

Przyjmuje pod wypłatem
formalnym 29.04.2025 r.

UŁ



Kraków, 23 kwietnia 2025 r.

Profesor dr hab. inż. Krystian Pyka
Wydział Geodezji Górniczej i Inżynierii Środowiska
AGH w Krakowie

Recenzja¹ pracy doktorskiej mgr inż. Pauliny Zachar

„Wykorzystanie sztucznej inteligencji do wykrywania obiektów na danych fotogrametrycznych”

Układ i zawartość pracy doktorskiej

Praca doktorska mgr inż. Pauliny Zachar ma formę książki, razem ze spisem literatury, spisem rycin i tabel zajmuje 167 stron, jest bogato ilustrowana. Dysertacja zawiera 11 rozdziałów poprzedzonych wstępem, zakończonych bibliografią, spisem rysunków i tabel. Dwa pierwsze rozdziały stanowią część teoretyczną, pozostałe - eksperymentalną.

Część teoretyczna, zajmująca blisko 1/3 objętości dysertacji, zawiera szerokie wprowadzenie do części badawczej. Wpierw przedstawia syntetycznie ujętą charakterystykę metod sztucznej inteligencji zorientowanych na dane obrazowe. Następnie jest pogłębiony opis dwóch modeli wraz z metodami przyspieszającymi proces uczenia, co później znalazło implementację w badaniach Doktorantki. Kolejnym elementem części teoretycznej jest odniesienie do miejsca danych fotogrametrycznych w dziedzinie głębokiego uczenia. Na tej podstawie Doktorantka podejmowała decyzję jakie dane wykorzysta w badaniach oraz jakie obiekty infrastruktury miejskiej będą przedmiotem wykrywania.

Wydźwięk części teoretycznej to wskazanie problemu głębokiego uczenia, związanego z danymi treningowymi. Konieczność trenowania modeli poprzez zaznaczenie ramek ograniczających z adnotacjami na dziesiątkach tysięcy zdjęć jest wciąż problemem spowalniającym wykorzystanie potencjału CNN. Repozytoria treningowe są bardzo rozległe ilościowo i tematycznie, ale głównie dla zdjęć naziemnych. Zdjęcia lotnicze mają do tej pory relatywnie niewiele zbiorów uczących. W tym zakresie Doktorantka wykonała dogłębną kwerendę. Jednocześnie, w reakcji na taką sytuację, zapoznała się z metodami które łagodzą problem braku danych uczących. Wiedziała bowiem, że w badaniach będzie się musiała z takim zagadnieniem zmierzyć. Moim zdaniem w części teoretycznej zabrakło przedstawienia miar jakości stosowanych w głębokim uczeniu. Są przywoływane w części eksperymentalnej, ale tylko niektóre zostały wystarczająco opisane.

Badania są przedstawione w rozprawie w sposób dobrze zorganizowany, przejrzysty. Kolejno są ujęte: plan eksperymentów, charakterystyka wykorzystanych danych, obiekty badawcze, analiza

¹ Recenzja została wykonana na podstawie uchwały 67/2025 Rady Naukowej Dyscypliny Inżynieria Lądowa, Geodezja i Transport Politechniki Warszawskiej

kompletności chmur punktów z gęstego dopasowania obrazów dla latarni, detekcja latarni z wykorzystaniem modelu YOLO, ocena wyników Segment Anything Model dla latarni, segmentacja SAM dla detekcji okien i finalne wyznaczanie lokalizacji obiektów. Doktorantka przedstawia w rozprawie badania skrupulatnie, dzieli się zarówno pozytywnymi wynikami jak i opisuje wątki badawcze w których nie uzyskała spodziewanych rezultatów.

Moim zdaniem charakterystyka wykorzystanych danych jest zbyt powierzchowna. Przykładowo nie ma danych pozwalających oszacować rozmiar plamki lasera na obiekcie. Zabrakło też inżynierskich aspektów wykrywanych latarni, takich jak materiały i rozmiary słupów nośnych (w tym zmiana średnicy wraz z wysokością), rozmiary wysięgników i opraw. Bez tych danych trudno jest podjąć próbę wyjaśnienia pewnych artefaktów, które pojawiły się w badaniach.

Byłoby lepiej, gdyby ostatni rozdział „Wnioski” nazywał się „Podsumowanie i wnioski”, gdyż jest to syntetyczne ujęcie wykonanych badań z odniesieniami do postawionych we wstępie celów i pytań badawczych. Według mnie zabrakło w dysertacji wyartykułowania problemów częściowo rozwiązanych, wymagających dalszych badań. Takie informacje zawsze ułatwiają planowanie badań potencjalnym kontynuatorom.

Praca jest napisana starannie, poprawnym językiem, zdarzające się od czasu do czasu błędy edycyjne i interpunkcyjne nie przeszkadzają w zrozumieniu treści.

Przedstawiona powyżej opinia na temat układu i zawartości pracy dowodzi, że Doktorantka posiada dobre podstawy teoretyczne z zakresu głębokiego uczenia i jego zastosowania w szeroko rozumianej fotogrametrii. Dobra baza teoretyczna pozwoliła rozsądnie zaplanować eksperymenty badawcze, prowadzić je w sposób systematyczny, a dzięki dociekliwości, wyciągać słuszne wnioski.

Ocena sposobu rozwiązania problemu naukowego

Temat badań wpłata się w bardzo aktualny trend automatyzacji rozpoznawania treści zdjęć przy pomocy sztucznej inteligencji. Skierowanie punktu ciężkości badań na rozwiązanie problemu o charakterze geoprzestrzennym ujawniło szereg przeszkód, nie znanych przy wykrywaniu obiektów zajmujących dużą część zdjęcia i mających wsparcie bogatego materiału uczącego. Sądzę, że użycie lotniczych zdjęć ukośnych do wykrywania obiektów topograficznych okazało się bardziej skomplikowane niż wyobrażało sobie Doktorantka na etapie planowania badań.

Rozpoczynając badania Doktorantka była świadoma, że musi rozwiązać problem braku danych uczących dla obiektów topograficznych odwzorowanych na zdjęciach ukośnych z lotniczej platformy załogowej wyposażonej w hybrydowy sensor. Wybrany sensor, ujmujący LiDAR i zespół kamer skierowanych pod różnymi kątami, jest dedykowany aplikacjom miejskim: do urbanistyczno- architektonicznego monitoringu rozwoju miast jak również dla potrzeb „zielono-niebieskiej” infrastruktury. Problem braku danych uczących Doktorantka postanowiła rozwiązać na drodze przeniesienia reprezentacji obiektów z chmury punktów na zdjęcia ukośne. Wcześniej przedstawiła pierwsze doświadczenia w zakresie transferu danych 3D na 2D w artykule

opublikowanym w ISPRS International Journal of Geo-Information w 2022 r. (jako pierwszy współautor). Przedstawione w tym artykule badania były oparte na zdjęciach pionowych i ukośnych z BSP. W rozprawie dane są pozyskiwane ze znacznie większych wysokości (dokładnie jakich nie wiem, bo w pracy brak informacji na ten temat), co stworzyło nowe uwarunkowania dla jakości gęstej chmury z dopasowania zdjęć (w stosunku do zdjęć z niskiego pułapu).

Dowodem na dobrą znajomość AI przez Doktorantkę, w tym rozumienie specyfiki poszczególnych metod, jest dobór algorytmu do rodzaju wykrywanych obiektów. Dostosowanie do typu obiektu topograficznego dotyczy też rozdzielczości przestrzennej zdjęć: latarnie były wyszukiwane na zdjęciach z wyższą rozdzielczością niż okna.

Cechą przedstawionych w rozprawie badań jest ich dociekliwość, polegająca na poszukiwaniu optymalnego rozwiązania w sytuacji, gdy możliwych jest kilka scenariuszy postępowania. Badanie różnych ścieżek prowadzących do jednego celu dotyczy wyboru modelu sieci neuronowej, stopnia jej wytrenowania oraz źródła danych do transferu obiektów 3D (dla latarni) na ramki ograniczające na zdjęciach. Ważnym elementem badań jest przestrzenna lokalizacja wykrywanych obiektów. Doktorantka stosowała wcięcia fotogrametryczne, metody cyfrowego przetwarzania obrazów i rzutowanie na powierzchnie 3D. Przestrzenna lokalizacja obiektów wiąże mocno badania z fotogrametrią, a przez to z geodezją.

Spory wysiłek badawczy poświęciła Doktorantka na wybór danych 3D do przyspieszenia procesu uczenia. Identyfikacja obiektów na chmurze punktów prowadziła do zdefiniowania ramek ograniczających na wielu zdjęciach ukośnych. Porównywane były chmury z DIM i z ALS. Moim zdaniem prawie równoległy rozwój obu metod pozyskiwania gęstej chmury pokazał, że chmura ze zdjęć jest dobra dla obiektów znacząco rozpiętych powierzchniowo, gorsza dla krawędzi i obiektów smukłych. Chmura ALS jest zdecydowanie lepszym medium dla reprezentacji obiektów „cienkich”, o czym świadczą operacyjnie stosowane wykrywanie i modelowanie linii energetycznych i konstrukcji ażurowych. W sytuacji kiedy Doktorantka dysponowała danymi z hybrydowego sensora, który poza zdjęciami synchronicznie dostarczał też bardzo gęstej chmury, można było od razu wybrać ALS jako źródło danych do projekcji położenia latarni na zdjęcia. Jest wielce prawdopodobne, że analiza rozmiarów latarni w kontekście rozmiaru piksela na obiekcie prowadziłyby do wniosku, że to są zbyt cienkie obiekty do odwzorowania a chmurze DIM. Tym niemniej badania potwierdziły wyższość chmury ALS do detekcji obiektów smukłych.

Do wykrycia latarni ulicznych został zastosowany model YOLOv8, wytrenowany stosunkowo małym kosztem dzięki rzutowaniu 3D na 2D. Uzyskano dobre wartości klasycznych metryk, co ważne metryki dla danych walidacyjnych i testowych były porównywalne. Dodatkowo sprawdzony został specyficzny model SAM, który dokonuje segmentacji obrazów. Identyfikacja segmentów odpowiadających latarniom następowała poprzez transfer skupień punktów z 3D na pozycję pikselową 2D. Uzyskano wiele wskazań obejmujących tylko sub-segmenty latarni. W związku z tym Doktorantka szukała rozwiązań poprawiających wieloznaczność poprzez

zabiegi nakładania obszarów. Zdaniem Doktorantki wyniki pokazały potencjał podejścia segmentacyjnego, z czym można się zgodzić, ale trzeba dodać, że podejście jest na początku ścieżki do sukcesu.

Dużo lepszym rezultatem zakończyło się wykorzystanie modelu SAM do wykrycia okien. Nie było to jednak prosta implementacja modelu lecz wsparta dodatkowymi operacjami usuwającymi lub łączącymi sub-segmenty. Ostatecznie Doktorantka uzyskała wartość Accuracy 78% dla centrum miasta Bordeaux oraz 63% i 92% odpowiednio dla dwóch budynków kampusu PW. Jeśli wziąć pod uwagę fakt, że są to wyniki uzyskane bez treningu na zbiorze uczącym, to w pełni należy zgodzić się z wnioskiem Doktorantki, że model SAM dobrze wykrywa okna, zwłaszcza, gdy przed nimi nie ma drzew, a elewacje z oknami nie są położone w cieniu innych budynków.

Geoprzestrzennym zwieńczeniem badań była lokalizacja wykrytych latarni i okien w terenowym układzie współrzędnych. Jeśli obiekty mają określone położenie na zdjęciach i znane są elementy orientacji wewnętrznej i zewnętrznej zdjęć, do obliczenia współrzędnych terenowych XYZ wystarczy wykonać wcięcie w przód. Ale w praktyce przejście od postaci oznaczającej wykrycie obiektu do prostego teoretycznie wcięcia jest bardzo złożone, i różni się w zależności od typu wykrytego obiektu. Potrzebne były dodatkowe operacje, aby określić położenie obiektów w układzie kamery/zdjęcia w którym zapisane może być równanie kolinearności. Nadto testowane było podejście oparte na rzutowaniu pozycji ze zdjęć na powierzchnie 3D, co często jest prostsze aplikacyjnie.

Przedstawiona powyżej opinia na temat sposobu rozwiązania problemu naukowego dowodzi, że Doktorantka posiada umiejętność samodzielnego prowadzenia pracy naukowej, wykazuje dostateczną dociekliwość badawczą i obiektywizm w zakresie wyciągania wniosków. Wielowątkowość rozwiązań niektórych problemów spowodowała, że wykryte artefakty czy anomalie nie doczekały się pełnego wyjaśnienia, ale były to elementy drugoplanowe na drodze do głównego celu.

Ocena stopnia oryginalności rozwiązanie problemu naukowego

Udowadniając przydatność zdjęć ukośnych do zasilania miejskich baz obiektów topograficznych Doktorantka skoncentrowała się na dwóch elementach: redukcji etapu uczenia oraz na ustaleniu lokalizacji przestrzennej wykrytych obiektów. Jest to cenne połączenie klasycznych efektów sztucznej inteligencji (gdzie na zdjęciu?) z rozwiązaniem zadania przestrzennego (gdzie w terenie?). Badania wykonane przez Doktorantkę wpisują się w kierunki rozwoju geoinformatyki. Doktorantka udowodniła, że potrafi wybrać i stosować najnowsze modele głębokiego uczenia, oraz umie wykorzystać rezultaty AI w zadaniach geoprzestrzennych.

Moim zdaniem rozwiązanie problemu lokalizacji ma charakter oryginalny. Raz, że jest to zagadnienie rzadko podejmowane, dwa, że rozwiązanie Doktorantki widzi zarówno konstrukcje stosowane od dawna w fotogrametrii oparte na kolinearności jak i ujmuje inną drogę polegającą na rzutowaniu pozycji ze zdjęć na wcześniej wymodelowane powierzchnie 3D. Cenna jest

poprawa wyników wykrywania przez zastosowanie dorobku analiz przestrzennych GIS i wykorzystanie metod przetwarzania obrazów. Doktorantka udowodniła, że rezultaty wykrywania wymagają „obróbki” do postaci przydatnej do zasilania baz danych, czy to bazy obiektów topograficznych, czy baz typu BIM.

Przedstawiona powyżej opinia na temat stopnia oryginalności rozwiązania problemu naukowego dowodzi, że Doktorantka pokazując możliwości wykorzystania AI do celów pomiarowych wniosła wkład w rozwój geoinformatyki.

Temat do dyskusji podczas publicznej obrony

W badaniach wykorzystane zostały dane hybrydowe obejmujące uprzednio zorientowane zdjęcia z pięciu kamer oraz bardzo gęstą chmurę z sensora LiDAR. To unikalny zestaw danych, który stworzył warunki do wielowariantowych eksperymentów badawczych. Zawsze jakość danych limituje jakość wyników uzyskanych na ich podstawie. Dlatego warto krytycznie podejść do problemu jakości danych źródłowych i ocenić, jaki może być wpływ jakości na końcowe wyniki. Uzyskane w badaniach dokładności lokalizacji latarni i okien zostały co prawda ocenione przez Doktorantkę jako zadowalające, ale myślę, że oczekiwała lepszych. Nie tylko finalne dokładności współrzędnych, ale też wyniki detekcji na zdjęciach ukośnych sugerują, że jakość danych wejściowych nie może być przemilczana. Znajduję w relacji z badań przesłanki wskazujące, że orientacja zewnętrzna zdjęć ukośnych mogła być była gorsza niż pionowych. Na str. 66-67 jest krótka informacja o jakości wyrównania sieci zdjęć oparta na fotopunktach i fotopunktach kontrolnych, ~~mająca dowodzić, że estymacja orientacji zdjęć jest bez zarzutu~~ (dlaczego tylko dla jednego z trzech zestawów danych?). Jednocześnie jest podany średni błąd reprojektacji wynoszący 0,585 piksela. Ta wartość sugeruje, że jednak było sporo punktów wiążących o kilku a może nawet kilkunasto pikselowym błędzie reprojektacji. Szkoda, że nie podano odchyłeń standardowych EOZ zdjęć ukośnych i nadirowych. Na str. 130 Doktorantka napisała: „Zaproponowane podejście z wykrywaniem linii i punktu przyziemia nie sprawdziłoby się ze względu na różnice w odwzorowaniu tego typu obiektów między zdjęciami ukośnymi a pionowymi”. Zdanie jest dla mnie nie do końca jasne, gdyż na zdjęciach nadirowych latarnie też kładą się radialnie i jest możliwość wyznaczenia przyziemia analogicznie jak dla ukośnych (co prawda z reguły jest mniejsza liczba promieni radialnych do przecięcia).

Pytania do Doktorantki: Z czego wynikają różnice w odwzorowaniu (w/w zdanie ze str. 130) latarni na zdjęciach ukośnych i nadirowych? W jakim stopniu błędy EOZ zdjęć ukośnych mogły wpłynąć na błędy średnie współrzędnych położenia latarni i okien?

Konkluzja

Uważam, że rozprawa doktorska mgr inż. Pauliny Zachar zawiera oryginalne rozwiązanie istotnego problemu z zakresu geodezji. Wyniki pracy dowodzą udowodnienia postawionej na wstępie tezy. Doktorantka wykazała w rozprawie, że posiada gruntowną wiedzę z fotogrametrii i potrafi dobrze ją wykorzystać w planowaniu i realizacji badań naukowych.

W związku z powyższym stwierdzam, że osiągnięcie naukowe mgr inż. Pauliny Zachar odpowiada wymaganiom określonym w art. 219 ust. 1 pkt 2 ustawy z dnia 20 lipca 2018 r. - Prawo o szkolnictwie wyższym i nauce.

W konkluzji wnioskuję do Rady Naukowej Dyscypliny Inżynieria Lądowa, Geodezja i Transport Politechniki Warszawskiej o dopuszczenie mgr inż. Pauliny Zachar do dalszych etapów przewodu doktorskiego.

