



dr hab. inż. Piotr A. Kowalski, prof. AGH
Katedra Informatyki Stosowanej i Fizyki Komputerowej,
Wydział Fizyki i Informatyki Stosowanej,
Akademia Górniczo-Hutnicza w Krakowie,
al. Mickiewicza 30, 30-059 Kraków
email: pkowal@agh.edu.pl



oraz
Centrum Informatycznych Metod Analizy Danych
Instytut Badań Systemowych
Polskiej Akademii Nauk

Kraków, 10.01.2025

RECENZJA

**rozprawy doktorskiej Pana mgra Jacka Kałużnego
pt. „Generowanie i walidacja syntetycznych zbiorów danych obrazów
do trenowania modeli sztucznej inteligencji w wizji komputerowej”
” Generating and Validating Synthetic Image Datasets for Training
AI Models in Computer Vision”**

1. Uwagi ogólne

Prawną podstawą przygotowania recenzji rozprawy doktorskiej Pana mgr. Jacka Kałużnego jest Umowa z Uniwersytetem im. Adama Mickiewicza w Poznaniu – Wydziałem Matematyki i Informatyki reprezentowaną przez Dziekana Pana Profesora UAM dr. hab. Krzysztofa Dyczkowskiego, którą otrzymałem 18 listopada 2024 r.

Recenzja została przygotowana na podstawie rozprawy doktorskiej. Przedmiotowa praca została zrealizowana pod kierunkiem Pana prof. UAM dr. hab. Krzysztofa Dyczkowskiego, z udziałem promotora pomocniczego – Pana dra Wojciecha Pałubickiego - na Wydziale Matematyki i Informatyki Uniwersytetu im. Adama Mickiewicza w Poznaniu. Recenzowana rozprawa doktorska przedstawiona została w postaci woluminu wydanego przez Doktoranta. Rozprawa doktorska jest rozpatrywana w dziedzinie nauk ścisłych i przyrodniczych w dyscyplinie informatyka.

Recenzowana praca doktorska koncentruje się na tworzeniu i ocenie syntetycznych zbiorów danych obrazowych, które mają służyć do trenowania modeli stosowanych w widzeniu komputerowym. Głównym celem badań jest rozwiązanie problemów

związanych z niedoborem danych oraz ich zmiennością, co utrudnia rozwój solidnych i precyzyjnych modeli sztucznej inteligencji. Autor zaproponował podejście obejmujące kilka kluczowych aspektów: opracowanie nowatorskich metod generowania danych syntetycznych, tworzenie realistycznych i różnorodnych zestawów danych oraz ich dokładną walidację opartą na szeroko zakrojonych eksperymentach.

Recenzowana rozprawa doktorska podejmuje niezwykle istotne i aktualne zagadnienie badawcze z obszaru widzenia komputerowego, koncentrując się na rekonstrukcji struktur 3D drzew oraz generowaniu i walidacji syntetycznych zbiorów danych obrazowych. Praca łączy aspekty teoretyczne i praktyczne, co czyni ją zarówno ambitną, jak i wartościową dla rozwoju informatyki i jej zastosowań.

W początkowej części badań autor skupił się na opracowaniu zaawansowanej metody rekonstrukcji botanicznych drzew na podstawie pojedynczych obrazów, wykorzystując innowacyjne algorytmy uczenia maszynowego, takie jak "Radial Bounding Volumes" oraz dwumodalne modele wzrostu. Rozwiązanie to zostało gruntownie zwalidowane na rzeczywistych danych, co wykazało jego wysoką skuteczność w odtwarzaniu precyzyjnych struktur 3D.

Kolejnym istotnym elementem pracy było stworzenie syntetycznych zbiorów danych z wykorzystaniem zaawansowanych technologii, takich jak ControlNet zintegrowany z Stable Diffusion, do generowania realistycznych obrazów roślin. Autor opracował kompleksowy proces tworzenia anotowanych zbiorów danych, które są optymalne do trenowania modeli głębokiego uczenia.

Ostatnia faza badań obejmowała rygorystyczną walidację syntetycznych zbiorów danych poprzez porównanie wydajności modeli trenowanych na danych syntetycznych i rzeczywistych. Wyniki jednoznacznie wskazują na potencjał danych syntetycznych, które w niektórych przypadkach przewyższają tradycyjne podejścia, co ma istotne znaczenie dla przyszłości trenowania modeli SI.

Dysertacja ta stanowi ważny wkład w rozwój widzenia komputerowego, dostarczając nowatorskich metod generowania i walidacji danych syntetycznych. Wyniki mają szerokie zastosowanie w takich dziedzinach jak automatyzacja rolnictwa, badania biologiczne czy inne obszary wykorzystujące zaawansowane technologie analizy danych wizualnych.

Praca została napisana w języku polskim, liczy 75 stron, składa się z wykazu rysunków, tabel, wstępu, trzech rozdziałów, wniosków. Pracę kończy spisu literatury obejmujący 99 pozycji (w tym kilku publikacji Doktoranta).

2. Ogólna charakterystyka rozprawy oraz aspektu badawczego

We wstępie (Rozdział 1) recenzowanej rozprawy doktorskiej jasno uzasadniono podjęcie tematyki badań, zwracając uwagę na jej znaczenie, aktualność oraz złożoność. Autor wyraźnie wskazuje na potrzebę opracowania nowych podejść, w tym zaawansowanych metod generowania i walidacji syntetycznych danych obrazowych, które mogą znacząco przyczynić się do rozwoju widzenia komputerowego. Rozprawa doktorska Pana mgr. Jacka KAŁUŻNEGO opiera się na dysertacji, w której określono cel pracy w postaci trzech hipotez badawczych, a to:

1. Sztuczne sieci neuronowe mogą być wykorzystywane do zbierania, wyodrębniania i analizy danych.
2. Modele proceduralne generujące zbiory danych syntetycznych są rozwiązaniem problemu małych zbiorów danych.
3. Integracja danych syntetycznych z rzeczywistymi zbiorami danych zwiększa odporność i uogólnialność modeli uczenia maszynowego.

We wstępie omówiono także strukturę rozprawy, w której kolejne rozdziały koncentrują się na różnych aspektach badań: od rekonstrukcji trójwymiarowych struktur botanicznych drzew, przez generowanie syntetycznych zbiorów danych, aż po ich walidację. Wszystkie podejścia zostały oparte na danych empirycznych i zweryfikowane w toku szczegółowych eksperymentów, co podkreśla rzetelność oraz praktyczny charakter przeprowadzonych badań.

Modele i metody zaprezentowane w pracy uwzględniają różne skale analizy – od precyzyjnego odwzorowania pojedynczych struktur 3D po zastosowanie wygenerowanych danych w szerokim kontekście trenowania modeli sztucznej inteligencji. Cel rozprawy, jej treść oraz układ zostały przedstawione w sposób przejrzysty i logiczny, co znacząco ułatwia odbiór pracy i podkreśla jej wkład w rozwój wiedzy w omawianym obszarze.

Rozdział 2 pracy doktorskiej skupia się na problemie rekonstrukcji geometrii pojedynczego drzewa na podstawie zaledwie jednej fotografii. Jest to zagadnienie kluczowe zarówno dla rozwoju technik wizualizacji 3D, jak i ich praktycznych zastosowań, takich jak rekonstrukcja urbanistyczna, grafika komputerowa w grach czy efektach specjalnych w filmach. Wyzwaniem w tej dziedzinie pozostają trudności wynikające z niepełności danych oraz złożoności struktur roślinnych, takich jak rozgałęzienia i ich wzajemne przesłanianie przez liście.

W pracy przedstawiono nowatorską metodę, która wykorzystuje sieci neuronowe do automatycznego generowania masek semantycznych drzew, identyfikacji ich gatunku oraz estymacji struktury 3D na podstawie dwuwymiarowych obrazów. W przeciwieństwie do tradycyjnych podejść, takich jak skanery laserowe czy metody oparte na wielu obrazach, nowa technika pozwala na automatyczną i efektywną rekonstrukcję bez potrzeby interwencji użytkownika.

Kluczowym elementem metody jest zastosowanie radialnych objętości ograniczających (RBV), które stanowią lekką, o stałej wielkości reprezentację kształtu drzew. RBV ułatwia naukę kształtu drzew oraz umożliwia precyzyjne modelowanie ich wzrostu, kierując się uzyskanymi maskami semantycznymi. Dzięki temu możliwe jest uchwycenie szczegółowych cech rzeczywistych drzew, takich jak złożone struktury rozgałęzień, co zostało potwierdzone poprzez wszechstronną ewaluację ilościową.

Eksperymenty wykazały, że metoda pozwala na skuteczną rekonstrukcję drzew różnych gatunków, a wygenerowane struktury mogą być bezpośrednio animowane za pomocą istniejących metod dynamiki prętów. Tym samym opracowano kompleksowe narzędzie, które umożliwia twórcom treści tworzenie gotowych do animacji modeli roślin z pojedynczych fotografii, otwierając nowe możliwości dla aplikacji w dziedzinach takich jak grafika komputerowa i wizualizacje 3D.

Rozdział 3 przedstawia opracowanie i zastosowanie nowego zbioru danych syntetycznych o nazwie LAESI (Leaf Area Estimation with Synthetic Imagery), który umożliwia precyzyjną analizę morfologii liści. Zbiór ten powstał przy użyciu proceduralnych modeli generatywnych oraz technologii ControlNet, które pozwalają na tworzenie realistycznych wizualizacji liści wraz z dokładnymi anotacjami. Rozdział szczegółowo omawia proces generowania danych oraz ich zastosowanie w trenowaniu modeli uczenia maszynowego do zadań takich jak szacowanie powierzchni liści i segmentacja semantyczna.

Główne wyzwanie, na które odpowiada LAESI, to kosztowne i czasochłonne zbieranie rzeczywistych danych treningowych, szczególnie w przypadku analizy liści w badaniach botanicznych. Zastosowanie syntetycznych danych w tym kontekście pozwala na obejście ograniczeń związanych z dostępnością dużych zbiorów oznakowanych obrazów z rzeczywistego świata.

W ramach pracy zaprezentowano dwa modele proceduralne: jeden służący do generowania tła przypominającego papier milimetrowy, a drugi do realistycznego odwzorowania kształtów liści. Metoda ControlNet została wykorzystana do wzbogacenia wizualnego realizmu wygenerowanych obrazów, co pozwoliło na stworzenie zbioru danych o dużej zmienności i wysokiej jakości. W efekcie zbiór LAESI dostarcza adnotacje w postaci masek semantycznych oraz etykiet powierzchni liści, co czyni go wszechstronnym narzędziem do trenowania modeli wizji komputerowej.

Przeprowadzone eksperymenty potwierdziły efektywność tej metody. Modele wizji komputerowej, trenowane na danych LAESI, osiągnęły wysoką dokładność w zadaniach przewidywania powierzchni liści oraz segmentacji semantycznej, przewyższając modele bazowe oparte na tradycyjnych, rzeczywistych danych. Co więcej, eksperymenty wykazały, że integracja filtrowania danych syntetycznych znacznie poprawia ich spójność i wydajność modeli.

Wnioski z badań podkreślają, że dane syntetyczne generowane proceduralnie, szczególnie w połączeniu z metodami generatywnymi, mogą stanowić wartościowe narzędzie w różnych dziedzinach, takich jak teledetekcja czy rolnictwo precyzyjne. Prezentowane podejście, które eliminuje konieczność ręcznego oznaczania obrazów, przyczynia się do znacznego obniżenia kosztów i czasu potrzebnego na zbieranie wyników empirycznych, jednocześnie umożliwiając rozwój nowych metod analizy w badaniach botanicznych i innych dziedzinach nauki.

Rozdział 4 kontynuuje prace nad tworzeniem syntetycznych zbiorów danych, rozszerzając wcześniejsze podejście o nowe techniki i bardziej złożone scenariusze. Tym razem uwzględniono analizę kilku roślin umieszczonych w lizymetrach, specjalnych donicach do kontrolowania klimatu. Celem badań jest przewidywanie map głębokości sadzonek drzew na podstawie obrazów 2D, co umożliwi ich dokładną rekonstrukcję w trzech wymiarach. W porównaniu do wcześniejszego rozdziału, proces generowania danych został znacznie rozbudowany. Wprowadzono nowe elementy, takie jak mapy głębokości oraz technikę

Low-Rank Adaptation (LoRA), które poprawiają dokładność i realizm generowanych danych. Obrazy generowane są w rozdzielczości 512x512 pikseli i odzwierciedlają złożoność układów roślin, uwzględniając m.in. przesłonięcia liści czy skomplikowane wzorce rozgałęzień, zachowując przy tym precyzyjną skalę.

Głównym celem tych badań jest dokładne przewidywanie map głębokości, co ma kluczowe znaczenie dla odtworzenia trójwymiarowej struktury sadzonek. Dzięki temu możliwe jest obliczenie powierzchni liści oraz monitorowanie wzrostu roślin. Do przewidywania map głębokości wykorzystano architekturę U-Net, która została dostosowana do specyficznych wymagań tego zadania. Modele zostały przeszkolone na syntetycznych danych wygenerowanych przy użyciu omawianych wcześniej technik, co pozwoliło na uzyskanie wysokiej precyzji w przewidywaniu głębokości i struktury roślin. Wyniki treningu przeprowadzonych na tych danych wykazały pozytywne rezultaty. Modele osiągnęły dobre wyniki w analizie zarówno zbioru treningowego, jak i walidacyjnego, przy użyciu miar takich jak średni błąd kwadratowy (MSE) czy średni błąd bezwzględny (MAE). Syntetyczne dane umożliwiły realistyczną reprezentację sadzonek, co poprawiło dokładność obliczeń powierzchni liści. Dzięki tym wynikom, otwierają się nowe możliwości badań, takie jak: rozwój modeli przewidujących maski instancyjne do identyfikacji poszczególnych liści; szacowanie rozmiaru wielu liści jednocześnie oraz rekonstrukcja trójwymiarowej struktury wszystkich sadzonek widocznych na obrazie.

Metoda generowania danych syntetycznych zaprezentowana w tym rozdziale ma potencjał do zastosowania w różnych dziedzinach, takich jak biologia, ekologia czy medycyna. Dzięki minimalnym wymaganiom dotyczącym danych rzeczywistych, może być stosowana w rozwiązywaniu problemów, które tradycyjnie wymagałyby dużych, kosztownych zbiorów danych.

Praca kończy się interesującymi rozważaniami na temat zastosowanych modeli, oceny, w jakim stopniu spełniły one początkowe hipotezy, oraz ich potencjału do dalszego rozwoju i doskonalenia. Zawiera również analizę ich niezaprzeczalnej użyteczności zarówno w kontekście teoretycznym, jak i praktycznym.

3. Ocena rozprawy

a. Uwagi krytyczno-polemiczne:

W niniejszej części pokrótce przedstawione zostaną główne mankamenty dysertacji. Ze względu na istotność tejże części recenzji niniejsze uwagi zostaną przedstawione w punktach, do których łatwiej będzie się odnieść Doktorantowi.

1. W pracy nie podjęto kwestii złożoności obliczeniowej zaproponowanych algorytmów, zarówno w kontekście pamięciowym, jak i czasowym. Wprowadzenie tej analizy pozwoliłoby na pełniejsze zrozumienie efektywności proponowanych rozwiązań, zwłaszcza w kontekście ich zastosowań w praktyce.
2. W dysertacji brakuje dokładniejszych informacji na temat procesu tworzenia zaproponowanych modeli. Należałoby wyjaśnić, czy modele te zostały wybrane na podstawie jakiejś ewolucji badawczej, czy też powstały ad hoc, bez pełnej analizy iteracyjnej.
3. Wyniki eksperymentalne w pracy słabo uwzględniają porównanie zaproponowanych metod z istniejącymi rozwiązaniami. Choćby uwzględniając specyfikę innych podejść, takie porównanie mogłoby znacznie wzbogacić pracę i pomóc w ocenie efektywności zaproponowanych rozwiązań.
4. Praca nie uwzględnia analizy granicznych przypadków, takich jak bardzo trudne warunki wejściowe, które mogłyby ukazać potencjalne ograniczenia zaproponowanych algorytmów w realnych aplikacjach.
5. Choć praca odnosi się do zastosowań w różnych dziedzinach, brakuje bardziej szczegółowego przedstawienia rzeczywistych, praktycznych przypadków użycia zaproponowanych rozwiązań. Wprowadzenie konkretnych przykładów mogłoby lepiej pokazać, jak modele sprawdzają się w różnych kontekstach.
6. W pracy brakuje omówienia potencjalnych ograniczeń zaproponowanych metod w kontekście różnych warunków rzeczywistych, takich jak zmienne oświetlenie, różne kształty liści czy obecność zakłóceń na obrazach. Dokładniejsza analiza tych ograniczeń mogłaby pomóc w ocenie, w jakim zakresie zaprezentowane algorytmy są uniwersalne.
7. W pracy nie pojawia się temat skalowalności zaproponowanych metod. Choćby zaprezentowane algorytmy mogły działać w kontekście testowych zbiorów danych, nie wiadomo, jak sprawdzają się one w przypadku dużych, rzeczywistych zbiorów danych o większej zmienności.

b. Uwagi szczegółowe i techniczne

W ramach tego punktu, muszę podkreślić bardzo przemyślany układ pracy oraz staranne przygotowanie całego woluminu z pracą doktorską. W szczególności z racji na obowiązki recenzenta wypunktuję ważniejsze mankamenty dysertacji:

- Na stronie 30, rysunki 2.22 są nieczytelne.
- Dość skąpy opis podsumowania całej pracy.
- Czy symbole M oraz M kreślone przypadkiem nie pokrywają się?

Oczywiście można tu wskazać jeszcze kilka uwag związanych z pojedynczymi niedociągnięciami stylistycznymi czy dość skąpym opisem pewnych części pracy. Doktorant również nie ustrzegł się nielicznych mankamentów natury technicznej takich jak błędy interpunkcyjne czy też pomyłki w oznaczeniach (o których nadmieniano wcześniej), itp. jednak w dużej mierze nie rzutują one na relatywnie wysoką ocenę pracy.

c. Ocena ogólna

Podsumowując, należy stwierdzić, że praca zawiera istotny i oryginalny wkład Doktoranta. Autor proponuje innowacyjne podejście do problemu, który stanowi istotne wyzwanie teoretyczne, a także ma duże znaczenie w kontekście praktycznych zastosowań. Hipotezy postawione na początku pracy zostały skutecznie zweryfikowane. Zaproponowane algorytmy zostały zaimplementowane, pomyślnie przetestowane na rzeczywistych danych. Doktorant efektywnie wykorzystał metody znane z literatury, przeanalizował je, dostosował oraz zaimplementował w kontekście swojej pracy. Praca świadczy o głębokiej wiedzy Autora, jego zdolnościach do rozwiązywania realnych problemów, programowania oraz zdolności do klarownego i precyzyjnego przedstawiania wyników. Jednym z głównych atutów pracy jest starannie opracowana grafika oraz przejrzyste wizualizacje zaproponowanych metod.

4. Podsumowanie

Recenzowana praca doktorska jest przykładem oryginalnego rozwiązania ciekawych zagadnień praktycznych. Do ich rozwiązania Autor rozprawy wykorzystał we właściwy sposób zaawansowane narzędzia technik informacyjnych jakimi niewątpliwie są algorytmy neuronowe. Świadczy to, o Jego dużej kompetencji w praktycznym

posługiwaniu się narzędziami współczesnej matematyki i informatyki. Uzyskane w pracy wyniki uważam za niewątpliwie oryginalną (nowatorską) propozycję rozwiązania trudnych zagadnień, które zostały zweryfikowane na drodze rzetelnie przeprowadzanych analiz walidacyjnych. Można zatem uznać, że recenzowana rozprawa doktorska ma charakter oryginalnej pracy projektowo-naukowej, o której mówią bieżące przepisy.

Konkludując uważam, że rozprawa doktorska mgr. Jacka Kałużnego zdecydowanie spełnia wymagania stawiane w odpowiednich przepisach rozprawom doktorskim i wobec tego stawiam wniosek o jej dopuszczenie do dalszych, przewidzianych przepisami, etapów postępowania o nadanie stopnia doktora.



dr hab. inż. Piotr A. Kowalski, prof. AGH