

Kraków, 19 listopada 2022 r.

dr hab. inż. Tomasz Owerko, prof. AGH
Wydział Geodezji Górniczej i Inżynierii Środowiska
AGH w Krakowie

**Recenzja rozprawy doktorskiej mgr. inż. Mateusza Żarskiego
„Wizja komputerowa i transfer learning w inspekcjach betonowych obiektów
mostowych”**

Temat rozprawy doktorskiej sformułowany jako „*Wizja komputerowa i transfer learning w inspekcjach betonowych obiektów mostowych*” obejmuje bardzo szeroką klasę zagadnień. Implikuje to oczekiwania szerokiej aktywności badawczej, która po części ma miejsce. Warto jednak mieć na uwadze, że wizja komputerowa, jak również transfer learning, są rozbudowanymi obszarami bieżącej aktywności naukowej wielu naukowców, również w innych dyscyplinach naukowych. Treść przedstawionej rozprawy zasadniczo odpowiada tytułowi. Warto jednak zauważyć, że *de facto* zawartość merytoryczna pracy obejmuje podzbiór technik i metod widzenia komputerowego oraz transfer learningu wybranych i zaadaptowanych do eksperymentów, które założył i przeprowadził autor rozprawy. Być może właściwe byłoby ujawnienie tego zawężenia w tytule pracy.

Autor definiuje problem naukowy, jakiego chce się podjąć, poprzez postawienie głównej tezy rozprawy brzmiącej:

„Wdrożenie algorytmów sztucznej inteligencji do analizy obrazów pozyskanych w trakcie wykonywania okresowego przeglądu obiektu mostowego podniesie efektywność pracy terenowego inspektora mostowego, a zastosowanie technik transfer learningu zwiększy wszechstronność i ogólnodostępność proponowanego rozwiązania w porównaniu do klasycznie uczonych algorytmów SI.”

Postawienie tezy zasadniczej jest oparte na trzech tezach pomocniczych:

- 1. Dzięki zastosowaniu techniki transfer learningu, w inżynierii budowlanej możliwe jest efektywne wykorzystanie sieci pierwotnie trenowanych na powszechnych, ogólnych zbiorach danych niezwiązanych z zagadnieniem uszkodzeń powierzchni betonowych.*
- 2. Uzyskany w wyniku pracy algorytm będzie miał zdolność do generalizowania wiedzy między różnymi przypadkami powierzchni betonowej w podobnym stopniu co algorytm wykorzystujący sieć neuronową z wagami wszystkich warstw inicjalizowanymi w sposób losowy.*
- 3. Metody wizji komputerowej oraz automatyzacja pomiaru długości celowej, umożliwiają automatyczne pozyskanie fizycznych wymiarów uszkodzenia na obrazie.*

W opinii recenzenta praktyczne wdrożenie i eksperymentalne potwierdzenie tezy pomocniczej numer 1 jest wartościowe naukowo. Po pierwsze eliminuje konieczność budowania bardzo dużych zbiorów uczących. Po drugie umożliwia w przyszłości stosowanie innych niż przyjęte pracy, potencjalnie bardziej efektywnych, klasyfikatorów oraz architektur sieci głębokich, które co do zasady nie będą trenowane na obrazach reprezentujących uszkodzenia betonowych obiektów mostowych.

Teza pomocnicza numer 2 sprowadza się do udowodnienia, że przyjęte w pracy rozwiązanie oparte na transfer learningu działa efektywnie, a co za tym idzie ma cechy rozwiązania uniwersalnego, czyli nie zawęża przestrzeń hipotez. To ważne, aczkolwiek nie jest to podejście przełomowe, wszak taka jest istota tej techniki. Osobną kwestią jest to, czy wybrana architektura rozwiązania ma takie cechy oraz czy została wytrenowana na wystarczająco uniwersalnym zestawie danych. Wartościowe byłoby zawarcie w pracy wyczerpującego uzasadnienia, dlaczego klasyfikatorem jest akurat regresja logistyczna, a nie np. sieci neuronowe, drzewa losowe lub rozwiązanie oparte o boosting. Być może przyjętym kryterium była prędkość obliczeń.

Teza pomocnicza numer 3 budzi pewne wątpliwości, ponieważ sformułowana w taki sposób nie jest rozwiązaniem przełomowym. Niemniej jednak samo rozwiązanie problemu technicznego stanowi wartość dodaną i mimo że w opinii recenzenta nie jest podejściem optymalnym zasługuje na docenienie.

Autor oprócz sformułowania tezy głównej oraz tez pomocniczych wskazuje cele, jakie zamierza osiągnąć w pracy. Jest ich siedem i mają charakter kamieni milowych. Odzwierciedlają one aktywność badawczą oraz wykonane eksperymenty mające na celu przeprowadzenie dowodu tezy głównej. Zaprezentowane formułowanie problemu naukowego nie jest przyjętym standardem, ale wspiera czytelnika w poprawnym odbiorze zawartości poszczególnych rozdziałów pracy. Trudno natomiast zrozumieć, dlaczego autor nie wskazuje w tezach i celach pracy zastosowania w swoich badaniach również algorytmów klasycznego uczenia maszynowego, mimo że to zrobił i udokumentował w pracy. Wszak aktywność autora nie ograniczyła się jedynie do uczenia głębokiego – to ważne, bo wskazuje na potencjalnie szerszy wachlarz aparatu badawczego niż sugeruje tytuł oraz cytowane tezy.

Podsumowując, uzasadnienie wyboru tematu pracy oraz podjętego problemu badawczego jest poprawne i nie budzi zastrzeżeń.

Praca składa się z siedmiu rozdziałów. Pierwszy z nich obejmuje poprawnie i wyczerpująco sformułowany wykaz ograniczeń i skrótów, uzasadnienie wyboru tematu rozprawy oraz omówienie wykorzystanych metod badawczych. Wątpliwość recenzenta budzi zawartość podrozdziału 1.4, w którym autor deklaruje użycie szeregu metod badawczych, a wymienia i przedstawia tylko dwie. Opisuje tylko dwie (pierwszą i ostatnią), co sprawia, że treść tego podrozdziału wydaje się niekompletna.

Rozdział drugi obejmuje dobrze wykonany, poprawny składniowo i strukturalnie przegląd literatury dotyczący problemu badawczego, którego rozwiązania podejmuje się doktorant. Mankamentem tego rozdziału jest brak szerszego odwołania się do tabel zawartych w *Wytycznych oceny stanu technicznego drogowych obiektów inżynierskich (WR-M-81)*, w których można dokładniej wskazać, jakiego typu uszkodzenia mostowych obiektów betonowych są przedmiotem zainteresowania rozpatrywanego problemu. To ważne, bo sam tytuł pracy jest dosyć ogólny, a realizacja algorytmiczna zawężona. Ponadto skorzystanie z zawartej w tym dokumencie klasyfikacji uszkodzeń byłoby potencjalnie bardzo wartościowym, oczekiwanym efektem pracy.

W rozdziale trzecim autor skupia się na omówieniu sztucznej inteligencji. Zdaniem recenzenta jego zawartość jest dosyć odtwórcza, ale poprawnie i wyczerpująco jak na przyjętą objętość sformułowana. Biorąc jednak pod uwagę, że domyślnym czytelnikiem rozprawy są inżynierowie reprezentujący dyscyplinę naukową Inżynieria Lądowa Geodezja i Transport rozdział ten ma uzasadnienie w przedstawionej formie.

Rozdziały czwarty, piąty i szósty stanowią, z merytorycznego punktu widzenia, fundamentalną część pracy. Rozdział czwarty obejmuje propozycję metody wspomagającej wykrywanie uszkodzeń powierzchni betonowej, ze szczególnym uwzględnieniem rys o rozwarości poniżej 0,2 mm. Autor pracy stosuje podejście znane w literaturze przedmiotu, a uzasadniając swój wybór argumentuje poprawnie. Natomiast część zalet rozwiązania deklarowanych przez autora nie została w pracy eksperymentalnie potwierdzona (w pracy nie ma przykładu zastosowania opracowanego rozwiązania do danych z rzeczywistego przeglądu obiektu mostowego). Ponadto autor deklaruje, że prezentowane rozwiązanie może być wykorzystane do detekcji innych uszkodzeń, takich jak erozja czy wykwity korozyjne. Na bazie ogólnej wiedzy, czyli budowy kolejnego zbioru uczącego oraz przeprowadzenia procesu ponownego trenowania sieci jest to możliwe, ale nie jest to wartością dodaną tej pracy, a jedynie cechą przyjętej techniki uczenia maszynowego. Analogicznie ogólna deklaracja

dotycząca możliwości uzyskania współrzędnych uszkodzeń nie została w pracy w pełni udowodniona.

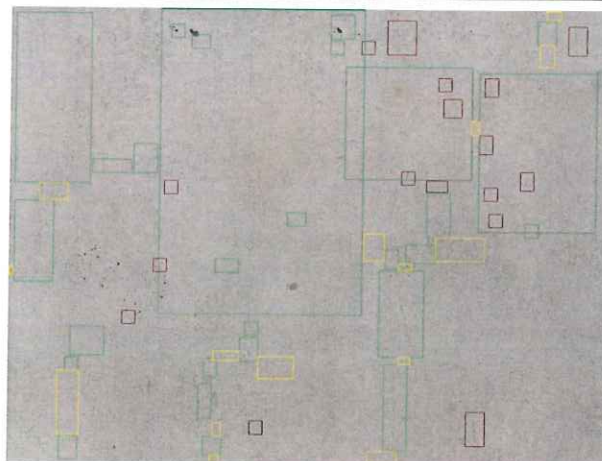
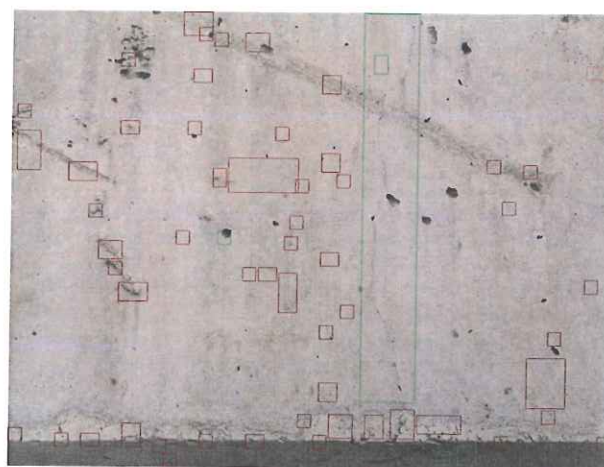
Autor przeprowadza dyskusję dotyczącą wyboru architektury sieci, natomiast trudno zrozumieć, dlaczego nie przeprowadził eksperymentów dla innych wariantów, mimo iż elastyczne podejście do implementacji rozwiązania potencjalnie to umożliwiałoby. Należy ocenić, że propozycja autora to ciekawa kombinacja prostych rozwiązań (transfer learning oparty na dobrze znanej architekturze sieci VGG oraz regresja logistyczna jako klasyfikator). Otwartym pytaniem pozostaje, dlaczego w ślad za deklarowaną modułowością rozwiązania nie poszła chęć przetestowania innych, bardziej wydanych metod jak np. MobileNet, dostosowanych do zadań mobilnych.

Istotną bardzo wartościową cechą pracy jest fakt, że autor sam przyjmuje ograniczenia środowiska obliczeniowego, co zawęży mu pole manewru co do wyboru rozwiązania. To cenne, bo nie każdy doktorat cechuje wrażliwość na praktyczne aspekty wdrożenia rozwiązania. Ma ona jednak zastosowanie praktyczne tylko do etapu używania sieci, bo taranować sieci głębokie można np. za pomocą serwisów takich jak Google Colab czy usług Akademickiego Centrum Komputerowego Cyfronet. Te możliwości nie są uwzględnione przez autora podczas rozpatrywania wariantu rozwiązania wymagającego większej mocy obliczeniowej.

W rozdziale 4.14 autor formułuje założenia do proponowanej metody oceny rozwiązania, w szczególności zwraca uwagę na konieczność stosowania adekwatnych metryk w przypadku zbiorów danych, w który występuje znacząca nierównomierność liczebności przypadków w poszczególnych klasach, a następnie to podejście konsekwentnie stosuje. To ważne, bo stanowi rzetelną ocenę rozwiązania w zakresie metryk numerycznych. Lekkim niedostatkim w zakresie ewaluacji rozwiązania pozostaje brak zastosowania w pracy rozwiązania typu Activation Maps, które pomogłoby ocenić czułość rozwiązania na poszczególne cechy zdjęć wykonywanych podczas prowadzonej inspekcji. Byłoby to szczególnie wartościowe, gdyby zakres eksperymentu, tak w fazie uczenia, jak i stosowania wytrenowanego modelu, poszerzyć o znacząco większą gamę zarysowań lub uszkodzeń. Można by wtedy udowodnić skalowalność rozwiązania na szerszy zakres problemów.

Na bazie dostarczonych algorytmów można sprawdzić działanie rozwiązania na zdjęciach powierzchni betonowych wykonywanych podczas przeglądów (w pracy autor nie dostarcza przykładów testów praktycznych). Jego działanie widać na zdjęciach poniżej.

Zielone ramki przedstawiają rezultaty prawdziwie dodatnie (rysa istnieje i algorytm ją wskazuje), ramki czerwone fałszywie dodatnie (rysa nie istnieje, a algorytm ją wskazuje np. myląc z zabrudzeniem lub fakturą powierzchni), natomiast ramki żółte fałszywie ujemnie (rysa istnieje, a algorytm jej nie odnajduje). Praktyczne działanie algorytmu należy ocenić jako działające ze zmiennym szczęściem. Na ich podstawie trudno znaleźć uzasadnienie dla deklarowanych przez autora metryk o wartościach 0,93 dla czułości i precyzji. Zastosowanie proponowanego przez autora rozwiązania na zdjęciach z przykładowych przeglądów obiektów mostowych na chwilę obecną nie daje podstaw do jego wdrożenia. Pomijając cechy rozwiązania np. przyjęty klasyfikator, prawdopodobnie zbiór uczący był zbyt mało urozmaicony – co zaskakujące, bo wg. autora jego rozwiązanie ma wspomagać w łatwym tworzeniu zbiorów uczących w sposób automatyczny. Ponadto czytelnik pracy nie ma jasnej informacji, jak powinien być przeprowadzany przegląd, aby opracowaną metodę zastosować np.: czy zdjęcia mogą być robione jako ukośne? Jaki będzie efekt pracy, jeżeli w kadrze na bliższym i dalszym planie jest filar?



Jedną z ciekawszych cech przedstawionej pracy jest przyjęta i zrealizowana metodyka automatycznej budowy zbiorów uczących. W opinii recenzenta stanowi ona być może największą, z punktu widzenia technicznego, wartość dodaną prezentowanej rozprawy. Jest dobrze zaprojektowana, rozsądnie zbudowana i daje wyniki praktyczne. Nie do końca zrozumiałe jest, dlaczego autor nie skorzystał z wytworzonego narzędzia i nie utworzył szerokiej bazy zdjęć, aby praktyczne działanie rozwiązania nieco poprawić.

Rozdział piąty przedstawia założenia i wyniki dwuetapowego eksperymentu porównawczego. Pierwszy z nich ma na celu pokazanie, że sieć o podobnej czułości uczona od wag losowych będzie wymagała znacząco dłuższego treningu niż ta, która podlega uczeniu z wykorzystaniem TL. Teza jest udowodniona, ale nie jest ona szczególnie przełomowa, wszak taki jest sens stosowania TL. Nie do końca wiadomo, w jakim celu taki eksperyment był przeprowadzony. Drugi z eksperymentów, porównanie do rozwiązań dostępnych w literaturze, jest zdecydowanie bardziej przekonujący i stanowi ciekawe, ważne uzasadnienie przyjętego rozwiązania. Trudno się odnieść do przedstawionych wyników, ponieważ w tym rozdziale zabrakło informacji jak autor przeprowadził preprocessing danych. Ten aspekt może znacząco wpłynąć na wyniki poszczególnych rozwiązań. Warto jednak zauważyć, że konkurencyjne rozwiązania obejmują prawdopodobnie szerszą przestrzeń hipotez. Przedstawione wyniki dla CrackNet i DeepCrack dla danych KrakN mogą wskazywać, że zdjęcia w zbiorze danych są mało urozmaicone.

Autor nie zdecydował się na udokumentowanie wielokrotnie deklarowanej uniwersalności przedstawionego rozwiązania poprzez przedstawienie wyników dla innych defektów. Szkoda, ponieważ bardzo znacząco podniosłoby to wartość naukową pracy. Zatem ta cecha rozwiązania, na chwilę obecną, może być traktowana jedynie jako hipoteza.

Rozdział szósty obejmuje zaprezentowanie założeń oraz budowę rozwiązania, które docelowo ma pomóc w ekstrakcji danych z obrazów. Treść rozdziału ma swoje silne i słabe strony. Podstawowym mankamentem metodycznym jest brak zastosowania do rozwiązania metod fotogrametrycznych o stałej bazie. Rozwiązanie takie jest z pewnością dokładniejsze oraz bardziej poręczne dla wykonującego przegląd inspektora, a ponadto umożliwia uzyskanie metrycznych wartości obrazu i cech bez pomiaru odległości i budowy dedykowanego urządzenia. Powyższe nie zmienia jednak faktu, że autor zaprojektował i zbudował poprawne, działające urządzenie, które spełnia swoją funkcję. Zdecydowanie zasługuje to na docenienie.

Ponadto rozwiązanie jest niskobudżetowe, a działa skutecznie, co jest ważne z praktycznego punktu widzenia.

Ciekawa cecha rozwiązania kryje się w zastosowanej metodzie ekstrakcji cech bazującej na wykorzystaniu maszyn wektorów nośnych – to odważna i obiektywnie mówiąc kontrowersyjna decyzja, gdyż zasadniczo ten algorytm może cechować się różnymi wynikami dla dużej liczby zróżnicowanych zdjęć. W opinii recenzenta ta cześć, obok opracowanej metody automatycznego tworzenia zbiorów uczących, stanowi istotną wartość dodaną prezentowanej pracy doktorskiej, ponieważ docelowo pozwala na automatyczną wektoryzację zarysowań.

Podsumowując, przedstawiona praca dowodzi, że mgr inż. Mateusz Żarski potrafi samodzielnie stawiać trafne, istotne dla otoczenia społeczno-gospodarczego hipotezy badawcze, projektować wydajne rozwiązania oraz prowadzić rzetelnie zaplanowane eksperymenty mające na celu weryfikację trafności przyjętych założeń. Na docenienie zasługuje fakt, że w pracy autor z dużą łatwością stosuje różne rozwiązania, nie ograniczając się do uczenia głębokiego. Co ważne potrafi prowadzić dogłębną i dobrze przemyślną analizę możliwości oraz sprawnie i efektywnie dobiera optymalne rozwiązanie do rozpatrywanego problemu, niejednokrotnie stawiając swoim rozwiązaniom wysokie wymagania techniczne i funkcjonalne. Opracowany komplet rozwiązań algorytmicznych wskazuje na łatwość nabywania przez doktoranta wiedzy również z innych niż Inżynieria Lądowa Geodezja i Transport dyscyplin naukowych. Ponadto opracowanie stanowi oryginalne rozwiązanie problemu naukowego dowodząc umiejętności samodzielnego prowadzenia pracy naukowej w przyszłości. Praca ma charakter twórczy oraz stanowi wartościowy zasób wiedzy, rzetelnie uporządkowanej oraz trafnie dobranej (co w rozpatrywanym problemie nie jest łatwe). Przedstawione uwagi odnoszą się do procesu badawczego oraz procesu dowodzenia tezy, ale nie umniejszają wartości poznawczej pracy. Należy je traktować jako wskazówki służące ewentualnemu uwzględnieniu w dalszym rozwoju naukowym doktoranta.

Recenzowana rozprawa doktorska spełnia wymagania stawiane pracom doktorskim zawarte w Ustawie z dnia 20 lipca 2018 roku Prawo o szkolnictwie wyższym i nauce (Dz. U. 2020, poz. 85). Składam wniosek o dopuszczenie mgr. inż. Mateusza Żarskiego do obrony publicznej.

Tomasz Owerko