

Olsztyn, 2025/12/10

Dr hab. inż. Piotr Zapotoczny, prof. UWM
Wydział Nauk Technicznych
Katedra Mechatroniki
Uniwersytet Warmińsko-Mazurski w Olsztynie

RECENZJA

Rozprawy doktorskiej (cyklu publikacji) mgr inż. Przemysława Dolaty pt. *Deep learning methods for computer vision-based industrial inspection applications (Zastosowanie metod uczenia głębokiego w wybranych problemach widzenia komputerowego w kontekście przemysłowej oceny jakości)*

Podstawa oceny: Recenzja wykonana na zlecenie Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacyjna Politechniki Wrocławskiej (uchwała nr 320/15/RDND03/2024-2028 z dnia 22/10/2025 roku).

Promotor pracy: dr hab. Maciej Zięba, prof. uczelni

Dziedzina: Nauki inżynieryjno-techniczne
Dyscyplina: Informatyka techniczna i telekomunikacja

Dane bibliograficzne dorobku doktoranta:

Wg. scholar.google.com

Cytowania: 65, H-indeks: 3

Przedmiot i cel recenzji

Rozprawa opisuje wyniki badań przedstawione w zbiorze sześciu tematycznie powiązanych artykułów naukowych i dotyczy systemów komputerowego widzenia (computer vision) wykorzystujących metody uczenia głębokiego (deep learning) w wybranych zadaniach inspekcji wizyjnej w środowisku przemysłowym. Forma pracy jest zgodna z art. 187 ust. 3 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce, zgodnie z którym rozprawę doktorską może stanowić m.in. zbiór opublikowanych i powiązanych tematycznie artykułów naukowych.

Rozprawa zawiera łącznie 120 stron, przy czym strony od 43-120 to załączniki z kopiami artykułów naukowych. Kolejne rozdziały od numeru 1 do 4 to

WPLYNEŁO
19-12-2025

ADN-IT / 242 / 2025

odpowiednio: 1. Wprowadzenie, 2. Klasyfikacja wielowidokowa, 3. Szybka segmentacja semantyczna, 4. Segmentacja instancji do wymiarowania obiektów 5. Podsumowanie oraz Podziękowanie i Bibliografia. Część polskojęzyczna ogranicza się do streszczenia; zasadnicza treść merytoryczna opracowana jest w języku angielskim, co jest standardem w dyscyplinie informatyka techniczna i telekomunikacja.

Cykl sześciu publikacji stanowią prace:

1. P. Dolata, M. Mrzygłód, J. Reiner. 2017. *Double-stream Convolutional Neural Networks for Machine Vision Inspection of Natural Products*. Applied Artificial Intelligence. 31(7-8), pp. 643-659, pkt. 20 MEiN, IF=0.587, 10 citations (WoS)
2. P. Dolata, J. Reiner. 2018. *Barley Variety Recognition with Viewpoint-aware Double-stream Convolutional Neural Networks*. Proceedings of the Federated Conference on Computer Science and Information Systems. DOI: 10.15439/2018F286. pkt. 70 MEiN, 7 citations (WoS)
3. P. Dolata, J. Reiner. *“Improving Real-Time Performance of U-Nets for Machine Vision in Laser Process Control”*. 2019. Proceedings of the Federated Conference on Computer Science and Information Systems DOI: 10.15439/2019F190. pkt. 70 MEiN, 70
4. P. Dolata, P. Wróblewski, M. Mrzygłód, J. Reiner. 2021. *“Instance Segmentation of Root Crops and Simulation-based Learning to Estimate their Physical Dimensions for On-line Machine Vision Yield Monitoring”*. 2021. Computers and Electronics in Agriculture. DOI: <https://doi.org/10.1016/j.compag.2021.106451>. pkt. 100 MEiN, 20 citations (WoS)
5. P. Dolata, P. Majewski, P. Lampa, M. Zięba, J. Reiner. 2025. *“Amodal Instance Segmentation for Mealworm Growth Monitoring Using Synthetic Training Images”* IEEE Access, DOI: [10.1109/ACCESS.2025.3550780](https://doi.org/10.1109/ACCESS.2025.3550780). pkt. 100 MEiN, IF-3,6. 1 citations (WoS)
6. P. Dolata, M. Zięba. 2025. *“Direct Detection of Elongated Objects Geometry via a Centerline-based Representation”* Recent Challenges in Intelligent Information and Database Systems ACIIDS 2025, DOI: [10.1007/978-981-96-5884-8_6](https://doi.org/10.1007/978-981-96-5884-8_6). pkt. 70 MEiN, 0 citations (WoS)

We wszystkich pracach Doktorant jest pierwszym autorem oraz zgodnie z treścią rozprawy jego autorski wkład w powstanie danej publikacji jest znaczący i stanowi: (a) opracowanie architektury sieci neuronowych oraz modeli regresyjnych do poszczególnych zastosowań, (b) przeprowadzenia

eksperymentu, (c) analizę wyników oraz (d) napisanie wstępnych wersji publikacji.

Publikacje obejmują lata 2017 -2025, co przy tak szybkim rozwoju systemów wizyjnych, a szczególnie metod głębokiego uczenia wspieranych sztuczną inteligencją wymuszają ocenę poszczególnych publikacji w kontekście stanu wiedzy na moment przygotowania eksperymentu, prowadzenia badań oraz czasu wydania publikacji.

Wykorzystanie analizy obrazu w zastosowaniach przemysłu rolno-spożywczego sięga lat 80. XX wieku. We wczesnym okresie rozwoju tej dziedziny systemy akwizycji obrazu były bardzo kosztowne, charakteryzowały się ograniczoną jakością (niską rozdzielczością, małą głębią bitową), a ponadto brakowało zaawansowanych narzędzi statystycznych do analizy pozyskiwanych danych. Obrazy analogowe obiektów musiały być uprzednio przekształcane do postaci cyfrowej, a następnie poddawane klasycznym metodom analizy obrazu. Z uwagi na brak wystarczająco szybkich komputerów oraz efektywnych algorytmów przetwarzania i analizy obrazu, praktyczne zastosowanie systemów wizyjnych było wówczas znacząco ograniczone.

W latach 90-tych oraz w kolejnych dekadach większość badań z wykorzystaniem systemów wizyjnych koncentrowała się na identyfikacji podstawowych cech fizycznych produktów rolniczych, takich jak geometria, kształt, barwa czy tekstura powierzchni. Badania te miały jednak w przeważającej mierze charakter laboratoryjny i stosunkowo ograniczone przełożenie na praktykę przemysłową. Sytuacja zaczęła się zmieniać wraz z dynamicznym rozwojem systemów akwizycji obrazu (kamer, aparatów, skanerów) oraz wzrostem mocy obliczeniowej procesorów i kart graficznych. Możliwości komputerowej analizy obrazu (KAO) uległy wówczas znacznemu poszerzeniu. Zaczęto konstruować systemy wizyjne, których zadaniem była nie tylko identyfikacja cech fizycznych, lecz także budowa modeli statystycznych umożliwiających określanie składu chemicznego, stopnia dojrzałości czy stanu fitosanitarnego (fitopatologicznego) produktów rolniczych.

Równoległe z rozwojem rozwiązań sprzętowych intensywnie rozwijały się algorytmy analizy obrazu oraz narzędzia statystyczne służące interpretacji wyników. Stało się możliwe przetwarzanie dużych zbiorów danych z wykorzystaniem m.in. metod analizy wielowymiarowej czy algorytmów genetycznych. Kolejnym kamieniem milowym było opracowanie i upowszechnienie sztucznych sieci neuronowych, które w istotny sposób przyspieszyły rozwój systemów wizyjnych. Połączenie szybkich, relatywnie tanich i wysokorozdzielczych systemów akwizycji obrazu z sieciami

neuronowymi umożliwiło projektowanie funkcjonalnych, niezawodnych systemów przemysłowych. Dalszy rozwój metod sztucznej inteligencji doprowadził do faktycznej rewolucji w zastosowaniach systemów wizyjnych w sektorze rolno-spożywczym.

Współcześnie systemy wizyjne są integralnym elementem szeroko rozumianego rolnictwa precyzyjnego: znajdują zastosowanie na polach uprawnych, w sadach, szklarniach oraz tunelach foliowych, a także w zakładach przetwórstwa rolno-spożywczego i w maszynach rolniczych przeznaczonych do uprawy roli, pielęgnacji roślin oraz zbioru plonów. W nowoczesnych gospodarstwach rolniczych testowane i wdrażane są roboty przeznaczone do selektywnego zbioru, sortowania owoców i warzyw, wykorzystujące zaawansowane algorytmy rozpoznawania obiektów i planowania trajektorii. Roboty udojowe, wyposażone w rozbudowane systemy wizyjne, stały się standardem w produkcji mleka, zapewniając wysoką powtarzalność procesu, oraz możliwość ciągłego monitorowania parametrów produkcyjnych. Coraz częściej systemy wizyjne integrowane są także z platformami bezzałogowymi (dronami) oraz z infrastrukturą Internetu Rzeczy (IoT), co pozwala na ciągłe śledzenie zmian w czasie i przestrzeni.

Należy jednak podkreślić, że nawet najbardziej zaawansowane systemy wizyjne, z implementacją metod sztucznej inteligencji, nie są w stanie w pełni zastąpić człowieka w wielu zadaniach związanych z rolnictwem, zwłaszcza na etapie budowy algorytmów rozpoznawania instancji. Wynika to z faktu, że obiekty biologiczne cechują się bardzo dużą zmiennością, zdeterminowaną wpływem warunków środowiskowych na ich fenotyp. Dobrym przykładem jest ziarno zbóż. W warunkach laboratoryjnych, na dobrze scharakteryzowanej populacji, możliwe jest opracowanie skutecznego modelu klasyfikacji odmianowej – czego liczne publikacje są dowodem. Jednak ze względu na zróżnicowanie warunków uprawy (miejsce, nawożenie, warunki pogodowe), jak również zmienność między sezonową, obraz ziaren tej samej odmiany może ulegać znacznym zmianom. Utrudnia to budowę systemów rozpoznawania, które byłyby odporne na tak szeroki zakres zmienności.

W tym kontekście szczególnie cenne i zarazem konieczne jest prowadzenie badań z wykorzystaniem zaawansowanych systemów wizyjnych, obejmujących zarówno nowoczesne komponenty sprzętowe (hardware), jak i wyspecjalizowane oprogramowanie (software), dedykowanych przemysłowi rolno-spożywczemu. Istotne znaczenie mają tu prace, które nie ograniczają się jedynie do pokazania możliwości metod w warunkach laboratoryjnych, lecz zmierzają do opracowania rozwiązań możliwych do wdrożenia w rzeczywistych procesach produkcyjnych, z uwzględnieniem niezawodności, powtarzalności,

kalibracji, odporności na zmienne warunki środowiskowe oraz opłacalności ekonomicznej. Mając na uwadze złożoność i wysoki stopień trudności opisanych zagadnień, tym bardziej należy podkreślić wagę podjęcia ich w ramach przedstawionej rozprawy doktorskiej. Doktorant podjął się problematyki o istotnym znaczeniu naukowym i aplikacyjnym, wymagającej zarówno dogłębnej znajomości metod komputerowej analizy obrazu, sztucznej inteligencji, jak i dobrej orientacji w praktyce przemysłu rolno-spożywczego. Zrealizowanie z powodzeniem wyznaczone cele badawcze, w tak wymagającym obszarze, zasługuje na szczególne podkreślenie w kontekście oceny dorobku naukowego.

Charakter pracy jest wyraźnie inżyniersko-aplikacyjny, ale przy silnym nacisku na metodykę. Doktorant opracował architektury dla dwustrumieniowych sieci w celu klasyfikacji wielowidokowej, punktowe modyfikacje U-Net z mechanizmem drop-path, zastąpienie ramek ograniczających reprezentacją „centerline” w detekcji oraz algorytmów przygotowania danych (symulacja 3D dla regresji wymiarów, syntetyczne dane do segmentacji instancji amodalnych AIS). Jednocześnie rozprawa nie jest skupiona na formalnym aparacie matematycznym – nacisk położony jest na eksperyment, analizę jakościową i ilościową oraz opisuje ograniczenia zastosowań. Całość tworzy spójną koncepcję: od klasyfikacji pojedynczych obiektów (ziarno jęczmienia) przez szybkie segmentowanie obiektów (owady) do sterowania procesem, aż po zaawansowane scenariusze wymiarowania wielu małych obiektów w gęstych scenach (rośliny okopowe).

Przedstawioną dysertację pozwoliłem sobie zrecenzować zgodnie z następującym planem:

1. Zakres i znaczenie pracy

- Czy zasadnicze zagadnienie badawcze stanowiące przedmiot rozprawy (teza rozprawy) zostało przez Autora sformułowane w sposób jednoznaczny i wystarczająco precyzyjny.
- Czy podjęta problematyka jest aktualna oraz posiada istotne znaczenie poznawcze i/lub aplikacyjne.

2. Charakter rozprawy

- Jaki charakter ma rozprawa (teoretyczny, doświadczalny, metodyczny, aplikacyjny lub inny).
- Czy charakter ten został wyraźnie określony przez Autora oraz konsekwentnie zrealizowany w pracy.

3. Przegląd literatury i stanu zagadnienia

- Czy w rozprawie przeprowadzono w sposób właściwy i dostatecznie wyczerpujący przegląd oraz krytyczną analizę źródeł (w tym literatury

światowej oraz stanu zagadnień w przemyśle), świadcząca o odpowiednio szerokiej i pogłębionej wiedzy Autora.

- Czy wnioski wynikające z przeglądu źródeł zostały sformułowane jasno, logicznie i przekonująco.

4. Rozwiązanie problemu badawczego

- Czy Autor rozwiązał postawione w rozprawie zagadnienie badawcze, stosując odpowiednie metody badawcze.
- Czy przyjęte założenia zostały należycie uzasadnione z punktu widzenia współczesnego stanu wiedzy.

5. Oryginalność i wkład Autora pracy w cykl publikacji

- Na czym polega oryginalność rozprawy oraz co stanowi samodzielny i oryginalny dorobek Autora.
- Jaka jest pozycja rozprawy na tle aktualnego stanu wiedzy i/lub poziomu techniki przedstawionego w literaturze światowej.

6. Prezentacja wyników i strona redakcyjna

- Czy Autor wykazał umiejętność poprawnego, zwięzłego i przekonującego przedstawienia uzyskanych wyników.
- Czy rozprawa odznacza się odpowiednią poprawnością językową i redakcyjną.

1. Zakres i znaczenie pracy.

Rozprawa dotyczy projektowania wyspecjalizowanych metod uczenia głębokiego dla zadań inspekcji wizyjnej w przemyśle, w sytuacjach, w których standardowe podejście oparte na *transfer learningu* nie daje zadowalających rezultatów. Autor formułuje cel pracy wprost w streszczeniu i we wstępie jako opracowanie metod głębokiego uczenia dla specyficznych zastosowań inspekcji wizyjnej w rolnictwie precyzyjnym, przetwórstwie żywności oraz wybranych zadaniach sterowania procesami przemysłowymi. Doktorant wraz z współautorami w swoich publikacjach zajął się klasyfikacją odmianową jęczmienia browarnego i pastewnego oraz identyfikacją ich cech fizycznych w procesie sortowania (praca nr. 1 i 2), określeniem w czasie rzeczywistym wielkości geometrycznej bulw ziemniaków w czasie ich zbioru (praca nr 4), identyfikacją pojedynczych osobników larw mącznika młynarka w hodowli przemysłowej (praca nr 5 oraz 6), Natomiast praca nr 3 dotyczyła wykorzystania inspekcji wizyjnej do kontroli w czasie rzeczywistym procesu obórki laserowej.

Doktorant sformułował trzy pytania badawcze dotyczące inspekcji wizyjnej:

Pytanie RQ1: Dotyczyło opracowania klasyfikatorów opartych o konwolucyjne sieci neuronowe do klasyfikacji wielowidokowej, w której obiekt podlegający klasyfikacji został opisany z dwóch stron obrazowania.

Pytanie RQ2: Dotyczyło optymalizacji architektury modelu segmentacji semantycznej opartego na U-Net, aby poprawić czas wnioskowania przy zachowaniu jakości segmentacji.

Pytanie RQ3: Dotyczyło budowy modelu widzenia komputerowego do identyfikacji i wymiarowania małych obiektów w gęstych scenach na podstawie pojedynczych obrazów

Pytania te sformułowano jasno i jednoznacznie, a ich zakres jest dobrze zarysowany zarówno we wstępie, jak i w rozdziałach 2-4.

Tematyka rozprawy jest wyraźnie aktualna i istotna. Autor porusza obszar intensywnie rozwijający się naukowo *deep learning*, *computer vision* i jest jednocześnie kluczowy dla praktyki przemysłowej (automatyczna kontrola jakości, kontrola procesów w czasie rzeczywistym, rolnictwo precyzyjne).

W kontekście okresu powstawania publikacji warto podkreślić, że pierwsze prace (2017-2019) powstawały w czasie, gdy otwarte źródła (open-source) bibliotek do uczenia maszynowego (*PyTorch* i *TensorFlow*) dopiero się rozwijały, a literatura dotycząca np. sieci konwolucyjnych była znacznie uboższa niż obecnie.

2. Charakter rozprawy

Rozprawa ma wyraźnie charakter metodyczno-doświadczalny i nie jest pracą czysto teoretyczną. Autor nie rozwija formalnych podstaw teorii uczenia głębokiego, lecz projektuje konkretne architektury sieci neuronowych oraz procedury uczenia, a następnie poddaje je systematycznej weryfikacji eksperymentalnej na danych pochodzących z rzeczywistych pomiarów (zboże, obróbka laserowa, bulwy ziemniaka, larwy mącznika młynarka).

Struktura rozprawy – obejmująca krótkie, syntetyzujące rozdziały oraz pełne teksty sześciu publikacji i jest typowa dla doktoratów realizowanych w formule „złożonych z cyklu prac”. Rozdziały 2-4 mają wyraźnie przeglądowo-interpretujący oraz integrujący charakter względem załączonych artykułów, umiejscawiają je w szerszym kontekście badawczym oraz ukazują spójność stawianych pytań badawczych i formułowanych odpowiedzi.

3. Przegląd literatury i analiza źródeł

Doktorant przedstawia poprawny i stosunkowo szeroki przegląd literatury (łącznie cytowanych jest 151 pozycji). Bibliografia obejmuje pozycje opublikowane w latach 2017-2024, w tym aktualne artykuły przeglądowe. Wnioski z przeglądu literatury są formułowane w sposób klarowny np.: wprost wskazywane są luki badawcze (brak metod CNN dostosowanych do klasyfikacji wielowidokowej w zadaniach inspekcji, niewystarczająca wydajność czasowa

klasycznych architektur typu U-Net, brak efektywnych sposobów pozyskiwania danych do AIS – *Segmentacja instancji amodalnych*, ograniczenia detektorów pudełkowych w scenach o dużej gęstości obiektów).

Na plus należy zaliczyć krytyczne podejście Autora, który nie tylko cytuje prace, ale też dość precyzyjnie wskazuje ich ograniczenia w kontekście swoich zastosowań np. trudności generatywnych metod AIS, niemożność wykorzystania ramek ograniczających w scenach z nakładającymi się elipsoidalnymi obiektami, czy ograniczoną reprezentatywność zbiorów obrazu z linii produkcyjnych.

4. Rozwiązanie postawionego zagadnienia – metody i założenia

Wszystkie trzy główne pytania badawcze zostały zrealizowane i rozwiązywane.

W odniesieniu do postawionego pytania RQ1 (klasyfikacja wielowidokowa) Doktorant projektuje i analizuje szereg dwustrumieniowych architektur konwolucyjnych sieci neuronowych (*double-stream CNN*), badając różne strategie fuzji informacji (*early fusion, view-pooling*). W dwóch pierwszych publikacjach Autor projektuje i bada dwustrumieniowe sieci konwolucyjne do klasyfikacji obiektów reprezentowanych przez parę obrazów (dwie strony ziarna). W pierwszej pracy porównuje różne strategie fuzji: wczesną, view-pooling, różne warianty późnej fuzji i fuzję na poziomie predykcji, na zbiorze 18 463 par obrazów jęczmienia w 7 klasach, z trójkrotną walidacją krzyżową. Autor wykazał, że dwustronne obrazowanie redukuje odsetek fałszywych oznaczeń przy porównywalnej czułości, co z punktu widzenia kontroli jakości jest istotne praktycznie.

W rozwiązaniu problemu badawczego oznaczonego RQ2 (przyspieszenie segmentacji semantycznej) proponowana jest procedura uczenia sieci U-Net z mechanizmem drop-path, umożliwiającym późniejszą swoistą „chirurgię” sieci – identyfikację i usunięcie bloków o niewielkim wpływie na jakość segmentacji przy jednocześnie wysokim koszcie obliczeniowym. Eksperymenty na zbiorze 250 obrazów z monitoringu procesu laserowego (4 klasy) pokazują, że usunięcie pierwszych trzech bloków w dół i ostatnich trzech w górę daje ok. pięciokrotne przyspieszenie (z ~155 ms do ~30 ms na obraz) przy spadku wartości miary F1 o 4–11% w zależności od klasy.

W celu realizacji pytania RQ3 (wymiarowanie małych obiektów w gęstych scenach) Autor podzielił problem na trzy podzadania: (3.1) estymację wymiarów 3D z pojedynczego obrazu przy braku adnotacji w przestrzeni fizycznej – zrealizowaną poprzez trening regresora na danych syntetycznych z zastosowaniem osobnych modeli segmentacji instancji dla domeny syntetycznej


i rzeczywistej; (3.2) pozyskiwanie danych do *segmentacji instancji amodalnych* (AIS) – rozwiązane przez zaprojektowanie kontrolowanego procesu akwizycji modalnych obiektów oraz deterministycznego generatora gęstych, częściowo zasłoniętych scen; (3.3) poprawę detekcji w obrazach o dużej gęstości obiektów – poprzez zastąpienie reprezentacji pudełkowej (bounding box) reprezentacją typu centerline oraz modyfikację modelu DETR (*DEtection Transformer*). Przyjęte założenia są jasno sformułowane i przekonująco uzasadnione np. możliwość kontrolowanej akwizycji danych dla Segmentacji Instancji Amodalnych (AIS), specyficzna geometria wydłużonych obiektów w zadaniu detekcji z użyciem linii środkowych.

Doktorant wyraźnie wskazuje również ograniczenia zastosowalności proponowanych metod (m.in. zależność od szczegółów domeny, brak uniwersalności reprezentacji centerline poza obiektami o wydłużonej geometrii), co należy uznać za istotny atut rozprawy. Zasadnicze zastrzeżenie krytyczne dotyczy zakresu weryfikacji ilościowej w niektórych zadaniach. W szczególności, w pracy poświęconej estymacji wymiarów bulw ziemniaka (publikacja 4) brak pełnych pomiarów referencyjnych na danych rzeczywistych ogranicza możliwość przeprowadzenia klasycznej oceny błędu regresji. Autor sam zwraca na to uwagę i częściowo kompensuje tę słabość analizą porównawczą rozkładów (histogramów), niemniej stanowi to pewne ograniczenie siły dowodu w tym fragmencie rozprawy.

Wszystkie zastosowane metody mają jasno opisane założenia: scena quasi-płaska w generowaniu danych AIS, możliwość zbudowania realistycznego symulatora 3D dla bulw, wydłużony kształt obiektów w metodzie centerline. Ogranicza to możliwość uogólniania analizy, jednak Doktorant nie ukrywa tych ograniczeń, przeciwnie podkreśla je jako naturalne kompromisy między ogólnością a użytecznością w konkretnych zastosowaniach.

5. Oryginalność pracy i wkład Autora

Rozprawa w przekonujący sposób dokumentuje oryginalny i samodzielny wkład Autora w rozwój metod uczenia głębokiego w zastosowaniu inspekcji wizyjnej. W obszarze klasyfikacji wielowidokowej Doktorant wprowadza i poddaje eksperymentalnej analizie dwustrumieniowe sieci splotowe z różnymi strategiami łączenia informacji pochodzącymi z obrazu dwóch stron ziarniaka. Należy podkreślić, że jest to jedno z wcześniejszych zastosowań tego typu modeli w rolnictwie i przetwórstwie żywności. Wcześniejsze systemy wizyjne oparte były na klasycznym przetwarzaniu obrazu wykorzystując metody *LDA*, *SVM*, *klasyczne sicc*, *k-NN* itp. Obraz obiektów pozyskiwany był tylko z jednej strony a w przypadku rejestracji wielowidokowej obrazu analizowane były oddzielnie. Sama metoda dwustrumieniowych sieci splotowych była w tamtym okresie już znana ale nie w zastosowaniach do produktów rolniczych,



które wymagają jednak indywidualnego podejścia i dostrajania takiej sieci do swoistych potrzeb.

W zakresie przyspieszenia modelu U-Net Doktorant proponuje nową, praktyczną procedurę trenowania, umożliwiającą późniejsze usuwanie części warstw bez konieczności ponownego trenowania sieci. Publikacja stanowi interesujący, choć przede wszystkim dodatkowy wkład w rozwój metod segmentacji opartej na architekturze U-Net z uwzględnieniem wymagań pracy w czasie rzeczywistym. Oryginalność pracy nie polega na wprowadzeniu nowej klasy modeli, lecz na twórczym połączeniu znanych technik (uproszczona architektura U-Net, Batch Normalization, modyfikacje połączeń skip) z koncepcją drop-path zastosowaną na poziomie całych skal sieci. Autorzy proponują spójną procedurę projektowania „odchudzonego” U-Neta, w której najpierw trenuje się pełny model z losowym wyłączeniem bloków, a następnie na tej bazie przeprowadza się optymalizację między jakością segmentacji a czasem działania. Szczególnie cenny jest aspekt aplikacyjny – pokazanie, że zaproponowane podejście pozwala znacząco skrócić czas przetwarzania przy niewielkim spadku dokładności. Ograniczeniem pozostaje wąska dziedzina zastosowania, niewielki zbiór danych oraz brak szerszego porównania z innymi architekturami segmentacyjnymi. Mimo to praca wnosi wartościowy wkład, pokazując praktyczną ścieżkę adaptacji klasycznych sieci segmentacyjnych do wymogów systemów wizyjnych czasu rzeczywistym w zastosowaniach przemysłowych.

Publikacja Doktoranta i współautorów nr 4 stanowi istotny i w momencie publikacji oryginalny wkład w rozwój systemów wizyjnych do monitoringu zbioru plonu upraw roślin okopowych. Autorzy podejmują problem szacowania wymiarów pojedynczych obiektów (tutaj bulw ziemniaka) bezpośrednio na kombajnie, na podstawie obrazów z standardowej kamery RGB, w warunkach silnego zagęszczenia obiektów oraz przy istotnych zniekształceniach perspektywicznych. W literaturze przywołanej w rozdziale „Related work” wskazują, że dotychczasowe rozwiązania: (1) opierały się na drogich i złożonych układach pomiarowych lub kontrolowanych warunkach laboratoryjnych, (2) zazwyczaj analizowały pojedyncze owoce/bulwy na kontrastowym tle, albo (3) ograniczały się do detekcji i zliczania plonu, bez precyzyjnego wymiarowania poszczególnych obiektów w warunkach polowych.

Oryginalność pracy przejawia się przede wszystkim w kompletnej integracji kilku elementów, które wcześniej nie występowały razem. Po pierwsze, autorzy adaptują framework Mask R-CNN do segmentacji poszczególnych bulw na przenośniku kombajnu, przy częściowych zakryciach oraz bez dodatkowego oświetlenia, uzyskując wysoką jakość segmentacji. Po drugie, proponują nieliniowy model regresyjny, który na podstawie parametrów elipsy dopasowanej do konturu obiektu (w przestrzeni kątowej) statystycznie szacuje

minimalną średnicę bulwy – parametr istotny handlowo, a jednocześnie niewidoczny bezpośrednio w obrazie 2D. Model ten jednocześnie koryguje zniekształcenia perspektywiczne i „odtworza” wymiar nieobserwowalny, wykorzystując założenia dotyczące rozkładu kształtu obiektu. Kolejny ważny element oryginalności dotyczy sposobu pozyskania danych uczących dla modelu regresyjnego. Autorzy trafnie zauważają, że praktycznie niemożliwe jest zebranie w warunkach rzeczywistych wiarygodnych pomiarów minimalnej średnicy dla każdej bulwy widocznej w obrazie. Proponują zatem środowisko symulacyjne oparte na Blenderze, w którym generują trójwymiarowe modele ziemniaków, odwzorowujące znane z literatury statystyki kształtu i rozkładu rozmiarów, a także realistyczne zachowanie bulw na przenośniku (kolizje, toczenie się, wzajemne zakrycia). Taka symulacja dostarcza jednocześnie generowanie masek segmentacyjnych oraz predykcje rzeczywistych wymiarów obiektów, co pozwala na nadzorowane uczenie regresji wymiarów. Istotnym wkładem koncepcyjnym jest również wprowadzenie tzw. podwójnego toru segmentacji (*dual segmentation pipelines*): osobny model Mask R-CNN trenowany jest na danych rzeczywistych, a osobny – na syntetycznych. To podejście stanowi wartościową propozycję metodologiczną, którą można potencjalnie przenosić na inne zadania wizji maszynowej w rolnictwie (różne uprawy, inne gatunki roślin). Wkład Doktoranta ma także wyraźny wymiar inżynierski i aplikacyjny. Praca opisuje kompletny, niedrogi system akwizycji obrazów na kombajnie (kamera RunCam 2 w obudowie drukowanej 3D, zintegrowany mikrokontroler i GPS), działający w rzeczywistych warunkach polowych oraz zweryfikowany na danych z dwóch partii ziemniaków o różnych klasach rozmiarowych. Autorzy pokazują, że ich system potrafi rozróżnić rozkłady przewidywanych średnic między partiami „small” i „large”, mimo ich częściowego nakładania się, co potwierdza praktyczną użyteczność podejścia do monitorowania plonów składających się obiektów o różnych wymiarach. Podsumowując, w okresie publikacji omawiana praca wyróżniała się na tle stanu badań jako pierwsza całościowo opisana metoda łącząca: (1) segmentację instancji upraw korzeniowych w warunkach polowych na kombajnie przy użyciu prostej kamery RGB, (2) szacowanie minimalnej średnicy pojedynczych bulw z jednego obrazu, oraz (3) uczenie modelu wielkości z wykorzystaniem symulacji 3D zamiast trudnych do pozyskania danych referencyjnych z pomiarów fizycznych. Tym samym publikacja wniosła znaczący wkład zarówno do metodologii wizji maszynowej w rolnictwie precyzyjnym, jak i do praktycznego rozwoju systemów on-line do monitoringu plonu upraw korzeniowych.

Doktorant jest pierwszym autorem wszystkich sześciu prac. W każdej z nich, w wydzielonym akapicie „*Contribution*”, szczegółowo opisuje swój wkład tj.: projektowanie architektury modelu, implementacje, prowadzenie

eksperymentów, analizie wyników, redagowanie pierwszej wersji manuskryptu. Na dzień złożenia rozprawy publikacje były cytowane łącznie 37 razy, co jak na specjalistyczną tematykę z pogranicza widzenia komputerowego i inspekcji przemysłowej stanowi dobry wskaźnik oddziaływania. Cztery prace powstały w ramach grantów NCBR, co pokazuje także praktyczne znaczenie i zdolność Doktoranta do pracy w dużych projektach badawczo-rozwojowych.

6. Prezentacja wyników i strona redakcyjna

Rozprawa jest napisana poprawnym językiem, z polskim streszczeniem. Tekst jest zwięzły i klarowny, każde z pytań badawczych ma wyraźnie wydzielony fragment (*context and research scope, motivation, content, conclusion, contribution*), co ułatwia ocenę wkładu Autora.

Przedstawienie wyników jest przekonujące i logicznie uporządkowane: na początku przedstawiony jest kontekst praktyczny, następnie sformułowanie problemu i luki badawczej, opis proponowanego rozwiązania, a na koniec omówienie rezultatów i ograniczeń. Dobrą praktyką jest jasne wyodrębnienie sekcji „*Contribution*” w streszczeniach rozdziałów – jasno pokazuje to stopień samodzielności Doktoranta.

Od strony redakcyjnej rozprawa jest uporządkowana, spełnia standardy przyjęte dla doktoratów opartych na cyklu publikacji: zawiera streszczenie, wprowadzenie, rozdziały syntetyzujące wyniki badań, podsumowanie, obszerną bibliografię oraz załączone pełne teksty artykułów. Bibliografia jest kompletna i poprawnie zredagowana.

7. Pytania do Doktoranta

1. Jakie są – w ocenie Doktoranta – ograniczenia możliwości uogólnienia zaproponowanych metod AIS na inne gatunki owadów lub inne zastosowania przemysłowe (np. w rolnictwie precyzyjnym), zwłaszcza w kontekście nieregularnych kształtów, wysokiej zmienności tekstur i różnic w warunkach oświetleniowych? Czy syntetyczne obrazy, mimo swojej skuteczności w uczeniu modeli AIS, mogą wprowadzać systematyczne błędy (bias) do modeli? Czy testowali Państwo skuteczność wytrenowanego modelu na danych pozyskanych w odmiennych warunkach środowiskowych (inne parametry oświetlenia, inny typ kamery, inne ustawienia akwizycji)?
2. Czy rozważał Pan rozszerzenie architektury double-stream o mechanizmy pozwalające sieci dynamicznie ważyć znaczenie cech pochodzących z poszczególnych widoków (np. z użyciem mechanizmów attention lub learnable view-weighting)? Jeżeli tak – z jakim skutkiem; jeżeli nie – czy

- dostrzeża Pan potencjał takiego podejścia w kontekście zadań klasyfikacji wielowidokowej w inspekcji przemysłowej?
3. Czy badał Pan możliwości adaptacji zoptymalizowanego modelu U-Net (z zastosowanym drop-path i uproszczoną architekturą) do innych procesów przemysłowych niż obróbka/napawanie lascrowe? Jakie widzi Pan ograniczenia przy rozszerzaniu tej metody na dane 3D (np. tomografia komputerowa) lub dane czasoprzestrzenne (sekwencje widoco, strumienie on-line)?
 4. W jaki sposób zaproponowany model poradzi sobie z adaptacją modelu do innych odmian roślin okopowych o bardziej zróżnicowanej morfologii niż ziemniaki, w szczególności kiedy ich kształt jest bardziej nieregularny np. marchew, pietruszka. Czy model jest odporny na obecność artefaktów w scenie, takich jak kamienie czy bryły ziemi, oraz jak można by w praktyce zwiększyć jego odporność na tego typu zakłócenia (np. przez dodatkowy moduł detekcji obiektów odstających lub etap wstępnej filtracji obiektów)?

Konkluzja:

Przedstawiona do oceny rozprawa doktorska stanowi spójną, metodycznie poprawną i oryginalną pracę badawczą z zakresu uczenia głębokiego i widzenia maszynowego w zastosowaniach przemysłowych. Mgr inż. Przemysław Dolata w sposób jasny zdefiniował cele oraz pytania badawcze, przeprowadził poprawną i aktualną analizę literatury, zaproponował własne rozwiązania metodyczne oraz poddał je rzetelnej weryfikacji doświadczalnej na danych pochodzących z rzeczywistych pomiarów. Osiągnięte wyniki mają niewątpliwą wartość poznawczą i aplikacyjną, a przedstawiony dorobek – udokumentowany cyklem publikacji – świadczy o samodzielności naukowej Autora oraz jego dobrej orientacji w aktualnym stanie wiedzy.

Biorąc pod uwagę całokształt przedstawionej pracy, stwierdzam, że rozprawa doktorska spełnia wymagania stawiane rozprawom doktorskim w art. 187 Ustawy Prawo o szkolnictwie wyższym i nauce z dnia 20 lipca 2018 roku i wnoszę o dopuszczenie mgr inż. Przemysława Dolatę do publicznej obrony pracy doktorskiej. Mając także na uwadze wysoki poziom naukowy pracy wnioskuję o wyróżnienie pracy zgodnie za zasadami przejętymi na Politechnice Wrocławskiej.

Dr hab. inż. Piotr Zapłocznny, prof. UWM

