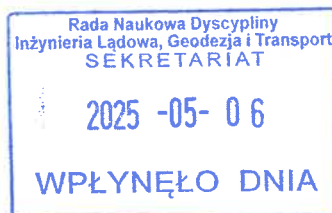


*Przyjmuję pod względem
formalnym
06.05.2025* *U87*

Wrocław, 28.04.2025 r.

Dr hab. inż. Grzegorz Józków
Instytut Geodezji i Geoinformatyki
Uniwersytet Przyrodniczy we Wrocławiu



RECENZJA

Rozprawy doktorskiej mgr inż. Pauliny Zachar

pod tytułem:

„Wykorzystanie sztucznej inteligencji do wykrywania obiektów na danych fotogrametrycznych”

1. Podstawa formalna

Formalną podstawą recenzji jest pismo RND.ILGiT.04.02.2025 Przewodniczącej Rady Naukowej Dyscypliny Inżynieria Lądowa, Geodezja i Transport Politechniki Warszawskiej, dr hab. inż. Katarzyny Osińskiej-Skotak, prof. uczelni, z dnia 20.02.2025 r.

2. Charakterystyka i ocena układu pracy

Praca doktorska mgr inż. Pauliny Zachar pt. „Wykorzystanie sztucznej inteligencji do wykrywania obiektów na danych fotogrametrycznych” przygotowana została na Politechnice Warszawskiej w dyscyplinie naukowej Inżynieria Lądowa, Geodezja i Transport pod kierunkiem prof. dr. hab. inż. Zdzisława Kurczyńskiego (promotor).

Oceniana rozprawa obejmuje 167 stron, wśród których zasadniczy tekst pracy spisany został na 140 stronach. Pozostałe części pracy stanowią: strona tytułowa, podziękowania, streszczenia (w języku polskim i angielskim), spisy treści, bibliografii, rysunków i tabel.

Zasadniczy tekst pracy podzielony został na 3 części: wstęp, część teoretyczna, część eksperymentalna. We wstępie Doktorantka nakreśla ogólnie problemy badawcze związane z wykorzystaniem sztucznej inteligencji w fotogrametrii, co stanowi motywację podjętych przez nią badań. W dalszej części wstępu podaje główny cel badawczy określony jako: „(...) opracowanie metodyki wykrywania obiektów na ukośnych zdjęciach lotniczych z wykorzystaniem metod głębokiego uczenia, a dokładniej konwolucyjnych sieci neuronowych w celu zasilania miejskich baz danych.”, a także definiuje 6 celów szczegółowych:

- Wybór klas obiektów kluczowych z perspektywy zarządzania przestrzenią miejską;
- Dobór danych fotogrametrycznych do zadania detekcji obiektów małej architektury miejskiej;
- Rozwiązanie problemu braku danych uczących;
- Porównanie różnych metod detekcji obiektów;
- Wykorzystanie aspektu przestrzennego – zasilanie baz obiektów miejskich;
- Ocena metod określania położenia obiektów w układzie terenowym;

oraz 4 pytania badawcze.

Następnie mgr inż. Paulina Zachar podaje tezę pracy: „Zastosowanie metod głębokiego uczenia wykazuje przydatność w wykrywaniu obiektów miejskich, a wykorzystanie potencjału lotniczych zdjęć ukośnych umożliwia określenie położenia obiektów w terenowym układzie współrzędnych i może być źródłem zasilania baz danych o obiektach miejskich.”. Wstęp kończy się krótką charakterystyką zakresu przeprowadzonych badań i wykorzystanych metod badawczych.

Druga część pracy nazwana przez Doktorantkę „Częścią teoretyczną” to przegląd literatury. W rozdziale 1 opisane zostały metody głębokiego uczenia maszynowego wykorzystywane w fotogrametrii i teledetekcji, w szczególności w odniesieniu do splutowych sieci neuronowych (ang. CNN – *Convolutional Neural Network*) i ich wykorzystania do detekcji obiektów na obrazach oraz segmentacji tych obrazów. Autorka szczegółowo opisuje również dwa modele takich sieci: SAM (*Segment Anything Model*) oraz YOLO (*You Only Look Once*), które wykorzystuje w swoich badaniach (podrozdział 1.1.2). W kolejnych podrozdziałach opisane zostały dwa aspekty związane z brakiem danych uczących służących trenowaniu sieci w konkretnym zadaniu. W podrozdziale 1.1.3 opisana została zasada działania mechanizmu sztucznego rozszerzenia danych uczących (ang. *data augmentation*), a w podrozdziale 1.1.4 technika wykorzystania w procesie trenowania jednego modelu „wiedzy zdobytej” w procesie trenowania innego modelu uczenia maszynowego, czyli tzw. transfer wiedzy (ang. *transfer learning*). Na końcu rozdziału 1 opisane zostały dane fotogrametryczne, które przetwarzane są z wykorzystaniem metod głębokiego uczenia maszynowego: obrazy 2D (podrozdział 1.2.1), chmury punktów 3D (podrozdział 1.2.2), modele siatkowe i bryłowe 3D (podrozdział 1.2.3). Rozdział 2 przeglądu literatury to opis uczenia maszynowego w kontekście przestrzeni miejskiej. Doktorantka opisała zastosowanie metod uczenia maszynowego do analizy przestrzeni miejskiej (rozdział 2.1), stan wiedzy dotyczący metod uczenia maszynowego do detekcji obiektów małej architektury (rozdział 2.2), a także opisała jakie dane mogą być wykorzystane do tego celu jednocześnie dokonując szczegółowego przeglądu dostępnych zbiorów obrazowych danych uczących, które potencjalnie mogłyby być wykorzystane w jej badaniach (rozdział 2.3). Na końcu części teoretycznej mgr inż. Paulina Zachar opisuje metodę przenoszenia informacji pomiędzy danymi fotogrametrycznymi różnego typu, tzw. żonglowanie reprezentacjami (ang. *juggling with representations*) (rozdział 2.4), automatyzacji tworzenia danych referencyjnych (podrozdział 2.4.1) oraz wykorzystania danych syntetycznych do tworzenia zbiorów uczących (podrozdział 2.4.2).

Trzecia część rozprawy zatytułowana została „Część eksperymentalna”. W rozdziale 3 (Autorka kontynuuje numerację rozdziałów z poprzedniej części) przedstawiony został plan eksperymentów. Doktorantka wymienia też rodzaje obiektów przestrzeni miejskiej, które będą wykrywane na danych fotogrametrycznych, a także określa jakie rodzaje danych będą wykorzystywane. Rozdział 4 stanowi ogólny opis wykorzystanych w badaniach danych, w szczególności danych lotniczych pozyskanych sensorem Leica CityMapper 2. W następnym rozdziale Doktorantka opisuje obiekty badawcze (okna i latarnie uliczne), w szczególności przedstawiając jak wyglądają latanie na różnych rodzajach danych fotogrametrycznych (chmurach punktów różnego typu, zdjęciach lotniczych, ortofotomapie). Dalej (rozdział 6) opisane zostały przeprowadzone analizy dotyczące gęstych chmur punktów pod kątem ich wykorzystania do detekcji latarni ulicznych. Ponadto w tym rozdziale przedstawiono również obszar opracowania i wykorzystane dane, a także sposób

przygotowania danych referencyjnych lotniczego skaningu laserowego (ALS – ang. *Airborne Laser Scanning*). W rozdziale 7 mgr inż. Paulina Zachar przedstawia eksperyment dotyczący detekcji latarni z wykorzystaniem modelu YOLO. Oprócz opisu sposobu podziału danych referencyjnych na 3 zbiory (treningowy, walidacyjny, testowy), opisana została tu również metoda oceny dokładności detekcji oraz zaprezentowane zostały wyniki wraz z ich dyskusją. Kolejny rozdział 9 to przedstawienie eksperymentu dotyczącego segmentacji zdjęć lotniczych w celu wydzielenia latarni z wykorzystaniem modelu SAM. Autorka wyróżniła tu dwa podrozdziały, w których opisała metodykę (rozdział 8.1) oraz wyniki segmentacji wraz z dyskusją (rozdział 8.2). W rozdziale 9 opisany został eksperyment dotyczący segmentacji zdjęć ukośnych w celu wydzielenia okien. W podrozdziałach opisane zostały wykorzystane dane (dla miast Bordeaux i Warszawa) oraz sposób ich przygotowania (rozdział 9.1), a także uzyskane wyniki wraz z ich analizą. Przedostatni rozdział rozprawy to opis eksperymentów dotyczących wyznaczania lokalizacji wykrytych obiektów. W rozdziale 10.1 przedstawiono metodykę, wyniki i dyskusję dla lokalizacji latarni ulicznych, a w rozdziale 10.2 dla okien. Ostatni rozdział rozprawy to wnioski z wykonanych badań.

Po zasadniczej części pracy znajduje się spis bibliografii, w którym umieszczono 147 pozycji, spis rysunków (65 pozycji) oraz spis tabel (25 pozycji).

Układ pracy w mojej ocenie jest generalnie poprawny. Doktorantka wyraźnie przedstawiła motywację prowadzonych badań, zdefiniowała cel ogólny oraz cele szczegółowe badań, a także podała tezę pracy. W mojej opinii cele pracy są dobrze zdefiniowane i wpisują się w tematykę aktualnie prowadzonych badań. Wyraźnie wydzielona została część teoretyczna, która opisuje obecny stan wiedzy dotyczący zagadnień sztucznej inteligencji oraz danych, mających związek z zakresem przedstawionych w rozprawie badań. W mojej opinii część teoretyczna została opisana prawidłowo (z pewnymi drobnymi uwagami) i prezentuje bardzo dobrze ogólną wiedzę Doktorantki w dyscyplinie naukowej, w której prowadzi badania, w szczególności w zakresie metod sztucznej inteligencji w fotogrametrii. Część rozprawy dotycząca pracy własnej opisuje prawidłowo wykonane badania począwszy od wyboru obiektów zainteresowania, danych testowych, metodyki badań, wyników oraz ich dyskusji. Umieszczone na końcu wnioski wynikają z przedstawionych wyników. Wykorzystana bibliografia jest również prawidłowa. Mgr inż. Paulina Zachar korzystała z aktualnych pozycji literatury anglojęzycznej, przede wszystkim z artykułów naukowych opublikowanych w czasopismach naukowych, artykułów będących materiałami konferencyjnymi lub manuskryptów wydanych przez publikacją w czasopismach naukowych, tzw. *preprint*. Pewne zastrzeżenia odnośnie układu pracy można mieć w kontekście niewielkich braków, podziału treści i ich ustrukturyzowania. Z tematem pracy wiążą się różne pojęcia używane przez Doktorantkę w rozprawie, np.: sztuczna inteligencja, uczenie maszynowe, uczenie głębokie, „tradycyjne” metody uczenia maszynowego, widzenie komputerowe, wizja komputerowa, widzenie maszynowe. Pojęcia te nie zostały wyjaśnione, co dla niektórych odbiorców może powodować niejasności jeśli definiują te terminy w nieco inny sposób niż Doktorantka. Ponadto niektóre terminy stosowane są w pracy synonimicznie (bez wyjaśnienia, że są to te same pojęcia), co również może powodować niejasności. Mimo tych niedoskonałości opisu, w mojej opinii, mgr inż. Paulina Zachar pojęciami tymi posługuje się prawidłowo, co potwierdza jej wiedzę w opisywanym temacie. Druga drobna uwaga dotyczy braku wydzielenia w części teoretycznej rozdziału dotyczącego sposobów oceny skuteczności metod uczenia maszynowego czy to w kontekście detekcji obiektów czy

segmentacji obrazów. Elementy te zostały wprowadzić opisane w części eksperymentalnej, ale wprowadza to trochę chaosu, bo te same wskaźniki (równania 1 i 4 oraz równania 2 i 5) opisane zostały dwukrotnie, a do tego użyto innych nazw dla tych samych wskaźników. W rozdziale 6 w tabelach podano wartości parametru IoU (s. 83), który wyjaśniony jest dopiero pod koniec rozdziału 8 (s. 102). W moim odczuciu, w „części eksperymentalnej” rozprawy znajduje się sporo treści, które są przeglądem literatury, np. rozdział 4.1 opisuje dane hybrydowe i odnosi się w dużej mierze do badań innych naukowców. Treści te wg mnie lepiej pasują do rozdziału 1.2. Innym przykładem jest opis badania rozkładu gęstości chmury punktów stosowany przez innych badaczy (rozdział 6.3, s. 74) – to jest również przegląd literatury więc powinien się znaleźć w „części teoretycznej” rozprawy. Kolejna uwaga dotyczy ustrukturyzowania opisu części eksperymentalnej. Moim zdaniem struktura ta jest trochę chaotyczna, np. rozdział 4 zatytułowany jest charakterystyka danych, a rozdział 5 obiekty badawcze. Szczegółowy opis danych oraz pola testowego dla eksperymentu dotyczącego latarni znajduje się w rozdziale 6.1, a dla okien w rozdziale 9.1. Ponadto w rozdziale 7 przedstawiono zarówno metodykę jak i wyniki bez wydzielenia podrozdziałów podczas gdy w rozdziale 8 metodyka i wyniki są opisane w różnych podrozdziałach. Podobnie tytuł podrozdziałów 10.1 i 10.2 sugeruje opis metodyki, a znajdują się tam również wyniki i ich dyskusja. Ostatnia uwaga dotyczy niejasności odnośnie zakresu wykorzystania literatury. Odwołania do literatury w tekście nie pokrywają się listą publikacji. Przykładowo, w tekście widnieje odniesienie do pracy Wang i in. (2024), ale w spisie nie ma publikacji z 2024 r., której pierwszym autorem byłby Wang. Doktorantka wspomina również o pracy doktorskiej Wojciecha Ostrowskiego, ale tej pracy nie ma przywołanej ani nie znajduje się ona w spisie bibliografii. Sytuacja, w której publikacja znajduje się na liście, a nie ma jej przywołanej ani razu w tekście jest częstsza, np. Weir i in. (2019), Xian i in. (2019), Xiao i in. (2015). Niejednoznacznych jest również kilka odwołań w przypadku gdy nazwisko pierwszego autora i roku publikacji powtarza się kilkukrotnie, np. odwołanie do Zachar i in. (2022) może odnosić się do dwóch publikacji.

Niezależnie do przedstawionych wyżej uwag uważam, że rozprawa zawiera wszystkie niezbędne treści zarówno dotyczące zdefiniowania celów badawczych, przedstawienia przeglądu literatury oraz opisu wykonanych badań. Zatem w tym zakresie oceniam ją pozytywnie.

3. Ocena merytoryczna pracy

Wykorzystanie metod sztucznej inteligencji w analizie danych fotogrametrycznych jest niewątpliwie aktualnym zagadnieniem badawczym i coraz częściej znajduje praktyczne zastosowanie. Mgr inż. Paulina Zachar w swojej pracy doktorskiej podjęła temat detekcji wybranych obiektów przestrzeni miejskiej na lotniczych zdjęciach ukośnych z wykorzystaniem splotowych sieci neuronowych oraz wyznaczania lokalizacji wykrytych obiektów z wykorzystaniem klasycznych metod fotogrametrii analitycznej. Lotnicze zdjęcia ukośne stanowią cenne źródło danych o obiektach znajdujących się w przestrzeni miejskiej, natomiast istnieje potrzeba automatyzacji ich przetwarzania by mogły skutecznie zasilać bazy danych. W tym zakresie metody sztucznej inteligencji wydają się mieć duży potencjał.

Spośród różnych obiektów przestrzeni miejskiej jako obiekty podlegające detekcji Doktorantka wybrała latarnie miejskie oraz okna ze względu na ich powszechność w tej przestrzeni, zasadność ich wykrywania w celu zasilania baz danych oraz sposób

odwzorowania w danych fotogrametrycznych, w szczególności na zdjęciach ukośnych. Motywacja wykonania badań w kontekście wybranych obiektów oraz analiza sposobu odwzorowania różnych obiektów na zdjęciach lotniczych i chmurach punktów potwierdzają osiągnięcie dwóch pierwszych szczegółowych celów badawczych.

Oba typy obiektów Doktorantka postanowiła wykrywać niezależnie opracowując metodykę dostosowaną do ich specyfiki. Podejście takie jest niewątpliwie prawidłowe ze względu na różny sposób przedstawienia tych obiektów na danych fotogrametrycznych, możliwości zastosowania różnych danych i dostosowanie metody detekcji aby osiągnąć lepszą skuteczność. Dla obu typów obiektów jednym z problemów było przygotowanie zbiorów danych referencyjnych mogących służyć uczeniu i testowaniu modeli sztucznej inteligencji.

W przypadku latarni ulicznych Doktorantka postanowiła wykorzystać metodę zonglowania reprezentacjami danych, aby wydzielić dane referencyjne na zdjęciach ukośnych korzystając z innego typu danych. Opracowana metoda polegała na wyznaczeniu lokalizacji poziomej latarni na podstawie ortofotomapy, automatycznym wydzieleniu fragmentów gęstych chmur punktów oraz chmur punktów ALS w promieniu 2 m od zlokalizowanych punktów, filtracji chmur punktów w celu pozostawienia wyłącznie punktów należących do latarni, zrzutowaniu punktów latarni na zdjęcia lotnicze i wyznaczeniu na tych zdjęciach ramek ograniczających zrzutowane punkty dla pojedynczych latarni. Ramki wyznaczone w ten sposób pozwoliły na określeniu na zdjęciach lotniczych obszarów, w których widoczne są latarnie, dzięki czemu utworzono dane referencyjne. Przedstawione w rozprawie wyniki tworzenia danych referencyjnych odnoszą się głównie do analiz i porównania dwóch typów chmur punktów, gdyż kompletność chmury punktów dla latarni warunkuje poprawność określenia ramek ograniczających na zdjęciach, a tym samym jakość danych referencyjnych. Do wykonania tych analiz Doktorantka postanowiła wykorzystać analizę rozkładu pionowego wydzielonych dla latarni chmur punktów. W mojej ocenie zaproponowana metodyka półautomatycznego przygotowania zbiorów uczących oraz analizy ich jakości poprzez analizę rozkładu punktów w celu eliminacji danych niskiej jakości, a także przedstawienie i dyskusja uzyskanych wyników nie budzą zastrzeżeń. Wyniki tego eksperymentu potwierdzają osiągnięcie trzeciego celu szczegółowego dla pierwszego typu obiektu.

W kolejnej części rozprawy mgr inż. Paulina Zachar opisała eksperyment dotyczący detekcji latarni z wykorzystaniem modelu YOLOv8 i przygotowanych danych referencyjnych. Opisany został tu sposób podziału kafelków na zbiór uczący, walidacyjny i testowy wraz z podaniem charakterystyki ilościowej. Doktorantka przedstawiła także informacje o trenowaniu modelu, opis metod oceny jego dokładności oraz uzyskane wyniki wraz z podaniem przykładów poprawnych i błędnych detekcji, a także przedstawiła dyskusję wyników. Dokładność modelu określona została na bardzo wysokim poziomie wynoszącym około 98%. Sposób przeprowadzenia badań w tym eksperymencie i uzyskane wyniki nie budzą zastrzeżeń merytorycznych. Podział danych na 3 zbiory wydaje się prawidłowy, chociaż podane są wartości ilościowe w odniesieniu do kafelków, a nie ramek, a także nie podano w jaki sposób dany kafelek kwalifikowano do tego a nie innego zbioru. Do oceny dokładności modelu wykorzystano typowo używane w tym celu wskaźniki, a wyniki skuteczności modelu przedstawiano również na wykresach.

Kolejna część badań wykonanych przez mgr inż. Paulinę Zachar dotyczyła segmentacji zdjęć w celu wydzielenia latarni z wykorzystaniem modelu SAM. Jest to zatem druga niezależna metoda detekcji latarni wykorzystująca segmenty obrazu. Opracowana przez Doktorantkę metodyka polegała na wyborze z chmury punktów ALS punktów reprezentatywnych dla każdej latarni w oparciu o analizę gęstości punktów (określoną w rozprawie jako liczbę punktów sąsiedztwa), a następnie wyznaczeniu współrzędnych tych punktów w układzie zdjęć i zasilenie takimi „podpowiedziami” gotowego, wytrenowanego wcześniej modelu SAM (ViT-B SAM). Dla wykrytych segmentów zdjęć przedstawiających latarnie utworzono następnie ramki ograniczające tak by wyniki segmentacji można było porównać do danych referencyjnych. Należy zaznaczyć, że w tym eksperymencie nie przygotowywano osobnych danych uczących i nie trenowano modelu, a jedynie wykorzystano podpowiedzi punktowe o lokalizacji latarni. Dzięki takiemu podejściu Autorka w znacznym zakresie ograniczyła ilość pracy potrzebnej na przygotowanie danych uczących i trenowanie modelu. Ocena skuteczności tej metody polegała na określeniu liczby wykrytych oraz niewykrytych latarni oraz wyznaczeniu parametru IoU dla ramek ograniczających. Wartości tego parametru autorka oszacowała na około 50-60% w zależności od kierunku zdjęcia. Wartość ta, mimo że niska, osiągnięta została z minimalnym wkładem użytkownika dotyczącym przygotowania danych do modelu. Zaplanowanie tego eksperymentu, jego przeprowadzenie oraz metoda oceny skuteczności nie budzą zastrzeżeń. Szkoda tylko, że wyznaczone w tym i poprzednim eksperymencie wskaźniki oceny skuteczności nie są identyczne co można było wykonać (w tym eksperymencie można było przyjąć, że cały zbiór ramek wyznaczonych w oparciu o dane referencyjne jest zbiorem testowym). Trudno jest porównywać parametry *Precision* i *Recall* z IoU. Niezależnie od różnego sposobu oceny dokładności, szczegółowy cel czwarty rozprawy w kontekście porównania metod detekcji należy uznać jako osiągnięty.

Kolejna część badań dotyczyła zastosowania modelu SAM do detekcji okien na ukośnych zdjęciach lotniczych. W pierwszej kolejności mgr inż. Paulina Zachar wykorzystała zdjęcia ukośne dla miejscowości Bordeaux oraz gotowe dane referencyjne w postaci masek okien. Przetestowała dwa modele SAM: ViT-B i ViT-H różniące się zestawem parametrów, aby skuteczniejszy dla danych z Bordeaux model wykorzystać dla drugiego pola testowego (Warszawy). Skuteczniejszy okazał się model ViT-H posiadający większą liczbę parametrów, który osiągnął dokładność (*Recall*) na poziomie 78%. Doktorantka wspomina, że wartość ta osiągnięta została bez dodatkowego trenowania modelu, jednak wcześniej pisze, że działanie modelu SAM powoduje powstanie wielu segmentów, nie tylko dla okien, dlatego wykonano czyszczenie uzyskanych masek poprzez eliminację zbyt dużych i zbyt małych masek oraz łączenie nakładających się masek. Poprawa wyników uzyskanych w wyniku automatycznej segmentacji jest prawidłowym podejściem, jednak w moim odczuciu, opis tej części badań, np. użytych parametrów, wyjaśnienie stosowanej metody poprawy są niewystarczające (podanie nazwy narzędzia z oprogramowania nie jest wystarczające). Jakość przeprowadzonej poprawy ma wpływ na uzyskane wartości wskaźników detekcji i segmentacji – niestety w pracy nie podano wartości tych wskaźników przed oraz po zastosowaniu czyszczenia masek. Eksperymenty dla drugiego pola testowego dotyczyły dwóch elewacji wybranych dwóch budynków w Warszawie, dla których dostępny był trójwymiarowy model klasy BIM, w którym wydzielone były okna. Metodyka przygotowania danych referencyjnych polegała na georeferencji modelu, wydzieleniu z niego fragmentów

przedstawiających okna oraz zrzutowaniu ich na zdjęcia ukośne. Pomysł przygotowania danych referencyjnych jest zatem podobny do tego zastosowanego w eksperymencie z detekcją latarni, ale wykorzystano inne dane (model 3D zamiast chmury punktów), bardziej odpowiednie dla tego typu obiektów. W wyniku segmentacji z wykorzystaniem modelu SAM ViT-H powstało wiele segmentów (Doktorantka zauważa, że pojawił się problem powstawania segmentów dla pojedynczych szyb a nie całości okna, czyli *over-segmentation*), które zostały przetworzone z wykorzystaniem filtrów morfologicznych, z których wykorzystano zamknięcie, a Autorka podaje, że najlepsze efekty uzyskano dla wielkości jądra (ang. *kernel*) wynoszącego 15 x 15. Do oceny skuteczności segmentacji zastosowano te same wskaźniki co dla pola Bordeaux, jednak obliczono je dla każdego zdjęcia osobno oraz obliczono średnią arytmetyczną (wynoszącą w zależności od budynku około 92% i 63% dla *Recall*). Doktorantka jednak nie wyjaśnia dlaczego obliczano średnią arytmetyczną, a nie np. ważoną (liczbą występujących okien) lub nie obliczono tej dokładności uwzględniając segmenty ze wszystkich zdjęć. Mimo drobnych uwag, całość tego eksperymentu poczynszy od jego zaplanowania, poprzez dobranie metod i danych, aż do oceny skuteczności i dyskusji wyników należy ocenić pozytywnie.

Ostatnia część badań przeprowadzonych przez mgr inż. Paulinę Zachar związana była z realizacją dwóch ostatnich szczegółowych celów badawczych dotyczących lokalizowania wykrytych metodami sztucznej inteligencji obiektów celem zasilania baz obiektów miejskich. Metodyka eksperymentów wykonanych w tej części nie wykorzystywała modeli sztucznej inteligencji, ale opierała się na wykorzystaniu metod analizy obrazu stosowanych w widzeniu maszynowym oraz zależności geometrycznych stosowanych w fotogrametrii. Dla latarni miejskich metoda lokalizacji polegała na analizie fragmentów zdjęć ukośnych zawartych w ramach ograniczających będących wynikami poprzednich eksperymentów, a następnie przetwarzaniu obrazu i wyznaczeniu linii, z których najdłuższa reprezentująca słup latarni przecinana była z ramką ograniczającą dzięki czemu wyznaczono na zdjęciach punkty będące potencjalnymi punktami dołu słupa latarni. Współrzędne terenowe tych punktów wyznaczono indywidualnie dla każdego zdjęcia korzystając z warunku kolinearności i wykorzystując NMT w celu wyznaczenia wysokości. Dzięki wielokrotnemu odwzorowaniu każdej latarni na różnych zdjęciach zbiór punktów w układzie terenowym został przeanalizowany w celu usunięcia obserwacji odstających. Do tego celu wykorzystany został algorytm klasteryzacji DBSCAN. Położenie latarni określono na podstawie średniej ze współrzędnych punktów przystających. Doktorantka wykorzystwała również inne podejście wyznaczenia współrzędnych lokalizacji latarni, które realizuje przestrzenne wcięcie w przód. Podejście to wymaga jednak wcześniejszej eliminacji punktów odstających, co zostało osiągnięte w wyniku zastosowania pierwszego podejścia wyznaczania współrzędnych terenowych latarni (pominięto jedynie ostatni krok polegający na obliczeniu średnich współrzędnych). Wyznaczone współrzędne porównane zostały z tymi określonymi na ortofotomapie. Dla latarni z poprawnie wyznaczoną lokalizacją średnia odchyłka kształtowała się na poziomie około 14 i 12 cm odpowiednio dla pierwszego i drugiego podejścia. W mojej opinii wyniki te należy uznać za satysfakcjonujące biorąc pod uwagę potencjalne niejednoznaczności w identyfikacji położenia latarni na ortofotomapie oraz średnicy słupów latarni. Koncepcja na wyznaczenie lokalizacji okien w układzie terenowym jest bardzo podobna do tej zastosowanej w przypadku latarni. Lokalizacja okien na zdjęciach ukośnych wyznaczona została dla pola testowego z Warszawy. Na podstawie zidentyfikowanych za pomocą modelu

SAM segmentów fragmenty zdjęć poddano przetworzeniu metodami analizy obrazu w celu wydzielenia krawędzi, a następnie punktów narożników okien. Wyznaczone na wielu zdjęciach punkty zostały zrzutowane na model siatkowy w celu uzyskania współrzędnych 3D. Kolejne etapy przetwarzania to odrzucenie punktów odstających, klasteryzacja punktów przynależnych do poszczególnych okien oraz wykreślenie wielokątów oznaczających okna. Szczegóły tych etapów nie zostały jednak w pracy opisane, w szczególności w kontekście zastosowanych metod. Ocena dokładności lokalizacji okien polegała na porównaniu współrzędnych ich narożników wyznaczonych ze zdjęć ukośnych oraz współrzędnych z modelu BIM przyjętych jako referencyjne. Uzyskano średnie odchyłki lokalizacji wynoszące 28 cm i 16 cm w zależności od analizowanego budynku. Badania wykonane w ramach tego eksperymentu, mimo że z pewnymi brakami w opisie merytorycznym pokazują, że szczegółowe cele badawcze 5 i 6 zostały osiągnięte.

Podsumowując cel badawczy, sposób zaplanowania badań, przyjęte metody, wykonane eksperymenty, prezentację wyników, ich analizę oraz wnioski, moja ocena merytorycznej części rozprawy jest zdecydowanie pozytywna mimo drobnych braków opisu szczegółów. W mojej opinii opisana w pracy metodyka ma duży potencjał do wykorzystania jej w praktyce, mimo że dla niektórych eksperymentów wskaźniki skuteczności mogą wydawać się niskie. W szczególności może być zaadaptowana do detekcji innych obiektów przestrzeni miejskiej, gdzie ze względu na liczebność tych obiektów ilość manualnej pracy musi być ograniczona. W tym zakresie szczególnie cenne w mojej opinii są pomysły wykorzystania innego rodzaju danych przestrzennych w celu łatwiejszego przygotowania zbiorów referencyjnych służących do trenowania modeli sztucznej inteligencji.

4. Pytania i uwagi szczegółowe

Podczas publicznej obrony proszę Doktorantkę o udzielenie odpowiedzi tylko na pytania, bez ustosunkowywania się do innych uwag i komentarzy; chyba że Doktorantka się z nimi nie zgadza – wówczas proszę uzasadnić swoje zdanie.

Pytania

- 1) W pierwszym pytaniu badawczym (s. 13) pisze Pani o tworzeniu zbiorów uczących o dużej dokładności, jednak w samej pracy nie odnalazłem analiz dotyczących dokładności zbiorów uczących. Proszę powiedzieć co Pani rozumie pod pojęciem „dużej dokładności” zbiorów uczących oraz czy analiza dokładności przygotowywanych zbiorów była wykonywana?
- 2) Detekcja latarni obejmowała 2 typy (pojedynczą oraz podwójną), które w ogólności są podobnej konstrukcji – wysoki słup z ramieniem, na końcu którego znajduje się lampa. Z przedstawionych rezultatów wynika, że inne typy latarni (np. bez ramienia, gdzie lampa znajduje się w osi słupa) nie były przedmiotem detekcji. Proszę powiedzieć, czy Pani zdaniem skuteczniejszą (dającą większość dokładność) detekcję różnych typów latarni dały by modele trenowane osobno dla każdego typu latarni, czy jeden wspólny model trenowany na różnych typach latarni jednocześnie? Pomijamy tu kwestie kosztów czy czasu trenowania takich modeli.
- 3) Proszę powiedzieć w jaki sposób rozwiązywano „problem brzegowy”, szczególnie podczas tworzenia kafelków, gdy krawędź kafelka przecinała referencyjną ramkę otaczającą latarnię?

- 4) W opisie danych referencyjnych dla eksperymentu dotyczącego detekcji latarni z wykorzystaniem modelu YOLO mowa jest o liczbie kafelków w poszczególnych zbiorach, a nie liczbie latarni. Czy przeprowadzano analizę pod kątem liczby obiektów (ramek) w każdym zbiorze, a jeśli tak, to czy dodatkowo wykonywana była selekcja tak by obrazy dla jednej latarni znajdowały się wyłącznie w jednym zbiorze? Jeśli nie, to czy może mieć to znaczenie w kontekście oceny dokładności modelu?
- 5) W kilku miejscach pracy pojawia się informacja o rozdzielczości terenowej zdjęć lotniczych – podawane są wartości GSD, np. na s. 111. Natomiast dla zdjęć lotniczych, nawet pionowych obrazujących teren poziomy, wielkość GSD nie jest stała, a dla zdjęć ukośnych zmienność ta jest jeszcze większa. Pisze Pani nawet o tej niejednorodności skali na s. 116. Proszę powiedzieć czym faktycznie są podane w pracy wartości GSD oraz jaka jest zmienność wielkości GSD na wykorzystanych zdjęciach ukośnych (dla wybranego obiektu testowego).
- 6) Przed wyznaczeniem lokalizacji latarni z wykorzystaniem fotogrametrycznego wcięcia w przód eliminowano obserwacje, które w pierwszej metodzie zostały zidentyfikowane jako odstające. Czy można usprawnić metodę opartą o wcięcie w przód tak by wcześniejsza eliminacja obserwacji odstających nie była potrzebna? Jeśli tak, to w jaki sposób?

Inne uwagi i komentarze

- 1) Tytuł pracy w mojej opinii jest zbyt szeroki w kontekście wykonanych badań. Zapowiada on raczej szeroką analizę stosowanych metod i różnych danych, podczas gdy rozprawa dotyczy szczególnego zastosowania – detekcji wybranych obiektów przestrzeni miejskiej na zdjęciach lotniczych, w szczególności ukośnych (inne dane nie są przetwarzane metodami sztucznej inteligencji), a spośród szerokiej gamy metod sztucznej inteligencji wykorzystano dwa modele CNN uczenia głębokiego. Tytuł powinien być bardziej szczegółowy i prezentować lepiej zakres wykonanych badań.
- 2) Niektóre stwierdzenia pojawiające się w tekście rozprawy nie są wiedzą powszechną i nie wynikają z wykonanych badań więc powinny zostać poparte literaturą, np. „Według Organizacji Narodów Zjednoczonych, do 2050 roku ok. 68% populacji ma zamieszkiwać miasta.” (s. 11, s. 58). W kontekście modelu SAM napisano „Model ten jest używany jest do segmentacji obrazów i można uznać, że obecnie jest najnowocześniejszym (ang. *state-of-the-art*) modelem segmentacji obrazu.” (s. 23), jednak nie przywołano żadnej analizy, która by to potwierdzała; ponadto w tym zdaniu zbędne jest jedno słowo „jest”. „Ponadto wyniki wielu badań naukowych z tego obszaru pokazują, że rozwiązania wykorzystujące CNNs przewyższają inne klasyczne klasyfikatory pod względem dokładności.” (s. 34) – mimo że jest mowa o wielu badaniach, to nie odwołano się do żadnego. „Powszechnie domyślną wartością progu IoU jest 0.50.” (s. 123) – nie podano żadnego odniesienia potwierdzającego tę powszechność.
- 3) Na s. 12 napisano „Ostatecznie analiza wykonalności zadania detekcji na zobrazowaniach ukośnych została przeprowadzona dla latarni ulicznych i okien. Wybrane obiekty są dobrą reprezentacją obiektów małej architektury miejskiej,(...)” – okna nie są obiektami małej architektury.
- 4) Źródło definicji „obiektów małej architektury” w postaci strony internetowej podmiotu komercyjnego oferującego lampy i meble ogrodowe (s. 47) nie jest źródłem

Y. S. S.

- wiarygodnym, tym bardziej, że podany link nie działa. Definicja ta jest podana w innych, lepszych źródłach, np. Ustawie Prawo budowlane.
- 5) Na początku podrozdziału 1.1.1 Doktorantka pisze o „bardziej tradycyjnych metodach uczenia maszynowego”, jednak nigdzie w rozprawie nie ma wyjaśnień odnośnie tego jakie metody uczenia maszynowego uważane są za nowoczesne, a jakie za tradycyjne.
 - 6) Na s. 12 napisano „Klasyfikacja obrazu polega na kategoryzacji całego obrazu do jednej klasy.”. W klasyfikacji (pikselowej) obrazu różne piksele (części obrazu) klasyfikowane są do różnych klas, a nie jednej.
 - 7) Na s. 35 napisano „Zastosowanie konwolucyjnych sieci neuronowych do klasyfikacji, segmentacji i detekcji zdjęć lotniczych pionowych jest tematem dość szeroko przebadanym, (...)”, co sugeruje, że CNN wykorzystywane są do detekcji zdjęć. O ile w przypadku klasyfikacji czy segmentacji możemy mówić w kontekście danych (np. segmentacja zdjęcia, chmury punktów), o tyle w kontekście detekcji mówimy o obiektach (np. wykrywamy okna, latarnie).
 - 8) Na s. 35 napisano o ortofotomapie i „true-orto” powstałych z gęstego dopasowania zdjęć, na s. 36, że teksturowane modele siatkowe tworzone są na podstawie gęstego dopasowania zdjęć ukośnych, a na s. 40, że gęste dopasowanie obrazu stosowane jest również do tworzenia modeli 3D miast. Takie stwierdzenia są zbyt dużymi uproszczeniami ponieważ gęste dopasowanie obrazu nie jest metodą tworzenia ortofotomapy czy modeli, ale metodą wykorzystywaną w tworzeniu gęstych chmur punktów, które mogą, ale nie zawsze muszą być wykorzystywane w procesie tworzenia wspomnianych wyżej produktów.
 - 9) W tabeli 2 dobrze by było umieścić odniesienia (linki) do poszczególnych modeli.
 - 10) Pisząc o częstotliwości skanerów (s. 57) należy podać o jakiej częstotliwości jest mowa (generowania impulsu, użyteczna, skanowania).
 - 11) Na s. 57 scharakteryzowano sensor Leica CityMapper-2, podsumowując stwierdzeniem, że „Można powiedzieć więc, że dane pozyskane tego typu sensorem są najbardziej dokładnymi danymi, jakie współcześnie można pozyskać z poziomu lotniczego.”. Z przedstawionego opisu i podanych parametrów nie wynika jednak jaką dokładność danych zapewnia ten sensor. Nie przywołano również żadnej literatury, która mogłaby by poprzeć wspomniane stwierdzenie.
 - 12) W tabeli 4 przedstawiono wygląd różnych elementów przestrzeni miejskiej na różnych danych (chmurach punktów, zdjęciach, ortofotomapie), ale nie podano parametrów tych danych (np. gęstość chmury punktów, rozdzielczość terenowa ortofotomapy). Bez podania tych parametrów wartość tych wizualizacji znacznie spada ponieważ na tym samym rodzaju danych, ale o innych parametrach obiekty te będą wyglądać inaczej.
 - 13) Wg mnie zbiory wykorzystanych danych powinny być nieco lepiej scharakteryzowane, np. nie wiadomo ile zdjęć wykorzystano dla pola testowego z Gdańska.
 - 14) Analizowana chmura punktów DIM (dla eksperymentu dotyczącego latarni), ani sposób jej tworzenia nie zostały w żaden sposób opisane. Różne algorytmy gęstego dopasowania obrazu, a nawet różne wartości ich parametrów mogą skutkować chmurami punktów o dosyć różnej charakterystyce. Cienkie elementy często są traktowane przez te algorytmy jako szumy i usuwane podczas filtracji map głębi.

- 15) Analizując histogram przedstawiony na Rysunku 13 napisano, że większość obiektów mieści się w zakresie 3,7-5,8 cm, natomiast na histogramie wysokość słupka dla tego przedziału pokazuje 86 obiektów, co nie jest większością z 389 obiektów.
- 16) Niektóre z podanych wyników, np. IoU (Tabela 13, Tabela 14) nie zostały omówione w tekście.
- 17) Wskaźniki służące ocenie skuteczności detekcji są niejednorodne w poszczególnych eksperymentach. Dla eksperymentu detekcji latarni z wykorzystaniem modelu SAM podano 3 parametry (liczbę wykrytych i niewykrytych latarni oraz IoU), podczas gdy dla analogicznego eksperymentu dla okien, dla danych z Bordeaux wskaźników tych jest 10.
- 18) Nie wyjaśniono w jaki sposób obliczane są wskaźniki: błąd lokalizacji (położenia) wyrażony w pikselach i spójność pikseli (s. 114).
- 19) Nie wyjaśniono w jaki sposób obliczano liczbę punktów sąsiednich (s. 93). W zależności od sposobu określania sąsiedztwa, użytych wartości parametrów wyniki mogą być bardzo różne.
- 20) We wnioskach napisano, że nie wykazano dokładności, które byłyby zadowalające w warunkach produkcyjnych, jednak nie określono jakie dokładności są wymagane w warunkach produkcyjnych.

Uwagi edytorskie i językowe

- Na s. 14 napisano „Kolejny aspekt badawczy będzie dotyczył problemu braku danych uczących i automatyzacji tego procesu.” – zdanie nie jest zrozumiałe. Wynika z niego że procesem jest problem brak danych uczących, który będzie automatyzowany.
- Sposób stosowania anglojęzycznych terminów w tekście jest niejednorodny. Często pisane są one kursywą, np. „(...) w oparciu o metody *data augmentation* (...)” (s. 28), czasami dodatkowo umieszczane są w cudzysłowie, np. „Wydajność, prostota i elastyczność modeli „*anchor-free*” (...)” (s. 26), czasami w cudzysłowie bez wyróżniania kursywą, np. „(...) stosowany jest szum „salt and pepper”, (...)” (s. 29), a czasami styl jest identyczny jak dla języka polskiego, np. „exposure – dostosowanie ekspozycji (...)” (s. 29). Zdarzają się też przypadki, gdzie słowa z języka angielskiego odmieniało stosując fleksję z języka polskiego, np. „(...) tzw. *anchor boxy*, (...)” (s. 25), „(...) scenariusze zastosowania transfer learningu obejmują:” (s. 32), „(...) liczba benchmarków (...)” (s. 35), „(...) wynikowe bounding-boxy okien (...)” (s. 125), „(...) Pythona (...)” (s. 140).
- Nazwy niektórych parametrów (szczególnie w tabelach, np. Tabele 6, 9, 21) pisane są po angielsku, mimo że istnieją ich polskojęzyczne odpowiedniki, stosowane też w tekście rozprawy w innych miejscach.
- W kilku miejscach zastosowano pewne skróty myślowe, np. „, gdzie funkcjami są liniowe połączenia (...)” (s. 19) – prawdopodobnie chodziło o funkcje liniowe. „(...) efekt działania morfologii (...)” (s. 120) – powinno być „filtru morfologicznego”; oraz sformułowanie potoczne „(...) obraz mógł zostać przepuszczony przez model.” (s. 92).
- Szczegóły niektórych rysunków nie są czytelne, nawet w wersji elektronicznej pracy (np. Rysunki 2, 14, 15, 21, 24).
- Rysunek 5 nie został przywołany w tekście.
- Rysunki z wieloma polami mieszczące się na jednej stronie (np. Rysunek 32) nie powinny być rozbijane na 2 strony.

- Rysunek 63 przywołany jest w tekście w kontekście transformacji współrzędnych punktu (s. 141), natomiast pokazuje on zupełnie coś innego.
- Na rysunkach 10, 11, 13, ani w ich podpisach nie podano jednostek.
- W kilku miejscach wartość i jednostka są pisane łącznie bez odstępów, np. „4km2” (s. 36), „2m” (s. 68), „5cm” i „2.5cm” (s. 111).
- W dwóch miejscach brakuje indeksu górnego: „4km2” (s. 36), „15 punktów/m2.” (s. 56).
- W kilku miejscach nie odmieniono słowa tabela podczas odwoływania się do podanej tabeli, np. „W Tabela 3 zestawiono (...)” (s. 50), „(...) w Tabela 5.” (s. 66).
- Separator liczb dziesiętnych jest niejednorodny, np. na s. 66 stosowano przecinek, a w tabeli na następnej stronie separatorem dziesiętnym jest kropka.
- W tabeli 5 występuje zarówno „Błąd RMS” jak i „RMSE” – stosowana nomenklatura powinna być jednolita.
- Znaleziono kilka błędów literowych, np. „ (...) smukłych obiektów a miastach (...)” (s. 42) – powinno być „w”, „(...) malej architektury, (...)” (s. 64) – powinno być „małej”, „(...) łącznia wyników (...)” (s. 139) – powinno być „łączenia”, „(...) a białe reszta zdjęcia (...)” (s. 14) – powinno być „reszty”, „(...) w terenowych układzie, (...)” (s. 140) – powinno być „terenowym”; oraz językowych, np. „Nieodłącznym elementem krajobrazu miejskiego stanowią budynki (...)” (s. 58) – powinno być „są”, „Drugim kryterium wyboru było jak najmniejsza ilość drzew (...)” (s. 109) – powinno być „była jak najmniejsza liczba”.
- Wyjaśnienia niektórych skrótów są niepotrzebnie powtarzane, np. wyjaśnienie DIM pojawia się na stronach 34 i 58, a GSD na stronach 56 i 116.
- Rysunki często są podpisywane jako tabele, np. Tabela 4 w swojej konstrukcji jest bardzo podobna do Rysunku 15, a Tabela 18 do Rysunku 16.
- W kilku miejscach (np. tytuł rozdziału 8.1, s. 65, s. 67) pojawia się określenie metodologia zamiast metodyka.

5. Wnioski końcowe

Przedstawione wyżej komentarze oraz uwagi krytyczne dotyczą w większości samej konstrukcji rozprawy i nie wpływają na jej wartość merytoryczną, którą oceniam pozytywnie. W mojej opinii rozprawa doktorska mgr inż. Pauliny Zachar stanowi oryginalne rozwiązanie problemu naukowego i pokazuje, że Kandydatka posiada wiedzę teoretyczną w dyscyplinie Inżynieria Lądowa, Geodezja i Transport, w szczególności w zakresie fotogrametrii, co umożliwia jej samodzielne prowadzenie pracy naukowej.

W związku z powyższym stwierdzam, że przedstawiona do oceny praca spełnia kryteria stawiane rozprawie doktorskiej określone w art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (t.j. Dz. U. z 2024 r. poz. 1571). Stawiam zatem wniosek o przyjęcie przedłożonej rozprawy doktorskiej i dopuszczenie mgr inż. Pauliny Zachar do jej publicznej obrony.

Grzegorz Józków

Dr hab. inż. Grzegorz Józków