

Szczecin, 11.05.2026 r.

prof. dr hab. inż. Krzysztof Okarma
Katedra Przetwarzania Sygnałów i Inżynierii Multimedialnej
Wydział Elektryczny
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

RECENZJA ROZPRAWY DOKTORSKIEJ
dla Rady Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne
Politechniki Gdańskiej

opracowana na podstawie uchwały Rady Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne Politechniki Gdańskiej z dnia 10.02.2026 r. oraz wynikającego z niej pisma Przewodniczącego Rady prof. dr. hab. inż. Piotra Jasińskiego z dnia 2.03.2026 r.

Tytuł rozprawy: **Efficient Deep Neural Networks for Video Denoising**
Autor rozprawy: **mgr inż. Piotr Kopa Ostrowski**
Dyscyplina naukowa: **Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne**
Promotor: **dr hab. inż. Tomasz Stefański, prof. uczelni**

I. TEMATYKA, TEZA NAUKOWA I CEL ROZPRAWY

Rozprawa doktorska mgr. inż. Piotra Kopy Ostrowskiego pt. „*Efficient Deep Neural Networks for Video Denoising*” została podzielona na pięć rozdziałów. Pierwszy z nich ma charakter wprowadzający, kolejny przedstawia ogólną tematykę badań oraz powiązane prace, identyfikując luki badawcze, podczas gdy rozdział 3. oraz 4. stanowią zasadniczy merytoryczny element rozprawy. Ostatni, piąty rozdział ma charakter podsumowujący, wskazano w nim także kierunki dalszych badań. Taka struktura rozprawy jest w pełni uzasadniona, zwrócić należy także uwagę na fakt, iż zawartość rozdziału 3. oraz 4. jest ściśle związana z dwoma publikacjami, których doktorant jest współautorem. Ukazały się one w prestiżowych czasopismach wydawanych przez IEEE (*IEEE Transactions on Circuits and Systems for Video Technology* – DOI: [10.1109/TCSVT.2025.3575717](https://doi.org/10.1109/TCSVT.2025.3575717) oraz *IEEE Signal Processing Letters* – DOI: [10.1109/LSP.2026.3681179](https://doi.org/10.1109/LSP.2026.3681179)).



Tematykę rozprawy, dotyczącą odsumiania sekwencji wideo z użyciem nowoczesnych sieci neuronowych, należy uznać za aktualną i istotną z punktu widzenia rozwoju technik wizyjnych, coraz powszechniej spotykanych urządzeniach elektronicznych wyposażonych w kamery, a także w systemach robotycznych, w szczególności w robotyce mobilnej. W tym kontekście na szczególną uwagę zasługuje podejście zaproponowane przez doktoranta oparte na zmniejszeniu złożoności obliczeniowej proponowanych algorytmów, co umożliwia ich efektywną implementację sprzętową na urządzeniach brzegowych oraz platformach stosowanych coraz powszechniej w przemyśle, jak na przykład NVIDIA Jetson Orin. Biorąc to pod uwagę, uzyskane rezultaty należy uznać za nowe na tle aktualnego stanu dyscypliny naukowej *automatyka, elektronika, elektrotechnika i technologie kosmiczne*.

Motywacją podjęcia badań z zakresu efektywnego odsumiania sekwencji wideo jest z jednej strony dynamiczny rozwój głębokich sieci neuronowych a z drugiej strony potrzeba opracowania małych modeli sieci, które byłyby w stanie wykorzystać informacje z sąsiednich klatek sekwencji wideo bez zbytniego wzrostu złożoności obliczeniowej algorytmu odsumiania. Kandydat postawił dwie hipotezy badawcze. Pierwszą z nich stanowi stwierdzenie, iż ograniczenie użycia informacji z klatki referencyjnej (pozbawionej szumu) na początku procesu trenowania sieci pozwolić może na poprawę jakości finalnej sekwencji wideo, dzięki zwiększeniu znaczenia i wpływu kontekstu czasowego na uzyskiwane rezultaty. W tym celu zaproponował on dwie metody określone jako *Frame Interpolation Pre-training* (FIP) oraz niezrównoważenie szumu (ang. *Noise Imbalancing – NI*). Druga postawiona hipoteza badawcza dotyczy zwiększenia efektywności estymacji ruchu dzięki znacznemu zmniejszeniu rozdzielczości wejściowych klatek sekwencji wideo, co pozwala na oszacowanie wektorów ruchu przy zachowaniu relatywnie niskiej złożoności obliczeniowej metod przepływu optycznego na potrzeby dalszych etapów odsumiania.

Sformułowanie problemu naukowego oraz postawione tezy należy uznać za trafne, a ich udowodnienie (jak wynika z dalszej części recenzji) świadczy o tym, iż kandydat posiada umiejętności związane z metodyką prowadzenia badań naukowych, przede wszystkim w zakresie sieci neuronowych oraz technik wizyjnych, w dyscyplinie naukowej *automatyka, elektronika, elektrotechnika i technologie kosmiczne*. Użyte do rozwiązania problemu naukowego metody uznaje się za właściwe i adekwatne do postawionego problemu, w szczególności biorąc pod uwagę założenia wynikające z postulowanego ograniczenia złożoności stosowanego modelu.

II. OCENA ZAWARTOŚCI MERYTORYCZNEJ ROZPRAWY I UWAGI DYSKUSYJNE

Zasadniczym osiągnięciem Doktoranta jest opracowanie uproszczonego modelu sieci neuronowej, określonego jako *Lightweight Video Denoiser* (LVD), którego efektywność została porównana z innymi metodami stosowanymi współcześnie do odsumiania sekwencji wideo. Realizacja tego celu poprzedzona została analizą innych rozwiązań, przedstawioną w rozdziale drugim. Ma on charakter przeglądu, co stanowi dobre wprowadzenie do tematyki rozprawy. Kandydat omówił w nim zastosowania głębokiego uczenia maszynowego z uwzględnieniem efektywności sieci neuronowych, a także techniki ich optymalizacji oraz trenowania. Ponadto, przedstawione zostały aktualne wyniki badań dotyczących poprawy jakości obrazów i sekwencji wideo ze szczególnym uwzględnieniem technik odsumiania. Trafnie wskazano luki badawcze, związane z zastosowaniem modeli o niskiej złożoności obliczeniowej oraz użyciem informacji z klatek sąsiednich. Dokonany przegląd literatury w tym zakresie należy uznać za rzetelny i adekwatny do tematyki rozprawy.

W rozdziale trzecim przeanalizowany został wpływ zastosowania techniki FIP na wyniki estymacji ruchu, co stanowiło bazę współautorskiego artykułu pt. „*Lifting Deep Image Denoisers to Video With Frame Interpolation Pre-Training*” opublikowanego w czasopiśmie *IEEE Transactions on Circuits and Systems for Video Technology*. Zaproponowane wstępne trenowanie sieci bez użycia informacji z aktualnej klatki w celu interpolacji sygnału na podstawie klatek sąsiednich jest oryginalnym podejściem, dzięki któremu możliwe jest poprawienie wyników uzyskiwanych w kolejnym etapie trenowania sieci służącej do odszumiania sekwencji wideo, w którym informacje z bieżącej klatki sekwencji wideo są już wprowadzone. Warto zauważyć, iż jego weryfikacja została przeprowadzona z użyciem czterech baz testowych, co potwierdziło słuszność zaproponowanej koncepcji.

Pewne wątpliwości budzi rysunek 3.4 przedstawiający histogramy ruchu dla dwóch baz testowych (DAVIS oraz Set8), na którym brak opisu poszczególnych osi. Prawdopodobnie na osi poziomej oznaczono przesunięcie wyrażone w pikselach natomiast na osi pionowej znajdują się wartości procentowe, ale nie zostało to wyjaśnione w tekście. Nieprawidłowo rozwinięty został także skrót SSIM (zgodnie z notacją używaną przez autorów wskaźnika powinno to być *Structural SIMilarity*, bez „*Index Measure*”).

Przedstawione w Tabeli 3.4 wartości kilku wskaźników jakości obrazu (PSNR, SSIM, LPIPS) potwierdzają skuteczność zaproponowanej metody, tym niemniej wybór akurat tych trzech wskaźników nie został wystarczająco umotywowany. Należy jednak nadmienić, iż poprawa jakości uzyskiwana dzięki zastosowaniu metody FIP nie jest znaczna. O dobrym opanowaniu warsztatu naukowego świadczy także wykonanie tzw. *ablation study*, chociaż zaprezentowane w Tabeli 3.1 wartości wskazują na relatywnie niewielki wpływ poszczególnych elementów na uzyskiwane wyniki. Nie jest również jasne, jakie dokładnie wartości znajdują się w Tabeli 3.5 – można domniemywać, iż są to wartości wskaźnika PSNR. Podobna uwaga dotyczy Tabeli 3.6.

Jak wspomniano w rozprawie, możliwe i dość powszechne jest stosowanie metod gradientowych do estymacji ruchu; w tym kontekście interesujące byłoby zweryfikowanie, w jaki sposób zaproponowane metody zachowują się w przypadku obecności „cięć” poszczególnych scen, czy też ujęć, w sekwencjach wideo.

W kolejnym rozdziale dysertacji przedstawiono ideę zastosowania metod przepływu optycznego dla klatek sekwencji wideo o obniżonej rozdzielczości wraz z niezbalansowaniem szumu. Metody te stały się podstawą kolejnej, również współautorskiej, publikacji pt. „*A Lightweight RAW Video Denoiser With Coarse Alignment and Imbalanced Noise Pre-Training*”, która ukazała się niedawno w czasopiśmie *IEEE Signal Processing Letters*. Idea estymacji przepływu optycznego dla obrazów o obniżonej rozdzielczości stanowi interesującą koncepcję, podobną nieco w swoich założeniach do jednej z efektywniejszych szybkich metod binaryzacji obrazów nierównomiernie oświetlonych, gdzie obniżenie rozdzielczości pozwala na estymację rozkładu oświetlenia, która ułatwia z kolei prognozowanie obrazu w kolejnym etapie przetwarzania. Ta koncepcja, zaproponowana przez Doktoranta, jest jednym z istotniejszych elementów przedstawionej rozprawy, gdyż bardzo dobrze wpisuje się założenia dotyczące możliwości implementacji opracowanej metody w urządzeniach brzegowych o ograniczonej mocy obliczeniowej. Również w tym rozdziale przedstawione zostały wyniki *ablation study* dla bazy CRVD, z których wynika stosunkowo niewielki wpływ dokonanych uproszczeń na uzyskiwane wyniki. Potwierdzenie przewagi zaproponowanej metody LVD nad innymi modelami jasno wynika z Tabel 4.7 oraz 4.8, choć jedynym wskaźnikiem jakości obrazu, który został tutaj zaprezentowany jest wspomniany już szczytowy stosunek sygnału do szumu (PSNR). Przedstawione rezultaty są lepsze od pozostałych metod, aczkolwiek należy zwrócić uwagę, iż różnice pomiędzy

rezultatami uzyskanymi dla poszczególnych metod są niewielkie. Pewną niekonsekwencją jest fakt wskazania tych wyników dla baz CRVD oraz ReCRVD, podczas gdy w Tabeli 4.6 znajdują się wyniki uzyskane tylko dla bazy CRVD.

W przedstawionej rozprawie zabrakło nieco porównania efektów zastosowania wybranej metody przepływu optycznego z innymi często stosowanymi metodami, jak chociażby znanym algorytmem Lucas-Kanade.

Ponieważ w założeniach dotyczących niezbalansowanego szumu określono, iż w sąsiadujących klatkach szum ten będzie zredukowany, nasuwa się także pytanie, w jaki dokładnie sposób jest to osiągnięte.

Dość istotnym mankamentem dotyczącym prezentacji wyników przeprowadzonych eksperymentów jest ograniczenie wyników do wskaźnika PSNR. Dla zachowania konsekwencji prezentacji wyników oraz komplementarności rozważań celowe byłoby wskazanie przynajmniej wartości tych samych wskaźników, które zostały użyte w rozdziale poprzednim. Jest to o tyle istotne, iż wskaźnik PSNR w żaden sposób nie bierze pod uwagę struktury obrazu a jedynie różnice wartości pomiędzy pojedynczymi pikselami. Między innymi z tego względu był on wielokrotnie krytykowany w literaturze dotyczącej oceny jakości obrazów cyfrowych, jak również ze względu na jego niską korelację z ocenami subiektywnymi dla wielu rodzajów zniekształceń i zakłóceń obrazu.

W przypadku zastosowania interpolacji dwuliniowej oraz metody najbliższego sąsiada, wskazanych na str. 72, nasuwa się pytanie, czy lepsze wyniki resamplingu nie mogłyby być osiągnięte przy zastosowaniu nieco wolniejszych, aczkolwiek zazwyczaj dokładniejszych, metod interpolacji, jak np. metoda Lanczosa.

Dodatkowym potwierdzeniem zalet zaproponowanego podejścia jest porównanie wyników implementacji metody LVD dla różnych konfiguracji sprzętowych stosowanych w urządzeniach brzegowych (procesor Intel Core i9-10900X, Jetson AGX Orin, smartfon Samsung S24 oraz procesor Qualcomm Dragonwing QCS 6490) w porównaniu z modelem UNet.

Pomimo wskazanych wyżej uwag natury dyskusyjnej, uważam, iż przedstawiona rozprawa zawiera istotne nowe elementy oraz metody, stanowiące wkład w rozwój dyscypliny naukowej *automatyka, elektronika, elektrotechnika i technologie kosmiczne*, spełniając wymogi stawiane dysertacjom doktorskim w tej dyscyplinie. Obejmuje ona najnowsze osiągnięcia nauki z zakresu zastosowania głębokich sieci neuronowych do odszumiania sekwencji wideo, co świadczy o znajomości aktualnej literatury z tego zakresu. Przedstawione w niej rozwiązania mają ponadto istotne znaczenie praktyczne, zwłaszcza biorąc pod uwagę możliwości ich zastosowania w urządzeniach brzegowych o relatywnie niskiej wydajności obliczeniowej.

Ze względu na fakt, iż zasadnicze elementy rozprawy doktorskiej są tożsame z dwoma wskazanymi wcześniej publikacjami, w których co prawda Kandydat jest pierwszym współautorem, ale nie jest autorem korespondencyjnym, a także fakt, iż mają one odpowiednio 4 oraz 5 współautorów, celowe jest precyzyjne wskazanie wkładu Doktoranta oraz wszystkich pozostałych współautorów w ich powstanie. Ze względu na fakt, iż rozprawa doktorska co do zasady jest opracowaniem samodzielnym, wyjaśnienia dotyczące udziału merytorycznego poszczególnych współautorów, powinny być złożone podczas publicznej obrony rozprawy.

III. UWAGI SZCZEGÓŁOWE

Rozprawa doktorska została napisana w całości w języku angielskim, a zatem praktycznie nieuniknione było popełnienie pewnych błędów językowych. Tym niemniej w pracy jest ich niezbyt wiele, co świadczy o staranności jej przygotowania. Wśród usterek językowych i technicznych zauważonych podczas sporządzania recenzji wymienić można m.in.

- nieprawidłowe sformułowanie na str. 9: „*The following abbreviations has been used*”,
- odwrócenie kolejności cytowanych publikacji, np. na str. 31: [18, 34],
- niestosowanie wielkich liter w nazwach własnych konferencji i czasopism w bibliografii,
- niestosowanie wielkich liter w akronimach oraz nazwach własnych występujących w tytułach publikacji w bibliografii, np. CNN w [32], VRT w [34], FastDVDnet w [45], BasicVSR++ w [51], czy też VGG w [76],
- używanie oznaczenia Unet zamiast częściej stosowanego U-Net lub UNet,
- pojedyncze usterki typograficzne, np. *Qualcomm Dragonwing CS649* (brak 0) na str. 84.

Usterki dotyczące formatowania bibliografii wynikają prawdopodobnie ze stosowania – niezbędnego przy blisko 250 pozycjach literaturowych – menedżera bibliografii, jednak użytego w niewłaściwej konfiguracji. W rezultacie niektóre pozycje bibliograficzne pozbawione są na przykład informacji o numerach stron, czy też miejscu odbywania się konferencji, co jest niespójne ze sposobem opisu innych pozycji tego rodzaju.

Warto zauważyć wysoką jakość ilustracji oraz ogólnie staranne formatowanie rozprawy. Zauważone nieliczne usterki świadczą wręcz na korzyść Doktoranta – są one obecne właściwie w każdej rozprawie, natomiast w ocenianej dysertacji ich liczba jest bardzo mała. Zauważone drobne usterki nie umniejszają w żadnej mierze wartości merytorycznej ocenianej rozprawy.

IV. WNIOSKI KOŃCOWE

Recenzowana rozprawa stanowi samodzielne (przy założeniu dominującego wkładu merytorycznego Kandydata w powstanie publikacji wskazanych w rozprawie jako ściśle z nią związane) rozwiązanie problemu naukowego, który mieści się w zakresie dyscypliny naukowej *automatyka, elektronika, elektrotechnika i technologie kosmiczne*. Należy zatem wysnuć wniosek, iż Doktorant wykazał się znajomością zagadnień dotyczących przetwarzania i analizy obrazów cyfrowych oraz sekwencji wideo, jak również wybranych metod sztucznej inteligencji, głównie z zakresu głębokich sieci neuronowych. Przedstawiona rozprawa potwierdza zatem również umiejętność zaplanowania i praktycznej realizacji badań naukowych oraz eksperymentów, także w powiązaniu z innymi dyscyplinami – w tym przypadku interdyscyplinarność rozważanego problemu wiąże się z pokrewną dyscypliną *informatyka techniczna i telekomunikacja*. Nie jest to jednak element podważający w jakikolwiek sposób związek z dyscypliną, w której kandydat wnioskuje o nadanie stopnia doktora nauk inżynieryjno-technicznych, tym bardziej, iż zagadnienia związane z przetwarzaniem i analizą obrazów cyfrowych, widzeniem maszynowym oraz

komputerowym, a także pewnymi aspektami sztucznej inteligencji są w znacznej mierze wspólne dla obu wymienionych dyscyplin naukowych.

Zakładając jedynie pomocniczy udział współautorów publikacji tożsamyh merytorycznie z rozprawą w ich powstaniu, stwierdzam, iż przedstawiona do recenzji rozprawa doktorska mgr. inż. Piotra Kopy Ostrowskiego pt. *„Efficient Deep Neural Networks for Video Denoising”* (w języku polskim *„Wydajne głębokie sieci neuronowe do usuwania szumów w sygnale wideo”*), której promotorem jest dr hab. inż. Tomasz Stefański, prof. uczelni, zrealizowana w dziedzinie nauk inżynieryjno-technicznych w dyscyplinie *automatyka, elektronika, elektrotechnika i technologie kosmiczne*, spełnia wymagania stawiane rozprawom doktorskim przez aktualnie obowiązującą ustawę *Prawo o szkolnictwie wyższym i nauce* z dnia 20 lipca 2018 roku (tekst jednolity Dz. U. z 2024 r. poz. 1571). **Stawiam zatem wniosek o jej przyjęcie i dopuszczenie do publicznej obrony.**



84213788

Ankieta dla celów statystyczno-analitycznych

Recenzent: prof. dr hab. inż. Krzysztof Piotr Okarma

Autor rozprawy: mgr inż. Piotr Kopa Ostrowski

Tytuł rozprawy: "Efficient Deep Neural Networks for Video Denoising"

Promotorzy:

1. dr hab. inż. Tomasz Stefański, profesor uczelni (promotor)

	tak	nie	1/
1. Czy teza rozprawy jest dobrze sformułowana?	☒	☐	☐
2. Czy tematyka rozprawy jest aktualna lub dostatecznie ważna?	☒	☐	☐
3. Czy wyniki są nowe na tle stanu dyscypliny, której rozprawa dotyczy?	☒	☐	☐
4. Czy rozprawa świadczy o tym, że autor posiadał:			
a) wiedzę na zaawansowanym poziomie?	☒	☐	☐
b) umiejętności związane z metodyką i metodologią prowadzenia badań naukowych?	☒	☐	☐
5. Jaki charakter ma rozprawa ^{2/}			
teoretyczny:	☒	☐	☐
doświadczalny:	☒	☐	☐
konstrukcyjny:	☐	☐	☐
technologiczny:	☒	☐	☐
6. Czy wyniki analityczne były weryfikowane?	☒	☐	☐
7. Czy wyniki pracy nadają się do bezpośredniego zastosowania w praktyce?	☐	☐	☒
8. Czy rozprawa zawiera błędy merytoryczne?	☐	☒	☐
9. Czy rozprawa jest poprawna pod względem językowym i redakcyjnym?	☒	☐	☐
10. Do której z następujących kategorii Recenzent zalicza rozprawę:			
a) niespełniająca wymagań,	☐	☐	☐
b) wymagająca wprowadzenia poprawek,	☐	☐	☐
c) zadowalająco spełniająca wymagania,	☒	☐	☐
d) wyraźnie wykraczająca ponad poziom przeciętny,	☐	☐	☐
e) wybitna.	☐	☐	☐

11.V.2026
data

podpis

1/ częściowo

2/ w przypadku charakteru mieszanego zaznaczyć jednocześnie kilka pozycji