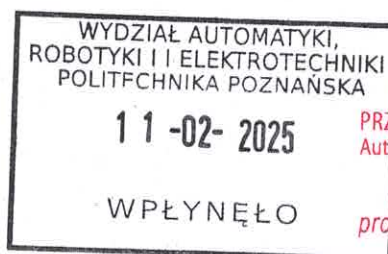


Tytuł, stopień, imię i nazwisko
prof. dr hab. inż. Bogdan Kwolek

data 4.02.2025 r.



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

RECENZJA ROZPRAWY DOKTORSKIEJ

przygotowana dla Rady Dyscypliny Automatyka, Robotyka, Elektrotechnika i Technologie
Kosmiczne Politechniki Poznańskiej

Tytuł rozprawy: Data-efficient and explainable machine learning in visual perception for autonomous vehicles

Autor rozprawy: mgr inż. Tomasz Nowak

Promotorzy: prof. dr hab. inż. Piotr Skrzypczyński, dr hab. inż. Michał Nowicki

1. Przedstawiona mi do recenzji rozprawa doktorska dotyczy percepcji wizualnej pojazdów autonomicznych. Zasadnicza część pracy dotyczy metod i rozwiązań, które zwiększają efektywność wykorzystania danych treningowych do uczenia modeli głębokich, a także rozwiązań umożliwiających wyjaśnialność decyzji systemów (modeli) percepcji wizualnej pojazdów autonomicznych. Zasadniczym celem prac badawczo-eksperymentalnych było opracowanie i praktyczne zweryfikowanie konkurencyjnych rozwiązań oraz architektur sieci neuronowych dla pojazdów autonomicznych. Postawiono oryginalne tezy badawcze, które wyrażają się z przekonaniem, iż (i) w oparciu o mapy ciepła z mechanizmami uwagi (ang. *attention heat maps*) oraz wizualizację znaczących cech pośrednich możliwe jest ukierunkowanie procesu uczenia sieci neuronowych, w tym poprzez aktywne uzupełnianie istniejących zbiorów danych, (ii) architektury głębokich sieci neuronowych oferują mechanizmy opisywania niepewności związanej z cechami geometrycznymi wyznaczanymi przez sieć, (iii) sieci neuronowe mogą uczyć się odkrywania cech geometrycznych na obrazach bez konieczności dokładnego etykietowania danych, w tym na ograniczonych rozmiarowo poetykietowanych danych, a także że (iv) włączenie wiedzy o ograniczeniach geometrycznych wynikających z modeli obiektów do architektur głębokiego uczenia zmniejsza liczbę fałszywych detekcji i jednocześnie zwiększa dokładność lokalizacji cech. Prawdziwość tez badawczych wykazano empirycznie. Przyjęty cel pracy zrealizowano proponując autorskie rozwiązania i metody do (i) wizualizacji przestrzennych obszarów zainteresowania w sieciach Faster R-CNN, która to wraz z procedurą uczenia kierowanego poprawia uczenie sieci, w szczególności w razie ograniczonej dostępności danych treningowych, (ii) metody szacowania niepewności charakterystycznych punktów 2D oraz 3D obiektów oraz propagacji wspomnianych oszacowań do estymat pozycji obiektu, (iii) metody szacowania współrzędnych 3D punktów charakterystycznych samochodu oraz reprezentacji kształtu pojazdu w formie gęstej siatki, (iv) metody do automatycznego etykietowania punktów charakterystycznych 3D pojazdów, (v) procedurę przetwarzania punktów charakterystycznych, która koryguje

niedokładne wykrycia wspomnianych punktów na podstawie ograniczeń geometrycznych, (vi) potok przetwarzania zaprojektowany do oszacowania pozy samochodów w bezpośrednim sąsiedztwie w stosunku do danego pojazdu w oparciu o pojedynczy obraz RGB z uwzględnieniem niepewności szacowania pozy. Autor zaproponował oryginalne rozwiązania, przygotował kompletne implementacje dla systemów autonomicznych, a także zweryfikował je w dwóch scenariuszach. Pierwszy scenariusz zakładał wykonanie manewru dokowania do stacji ładowania przez elektryczny autobus miejski. W omawianym eksperymencie proponowane metody skutecznie wyznaczały pozę (pozycję i orientację) autobusu względem stacji ładowania, co pozwalało na precyzyjne dokowanie z marginesem błędu nieprzekraczającym 30 cm na dystansie ponad 30 metrów. Drugi scenariusz obejmował estymację pozy pojazdów znajdujących się w otoczeniu w środowisku miejskim. Dokładność szacowania pozy tych pojazdów została zweryfikowana przy użyciu zbioru danych ApolloCar3D, który zawiera obrazy zarejestrowane w przestrzeni miejskiej. Na wyżej wymienionym zbiorze danych zaproponowana metoda uzyskiwała wyniki konkurencyjne w porównaniu do wyników uzyskiwanych przez metody state-of-the-art. Niepewność oszacowania pozy oceniano w zadaniu dokowania do stacji ładowania. Oceniano także niepewność szacowania pozycji otaczających pojazdów. Tematyka podjęta w rozprawie jest w pełni uzasadniona pod względem poznawczym, a na płynące z niej wnioski istnieje zapotrzebowanie w obszarze robotyki i sztucznej inteligencji. Wyniki badań eksperymentalnych są rzetelne i przekonujące.

2. Rozprawa napisana jest w języku angielskim. Praca składa się z pięciu zasadniczych rozdziałów, wstępu oraz podsumowania. Rozdziały pracy poświęcono detekcji obiektów, estymacji punktów charakterystycznych 2D, wyznaczaniu trójwymiarowego kształtu obiektu, estymacji pozy oraz wykorzystaniu percepcji wizualnej przez pojazdy autonomiczne. Rozprawa liczy łącznie 112 stron. Bibliografia liczy łącznie 125 pozycji. Streszczenia pracy przedstawiono zarówno w języku angielskim jak i polskim. W części wstępnej zamieszczono wykaz skrótów oraz wykaz oznaczeń i symboli. We wprowadzeniu przedstawiono motywację do podjęcia badań, sformułowano problematykę badań oraz postawiono oryginalne tezy badawcze, w dalszej części omówiono zawartość rozprawy, zagadnienia badawcze i proponowane rozwiązania, a następnie wyszczególniono współautorskie publikacje. W rozdziale 2 zaproponowano metodę wizualizacji obszarów zainteresowania w sieci neuronowej do detekcji obiektów. Proponowane rozwiązanie umożliwia lepsze zrozumienie i optymalizację modelu sieci poprzez wskazanie, które regiony obrazu są istotne. Proponowana procedura uczenia kierowanego poprawia uczenie sieci i uzyskiwane wyniki, w tym w przypadku ograniczonej ilości danych treningowych. Rozdział 3 dotyczy szacowania punktów charakterystycznych na obrazach, koncentrując się na rozwiązaniach dla stacji ładowania autobusów elektrycznych. Przedstawiono trzy architektury sieci oraz nową funkcję straty, a także metodę szacowania niepewności predykcji sieci. W rozdziale 4 zaprezentowano architekturę sieci neuronowej do szacowania współrzędnych 3D punktów charakterystycznych na podstawie pojedynczego obrazu oraz funkcję straty do uczenia sieci w oparciu o reprojektację. Zaproponowano metodologię automatycznego pozyskiwania rzeczywistych lokalizacji punktów charakterystycznych 3D w oparciu o punkty charakterystyczne 2D oraz modele CAD obiektów. Zaprezentowano także dwie architektury sieci do kodowania i szacowania kształtów samochodów. W rozdziale 5 zaprezentowano potok przetwarzania do estymacji pozy, który łączy detekcję punktów charakterystycznych 2D z estymacją współrzędnych 3D punktów charakterystycznych. W proponowanym potoku wykorzystano algorytm PnP (Perspective-n-

Point), który opiera się na korespondencjach 2D-3D. Ponadto, celem dostarczenia miar niepewności związanych z estymacjami pozy zaproponowano mechanizm szacowania niepewności, który propaguje niepewności z wykrytych punktów charakterystycznych 2D oraz estymat punktów 3D. W rozdziale 6 zademonstrowano praktyczne zastosowania metod omówionych w rozdziałach od 2 do 5, ilustrując je studiami i wynikami prac eksperymentalnych na danych pozyskanych w rzeczywistych scenariuszach. W rozdziale zamieszczono studium przypadków oceniające wpływ poszczególnych komponentów na uzyskiwane wyniki. Wyniki uzyskiwane przez proponowane metody odniesiono do wyników uzyskiwanych przez najbardziej konkurencyjne rozwiązania i systemy percepcji pojazdów autonomicznych. Rozdział 7 podsumowuje pracę doktorską i prezentuje wkład rozprawy w rozwój metod uczenia maszynowego w obszarach wiedzy związanymi z pojazdami autonomicznymi. W części końcowej nawiązano do tych części pracy, które najściślej związane są z wykazaniem prawdziwości postawionych tez badawczych, a także zasugerowano przyszłe kierunki badań. Od strony formalnej rozprawa zredagowana jest poprawnie, z właściwą systematyką podjętych zagadnień naukowych. Kolejność rozdziałów jest logiczna, a ich tytuły oraz podtytuły w syntetyczny sposób odzwierciedlają zawartość merytoryczną. Na podkreślenie zasługują układ pracy, w tym przede wszystkim spójna chronologia omawianych zagadnień i bogata szata graficzna. Prezentacja podjętej problematyki badawczej jest kompleksowa i wyczerpująca. Autor wykazał się głębokim zrozumieniem podjętej problematyki badawczej. Zastosowana metodologia badawcza jest adekwatna do postawionych celów pracy i pozwala na uzyskanie rzetelnych wyników prac eksperymentalnych. Doktorant precyzyjnie oraz poprawnie formuje wnioski wynikające z przeprowadzonych badań empirycznych. Autorskie rozwiązania i uzyskane wyniki przeanalizowano w kontekście literatury przedmiotu, co pozwala na umiejscowienie własnego wkładu w rozwój dziedziny w szerokim kontekście naukowym. Wnioski końcowe pracy znajdują solidne uzasadnienie zarówno w proponowanych rozwiązaniach, jak i w uzyskanych wynikach empirycznych. Bogaty warsztat metodologiczny, połączony z kreatywnością Autora, zaowocowały wartościowym opracowaniem naukowym. Problematyka podjęta w recenzowanej rozprawie ma istotne znaczenie nie tylko z perspektywy naukowej, ale również praktycznej. Wyniki badań mają potencjalne zastosowanie praktyczne, co zwiększa wartość dysertacji dla środowiska akademickiego oraz poza nim. Praca stanowi cenny materiał źródłowy dla przyszłych badań nad pokrewnymi zagadnieniami.

3. Na oryginalny dorobek Doktoranta składa się metoda sterowanego uczenia sieci, która usprawnia uczenie modelu poprzez wykorzystanie informacji zwrotnej z bloku odpowiedzialnego za wizualizację istotnych obszarów obrazu dla sieci. Metodę projektowano w kontekście minimalizacji fałszywie pozytywnych detekcji. Obszary obrazu do których sieć przywiązuje istotną uwagę ilustrowane są mapami ciepła. Wspomniany mechanizm uwagi analizowano dla sieci Faster R-CNN. Wykorzystano cechy z warstwy RPN (ang. *Region Proposal Network*) do tworzenia tzw. anchors (A), które są standardowym mechanizmem w pokrewnych sieciach, spośród których połowę zarezerwowano do reprezentacji tła, zaś pozostałą liczbę do reprezentacji obiektów. Mapy ciepła wyznaczano w oparciu o cechy RPN oraz wagi wyznaczane w oparciu o wagi wykorzystywane do wyznaczenia (A). Mapy ciepła posłużyły do ukierunkowania uczenia sieci i poprawy wyników uzyskanych podczas wstępnej fazy uczenia modelu poprzez wybór przykładów negatywnych i sterowane rozszerzenie zbioru treningowego. Ewaluację proponowanej metody zrealizowano na własnym zbiorze PUT oraz zbiorze SBC (Solaris Bus and Coach), które zawierają 2000 i 1000 obrazów oraz na

dodatkowych obrazach, które nie zawierają treści związanych ze stacjami ładowania pojazdów, a także na wybranej sekwencji z powszechnie wykorzystywanego w badaniach naukowych zbioru KITTI. Eksperymenty i prezentowane wizualizacje pokazują znaczny potencjał metody, w szczególności w kontekście uczenia sieci na ograniczonych zbiorach danych. Na wartościowy i oryginalny element pracy składają się sieci RKN (Regression Keypoint Network), MRHKN (Max Resolution Heatmap Keypoint Network) oraz GAKN (Geometry-Aware Keypoint Network). Sieć GAKN w trakcie uczenia uwzględnia kontekst geometryczny dzięki uwzględnieniu w rozszerzonej funkcji celu błędów reprojektacji. Wspomniany błąd reprojektacji został zaprojektowany celem nakładania kar na wszelkie przestrzenne układy punktów charakterystycznych, które naruszają dopuszczalne konfiguracje geometryczne, zniechęcając tym samym model do przewidywania nierealistycznych konfiguracji punktów. W proponowanej architekturze sieci GAKN za uwagę zasługuje blok UEH (Uncertainty Estimation Head), który umożliwia oszacowanie niepewności współrzędnych punktów 2D. Warto także wspomnieć, że obliczenia związane z wyznaczeniem błędu reprojektacji zajmują około 4 milisekundy na obraz. Na oryginalny element pracy składa się architektura sieci neuronowej do szacowania współrzędnych 3D punktów charakterystycznych na podstawie pojedynczego obrazu oraz funkcja straty do uczenia sieci w oparciu o reprojektację. W proponowanym podejściu do estymacji współrzędnych 3D punktów charakterystycznych wyznaczane są dwie oddzielne mapy cech. Omawiane podejście z dwiema mapami, które niezależnie uwzględnia dwie ortogonalne płaszczyzny, pozwala na zastosowanie podobnego potoku przetwarzania jak w przypadku estymacji punktów charakterystycznych 2D. Ponadto dzięki takiemu podejściu, niepewność punktów charakterystycznych 3D można oszacować analogicznie jak dla punktów charakterystycznych 2D. Proponowana metodologia automatycznego pozyskiwania rzeczywistych lokalizacji punktów charakterystycznych 3D w oparciu o punkty charakterystyczne 2D oraz modele CAD obiektów jest wartościowa, w szczególności w kontekście trenowania sieci dla zastosowań związanych z percepcją wizualną pojazdów autonomicznych. Estymacja współrzędnych punktów charakterystycznych 3D realizowana jest w oparciu o punkty przecięcia siatki modelu 3D przez tory promieni przyjętego modelu kamery. Do określenia miejsc przecięcia promieni z powierzchniami modelu CAD wykorzystano algorytm Möllera-Trumbore'a. Omawiane rozwiązanie dopełnia autorska metoda reprezentacji kształtu pojazdu w formie gęstej siatki. W pierwszym etapie trenowany jest autoenkoder wariacyjny (VA) odpowiedzialny za kodowanie kształtu samochodu w przestrzeni ukrytej (ang. *latent space*), zaś w drugim etapie trenowana jest sieć zawierająca wytrenowany wcześniej moduł VA, która odpowiedzialna jest za wyznaczanie siatki pojazdu na podstawie obrazu wejściowego. Architektura sieci VA wykorzystuje sieci GCN (Graph Convolutional Networks) do kodowania i dekodowania trójwymiarowego kształtu samochodów. W części eksperymentalnej pracy porównywano dokładność wyznaczania punktów 3D w oparciu o obydwie metody 3D w oparciu o metrykę MPJPE (Mean Per Joint Position Error) oraz metrykę A3DP-Abs (Absolute Average 3D Precision). Na oryginalny dorobek Doktoranta składa się także potok przetwarzania do estymacji pozy, który łączy detekcję punktów charakterystycznych 2D z estymacją współrzędnych 3D punktów charakterystycznych. Godna uwagi jest także metoda umożliwiająca dostarczanie miar niepewności związanych z estymacjami pozy, która propaguje niepewności z wykrytych punktów charakterystycznych 2D oraz estymat punktów 3D w oparciu o transformację UT (Unscented Transform) – znaną z bezśladowego Filtru Kalmana UKF (Unscented Kalman Filter). Omawiane metody przebadano eksperymentalnie pod kątem rzeczywistych zastosowań, w tym zastosowaniach

związanych z dokowaniem elektrycznych autobusów do stacji ładowania oraz estymacji pozy pojazdów znajdujących się w sąsiedztwie, w tym także na danych rzeczywistych. Metody te z powodzeniem zastosowano do estymacji pozy autobusu w układzie współrzędnych ładowarki autobusu z odległości przekraczającej 30 metrów, wykorzystując pojedynczą kamerę RGB. Proponowane rozwiązania umożliwiają zaplanowanie i wykonanie manewru dokowania z marginesem błędu mniejszym niż 30 cm. Rozwiązania opracowane na potrzeby estymacji pozy pojazdu przebadano w rzeczywistych warunkach drogowych. Na ogólnodostępnym zbiorze danych ApolloCar3D zaproponowana metoda osiągnęła wyniki konkurencyjne z wynikami uzyskiwanymi przez rozwiązania state-of-the-art.

4. Poniższe uwagi odnoszą się do bardziej szczegółowych kwestii związanych z realizacją celu rozprawy i postawionych tez badawczych, nie są zarzutami i mają na ogół polemiczny, a nie krytyczny charakter, nie ujmujący przy tym wartości pracy. Słabych stron rozprawy jest niewiele i są one mniej istotne. W kontekście proponowanych rozwiązań dla uczenia kierowanego z pewnością nie zaszkodziłoby odniesienie się do technik uczenia sieci w oparciu o podejście Active learning, w tym w kontekście proponowanego uzupełniania zbiorów trenujących. Od strony poznawczej, w tym także w kontekście ogólności prezentowanych rozwiązań warto byłoby odnieść się do możliwości zastosowania kamery aktywnej zamiast wysokorozdzielczych kamer RGB. Interesujące byłoby szersze przeanalizowanie technik znanych pod nazwą non-maximum suppression w kontekście pełniejszego wykorzystania RPN. Pracy z pewnością nie zaszkodziłoby przeanalizowanie użyteczności w proponowanych architekturach sieci technik regularyzacyjnych typu dropout, BatchNormalization, a w szczególności przytoczenie szczegółów związanych z uczeniem sieci, np. lr, parametrów optymalizatorów, w także zagadnień związanych z implementacją, np. języka programowania, API, itp. Przy starannym przejściu rozprawy dostrzegłem szereg błędów o charakterze edycyjnym – niemniej liczba potknięć nie odbiega od normy. Przykładowe niedociągnięcia edycyjne: Fig.6.2 (brak odstępu), niespójny z resztą pracy podpis tabeli 6.1 (pod tabelą), HRNet estimates (top) – powinno być (left), Algorithms that solves ..., Median of all ..., Figures 6.28 illustrates ..., The results of experimental evaluations presented in Section 6.7 supports ..., błędy typu: pose translation, and 3D dimension estimation ..., by outlier characteristic point estimations ..., The table 6.9 compares an approach described in Section ..., niespójna pisownia HRNet – HrNet, ApolloCar 3D – ApolloCar3D. W kilku miejscach w tekście pracy niepoprawnie użyto "position" zamiast "pose". W kilku zdaniach nie zaszkodziłoby zacytowanie prac, np.: This method, while generally slower than one-stage detectors ..., the Nelder-Mead optimization algorithm ..., W dysertacji w kilku miejscach nie wyjaśniono na jakich CPU / GPU realizowano pomiary czasu wykonania, np.: ... takes approximately 4 milliseconds per image.

Powyższe uwagi mają przede wszystkim charakter dyskusyjny i nie umniejszają w najmniejszym stopniu wartości pracy, która stanowi oryginalny wkład w dziedzinę robotyki i inteligencji maszynowej.

5. Przetawiona mi do recenzji rozprawa doktorska mgr. inż. Tomasza Nowaka proponuje nowatorskie i efektywne rozwiązania dla percepcji wizualnej pojazdów autonomicznych. Rozprawa potwierdza znaczącą wiedzę Doktoranta w obszarze robotyki oraz szeroko rozumianej inteligencji maszynowej, umiejętność samodzielnego prowadzenia badań naukowych, a także opanowaniem warsztatu badawczego, zarówno od strony teoretycznej, jak i eksperymentalnej. Pan mgr. inż. Tomasz Nowak wykazał się znajomością matematycznych

podstaw wymaganych do projektowania metod uczenia głębokich sieci neuronowych w oparciu o małe repozytoria danych trenujących. Rozprawa jest ważnym przyczynkiem w zakresie wiedzy dotyczącej projektowania i realizacji rozwiązań mających na celu rozszerzenie możliwości interpretacyjnych sieci neuronowych oraz zwiększenie skuteczności głębokich sieci neuronowych w zadaniach percepcji wizualnej pojazdów autonomicznych. W podsumowaniu stwierdzam, że recenzowana rozprawa mgr inż. Tomasza Nowaka spełnia wszystkie wymagania ustawowe i zwyczajowe stawiane rozprawom doktorskim. Stwierdzam niniejszym, że przedstawiona mi do recenzji rozprawa doktorska spełnia kryteria określone w art. 13 ust. 1 Ustawy. Wnoszę o dopuszczenie Pana mgr. inż. Tomasza Nowaka do kolejnych etapów postępowania o nadanie stopnia doktora nauk inżynieryjno-technicznych. Ponadto, wysoki poziom merytoryczny pracy, potencjał praktyczny zaprojektowanych i zrealizowanych rozwiązań, a w szczególności nietrywialnie i nowatorskie rozwiązania zaproponowane w pracy, poparte współautorską publikacją w czasopiśmie z IF są czynnikami przemawiającymi za wyróżnieniem rozprawy doktorskiej Pana mgr. inż. Tomasza Nowaka.

Bohdan Kwilek