



PRZEWODNICZĄCY RADY DYSCYPLINY
Automatyka, Elektronika, Elektrotechnika
Technologie Kosmiczne
prof. dr hab. inż. Wojciech Szelaąg

Warszawa, dn. 13 lutego 2025

dr hab. inż. Marcin Iwanowski, prof. uczelni
Politechnika Warszawska, Wydział Elektryczny
Instytut Sterowania i Elektroniki Przemysłowej

Recenzja rozprawy doktorskiej
mgr. inż. Tomasza Nowaka
pt. „Data-efficient and explainable machine learning in visual
perception of autonomous vehicles”

1 Wprowadzenie

Niniejsza recenzja rozprawy doktorskiej została przygotowana na zlecenie Rady Naukowej Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne Politechniki Poznańskiej (uchwała nr 42/2023-2024 z dnia 18 czerwca 2024) na podstawie zawiadomienia o wyznaczeniu recenzentów podpisanego przez Przewodniczącego ww. Rady prof.dr.hab.inż. Wojciecha Szelaąg z dnia 27 listopada 2024. Promotorem recenzowanej rozprawy jest prof.dr hab.inż. Piotr Skrzypczyński, zaś promotorem pomocniczym, dr inż. Michał Nowicki.

2 Zawartość rozprawy

Recenzowana rozprawa liczy 112 stron. Została napisana w języku angielskim, składa się ze streszczenia, podziękowań, list skrótów i oznaczeń, siedmiu rozdziałów oraz bibliografii.

Rozdział pierwszy – wprowadzenie – rozpoczyna się od przedstawienia motywacji do podjęcia tematyki uczenia maszynowego w pojazdach autonomicznych. Przedstawia problemy badawcze, których praca dotyczy, zawiera opis zawartości pracy oraz omówienie osiągnięć Autora, potwierdzających Jego dorobek badawczy.

Rozdziały od drugiego do szóstego są zasadniczymi rozdziałami pracy i dotyczą kluczowych elementów składowych (rozdziały 2-5) oraz całościowej koncepcji przykładowych systemów wizyjnej do percepcji otoczenia pojazdu autonomicznego. Rozdziały te mają podobną strukturę, co ułatwia ich lekturę. Pierwszy punkt każdego z nich stanowi wprowadzenie zakończone listą zagadnień, które Autor uważa za swoje największe osiągnięcia. W drugim punkcie znajduje się omówienie pokrewnych prac z danego zakresu tematycznego. Kolejne punkty omawiają szczegółowo dane zagadnienie.

Rozdział drugi jest poświęcony detekcji obiektów. W szczególności omówione jest autorskie podejście Doktoranta, oparte na modelu Faster R-CNN, polegające na ekstrakcji mapy obszarów istotnych w procesie detekcji (mapy uwagi), jej wizualizacji, oraz sposób wykorzystania w procesie treningu sterowanego (guided learning).

Rozdział trzeci dotyczy metod ekstrakcji punktów charakterystycznych na obrazach dwuwymiarowych. Autor omawia trzy modele detekcji punktów: Regression Keypoint Network, Max Resolution Heatmap Keypoint Network oraz Geometry-aware Keypoint Network oraz przedstawia metodę estymacji stopnia niepewności detekcji. Rozdział zamyka omówienie wizualizacji gradientu.

W rozdziale czwartym, Autor omawia zagadnienie estymacji parametrów trójwymiarowego modelu obiektu sceny wizyjnej. Kolejne punkty są poświęcone estymacji współrzędnych trójwymiarowych, przygotowaniu zbioru danych uczących, funkcji straty oraz finalnej estymacji parametrów modelu trójwymiarowego.

Rozdział piąty dotyczy estymacji położenia obiektów w przestrzeni trójwymiarowej na podstawie obrazów dwuwymiarowych. Są w nim rozróżnione są dwa podejścia, omówione w osobnych podrozdziałach: dla obiektów o znanym kształcie i rozmiarach (punkty ładowania) oraz o kształcie nieznanym (inne pojazdy).

Przedstawione w rozdziałach 2-5 modele uczenia maszynowego zostały wykorzystane do stworzenia przykładowych podsystemów wizyjnych pojazdów autonomicznych na potrzeby lokalizacji punktu ładowania autobusu i autobusu względem punktu ładowania oraz detekcji i lokalizacji sąsiadujących pojazdów. Zostały one przedstawione w rozdziale szóstym. Rozdział ten zawiera wyniki eksperymentalne uzyskane przez Autora, zarówno w odniesieniu do pojedynczych metod, jak i do metod złożonych. Na początku przedstawiono wyniki eksperymentalnej ewaluacji metody wizualizacji uwagi i jej wykorzystania do treningu sterowanego. W kolejnym punkcie przedstawiono kompletną metodę estymacji położenia w systemie wspomagania zbliżania autobusu do stacji ładowania. Następny punkt, piąty jest poświęcony estymacji położenia samochodów. Ostatnie dwa punkty poświęcono odpowiednio estymacji kształtu pojazdów oraz szacowaniu niepewności.

W zakończeniu podsumowano pracę, przedstawiając wnioski z przeprowadzonych prac, wkład własny Autora oraz możliwe kierunki dalszych badań.

Na końcu pracy zamieszczono bibliografię zawierającą 125 pozycji literatury.

3 Ocena merytoryczna pracy

3.1 Problem naukowy i teza pracy

Zakres tematyczny recenzowanej pracy obejmuje systemy wizyjne pojazdów autonomicznych. Jest to obszar bardzo aktualny z uwagi na intensywnie rozwijające się metody sztucznej inteligencji na potrzeby systemów wspomagania kierowcy (ADAS, advanced driver assistance systems). Podjęcie się przez Doktoranta badań w tym obszarze uważam zatem za wybór trafny. Z uwagi na zróżnicowanie informacji wizyjnej niezbędnej do skutecznej nawigacji pojazdu, oraz na dużą i ciągle rosnącą różnorodność modeli uczenia maszynowego, pojęcie prac w tej dziedzinie w zakresie wymaganym w przypadku rozprawy doktorskiej, wymaga zawężenia tego bardzo szerokiego spektrum zagadnień do jednego, co najwyżej kilku wątków. Tak też i zrobił Doktorant, i to także należy ocenić pozytywnie. Z jednej strony przedmiotem badań uczynił wątki takie jak detekcja obiektów na potrzeby nawigacji, estymacja punktów charakterystycznych obiektów czy estymację trójwymiarowych modeli obiektów. Z drugiej strony w pracy można wyróżnić jeden wątek zrealizowany w szerszym zakresie – identyfikacja położenia (lokalizacja) punktów ładowania autobusu elektrycznego z użyciem modeli uczenia głębokiego. Wątek ten został zaawansowany w większym stopniu niż pozostałe, które skądinąd zostały także w nim skutecznie wykorzystane.

Sformułowana przez Autora w punkcie 1.2 teza badawcza została podzielona na cztery tezy pomocnicze, które brzmią następująco:

1. Architektury głębokich sieci neuronowych służących do ekstrakcji cech obrazu umożliwiają sterowanie treningiem poprzez polepszanie zbiorów uczących.
2. Architektury tego typu pozwalają na opis niepewności detekcji cech geometrycznych generowanych przez modele głębokie.
3. Zastosowanie trójwymiarowych modeli znanych obiektów pozwala na polepszenie jakości detekcji punktów charakterystycznych na dwuwymiarowych obrazach, stanowiących bazę do ekstrakcji tych modeli.
4. Znajomość ograniczeń geometrii znanych modeli obiektów umożliwia minimalizację błędnych detekcji cech i zwiększa zarazem dokładność ich lokalizacji.

Tak postawione tezy, choć mogłyby być sformułowane nieco bardziej precyzyjnie, dobrze lokalizują tematykę i rysują plan badawczy Autora.

3.2 Analiza stanu wiedzy

Autor wykorzystał w pracy 125 źródeł literaturowych. Zważywszy na rozległość tematyki, nawet z uwzględnieniem jej niezbędnego zawężenia do wymienionych wyżej wybranych wątków, kluczowym aspektem pracy był dobór metod, zatem i referujących ich artykułów, które należało uwzględnić w pracy. Dokonany przez Doktoranta wybór należy w tym kontekście ocenić pozytywnie. Wprawdzie jest zauważalna koncentracja Autora na źródłach sprzed kilku lat, jednak zważywszy na tempo rozwoju dziedziny oraz cykl przygotowawczy rozprawy doktorskiej, problem ten można uznać za typowy dla aktualnie realizowanych prac z zakresu uczenia maszynowego.

Autor słusznie rozdzielił przegląd literatury na poszczególne rozdziały pracy 2-6, gdzie każdy drugi podrozdział ma właśnie charakter przeglądowy i prezentuje przegląd rozwiązań z danego zakresu. Ponadto, we wprowadzeniu (rozdział 1) znajduje się omówienie wybranych aspektów wizyjnej percepcji pojazdów z niezbędnymi odniesieniami do literatury.

3.3 Proponowane rozwiązania

Zakres tematyczny pracy, zgodnie z jej tytułem obejmuje zagadnienia efektywności oraz objaśnialności modeli uczenia maszynowego w wizyjnej percepcji otoczenia pojazdu. Wyniki uzyskane przez Autora można zgrupować w cztery wątki tematyczne.

Wątek pierwszy (rozdział 2) obejmuje detekcję obiektów. Przedmiotem badań była metoda detekcji obiektów oparta na klasycznym modelu Faster R-CNN. Wprowadzona przez Autora modyfikacja polegała na wyprowadzeniu z modelu warstwy atencyjnej wizualizującej aktywne obszary obrazu w kontekście uwidocznionym na nim obiektów. Informacja o tych obszarach skutecznie objaśnia zachowanie sieci dla konkretnych obrazów, co zostało wykorzystane do zwiększenia efektywności uczenia sieci. Na bazie tej uwagi zaproponowano metodę sterowanego treningu sieci zwiększającą skuteczność i dokładność modelu, optymalizując wykorzystanie danych. Opracowane podejście zostało zweryfikowane eksperymentalnie, co opisano w punkcie 6.3.

Wątek drugi (rozdział 3) dotyczy detekcji punktów charakterystycznych na obrazach dwuwymiarowych. W ramach badań opracowano trzy modele służące do precyzyjnej ekstrakcji punktów

charakterystycznych. Metody te zostały opracowane z wykorzystaniem danych uczących ukazujących punkty ładowania autobusów elektrycznych. Dodatkowo zaproponowano także nową funkcję straty uwzględniającą informacje o trójwymiarowym modelu obiektu i ograniczeń geometrycznych. W ramach tego wątku zaproponowano także metodę szacowania niepewności detekcji punktów charakterystycznych wraz z możliwością zwiększenia precyzji ich lokalizacji na podstawie modelu trójwymiarowego. Opracowane podejścia zostało wykorzystane w metodzie lokalizacji punktu ładowania co opisano w punkcie 6.5.1 rozprawy.

Wątek trzeci (rozdział 4) jest poświęcony estymacji parametrów trójwymiarowego modelu obiektu sceny wizyjnej w zastosowaniu do detekcji pojazdów na obrazie dwuwymiarowym. Opracowano w tym celu model uczenia maszynowego estymujący punkty charakterystyczne samochodów w przestrzeni trójwymiarowej na podstawie pojedynczego zdjęcia. Metoda estymuje położenie kilkudziesięciu punktów charakterystycznych pojazdu. Dzięki temu może być wykorzystana do automatycznego etykietowania pojazdów polegającego na wskazaniu tych punktów. Informacja o wynikach eksperymentalnych została zamieszczona w punkcie 6.5.2. Dodatkowo opracowano metodę ekstrakcji siatki opisującej kształt pojazdu na podstawie obrazu. Weryfikacja eksperymentalna opracowanego podejścia w tym zakresie została opisana w punkcie 6.6.

Wątek czwarty (rozdział 5) dotyczy estymacji położenia kamery na podstawie obrazów dwuwymiarowych wraz z oceną jej dokładności. Zweryfikowano w nim praktycznie metody estymacji parametrów obiektu – w tym przypadku punktu ładowania autobusu – w celu określenia położenia kamery względem tego obiektu. Dzięki temu możliwe było oszacowanie pozycji autobusu względem punktu ładowania co umożliwiło dokowanie autobusu do punktu ładowania. Weryfikacja eksperymentalna opracowanego podejścia została opisana w punkcie 6.4. Oprócz powyższego zagadnienia lokalizacji obiektu o znanych cechach geometrycznych: kształcie i rozmiarze (punkty ładowania), w ramach tego wątku opracowano także metodę określania położenia obiektów o kształcie nieznanym w zastosowaniu do lokalizacji pojazdów.

3.4 Oryginalność rozprawy

Rozwiązania zaproponowane przez Autora są oryginalne i konkurencyjne wobec rozwiązań znanych z literatury. Wyniki eksperymentalne uzyskane w wyniku badań dowodzą, postawionych w pierwszym rozdziale, czterech tez. Oprócz wyników cząstkowych, ważnym efektem prac jest metoda lokalizacji punktu ładowania autobusu elektrycznego, oparta o rozwiązania zaproponowane przez Autora, która została zweryfikowana podczas rozbudowanych badań eksperymentalnych, realizowanych w ramach projektu badawczego, w którym Doktorant brał aktywny udział. Wyniki prac badawczych przedstawionych w pracy zostały opublikowane w jednym artykule w czasopiśmie naukowym (International Journal of Applied Mathematics and Computer Science) oraz w siedmiu artykułach opublikowanych w materiałach międzynarodowych konferencji naukowych, co także potwierdza ich stopień zaawansowania.

4 Dyskusja nad rozprawą

4.1 Uwagi merytoryczne

Lektura pracy pozwoliła na sformułowanie następujących merytorycznych uwag dyskusyjnych:

1. **Wykorzystanie alternatywnych komponentów w proponowanych modelach** Opracowane rozwiązanie wykorzystujące detekcję obiektów obejmuje modyfikację klasycznej sieci Faster R-CNN wykorzystującej do ekstrakcji cech obrazu model Resnet. Zarówno jednak Faster R-CNN jak i Resnet są modelami dość wiekowymi, opracowanymi i udostępnionymi niemal 10 lat temu. Od tego czasu pojawiła się duża liczba modeli nowszych, zarówno przeznaczonych do ekstrakcji cech obrazu, jak i modeli atencyjnych generujących mapy uwagi wskazujące na istotne części obrazu. Jakie nowsze modele mogłyby zostać użyte jako alternatywa do metody użytej w pracy ? Jak mogłyby one wpłynąć na finalne wyniki pod względem efektywności obliczeniowej i dokładności ?
2. **Wykorzystanie znaczników do lokalizacji.** Zagadnienie lokalizacji obiektów w przestrzeni na podstawie zadanych dwuwymiarowych obrazów cyfrowych przedstawiających te obiekty, w ogólnym przypadku, jest zadaniem trudnym w związku z licznymi możliwymi zakłóceniami i zniekształceniami. Dlatego często, gdy jest taka możliwość, stosuje się znaczniki ułatwiające lokalizację. Punkt ładowania pojazdu wydaje się być obiektem, dla którego taka możliwość zachodzi – są to obiekty powtarzalne, z możliwością umieszczenia na nich niezbędnego znakowania np. znaczników typu ARUCO. Znaczniki takie mogą być stosunkowo efektywnie i precyzyjnie lokalizowane. Czy zastosowanie tego typu podejścia pozwoliłoby na zwiększenie dokładności i szybkości lokalizacji punktu ładowania ?
3. **Uniwersalność opracowanej metody detekcji punktów charakterystycznych.** Zaproponowane metody detekcji punktów charakterystycznych zostały przygotowane pod konkretne zastosowanie detekcji punktów ładowania autobusów. W jakimś zakresie podejście to mogłoby zostać rozszerzone na inne rodzaje obiektów w obszarze zainteresowania systemu wizyjnego pojazdu ? Czy metody wymagałyby modyfikacji i jeśli tak – to jakich ?
4. **Inne przypadki użycia.** Uwaga jest rozwinięciem uwagi poprzedzającej i dotyczy innych przypadków użycia zaproponowanego podejścia. Takimi przypadkami mogą być zadania takie jak podjazd autobusu do przystanku, sygnalizacji świetlnej, innego pojazdu itp. Każde z tych zadań może wiązać się z koniecznością użycia – całkowicie lub częściowo – innego podejścia do analizy sceny wizyjnej. W jakim stopniu zaproponowane podejście mogłoby zostać wykorzystane w innych typowych (takich np. jak wymienione powyżej) przypadkach użycia ?

4.2 Uwagi redakcyjne

Praca została napisana w poprawnym języku angielskim. Jest staranna pod względem edycyjnym. W pracy widać jednak trudności jaki sprawiał dość jednak zróżnicowany zakres tematyczny pracy. Opis poszczególnych metod w rozdziałach od drugiego do piątego jest spójny – każdy rozdział ma jasną i zrozumiałą strukturę, co znakomicie ułatwia ich lekturę. Jednak już konstrukcja rozdziału 6 jest bardziej skomplikowana i wymaga od czytającego precyzyjnego połączenia opisanych wyników z konkretnymi metodami opisanymi we wcześniejszych rozdziałach. Wprawdzie niekiedy Autor wskazuje stosowne powiązania (wp. w punkcie 6.3 czy 6.6), jednak w innych przypadkach takiej informacji brak (np. punkt 6.7). Podobna uwaga dotyczy przyporządkowania wyników raportowanych w podrozdziałach rozdziału 6 do odpowiednich tez pracy. Tutaj wprawdzie to przyporządkowanie zostało wskazane wprost w liście umieszczonej w punkcie 7.2, jednak w przypadku trzeciej i czwartej tezy miejsca, w których są one udowodnione są rozrzucone po kilku częściach

pracy, co utrudnia analizę. Rozwiązaniem tych problemów mogłoby być przeformułowanie tez lub zmiana struktury tej części pracy.

Z uwagi na staranną edycję pracy, liczba znalezionych podczas lektury pracy drobnych błędów czy nieścisłości jest krótka:

- str.32, punkt 3.5 – nie jest jasny tytuł "gradient visualization". Celem wizualizacji wykorzystanej w tym miejscu jest prezentacja obszarów aktywnych dla konkretnego obrazu wejściowego. Klasyczna metoda używana w tym celu istotnie używa informacji o gradientach. Jednak Autor skorzystał (co nie budzi zastrzeżeń) z metody Score-CAM, która właśnie z niej nie korzysta. Wobec tego tytuł podrozdziału nie odpowiada istocie zadania w nim opisanego.
- str. 42, rysunek 4.5 – "Median of all intersection points- pierwsze słowo zawiera literówkę, powinno być raczej "Median ...". Czy jednak na pewno w tym miejscu obliczana jest mediana? To nie wynika z tekstu.
- str.59, wzór 5.5 jest $\hat{p}_i^{2d}$, powinno być $\hat{p}_i^{3d}$.
- str.86-88, rysunki 6.21,6.22,6.23 w podpisach jest mowa o słupkach czarnych i szarych, tymczasem na wykresach są czerwone i niebieskie.
- str. 95, rysunek 6.25 podpis, jest "3 sigma", powinno być " 3σ ".

5 Podsumowanie

Podsumowując, niezależnie od zawartych w niniejszej opinii uwag krytycznych, które w większości są elementem dyskusji nad pracą, uważam, że Autor pracy wykazał wysokie kompetencje w wybranym przez siebie obszarze badawczym. Recenzowana rozprawa zawiera oryginalne rozwiązania problemów naukowych oraz pokazuje umiejętności Autora w zakresie samodzielnego prowadzenia pracy naukowej. **Całościowa ocena rozprawy jest pozytywna.** Uważam, że rozprawa spełnia wymagania stawiane rozprawom doktorskim przez stosowne, aktualnie obowiązujące, akty prawne. Dlatego wnioskuję o dopuszczenie rozprawy doktorskiej mgr. inż. Tomasza Nowaka do publicznej obrony.