

Łódź, 11. grudnia 2025 roku

prof. dr hab. inż. Piotr M. Szczypiński
Instytut Elektroniki Politechniki Łódzkiej
Al. Politechniki 8, 93-590 Łódź

Recenzja rozprawy doktorskiej

mgr. inż. Przemysław Dolaty

p.t. Zastosowanie metod uczenia głębokiego w wybranych problemach widzenia komputerowego w kontekście przemysłowej oceny jakości

ang. Deep learning methods for computer vision-based industrial inspection applications

przygotowana pod opieką dr. hab. Macieja Zięby, prof. uczelni

Recenzja przygotowana na prośbę prof. dr hab. Wojciecha Bożejko, Przewodniczącego Rady Dyscypliny Naukowej Informatyki Technicznej i Telekomunikacji Politechniki Wrocławskiej (pismo RDN-IT/201/2025 z dnia 30.10.2025 r.)

Zawartość pracy

Rozprawę doktorską napisano w języku angielskim. Obejmuje ona cykl sześciu publikacji (trzy w indeksowanych czasopismach oraz trzy doniesienia pokonferencyjne) oraz autoreferat, w którym autor wyjaśnia problem naukowy, przedstawia cel badań oraz streszcza zawartość tych publikacji. Dostarczona dokumentacja zawiera również streszczenia w językach polskim i angielskim. Mgr inż. Przemysław Dolata jest pierwszym autorem wszystkich artykułów tworzących cykl.

Problem badawczy, cel i tezy rozprawy

Autor w autoreferacie (str. 12) formułuje trzy główne pytania badawcze:

1. *Jak skonstruować i wytrenować oparty na splotowych (konwolucyjnych) sieciach neuronowych system klasyfikacji wielowidokowej przeznaczony do przemysłowej oceny jakości?*

(How to construct and train a convolutional neural network-based multi-view classification system for industrial quality inspection?)

2. *Jak zoptymalizować architekturę modelu segmentacji semantycznej typu U-Net, aby poprawić jego czas wnioskowania przy jednoczesnym zachowaniu jakości segmentacji?*

(How to optimize the architecture of a U-Net-based semantic segmentation model in order to improve its inference time while retaining the segmentation quality?)

WPLYNĘŁO

19-12-2025

1

RDN-IT/243/2025

3. Jak zbudować modele wizji komputerowej do wymiarowania małych obiektów w gęstych scenach na podstawie pojedynczych obrazów?

(How to build computer vision models for the downstream task of dimensioning of small objects in dense scenes from single images?)

W odniesieniu do pierwszego pytania autor zajmuje się zagadnieniem analizy wielowidokowej (ten sam klasyfikowany obiekt obrazowany jest z wielu stron). We wstępie autoreferatu, wskazano na ograniczone możliwości gromadzenia odpowiedniej liczby przykładów uczących do uczenia sieci głębokich od zera a z drugiej strony na problem z wykorzystaniem istniejących, wstępnie nauczonych sieci neuronowych i zastosowania metod uczenia transferowego.

W odniesieniu do drugiego pytania, autor wskazuje na opóźnienia w przetwarzaniu obrazów i ich segmentacji jakie wprowadza sieć U-Net. Wskazuje, że takie opóźnienia uniemożliwiają monitorowanie procesów napawania laserowego w czasie rzeczywistym.

W odniesieniu do trzeciego pytania autor wskazuje na problem wymiarowania obiektów, które wzajemnie się przesłaniają, i część z nich w analizowanym obrazie nie jest widoczna w całości. Rozwiązywane są dwa problemy, jeden związany z przygotowaniem wiarygodnych danych uczących dla uczenia nadzorowanego i oznaczeniem etykiet (masek obszarów zainteresowania) obejmujących przesłonięte fragmenty obiektów. Drugi problem to właściwa reprezentacja przesłaniających się obiektów w procesie ich identyfikacji – reprezentacja za pomocą ramek ograniczających (*bounding box*) nie umożliwia jednoznacznej lokalizacji poszukiwanych obszarów.

W dalszej części autoreferatu pytania 1 i 3 uzupełnione są kolejnymi pięcioma pytaniami uszczegóławiającymi. Pytania te również w sposób pośredni wskazują na problemy badawcze takie jak łączenie i wydobywanie istotnych informacji z wielu obrazów przedstawiających ten sam obiekt, oraz niejednoznaczność identyfikacji obiektów trójwymiarowych, wzajemnie się przesłaniających i obrazowanych jedynie w dwóch wymiarach.

Należy zwrócić uwagę, że autor formułuje problemy w sposób ogólny, nie wyjaśnia w sposób szczegółowy na czym one polegają w praktyce i w jakim kontekście są rozważane. Aby to zrozumieć należy zapoznać się z poszczególnymi publikacjami cyklu. Okazuje się, że trzy główne pytania badawcze sformułowane są w autoreferacie w sposób ogólny i uniwersalny, jednak przeprowadzone badania i ich wyniki odnoszą się jedynie do szczególnych przypadków danych obrazowych a uzyskane wyniki umożliwiają wyciągnięcie wniosków w odniesieniu do tych szczególnych przypadków. I tak, obrazowanie wielowidokowe sprowadza się do analizy par obrazów dwóch stron ziemiaków jęczmienia, a optymalizacja sieci U-Net prowadzona jest jedynie w kontekście specyficznych obrazów procesu napawania laserowego.

Należy uznać, że problemy badawcze podjęte w pracy obejmują zagadnienia analizy obrazów cyfrowych mieszczące się w dyscyplinie informatyka techniczna i telekomunikacja. W mojej ocenie rozwiązania przedstawionych problemów badawczych nie są oczywiste. Wymagały one od doktoranta opracowania oryginalnych algorytmów, przemyślnego połączenia znanych lub znaczącego udoskonalenia istniejących metod, co stanowi istotny wkład doktoranta do dyscypliny naukowej.

Głównym celem rozprawy doktorskiej było opracowanie wyspecjalizowanych metod uczenia głębokiego dedykowanych wybranym zastosowaniom inspekcji wizyjnej w kontekście przemysłowej oceny jakości w rolnictwie precyzyjnym i przetwórstwie żywności, gdzie konieczne było zaprojektowanie dedykowanych, często od podstaw tworzonych modeli głębokiego uczenia. Aby osiągnąć cel główny przeprowadzono badania w trzech komplementarnych obszarach. Celem badań w pierwszym z nich było opracowanie splotowych klasyfikatorów obrazu przeznaczonych do klasyfikacji odmianowej ziaren jęczmienia, w której obiekt jest reprezentowany za pomocą dwóch obrazów przedstawiających brzuszną i grzbietową stronę

ziarniaka. Drugim celem była optymalizacja architektury sieci U-Net w monitorowaniu procesów laserowych wymagających przetwarzania w czasie rzeczywistym. Trzecim celem było opracowanie modeli przeznaczonych do estymacji wymiarów małych i przesłaniających się obiektów, przy wykorzystaniu dwuwymiarowych obrazów RGB. Obejmował on opracowanie metody przygotowania zbioru danych uczących, szacowania wymiarów bulw ziemniaków obrazowanych za pomocą kamery wizyjnej na pasie transmisyjnym, zaprojektowanie procedur trenowania modeli amodalnej segmentacji instancyjnej, a także poprawę skuteczności detekcji larw mącznika poprzez wprowadzenie nowej reprezentacji geometrycznej opartej na linii środkowej. W efekcie uzyskano zestaw użytecznych narzędzi, odpowiadających na potrzeby nowoczesnej inspekcji przemysłowej.

Cele rozprawy odpowiadają na realne wyzwania współczesnej inspekcji wizyjnej, szczególnie tam, gdzie istniejące metody głębokiego uczenia okazują się niewystarczające. Każdy z trzech obszarów badawczych wnosi wyraźną wartość zarówno naukową, jak i aplikacyjną. Cele te chociaż realizowane dla wybranych i specyficznych obiektów zainteresowania (obrazowanie procesów laserowych, ziarniaków jęczmienia, bulw ziemniaków i larw) wykazują się potencjałem wykorzystania również w innych dziedzinach przemysłu.

Autor nie formułuje explicite tez ani hipotez badawczych, jednak analiza wniosków zawartych w sześciu publikacjach naukowych pozwala jednoznacznie stwierdzić, że przeprowadzone badania doprowadziły do szeregu ważnych i dobrze udokumentowanych obserwacji. W szczególności wykazano, że:

1. architektura dwustrumieniowej splotowej sieci neuronowej z późnym łączeniem danych na poziomie predykcji zapewnia największą skuteczność klasyfikacji w zadaniach obrazowania dwustronnego ziarniaków;
2. wstępne rozpoznawanie orientacji ziarniaków (stron brzusznej i grzbietowej) istotnie zwiększa dokładność klasyfikacji odmianowej;
3. możliwe jest uproszczenie architektury U-Net poprzez usunięcie wybranych warstw w sposób, który przyspiesza działanie modelu bez wyraźnej utraty jakości segmentacji;
4. nieliniowy model regresji trenowany na syntetycznych danych geometrycznych stanowi efektywne narzędzie do estymacji trzech wymiarów bulw ziemniaka na podstawie pojedynczych obrazów dwuwymiarowych;
5. modele amodalnej segmentacji instancyjnej mogą być skutecznie trenowane na danych syntetycznych generowanych z wykorzystaniem modalnych instancji obiektów;
6. zastąpienie klasycznych ramek ograniczających reprezentacją w postaci linii środkowej definiowanej przez trzy punkty znacząco poprawia skuteczność identyfikacji larw mączniaka.

Uzyskane wnioski cechują się wysoką wartością naukową i aplikacyjną, gdyż nie tylko poszerzają współczesne rozumienie możliwości i ograniczeń metod uczenia głębokiego w inspekcji wizyjnej, lecz także prowadzą do opracowania skutecznych, zweryfikowanych eksperymentalnie rozwiązań o potencjale wdrożeniowym. Otrzymane rezultaty stanowią istotny wkład w rozwój metod klasyfikacji, identyfikacji obiektów i segmentacji obrazów, a równocześnie mogą służyć jako fundament dla dalszych badań i innowacyjnych zastosowań w przemyśle.

Przegląd aktualnego stanu wiedzy

Zestawienie bibliograficzne przedstawione w rozprawie obejmuje znaczącą liczbę źródeł, świadcząca o szerokim i systematycznym przeglądzie literatury. Autoreferat opiera się na 42 pozycjach bibliograficznych, natomiast publikacje wchodzące w skład cyklu badawczego zawierają od 11 do 40

cytowań. Cytowana literatura reprezentuje pełne spektrum materiałów typowych dla badań z obszaru informatyki w zagadnieniach wizji komputerowej: od prac fundamentalnych dotyczących architektur głębokiego uczenia, przez artykuły aplikacyjne z dziedziny inspekcji wizyjnej, po zaawansowane przeglądy metod oraz dokumentację techniczną narzędzi programistycznych. Tak kompleksowy dobór źródeł wskazuje na bardzo dobre rozeznanie autora w zakresie współczesnego stanu wiedzy.

Pod względem merytorycznym literatura została dobrana w sposób jednoznacznie powiązany z celami rozprawy i poszczególnymi problemami badawczymi. Autor przedstawia logiczne uzasadnienie wybranej metodyki, rozpoczynając od kontekstu automatyzacji przemysłowej i przechodząc do współczesnych wyzwań w rolnictwie i przetwórstwie żywności. Przegląd literatury odzwierciedla trzy główne obszary badań: mechanizmy fuzji informacji w klasyfikacji wielowidokowej, optymalizację i redukcję złożoności architektur segmentacyjnych oraz problemy detekcji i segmentacji instancyjnej w scenach o dużej gęstości i licznych okluzjach. Dobór źródeł stanowi więc bezpośrednie wsparcie metodyczne dla prowadzonych analiz i jednocześnie służy do trafnej identyfikacji podjętych problemów badawczych.

Istotną cechą przeglądu jest jego aktualność oraz umiejętne osadzenie prac w kontekście zmieniającego się stanu wiedzy w ciągu ostatnich kilkunastu lat. Autor nie tylko korzysta z literatury najnowszej, lecz także świadomie wskazuje historyczne ograniczenia wcześniejszych własnych prac, odnosząc je do dominujących wówczas narzędzi, takich jak Caffè czy wczesnych metod przycinania sieci neuronowych. Takie spojrzenie retrospektywne podkreśla zdolność do krytycznej analizy zmian w metodyce przetwarzania i analizy obrazów i uczenia maszynowego.

Podsumowując, autor przedstawił literaturę aktualną, adekwatną i istotną z punktu widzenia tematyki rozprawy, wykazując przy tym umiejętność krytycznej oceny cytowanych prac. Sposób doboru i analizy źródeł potwierdza jego szeroką wiedzę teoretyczną, bardzo dobre przygotowanie merytoryczne i umiejętność identyfikowania problemów naukowych w dyscyplinie informatyka techniczna i telekomunikacja.

Zastosowana metodyka badawcza

Metodyka badawcza w rozprawie została ukształtowana wokół trzech kluczowych zagadnień badawczych, zorientowanych na opracowanie metod głębokiego uczenia dla złożonych problemów inspekcji wizyjnej, w których istniejące techniki głębokiego uczenia maszynowego okazały się niewystarczające.

Metodyka została oparta na pięcioetapowym schemacie prowadzenia badań naukowych. W pierwszym etapie autor przeprowadził szczegółowy przegląd aktualnego stanu wiedzy w obszarach klasyfikacji wielowidokowej, szybkiej segmentacji semantycznej oraz segmentacji instancyjnej w gęstych scenach. Autor poddał istniejące rozwiązania krytycznej ocenie, która pozwoliła na identyfikację istotnych problemów badawczych i pozwoliła wskazać kierunki dla dalszych prac.

W kolejnym etapie opracowane zostały oryginalne metody przygotowania danych uczących oraz pozyskiwania materiału badawczego. Autor brał udział w przygotowaniu zbiorów obrazów dwóch stron ziarniaków jęczmienia, przygotował obrazy procesów laserowych do uczenia nadzorowanego, a także opracował metodę syntetycznego generowania obrazów przesłaniających się bulw ziemniaków i larw mącznika do uczenia regresyjnych i segmentacyjnych sieci neuronowych.

Następny etap badań obejmował opracowanie nowych algorytmów oraz modyfikację istniejących modeli uczenia głębokiego. Autor zaprojektował m.in. dwustrumieniowe architektury CNN o zróżnicowanych strategiach fuzji, warianty architektur *viewpoint-aware*, metodę „przycinania” sieci U-Net opartą na mechanizmie usuwania zbędnych warstw, a także nowe podejście do detekcji larw mącznika wykorzystujące ich reprezentację za pomocą linii środkowej i wdrożył je w architekturze DETR. W każdym przypadku

modyfikacje miały charakter oryginalny — dotyczyły zarówno struktury sieci, funkcji kosztu, jak i strategii treningowych, tworząc modele dostosowane do specyficznych danych wejściowych.

Czwarty etap obejmował planowanie i realizację eksperymentów badawczych, a także weryfikację i walidację opracowanych metod. Autor przeprowadził liczne eksperymenty porównawcze, studia ablacyjne, testy przenoszenia wiedzy między domenami syntetyczną i rzeczywistą, a także oceny jakości detekcji i segmentacji w warunkach obrazowania częściowo przesłoniętych obiektów. W porównaniach wykorzystał zarówno miary ilościowe, jak i oceny jakościowe.

Ostatni etap stanowiło ogólne wnioskowanie, obejmujące krytyczną ocenę opracowanych metod, wskazanie ich ograniczeń oraz potencjalnych ryzyk zastosowań, a także określenie kierunków dalszych badań. Autor wykazał świadomość zarówno technologicznych granic stosowanych architektur, jak i wymagań poszczególnych sektorów przemysłu.

Podsumowując, metodyka badawcza zastosowana w rozprawie świadczy o dojrzałości naukowej autora, jego umiejętności systematycznego planowania eksperymentów, krytycznego doboru narzędzi oraz twórczego rozwiązywania problemów. Autor wykazał się kompetencjami niezbędnymi do prowadzenia samodzielnych badań naukowych — od krytycznej analizy literatury, poprzez przygotowanie materiału badawczego, opracowanie nowych metod, ich walidację, umiejętność formułowania wniosków i wskazywania dalszych kierunków badań.

Oryginalność i wkład do dyscypliny

Ważnym osiągnięciem doktoranta są prace nad dwustrumieniową architekturą sieci spłotowych przeznaczoną do klasyfikacji odmian jęczmienia na podstawie par obrazów pojedynczych ziarniaków. Zaproponował on rozwiązanie oparte na połączeniu dwóch sieci spłotowych, z których każda niezależnie wydobywa cechy z jednego z obrazów wejściowych. Doktorant opracował i szczegółowo przebadał sześć wariantów takiej architektury, różniących się sposobem i miejscem integracji informacji z obu strumieni. Wykazał, że największą dokładność klasyfikacji uzyskuje się poprzez późną fuzję — na poziomie warstw w pełni połączonych odpowiedzialnych za końcową predykcję klasy. Wyniki te zostały jednoznacznie udokumentowane przy użyciu miar jakości klasyfikacji oraz krzywych ROC. W tym etapie badań autor zaproponował również współdzielenie wag i współczynników filtrów pomiędzy obiema gałęziami bliźniaczej sieci, co było uzasadnione faktem, że każda z nich analizowała zarówno obrazy strony brzusznej, jak i grzbietowej.

W kolejnej publikacji doktorant znacząco rozwinął wcześniejszą koncepcję poprzez wprowadzenie dodatkowej sieci neuronowej o architekturze AlexNet odpowiedzialnej za automatyczne rozpoznawanie, który z obrazów w parze przedstawia stronę brzuszną, a który grzbietową. Po klasyfikacji orientacji obrazy są kierowane do odpowiednich wyspecjalizowanych gałęzi sieci, których wagi były trenowane niezależnie na dedykowanych podzbiorach danych. W efekcie uzyskano architekturę *viewpoint-specific*, istotnie poprawiającą dokładność klasyfikacji dzięki uchwyceniu subtelnych różnic wizualnych charakterystycznych dla poszczególnych stron ziarniaka. **Takie połączenie sieci neuronowych, jednej do wstępnej analizy orientacji ziarniaka oraz drugiej dwustrumieniowej do rozpoznawania jego odmiany należy uznać za oryginalne i wartościowe rozwiązanie.**

W trzeciej publikacji z cyklu autor zaproponował modyfikację architektury U-Net, która polega na uproszczeniu jej struktury w celu skrócenia czasu wnioskowania, co jest kluczowe w zastosowaniach wymagających przetwarzania w czasie rzeczywistym. Podejście to opiera się na identyfikacji warstw, które w danym zadaniu — takim jak monitorowanie procesów laserowych — nie wnoszą istotnego wkładu do jakości segmentacji, mimo że ich działanie okupione jest znacznym kosztem obliczeniowym. Warstwy te są zastępowane prostymi operacjami resamplingu (pomniejszenia lub powiększenia obrazu), co znacząco

skraca czas przetwarzania. Istotną rolę odgrywa zastosowanie regularyzacji drop-path, która w trakcie treningu losowo wyłącza różne kombinacje warstw, ucząc sieć odporności na brak fragmentów swojej architektury. W badaniach wykazano, że eliminacja pierwszych trzech warstw enkodera oraz trzech ostatnich dekodera pozwala uzyskać pięciokrotne przyspieszenie działania modelu.

Badania nad modyfikacją sieci U-Net przeprowadzono na bardzo specyficznych danych, obejmujących obrazy monitorowania procesu napawania laserowego. Jedyny zaprezentowany przykład przedstawia jasną, nieostrą plamę na ciemnym tle – obraz pozbawiony drobnych szczegółów i komponentów o wysokich częstotliwościach. W takim kontekście redukcja warstw odpowiedzialnych za analizę szczegółów i wysokoczęstotliwościowych składowych obrazu nie jest zaskakująca, lecz wynika bezpośrednio z charakteru danych. Można oczekiwać, że dla bardziej złożonych wizualnie obrazów wykorzystywanych w innych publikacjach doktoranta (np. przedstawiających ziarna, bulwy ziemniaków czy larwy owadów) taka redukcja nie byłaby możliwa bez wyraźnej utraty jakości segmentacji. **Z tego względu należy uznać, że zaprezentowane wyniki dotyczące optymalizacji architektury U-Net mają ograniczoną wartość uogólniającą i wnoszą jedynie nieznaczny wkład do dyscypliny.**

W zakresie amodalnej segmentacji instancyjnej (AIS) doktorant opracował oryginalne metody generowania syntetycznych danych treningowych, w których eliminuje się potrzebę manualnej anotacji amodalnej. Zaproponowane podejście wykorzystuje kontrolowany proces pozyskiwania obrazów w warunkach pełnej widoczności obiektów, a następnie generuje złożone, gęsto okludowane sceny poprzez wykorzystanie narzędzi przetwarzania obrazów lub renderowania trójwymiarowej grafiki komputerowej. W trakcie syntezy obrazu automatycznie powstają również odpowiadające im anotacje amodalne. Taka metodyka upraszcza proces przygotowania danych, umożliwia skalowalne generowanie dużych zbiorów treningowych i otwiera drogę do stosowania AIS w domenach, w których pozyskanie klasycznych anotacji jest niepraktyczne lub utrudnione.

Kolejnym osiągnięciem jest propozycja nowej reprezentacji geometrycznej opartej na linii środkowej, zaprojektowanej specjalnie z myślą o poprawie detekcji wydłużonych obszarów larw owadów w warunkach ich dużej gęstości i okluzji. Standardowe ramki ograniczające, które są fundamentem współczesnych metod detekcji, okazują się w takich scenach nieadekwatne, prowadząc do błędów lokalizacji i nakładania się predykcji. Zaproponowana reprezentacja w postaci linii środkowej definiowanej przez trzy punkty pozwala jednoznacznie odwzorować geometrię nawet wtedy, gdy widoczny jest jedynie fragment larwy. Doktorant w oryginalny i pomysłowy sposób zaadaptował architekturę DETR, modyfikując jej warstwę wyjściową tak, aby model operował reprezentacją w postaci linii środkowej, co pozwoliło skutecznie uchwycić geometrię wydłużonych i częściowo przysłoniętych instancji.

Dokonania doktoranta stanowią znaczący wkład w rozwój metod AIS, szczególnie w zadaniach wymagających precyzyjnego szacowania wymiarów obiektów w gęstych i złożonych scenach. Zaproponowane metody skutecznie rozwiązują kluczowe problemy praktyczne i tworzą solidne podstawy do dalszego rozwoju technik segmentacji instancyjnej w warunkach wzajemnego przesłaniania się obiektów.

Należy zauważyć, że w swoich badaniach doktorant nie ograniczał się do kosmetycznych modyfikacji istniejących rozwiązań — konsekwentnie formułował własne koncepcje i wdrażał je w postaci oryginalnych algorytmów oraz istotnych usprawnień architektur głębokich sieci neuronowych. Jego prace cechuje samodzielność oraz gotowość do podejmowania ryzyka naukowego związanego z testowaniem nowych, wcześniej niezwyfikowanych pomysłów. Opracowane przez autora metody budowania zbiorów uczących, architektury dwustrumieniowe, metoda automatycznego

rozpoznawania orientacji ziarniaków i integracji tej informacji w systemie klasyfikacyjnym, oraz metody amodalnej segmentacji instancji stanowią niewątpliwie oryginalny i znaczący wkład w rozwój dyscypliny informatyki technicznej i telekomunikacji.

Formalna i prezentacyjna strona rozprawy

Rozprawa została przygotowana w formie cyklu sześciu publikacji, których doktorant jest pierwszym autorem, oraz autoreferatu stanowiącego ich spójne i syntetyczne omówienie. Autoreferat obejmuje abstrakt w języku angielskim oraz streszczenie po polsku, a także podsumowanie, podziękowania, listę publikacji tworzących cykl oraz zestaw 42 źródeł wykorzystanych w przeglądzie literatury. Taka struktura jest zgodna z wymaganiami stawianymi rozprawom opartym na publikacjach i w przejrzysty sposób porządkuje zarówno kontekst badań, jak i ich wkład do dyscypliny.

Zarówno autoreferat, jak i wszystkie publikacje wchodzące w skład cyklu zostały przygotowane w języku angielskim. Każdą z sześciu publikacji omówiono w autoreferacie według jednolitego, klarownego schematu: przedstawiono motywację i cel badań, kluczowe elementy metodyki, najważniejsze wyniki, wnioski końcowe oraz precyzyjnie określony wkład doktoranta i współautorów. Dzięki temu dokumentacja pozwala czytelnikowi zrozumieć logikę prowadzonych badań oraz związki pomiędzy poszczególnymi etapami pracy.

Pewnym mankamentem autoreferatu jest jego nadmiernie ogólnikowy charakter. W wielu miejscach autor unika jednoznacznych sformułowań i przedstawia zagadnienia w sposób zbyt uogólniony, co utrudnia uchwycenie właściwej istoty omawianych problemów badawczych. Niektóre stwierdzenia są na tyle nieprecyzyjne, że ich pełne znaczenie staje się zrozumiałe dopiero po szczegółowej analizie treści poszczególnych publikacji oraz poznaniu faktycznego materiału badawczego. Taka niejednoznaczność osłabia autonomiczną wartość autoreferatu i ogranicza jego funkcję jako syntetycznego, samodzielnego podsumowania całego cyklu badawczego.

Pod względem prezentacyjnym zarówno autoreferat, jak i publikacje są przygotowane rzetelnie i z dbałością o estetykę. Całość wzbogacają liczne ilustracje przedstawiające materiał badawczy, konstrukcje stanowisk akwizycji obrazów, schematy architektur sieci neuronowych oraz wizualizacje przebiegu uczenia i rezultatów eksperymentów. Wyniki badań zaprezentowano w formie wykresów metryk wydajności, tabel z danymi ilościowymi oraz dodatkowych zestawień ułatwiających interpretację rezultatów.

Wszystkie publikacje wchodzące w skład cyklu charakteryzują się właściwą, zgodną ze standardami naukowymi strukturą, obejmującą jasno określoną motywację, przegląd literatury, opis metodyki, prezentację wyników oraz wnioski. Dzięki tej przejrzystości i konsekwencji redakcyjnej każda z prac umożliwia czytelnikowi pełne zrozumienie kontekstu badań, sposobu rozwiązania postawionego problemu oraz znaczenia uzyskanych rezultatów, a także ułatwia ocenę wkładu doktoranta i wartości naukowej prezentowanych osiągnięć.

Podsumowując, pomimo wskazanych mankamentów, zarówno formalna, jak i prezentacyjna strona rozprawy spełnia wymagania stawiane pracom doktorskim w dyscyplinie informatyka techniczna i telekomunikacja. Dokument jest logicznie uporządkowany, poprawny pod względem merytorycznym i stylistycznym, a wysoka jakość wizualna i strukturalna sprzyja czytelności oraz pełnemu zrozumieniu przedstawionych wyników.

Uwagi krytyczne i zagadnienia do dyskusji

1. Double-stream Convolutional Neural Networks for Machine Vision Inspection of Natural Products

W zaproponowanym rozwiązaniu autor zastosował fuzję danych na poziomie cech już po ich wydobyciu przez warstwy splotowe. Zasadne jest pytanie, czy rozważano alternatywę w postaci wczesnej fuzji

na poziomie danych wejściowych. Jeśli nie – z jakich powodów takie podejście zostało odrzucone? Czy możliwe i celowe byłoby scalenie dwóch obrazów w jeden i wykorzystanie jednocanalowej sieci spłotowej, zamiast architektury dwustrumieniowej?

2. Barley Variety Recognition with Viewpoint-aware Double-stream Convolutional Neural Networks

Różnice dokładności klasyfikacji między wersjami *viewpoint-ignorant* i *viewpoint-aware* są stosunkowo niewielkie. Czy zaobserwowane korzyści ze wstępnego rozpoznania orientacji ziarniaków wynikają z istotnych właściwości modelu, czy też mogą być efektem losowych wahań procesu uczenia? Czy autor przeprowadził wielokrotne próby uczenia oraz testy statystyczne potwierdzające istotność uzyskanych wyników? Jeśli nie to czy przeprowadzenie takich testów byłoby zasadne?

W pracy wspomniano o zastosowaniu sieci AlexNet do rozpoznawania orientacji brzuszno–grzbietowej ziarniaka. Brakuje jednak informacji o procedurze uczenia tej sieci oraz o uzyskanych dokładnościach klasyfikacji obrazów ze względu na orientację ziarniaków. Czy analizowano wpływ ewentualnych błędów rozpoznawania orientacji na końcową skuteczność rozpoznawania odmian?

3. Improving Real-Time Performance of U-Nets for Machine Vision in Laser Process Control

Rysunek 3 przedstawia pojedynczy obraz z procesu napawania laserowego. Czy jest on reprezentatywny dla całego zbioru danych? Proszę o doprecyzowanie znaczenia i interpretacji obszarów oznaczonych kolorami białym, zielonym i czerwonym. Jak wykorzystywano wyniki segmentacji i analizy tych obrazów do sterowania procesem napawania laserowego?

Wyniki segmentacji uzyskane zarówno dla pełnej, jak i uproszczonej wersji U-Net odbiegają od obszarów referencyjnych. Wymaga to komentarza: na ile takie rezultaty są praktycznie użyteczne? Ponadto wyniki segmentacji uzyskane przez oba modele różnią się między sobą – uproszczona sieć wygenerowała dwa obszary oznaczone kolorem zielonym zamiast jednego. Jakże ma to znaczenie aplikacyjne i czy tę różnicę można uznać za nieistotną dla kontrolowanego procesu laserowego?

4. Instance segmentation of root crops and simulation-based learning...

Na rysunku 3 przedstawiono geometrię systemu akwizycji oraz minimalną średnicę bulwy d_{min} zaznaczoną ukośnie względem podłoża. Pod jakim kątem należy dokonać pomiaru tej średnicy? Czy nie należałoby mierzyć tej średnicy w kierunku prostopadłym do powierzchni, przyjmując uproszczony elipsoidalny model kształtu bulwy i równomierny rozkład masy?

W artykule opisano metodę estymacji najmniejszej średnicy, lecz nie wyjaśniono, jaką metodą pozyskiwano rzeczywiste wartości referencyjne na potrzeby budowania zbioru uczącego i jego etykietowania. Stwierdzenie, że „physical space annotations for the training of the regression model could not be acquired” sugeruje brak takich danych. Później wskazano, że uzyskano je w symulacji 3D. Na ile wiarygodne są tak uzyskane dane i czy w modelowaniu uwzględniono zjawiska fizyczne, takie jak grawitacja czy wzajemne oddziaływania między bulwami ziemniaków?

6. Direct detection of elongated objects geometry via a centerline-based representation

W artykule połączono reprezentację obiektu w postaci linii środkowej z architekturą DETR, jednak forma publikacji – krótka nota konferencyjna – skutkuje niedostatkami szczegółów technicznych. Rysunek 2 przedstawia architekturę DETR jedynie w formie pojedynczego symbolu, a opis modyfikacji sieci jest nazbyt zwięzły. Na czym dokładnie polegała adaptacja architektury DETR do zadania predykcji linii środkowej? Czy rozważano zastosowanie bardziej szczegółowej reprezentacji linii środkowej, opartej na większej liczbie punktów niż trzy i czy taką reprezentację również można wbudować w sieć DETR?

W odniesieni do wszystkich publikacji

Ponieważ istotnym elementem badań są opracowania algorytmiczne i implementacyjne, warto zapytać, czy kody źródłowe implementujące opracowane algorytmy i modele neuronowe zostały publicznie udostępnione? Ich publikacja mogłaby zwiększyć transparentność badań oraz ułatwić ich dalsze wykorzystanie i rozwój w środowisku naukowym.

Wniosek końcowy

Podsumowując stwierdzam, że rozprawa doktorska mgr. inż. Przemysława Dolaty zatytułowana *Zastosowanie metod uczenia głębokiego w wybranych problemach widzenia komputerowego w kontekście przemysłowej oceny jakości* (ang. *Deep learning methods for computer vision-based industrial inspection applications*) prezentuje oryginalne rozwiązanie problemu naukowego polegające na opracowaniu metod uczenia głębokiego w analizie obrazów cyfrowych. Pan mgr. inż. Przemysław Dolata wykazał się umiejętnością samodzielnego prowadzenia pracy badawczej, wiedzą w zakresie informatyki technicznej i telekomunikacji odpowiednią do rozwiązania rozważanych problemów naukowych oraz znajomością aktualnego stanu wiedzy i literatury światowej. Rozprawę doktorską stanowi zbiór sześciu opublikowanych i powiązanych tematycznie artykułów naukowych, uzupełnionych o autoreferat oraz streszczenia w języku polskim i angielskim. Taka forma jest w pełni właściwa dla rozpraw opartych na cyklu publikacji i spełnia formalne wymagania stawiane pracom doktorskim przygotowywanym w tym trybie.

W mojej opinii przedstawiona do recenzji rozprawa spełnia wymagania stawianych rozprawom doktorskim w Art. 187 Ustawy Prawo o szkolnictwie wyższym i nauce z dnia 20 lipca 2018 roku. Wnoszę o jej przyjęcie i dopuszczenie do dalszych etapów postępowania doktorskiego.

Wniosek o wyróżnienie

W świetle przedstawionych osiągnięć, obejmujących zarówno opracowanie oryginalnych metod pozyskiwania i przetwarzania danych, jak i konstrukcję nowatorskich modeli głębokiego uczenia dostosowanych do złożonych problemów analizy obrazów, rozprawę należy uznać za wyróżniającą się pod względem merytorycznym. Doktorant wykazał się wysokim poziomem samodzielności badawczej, umiejętnością formułowania i rozwiązywania trudnych problemów naukowych oraz zdolnością do tworzenia rozwiązań o istotnej wartości poznawczej i aplikacyjnej. Dwie publikacje, których doktorant jest pierwszym (wiodącym) autorem, zostały zamieszczone w czasopiśmie *Computers and Electronics in Agriculture* (w 2021 roku IF = 6,757, aktualnie IF = 8,9) oraz *IEEE Access* (IF = 3,6) o sumarycznym współczynniku wpływu większym niż 10. Według Scopus doktorant uzyskał wskaźnik Hirscha $h = 3$ a jego prace zacytowano w 41 publikacjach (bez uwzględnienia autocytowań). Wyniki pracy Pana mgr. inż. Przemysława Dolaty stanowią znaczący wkład w rozwój metod wizji komputerowej, a całość rozprawy spełnia kryteria uzasadniające przyznanie jej wyróżnienia o co niniejszym wnioskuje.

Prof. Szymon