy wydajnościowe na różnych platformach sprzętowych.

Z drugiej strony, rozprawa pozostaje w dużej mierze osadzona w podejściu eksperymentalno-inżynierskim. Proponowane rozwiązania opierają się w istotnym stopniu na decyzjach heurystycznych, takich jak dobór parametrów skalowania, harmonogramów treningowych czy strategii przetwarzania danych, które nie wynikają z jednolitej ramy teoretycznej. Ograniczona jest również formalna analiza mechanizmów działania zarówno LROF, jak i NI, a obserwowane zależności pomiędzy parametrami a jakością rekonstrukcji są w dużej mierze uzasadniane empirycznie.

Ponadto, mimo szerokiej ewaluacji, analiza odporności metod w warunkach silnego przesunięcia domenowego oraz w scenariuszach ekstremalnych degradacji pozostaje częściowo ograniczona. W rezultacie, choć zaproponowane podejście wykazuje dobrą stabilność w badanych warunkach eksperymentalnych, jego pełna ogólność nie została w pełni wykazana na poziomie teoretycznym.

W konsekwencji należy stwierdzić, że rozprawa stanowi bardzo solidne i nowoczesne opracowanie problemu efektywnego odszumiania sekwencji wideo, wyraźnie przewyższające poziom przeciętnych prac doktorskich. Jednocześnie jednak jej charakter pozostaje przede wszystkim systemowo-eksperymentalny, bez elementu przełomowej nowości teoretycznej, który uzasadniałby zakwalifikowanie jej jako pracy wybitnej.

4 Ocena końcowa

Przedłożona rozprawa doktorska stanowi wartościowe i oryginalne opracowanie naukowe dotyczące efektywnego odszumiania sekwencji wideo z wykorzystaniem metod głębokiego uczenia.

Autor wykazał się:

- Bardzo dobrą znajomością literatury,
- Umiejętnością prowadzenia samodzielnych badań naukowych,
- Wysoką kulturą metodologiczną,
- Zdolnością krytycznej analizy wyników,
- Sporą znajomością praktycznych aspektów związanych z implementacją uzyskanych wyników badawczych.

Rozprawa zawiera oryginalne rozwiązania naukowe oraz szeroką i wiarygodną analizę eksperymentalną. Szczególnie cenne są praktyczne aspekty pracy związane z wdrażaniem modeli na platformach edge AI.

Stwierdzam, że rozprawa spełnia wymagania stawiane rozprawom doktorskim określone w art. 187 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce, i wnoszę o dopuszczenie Autora rozprawy, mgr. inż. Piotra Kopę Ostrowskiego do dalszych etapów postępowania doktorskiego.

dr hab. inż. Rafał Długosz, prof. PBŚ