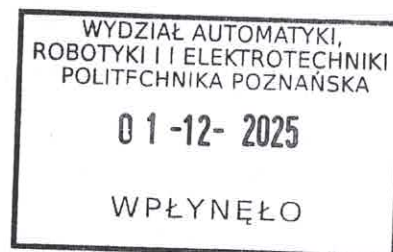


Prof. dr hab. inż. Michał Kawulok
Katedra Algorytmiki i Oprogramowania
Wydział Automatyki, Elektroniki i Informatyki
Politechnika Śląska

Gliwice, 27 listopada 2025



RECENZJA

rozprawy doktorskiej pt.

„A Multi-Module Object Detection System Based on Deep Neural Networks”

autorstwa Pani mgr inż. Aleksandry Kos

1 Podstawa wykonania recenzji

Rozprawa doktorska Pani magister inżynier Aleksandry Kos zatytułowana „A Multi-Module Object Detection System Based on Deep Neural Networks” została opracowana w roku 2025 na Wydziale Automatyki, Robotyki i Elektrotechniki Politechniki Poznańskiej. Promotorem pracy jest Pan dr hab. inż. Dominik Belter, prof. PP, a promotorem pomocniczym – Pan dr inż. Karol Majek. Wykonanie recenzji zostało zlecone dnia 19 września 2025 r. przez Przewodniczącego Rady Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne, Pana prof. dr. hab. inż. Wojciecha Szelaąg.

2 Tematyka rozprawy i ocena jej aktualności

Rozprawa doktorska Pani Aleksandry Kos dotyczy wykrywania obiektów zajmujących niewielki obszar w obrazach o wysokiej rozdzielczości, z uwzględnieniem zastosowań wymagających działania w czasie rzeczywistym. Zadanie to stanowi poważne wyzwanie z dwóch powodów. Po pierwsze, na skuteczność detektora niekorzystnie wpływa niewielki rozmiar obiektów (rozumiany jako wielkość zajmowanego przez nie obszaru w pikselach, a niekoniecznie wynikający z ich wymiarów fizycznych). Po drugie, analiza obrazów o wysokiej rozdzielczości (czyli złożonych z dużej liczby pikseli) wymaga czasochłonnej analizy, co prowadzi do problemów wydajnościowych, zwłaszcza w systemach czasu rzeczywistego. Przedstawione w pracy rozwiązania skutecznie adresują obydwie kwestie.

Rozwiązania opisane w pracy doktorskiej dobrze odpowiadają na problemy występujące w rzeczywistych zastosowaniach. Stały wzrost dostępności danych wysokiej rozdzielczości prowadzi do problemów związanych z ich przesyłaniem, gromadzeniem i przetwarzaniem, natomiast skuteczna analiza danych realizowana w czasie rzeczywistym może pozwalać na ich odpowiednią selekcję i indeksację, ułatwiającą dalsze kroki ich przetwarzania. Ma to szczególne znaczenie w przypadku obrazowania realizowanego przez urządzenia cechujące się silnymi ograniczeniami transferu oraz składowania danych, takimi jak drony czy satelity. Rozwiązania opracowane w recenzowanej pracy doktorskiej między innymi przybliżają możliwość prowadzenia wydajnej analizy danych realizowanej na urządzeniach końcowych bezpośrednio po ich akwizycji. Podsumowując, stwierdzam, że **podjęta tematyka badawcza jest aktualna, a Autorka rozprawy we właściwy sposób ukierunkowała swoje prace badawcze.**

3 Zawartość rozprawy i ocena jej układu

Rozprawa została napisana w języku angielskim, składa się z ośmiu rozdziałów, bibliografii oraz spisu symboli. Ponadto pracę otwiera strona tytułowa, streszczenie w języku polskim i angielskim, a także strona z podziękowaniami. Łącznie praca obejmuje 185 stron maszynopisu.

Rozdział 1 przedstawia problematykę wykrywania małych obiektów w obrazach o wysokiej rozdzielczości, uzasadnienie podjęcia tej tematyki, a także zarys opracowanego rozwiązania. Ponadto sformułowana została główna teza pracy oraz trzy tezy pomocnicze. Następnie Autorka opisuje zawartość poszczególnych rozdziałów oraz przedstawia artykuły, w których zostały opublikowane wyniki zawarte w dysertacji.

Rozdział 2 opisuje istniejące rozwiązania w zakresie detekcji obiektów, ze szczególnym uwzględnieniem technik wykrywania obiektów zajmujących niewielki obszar w obrazach o wysokiej rozdzielczości. Przedstawione są tam również metryki służące do oceny jakości detekcji oraz śledzenia obiektów w sekwencjach wideo. Przegląd ten jest wyczerpujący, przy czym szkoda, że Autorka nie zdecydowała się na przedstawienie bardziej przekrojowych analiz, na przykład w postaci schematów lub tabel obrazujących taksonomię metod i wskazujących na ich podobieństwa oraz różnice. Ponadto w rozdziale 2.5 przedstawione są główne kategorie metod detekcji małych obiektów (str. 24) – jedną z nich jest augmentacja danych, która moim zdaniem nie powinna się w takim zestawieniu znaleźć.

Rozdział 3 zawiera wysokopoziomowy opis opracowanego rozwiązania z uwzględnieniem kolejnych etapów przetwarzania danych – oszacowania lokalizacji regionów zainteresowania (ang. *region of interest*, ROI), ich predykcji na podstawie śledzenia, właściwego wykrywania obiektów oraz filtracji wykrytych lokalizacji. Ponadto w rozdziale przedstawione zostały zbiory benchmarkowe, które zostały wybrane do walidacji eksperymentalnej.

Trzy kolejne rozdziały (4, 5 i 6) są poświęcone poszczególnym etapom opracowanego podejścia. W rozdziale 4 opisane są techniki wykrywania ROI w pojedynczych obrazach, natomiast rozdział 5 przedstawia wykorzystanie informacji zawartej w sekwencjach wideo. Rozdział 6 opisuje metody filtracji wykrytych obiektów ukierunkowane na odpowiednie scalanie wyników pracy detektora w wielu analizowanych oknach. Każdy z tych rozdziałów składa się z krótkiego wprowadzenia, w którym w większości powtarzane są informacje opisane w rozdziale 3, następnie pojawia się uzasadnienie nowości opracowanych rozwiązań, ich szczegółowy opis, przedstawienie wyników badań eksperymentalnych, w tym analizy ablacyjnej. Każdy z rozdziałów podsumowany jest punktem zawierającym wnioski wpływające z przedstawionych analiz.

W rozdziale 7 Autorka przedstawiła szczegóły dotyczące implementacji opracowanego systemu oraz jego przystosowania do uruchomienia na systemie wbudowanym. Pracę podsumowuje rozdział 8, w którym uzasadniono potwierdzenie postawionych tez oraz przedstawione zostały kierunki dalszych prac badawczych.

Struktura pracy jest logiczna i typowa dla dysertacji doktorskich – rozpoczyna się od wprowadzenia zawierającego omówienie tez pracy, przedstawia stan wiedzy w dziedzinie, a następnie prezentuje poszczególne elementy opracowanego systemu. W mojej ocenie istotną wadą pracy, wynikającą częściowo z przyjętego układu, jest bardzo duża ilość powtórzeń prezentowanych informacji, które wprowadzają pewien chaos, utrudniają czytanie pracy i niepotrzebnie zwiększyły jej objętość. Moim zdaniem dałoby się uniknąć tych sytuacji bez zmieniania układu, przy czym uważam, że lepszym rozwiązaniem byłoby przedstawienie informacji zawartych w rozdziałach 3–6 w dwóch rozdziałach – pierwszy z nich opisywałby opracowane rozwiązania, natomiast drugi opisywałby walidację eksperymentalną. Ponadto rozważałbym przeniesienie informacji zawartych w rozdziale 7

do załącznika – zdecydowanie doceniam opublikowanie implementacji opracowanych metod, przy czym instrukcja używania kodu nie jest kluczowym elementem dysertacji. Na koniec chciałbym odnotować brak spisu rysunków oraz tabel, które standardowo są w pracach umieszczane i ułatwiają nawigację. Pomimo wszystkich wspomnianych uwag **układ pracy oceniam jako poprawny**.

4 Najważniejsze osiągnięcia rozprawy

W mojej ocenie najważniejszym osiągnięciem Autorki rozprawy jest opracowanie, implementacja, oraz walidacja systemu realizującego cały proces wykrywania obiektów z uwzględnieniem całego spektrum kwestii praktycznych związanych z wydajnością obliczeniową. Praca nie zawiera jednoznacznych informacji dotyczących wdrożenia systemu w praktyce, jednak lektura pracy wskazuje na dogłębne zrozumienie wymagań stawianych rzeczywistym systemom. Sądzę, że opracowane rozwiązanie może być wdrożone w praktyce, a na pewno stanowi ważny krok w tym kierunku.

Na uwagę zasługuje także szeroka analiza stanu wiedzy, która pozwoliła na wskazanie luk badawczych związanych ze stosowaniem istniejących rozwiązań w praktyce. Większość metod opisywanych w literaturze naukowej opracowywana jest w oderwaniu od konkretnych zastosowań – nawet jeżeli do walidacji są wykorzystywane dane rzeczywiste, to wiele faktycznych problemów inżynierskich pozostaje niezaadresowanych. Kompleksowy system opisany w pracy skutecznie tę lukę zapełnia, integrując metody znane z literatury z nowymi technikami bazującymi na regułach heurystycznych, które są ukierunkowane na podniesienie wydajności obliczeniowej. Dobrze obrazują to testy skoncentrowane na kwestiach wydajnościowych, przedstawiające czułość algorytmów na wartości hiperparametrów oraz analizujące sytuacje błędnej detekcji – przykładowo opisane w rozdziałach 4.5.2, 5.4.2, i 5.5.2.

Do osiągnięć rozprawy zaliczam także upowszechnienie uzyskanych wyników w ramach czterech artykułów opublikowanych w czasopiśmie z listy JCR oraz pięciu artykułów w materiałach konferencyjnych. Niezwykle ważne jest także opublikowanie kodu implementującego opracowane techniki – istotnie zwiększa to szanse na to, że prace te zostaną zauważone przez środowisko naukowe i przyczynią się do dalszego rozwoju metod detekcji obiektów w obrazach wysokorozdzielczych.

5 Uwagi krytyczne i dyskusyjne

Pierwsza część uwag dotyczących rozprawy ma charakter dyskusyjny i spodziewam się, że zostanie omówiona w czasie publicznej obrony:

1. Teza pracy (oraz trzy tezy pomocnicze) sformułowana została bardzo ogólnie. Moim zdaniem postawioną tezę można potwierdzić już na podstawie analizy literatury naukowej, bez potrzeby przeprowadzenia prac badawczych przedstawionych w rozprawie doktorskiej. W zasadzie przegląd stanu wiedzy zaprezentowany w rozdziale 2 już tę tezę potwierdza. Wobec tego prosiłbym Doktorantkę o wskazanie, czy Jej zdaniem któreś elementy tezy nie znajdują jednoznacznego potwierdzenia w literaturze przedmiotu, względnie o przeformułowanie tezy w taki sposób, żeby jej potwierdzenie uzasadniało wykonanie badań przedstawionych w pracy.
2. Przedstawione rozwiązanie ma budowę modułową, łączącą elementy oparte o sieci głębokie z technikami heurystycznymi. W pracy przedstawione są argumenty uzasadniające zalety takiego podejścia, jednak w mojej opinii realizacja całego procesu przez sieć neuronową na

zasadzie *end-to-end* powinna przynieść wymierne korzyści, przekładając się na poprawę skuteczności. Czy Doktorantka brała pod uwagę opracowanie tego rodzaju rozwiązania i zastąpienie szeregu reguł heurystycznych rozwiązaniami w pełni trenowalnymi?

3. Wielokrotnie pojawia się nawiązanie do kwestii przetwarzania danych w czasie rzeczywistym, jednak pojęcie to nie jest jednoznacznie w pracy zdefiniowane. Wymagania takie zwykle wynikają z dynamiki procesu fizycznego, dla którego system jest projektowany. W pracy pojęcie to jest używane w kontekście szybkości przetwarzania mierzonej liczbą klatek na sekundę (ang. *frames per second*, FPS), bez wskazania rzeczywistych ograniczeń czasowych narzuconych przez aplikację. Podanie wartości FPS bez odniesienia do tempa zmian zjawiska nie pozwala jednoznacznie ocenić, czy system spełnia kryteria pracy w czasie rzeczywistym. Ponadto w systemach czasu rzeczywistego istotna jest nie tylko wartość FPS, ale również opóźnienie, czyli czas potrzebny od momentu akwizycji obrazu do przedstawienia wyniku detekcji.
4. Jedną z ważniejszych kwestii poruszonych w pracy są problemy dotyczące czasu przetwarzania obrazów. Zaproponowane zostały w tym celu między innymi algorytmy SegTrackDetect, OBS, OBM. Jaka jest złożoność obliczeniowa tych algorytmów? Jak się ona ma do złożoności innych metod opisanych w literaturze? Czy przewaga w zakresie wydajności wykazana w pracy może się utrzymać w przypadku wykrycia większej liczby obiektów?
5. Jakie są ograniczenia opracowanych algorytmów? Sporo uwagi poświęcono ograniczeniom istniejących metod, a także szeroko omówione zostały ograniczenia algorytmu TinyROI, które są adresowane przez jego ulepszoną wersję (nazwaną SegTrackDetect). Wskazana jednak byłaby szersza analiza dotycząca ograniczeń algorytmów SegTrackDetect, OBM i OBS.
6. Autorka w tekście operuje pojęciami obiektów małych (ang. *small objects*) lub drobnych (ang. *tiny objects*), jednak terminy te nie są jednoznacznie zdefiniowane. Z tabel 2.1 i 2.2 wynika, że pojęcia te są różnie rozumiane w zależności od kontekstu; dalej w pracy czytamy między innymi, że drobne obiekty zajmują obszar kilku pikseli (str. 29), a na str. 24 pojawia się informacja, że małe obiekty cechują się małym rozmiarem względnym – nie są to wystarczająco precyzyjne określenia. Ponadto na str. 107 opisywany jest rysunek 5.9 w odniesieniu do jakości segmentacji obiektów drobnych i bardzo drobnych (ang. *very tiny and tiny*) – nie dostrzegam takiego rozróżnienia na tym wykresie. Prosiłbym o jednoznaczne zdefiniowanie tych pojęć w trakcie obrony.
7. W rozdziałach 1, 2, 4, i 5 (str. 2, 14, 18, 24, 55, 82) pojawia się stwierdzenie, że małe obiekty mogą przestać być widoczne w obrazach pomniejszonych. W literaturze znane są podejścia pomniejszania adaptacyjnego, które zachowuje cechy lokalne drobnych obiektów. Zagadnienie to jest analizowane przykładowo w pracy:

- Marin, Dmitrii, et al. „Efficient segmentation: Learning downsampling near semantic boundaries.” Proc. IEEE/CVF ICCV. 2019.

Czy tego rodzaju podejścia mogłyby znaleźć zastosowanie w analizowanym przypadku?

8. Opracowane metody zostały porównane z innymi rozwiązaniami znanymi z literatury, jednak ich dobór nie jest w pełni uzasadniony. W szczególności w rozdziale 4 do porównań została wykorzystana wyłącznie metoda SAHI – porównanie zdecydowanie zyskałoby, gdyby została uwzględniona jakaś metoda z grupy *focus-and-detect*. W rozdziale 6 algorytmy OBS i OBM

są porównane wyłącznie z podstawowym wariantem metody NMS, podczas gdy istnieje wiele jej rozszerzeń – niestety z przedstawionych w pracy badań nie wynika, czy OBS i OBM posiadają istotną przewagę nad ulepszonymi wariantami metody NMS (np. soft-NMS). Dla jasności – nie oczekuję, że zostaną wykonane jakieś dodatkowe testy, jednak prosiłbym, żeby Doktorantka odniosła się do tej uwagi w trakcie obrony.

9. Na str. 95 pojawia się stwierdzenie, że w zbiorze danych pojawiają się nieścisłości w adnotacjach dla obiektów średniej wielkości. Szkoda, że ta myśl nie została rozwinięta – dobrze by było przedstawić przykłady takich błędów i przeanalizować ich wpływ. W dalszej części paragrafu pojawia się teza, że model DiffusionVID lepiej radzi sobie z zaszumionymi etykietami – taki wniosek może być prawidłowy, jeżeli szum dotyczy zbioru treningowego, natomiast pytanie czy tego rodzaju wnioski są uzasadnione w przypadku błędnych etykiet dla zbioru testowego.

Druga część uwag krytycznych dotyczy przede wszystkim kwestii językowych i redakcyjnych, a także jest związana ze sposobem prezentacji informacji zawartych w rozprawie:

1. W tekście pracy Autorka wielokrotnie powraca do kwestii, które zostały wcześniej omówione. W wielu przypadkach informacje są powtarzane, czasami uzupełniane, jednak takie podejście zdecydowanie utrudnia lekturę pracy oraz odnalezienie w niej istotnych informacji. Parę przykładów (w żadnym razie nie jest to wyczerpująca lista):
 - Zaproponowany proces analizy obrazów przedstawiony jest na rysunkach 3.1, 4.2, 5.2 i 7.1. Dwa ostatnie rysunki są niemal identyczne. Również towarzyszące rysunkom opisy powielają te same informacje.
 - Moduł szacujący położenie ROI (ang. *ROI Estimation Module*) omówiony jest w rozdziałach 3.1.1, 4.3.1 i 5.3.1 i opisy te niewiele się od siebie różnią – przykładowo we wszystkich przypadkach pojawia się informacja, że jest on oparty na architekturze sieci U-Net i wykorzystuje sieć ResNet-18 (wzmianka o tym pojawia się jeszcze w innych miejscach, np. w rozdziale 5.4).
 - Opis zbioru MTSD przedstawiony jest w rozdziale 2.4 oraz później w rozdziale 3.2.
 - Wspomniany wcześniej w punkcie 7 niniejszej recenzji problem wykrywania obiektów w obrazach pomniejszych również jest dyskutowany wielokrotnie.
2. W tekście pracy występuje znacząca i nieuzasadniona redundancja definicji skrótów. Przykładowo, skrót ROI został zdefiniowany 29 razy, NMS – 34 razy, a OBS – aż 47 razy! Niestety takich przykładów jest w pracy o wiele więcej, co negatywnie wpływa na czytelność pracy.
3. W pracy wyróżnione są metody *anchor-based* i *anchor-free*, jednak nie zostało wyjaśnione co te pojęcia oznaczają.
4. W tekście zbyt często pojawiają się określenia opisowe i mało konkretne, na przykład „substantial computational resources,” „very slow processing times,” „works well.” Oczywiście ich stosowanie jest dopuszczalne i często uzasadnione, jednak w mojej ocenie takich sformułowań jest zbyt wiele. Ponadto Doktorantka często operuje pojęciem „significant” w odniesieniu do wyników liczbowych, co sugeruje istotność potwierdzoną testami statystycznymi, które jednak nie zostały w pracy zastosowane.

5. Przedstawione na rysunkach przykłady są trudne w interpretacji – nie są one w sposób wystarczający opisane, a kolory użyte do oznaczania poszczególnych obiektów często pozostają niewyjaśnione. Przykładowo dotyczy to rysunków 4.6, 4.8, 5.3, 6.5 – nie wiadomo, które wyniki detekcji Autorka uznaje za prawidłowe, które za błędne i dlaczego. Szkoda, że prezentacja tych przykładów nie została dopracowana, bo z pewnością byłyby one pomocne w zrozumieniu zasady działania poszczególnych technik.
6. Różnice pomiędzy metodami SegTrackDetect i TinyROI zobrazowane na Alg. 1 są niewielkie – moim zdaniem należałoby TinyROI traktować jak zubożoną wersję SegTrackDetect, a różnice między nimi mogłyby zostać zbadane w ramach analizy ablacyjnej.
7. Zauważyłem również parę innych drobnych błędów i nieścisłości:
 - We wzorze (2.7) sumowane są elementy o indeksach od 0 do T , gdzie T to liczba klatek w sekwencji. W takim układzie klatek jest jednak $T+1$ (względnie indeksowanie powinno zaczynać się od 1).
 - Na str. 90 pojawia się literówka („ths” → „thus”).
 - Na str. 101: „dilation, based” → „dilation-based”?
 - Na str. 106 w opisie tabeli 5.4 pojawia się informacja, że w nawiasach podany jest zysk w zakresie wydajności (ang. *performance gain*), przy czym w przypadku obniżenia wartości FPS bardziej adekwatnym stwierdzeniem byłby spadek wydajności (ang. *performance drop*).
 - Na str. 109: „These findings validate the third thesis...” – chyba nie chodziło tutaj o trzecią tezę, bo dotyczy ona czegoś innego.
 - Skrótoby OBS i OBM są w pracy napisane inną czcionką, co wprowadza pewien chaos typograficzny. Ujednolicenie formatowania poprawiłoby czytelność i estetykę pracy.
 - Tabela 4.1 i rysunek 4.5 przedstawiają te same wyniki w różnej postaci – tabelarycznej i graficznej. Powinno się takiej redundancji unikać.

6 Wniosek końcowy

Przedstawiona przez mgr inż. Aleksandrę Kos rozprawa pt. „A Multi-Module Object Detection System Based on Deep Neural Networks” stanowi w mojej ocenie **oryginalne rozwiązanie problemu naukowego** i potwierdza **ogólną wiedzę teoretyczną** Doktorantki w dyscyplinie *automatyka, elektronika, elektrotechnika i technologie kosmiczne*. Przedstawione w recenzji uwagi krytyczne nie wpływają na moją jednoznacznie pozytywną ocenę rozprawy. Tym samym stwierdzam, że spełnia ona wymagania określone w art. 187 ustawy z dnia 20 lipca 2018r. Prawo o szkolnictwie wyższym i nauce.

Biorąc powyższe pod uwagę, wnioskuję o dopuszczenie Pani mgr inż. Aleksandry Kos do dalszych kroków procedury uzyskania stopnia doktora nauk technicznych, w tym do publicznej obrony.

