



УНИВЕРЗИТЕТ У НОВОМ САДУ

ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА
НОВИ САД



**Семантичка сегментација дубоким
неуронским мрежама са применом на
сликама различитих модалитета у
прецизној пољопривреди**
ДОКТОРСКА ДИСЕРТАЦИЈА

Ментори:
др Марко Панић
Проф. др Татјана Лончар-Турукало

Кандидат:
Жељана Грбовић

Нови Сад, 2026



UNIVERZITET U NOVOM SADU
FAKULTET TEHNIČKIH NAUKA



**Semantička segmentacija dubokim
neuronskim mrežama sa primenom na
slikama različitih modaliteta u preciznoj
poljoprivredi**
DOKTORSKA DISERTACIJA

Mentori:
Dr Marko Panić
Prof. dr Tatjana Lončar-Turukalo

Kandidat:
Željana Grbović

Novi Sad, 2026

УНИВЕРЗИТЕТ У НОВОМ САДУ
ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА
ОБРАЗАЦ – 5a

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА¹

Врста рада:	Докторска дисертација
Име и презиме аутора:	Жељана Грбовић
Ментор (титула, име, презиме, звање, институција):	Проф. Др. Татјана Лончар-Турукало, редовни професор, Факултет техничких наука, Универзитет у Новом Саду др. Марко Панић, виши научни сарадник, Института Биосенс-Истраживачко-развојни институт за информационе технологије биосистема
Наслов рада:	Семантичка сегментација дубоким неуронским мрежама са применом на сликама различитих модалитета у прецизној пољопривреди
Језик и писмо рада:	Српски, латиница
Физички опис рада:	Унети број: Страница 173 Поглавља 6 Референци 328 Табела 27 Слика 41 Графикона Прилога
Научна област:	Електротехничко и рачунарско инжењерство
Ужа научна област (научна дисциплина):	Обрада сигнала
Кључне речи / предметна одредница:	Семантичка и сегментација на нивоу инстанце, хиперспектралне и РГБ слике, дубоко учење, 3Д реконструкција (НеРФ), фенотипизација биљака, прецизна пољопривреда

¹ Аутор докторске дисертације потписао је и приложио следеће Обрасце:

5б – Изјава о ауторству;

5в – Изјава о истовестности штампане и електронске верзије докторског рада и дозвола за објављивање личних података;

5г – Изјава о коришћењу.

Ове Изјаве се чувају у институцији у штампаном и електронском облику и не кориче се са радом.

Апстракт на језику рада:	У овој докторској дисертацији обрађен је проблем семантичке сегментације комплексних и међусобно преклопљених региона од интереса кроз различите модалитете слика са применом у прецизној пољопривреди. Посебан акценат стављен је на сегментацију класова пшенице на термалним и РГБ снимцима, као и на сегментацију круничних листова парадајза на хиперспектралним снимцима, при чему су оба модалитета обухваћена сегментацијом на нивоу пиксела и инстанци. Ови проблеми представљају један од кључних изазова у области фенотипизације биљака и процене приноса, услед сложене геометрије биљних структура, њихове хомогености и варијабилних услова осветљаја и позадине. Посебан изазов уочен је код сегментације класова пшенице у 2Д простору, где проблем преклапања и хомогених морфолошких карактеристика доводи до значајних ограничења у постизању највишег нивоа тачности (изнад 95%). Да би се превазишла ова ограничења, сагледана је и методологија која комбинује савремене архитектуре дубоког учења са 3Д методама реконструкције сцене. Кључни елемент методологије за превазилажење проблема преклапања објеката од интереса у простору, представља алгоритам енгл. \textit{Neural Radiance Fields, NeRF}, који омогућава формирање тродимензионалне реконструкције жељених структура на основу ограниченог броја дводимензионалних снимака или видео записа, чиме је омогућено раздвајање преклопљених региона и поуздано одређивање броја класова по јединици површине. За потребе истраживања креирани су различити скупови података: за детекцију класова пшенице: (1) вишегодишњи теренски скуп података, као и (2) лабораторијски скуп података посебно креиран за 3Д реконструкцију. Поред тога, за студију са фокусом на детекцију круничних листова парадајза дефинисана процедура која укључује сарадњу са произвођачима, лабораторијско снимање и специјално дизајниране експерименте за прикупљање хиперспектралних података. Методолошки приступ заснива се на развоју нове архитектуре и модела за семантичку сегментацију заснованих на дубоким неуронским мрежама, које су на прикупљеним подацима показале резултате упоредиве, а у појединим случајевима и супериорне у односу на \textit{the state-of-the-art} архитектуре. Даљим развојем истраживања, фокус је проширен на сегментацију на нивоу инстанце, где су коришћени комплекснији модели дубоких неуронских мрежа, укључујући архитектуру \textit{Mask Region-based Convolutional Neural Network (Mask R-CNN)}, који су омогућили прецизну сегментацију појединачних класова пшенице у РГБ снимцима и круничних листова парадајза у хиперспектралним снимцима. Исте методе за семантичку сегментацију су коришћене за добијање маски на лабораторијском скупу података. Добијене маске су затим интегрисане у процес адаптације NeRF методе, омогућавајући
---------------------------------	---

	тродимензионалну реконструкцију жељених сегмената пшенице, одн. класова у простору. На овај начин значајно је олакшано раздвајање преклопљених биљних структура, које су у 2Д простору представљале готово нерешив проблем, чиме је омогућен прецизнији увид у број и дистрибуцију класова, као кључне параметре за финалну процену приноса. Резултати ове дисертације показују да интеграција семантичке и сегментације на нивоу инстанце са напредним методама реконструкције 3Д сцена омогућава прецизно раздвајање сложених и преклапајућих региона од интереса. Креирани скупови података и универзална архитектура за семантичку сегментацију пружају поуздану основу за примену у фенотипизацији биљака и процени приноса, наглашавајући важност решавања оваквих проблема за даљи развој прецизне пољопривреде у домену дигиталне прецизне пољопривреде.
Датум прихватања теме од стране надлежног већа:	24.06.2021.
Датум одбране: (Попуњава накнадно институција)	
Чланови комисије: (титула, име, презиме, звање, институција)	Председник: проф. др. Милан Сечујски, редовни професор, Факултет техничких наука, Универзитет у Новом Саду Члан: др. Никша Јаковљевић, ванредни професор, Факултет техничких наука, Универзитет у Новом Саду Члан: др. Јованка Пантовић, редовни професор, Факултет техничких наука, Универзитет у Новом Саду Члан: проф. др. Милош Пајић, редовни професор, Пољопривредни факултет, Универзитет у Београду
Напомена:	

**UNIVERSITY OF NOVI SAD
FACULTY OF TECHNICAL SCIENCES**

KEY WORD DOCUMENTATION²

Document type:	Doctoral dissertation
Author:	Željana Grbović
Supervisor (title, first name, last name, position, institution)	Prof. Dr. Tatjana Lončar-Turukalo, Faculty of Technical Sciences, University of Novi Sad Dr. Marko Panić, Research and Development Institute for Information Technologies in Biosystems, BioSense Institute, University of Novi Sad
Thesis title in English:	Semantic segmentation using deep neural networks across different image modalities in precision agriculture
Language and script:	Serbian language, latin
Physical description:	Number of: Pages 175 Chapters 6 References 328 Tables 27 Illustrations 41 Graphs Appendices
Scientific field:	Electrical and computer engineering
Scientific subfield (scientific discipline):	Signal Processing
Subject, Key words:	semantic and instance segmentation, hyperspectral and RGB imaging, deep learning, 3D reconstruction, NeRF, plant phenotyping, precision agriculture.
Abstract in English:	This doctoral dissertation addresses the problem of semantic segmentation of complex and overlapping structures across multiple imaging modalities with applications in precision agriculture. Particular emphasis is placed on segmenting wheat ears in thermal and RGB images, as well as segmenting tomato crown leaves in hyperspectral images, encompassing both pixel-level and instance-level approaches. These challenges represent a critical bottleneck

² The author of the doctoral dissertation has signed the following Statements:

56 – Statement on the authorship,

5B – Statement that the printed and e-version of the doctoral dissertation are identical and authorization to use personal data,

5r – Copyright statement.

The paper and e-versions of Statements are held at the institution and are not included into the printed thesis.

	in plant phenotyping and yield estimation due to the complex geometry of plant structures, their inherent homogeneity and variability, as well as the complexities of illumination and background field conditions. A major difficulty was observed in the 2D segmentation of wheat ears, where overlapping structures and homogeneous morphological features impose significant limitations on achieving the highest accuracy levels (above 95\%). To overcome these limitations, a methodology combining the state-of-the-art deep learning architectures with 3D scene reconstruction techniques was explored. A key element of this approach is Neural Radiance Field (NeRF), which enables three-dimensional reconstruction of targeted plant structures from a limited number of 2D images or video sequences, facilitating the separation of overlapping structures and the more reliable estimation of the number of wheat ears per unit area. For this research, multiple datasets were created: (1) a multi-year field RGB dataset collected using smartphones and (2) a laboratory-created dataset specifically tailored for 3D reconstruction for wheat ears. Additionally, a dedicated procedural pipeline was developed involving collaboration with growers, controlled laboratory imaging, and specially designed experiments for hyperspectral data acquisition on tomato fruit. The methodological approach is based on the development and application of neural networks for semantic segmentation, which achieve comparable and, in some cases, even superior performance relative to state-of-the-art architectures on the collected datasets. Further research extended the approach to instance segmentation, employing advanced deep neural network models, including Mask Region-based Convolutional Neural Networks (Mask R-CNN), which enabled the precise delineation of individual wheat ears in RGB images and tomato crown leaves in hyperspectral images. The same methodology was applied to generate masks in the controlled laboratory conditions, which were subsequently integrated into a NeRF-based pipeline, enabling three-dimensional reconstruction of individual wheat ears. This significantly facilitated the separation of overlapping structures, which were nearly inseparable in 2D space, thereby allowing more accurate quantification and spatial distribution of wheat ears, critical parameters for the final yield estimation. Overall, the results demonstrate that integrating semantic and instance segmentation with advanced 3D reconstruction techniques enables precise delineation of complex and overlapping regions of interest across multiple image modalities. The created datasets and universal architecture provide a robust foundation for plant phenotyping and yield estimation, highlighting the importance of addressing such challenges for the further development of precision agriculture and plant phenotyping technologies.
Date of endorsement by the scientific board:	24.06.2021.

Date of defence: (Filled in by the institution)	
Thesis defence board: (title, first name, last name, position, institution)	Chair: Prof. dr. Milan Sečujski, Faculty of Technical Sciences, University of Novi Sad Member: dr Nikša Jakovljević, Faculty of Technical Sciences, University of Novi Sad Member: Prof. dr. Jovanka Pantović, Faculty of Technical Sciences, University of Novi Sad Member: Prof. dr. Miloš Pajić, Faculty of Agriculture, University of Belgrade
Note:	

Predgovor

Perpetuum mobile – snaga koja ne miruje, već se stalno preobražava iz ideje u akciju, iz izazova u rešenje, iz teorije u primenu. Iako u fizici ovakav sistem ostaje nemoguć, za mene on predstavlja životni i istraživački ideal: energiju koja se ne gubi, već stalno kruži i obnavlja se. Ova disertacija je rezultat upravo takve energije, velike znatiželje i naučne upornosti.

Na prvom mestu, dugujem ogromnu zahvalnost svojoj velikoj porodici **Šarić, Grbović, Delić i Veberović**. Najviše hvala **mom suprugu Danilu i deci, Dimitriju i Mitri**, na svojoj neizmernoj ljubavi, podršci i razumevanju. Hvala mojim **roditeljima i sestri** na podršci i veri u mene tokom čitavog školovanja. Veliko hvala **treneru Miodragu Tomanoviću i Health Club-u** na pomoći u održavanju energije i motivacije. Hvala **svim prijateljima i najdražim kolegama sa Instituta BioSens**, (dr Bojani Ivošević, Maši Buđen, Branislavu Pejaku, dr Oskaru Marku, dr Sanji Brdar, Vladanu Filipoviću, Nini Pajević i Dimitriju Stefanoviću). Veliku zahvalnost dugujem dugogodišnjim saradnicima sa Wageningen Univerziteta, **dr Aneeshu Chauhanu i dr Hendriku de Villiersu**. Takođe zahvaljujem **dr Nemanji Đuriću** na korisnim savetima, kao i **dr Slobodanu Iliću** čije su sugestije značajno uticale na završni deo ove disertacije. Posebnu zahvalnost dugujem **svojoj mentorki, dr Tatjani Lončar-Turukalo**, na strpljenju i poverenju koje mi je pružila tokom ovog istraživačkog puta. Za kraj, najveću zahvalnost dugujem **svom mentoru i kolegi dr Marku Paniću**, na velikom strpljenju i vođenju na ovom zajedničkom putu.

Ovu doktorsku disertaciju posvećujem svim studentima, profesorima i istraživačima koji svojim znanjem, posvećenošću i istrajnošću doprinose razvoju nauke i akademske zajednice, neprestano usmeravajući društvo ka napretku i boljoj budućnosti.

Sažetak

U ovoj doktorskoj disertaciji obrađen je problem semantičke segmentacije kompleksnih i međusobno preklapljenih regiona od interesa kroz različite modalitete slika sa primenom u preciznoj poljoprivredi. Poseban akcenat stavljen je na segmentaciju klasova pšenice na termalnim i RGB snimcima, kao i na segmentaciju kraničnih listova paradajza na hiperspektralnim snimcima, pri čemu su oba modaliteta obuhvaćena segmentacijom na nivou piksela i instanci. Ovi problemi predstavljaju jedan od ključnih izazova u oblasti fenotipizacije biljaka i procene prinosa, usled složene geometrije biljnih struktura, njihove homogenosti i varijabilnih uslova osvetljaja i pozadine.

Poseban izazov uočen je kod segmentacije klasova pšenice u 2D prostoru, gde problem preklapanja i homogenih morfoloških karakteristika dovodi do značajnih ograničenja u postizanju najvišeg nivoa tačnosti (iznad 95%). Da bi se prevazišla ova ograničenja, sagledana je i metodologija koja kombinuje savremene arhitekture dubokog učenja sa 3D metodama rekonstrukcije scene. Ključni element metodologije za prevazilaženje problema preklapanja objekata od interesa u prostoru, predstavlja algoritam engl. *Neural Radiance Fields*, *NeRFs*, koji omogućava formiranje trodimenzionalne rekonstrukcije željenih struktura na osnovu ograničenog broja dvodimenzionalnih snimaka ili video zapisa, čime je omogućeno razdvajanje preklapljenih regiona i pouzdano određivanje broja klasova po jedinici površine.

Za potrebe istraživanja kreirani su različiti skupovi podataka: za detekciju klasova pšenice: (1) višegodišnji terenski skup podataka, kao i (2) laboratorijski skup podataka posebno kreiran za 3D rekonstrukciju. Pored toga, za studiju sa fokusom na detekciju kraničnih listova paradajza definisana procedura koja uključuje saradnju sa proizvođačima, laboratorijsko snimanje i specijalno dizajnirane eksperimente za prikupljanje hiperspektralnih podataka.

Metodološki pristup zasniva se na razvoju nove arhitekture i modela za semantičku segmentaciju zasnovanih na dubokim neuronskim mrežama, koje su na prikupljenim podacima pokazale rezultate uporedive, a u pojedinim slučajevima i superiorne u odnosu na *state-of-the-art* arhitekture. Daljim razvojem istraživanja, fokus je proširen na segmentaciju na nivou instance, gde su korišćeni kompleksniji modeli dubokih neuronskih mreža, uključujući arhitekturu *Mask Region-based Convolutional Neural Network* (*Mask R-CNN*), koji su omogućili pre-

ciznu segmentaciju pojedinačnih klasova pšenice u RGB snimcima i kruničnih listova paradajza u hiperspektralnim snimcima.

Iste metode za semantičku segmentaciju su korišćene za dobijanje maski na laboratorijskom skupu podataka. Dobijene maske su zatim integrisane u proces adaptacije NeRF metode, omogućavajući trodimenzionalnu rekonstrukciju željenih segmenata pšenice, odn. klasova u prostoru. Na ovaj način značajno je olakšano razdvajanje preklopljenih biljnih struktura, koje su u 2D prostoru predstavljale gotovo nerešiv problem, čime je omogućen precizniji uvid u broj i distribuciju klasova, kao ključne parametre za finalnu procenu prinosa.

Rezultati ove disertacije pokazuju da integracija semantičke i segmentacije na nivou instance sa naprednim metodama rekonstrukcije 3D scena omogućava precizno razdvajanje složenih i preklapajućih regiona od interesa. Kreirani skupovi podataka i univerzalna arhitektura za semantičku segmentaciju pružaju pouzdanu osnovu za primenu u fenotipizaciji biljaka i proceni prinosa, naglašavajući važnost rešavanja ovakvih problema za dalji razvoj precizne poljoprivrede u domenu digitalne precizne poljoprivrede.

Ključne reči: semantička i segmentacija na nivou instance, hiperspektralne i RGB slike, duboko učenje, 3D rekonstrukcija (NeRF), fenotipizacija biljaka, precizna poljoprivreda.

Abstract

This doctoral dissertation addresses the problem of semantic segmentation of complex and overlapping structures across multiple imaging modalities with applications in precision agriculture. Particular emphasis is placed on segmenting wheat ears in thermal and RGB images, as well as segmenting tomato crown leaves in hyperspectral images, encompassing both pixel-level and instance-level approaches. These challenges represent a critical bottleneck in plant phenotyping and yield estimation due to the complex geometry of plant structures, their inherent homogeneity and variability, as well as the complexities of illumination and background field conditions.

A major difficulty was observed in the 2D segmentation of wheat ears, where overlapping structures and homogeneous morphological features impose significant limitations on achieving the highest accuracy levels (above 95%). To overcome these limitations, a methodology combining the state-of-the-art deep learning architectures with 3D scene reconstruction techniques was explored. A key element of this approach is Neural Radiance Field (NeRF), which enables three-dimensional reconstruction of targeted plant structures from a limited number of 2D images or video sequences, facilitating the separation of overlapping structures and the more reliable estimation of the number of wheat ears per unit area.

For this research, multiple datasets were created: (1) a multi-year field RGB dataset collected using smartphones and (2) a laboratory-created dataset specifically tailored for 3D reconstruction for wheat ears. Additionally, a dedicated procedural pipeline was developed involving collaboration with growers, controlled laboratory imaging, and specially designed experiments for hyperspectral data acquisition on tomato fruit.

The methodological approach is based on the development and application of neural networks for semantic segmentation, which achieve comparable and, in some cases, even superior performance relative to state-of-the-art architectures on the collected datasets. Further research extended the approach to instance segmentation, employing advanced deep neural network models, including Mask Region-based Convolutional Neural Networks (Mask R-CNN), which enabled the precise delineation of individual wheat ears in RGB images and tomato crown leaves in hyperspectral images.

The same methodology was applied to generate masks in the controlled laboratory conditions, which were subsequently integrated into a NeRF-based pipeline, enabling three-dimensional reconstruction of individual wheat ears. This significantly facilitated the separation of overlapping structures, which were nearly inseparable in 2D space, thereby allowing more accurate quantification and spatial distribution of wheat ears, critical parameters for the final yield estimation.

Overall, the results demonstrate that integrating semantic and instance segmentation with advanced 3D reconstruction techniques enables precise delineation of complex and overlapping regions of interest across multiple image modalities. The created datasets and universal architecture provide a robust foundation for plant phenotyping and yield estimation, highlighting the importance of addressing such challenges for the further development of precision agriculture and plant phenotyping technologies.

Keywords: semantic and instance segmentation, hyperspectral and RGB imaging, deep learning, 3D reconstruction, NeRF, plant phenotyping, precision agriculture.

Sadržaj

Popis slika	11
Popis tabela	13
1 Uvod	14
2 Metodologija	20
2.1 Metrike performansi za semantičku segmentaciju	23
2.1.1 Metrike performansi za segmentaciju na nivou instance	25
2.2 Hardverska infrastruktura	26
2.3 Analiza tradicionalnih metoda i motivacija za primenu složenijih modela	27
2.4 Semantička segmentacija	31
2.4.1 Pregled literature za semantičku segmentaciju klasova pšenice	32
2.4.2 Predlog arhitekture za semantičku segmentaciju klasova pšenice iz termalnih i RGB slika	35
2.4.3 Evaluacija predložene arhitekture na hiperspektralnim podacima	40
2.5 Segmentacija instanci	45
2.5.1 Pregled literature za segmentaciju na nivou instance u hiperspektralnim podacima	46
2.5.2 Izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama	51
2.5.3 Pregled literature za segmentaciju na nivou instance za detekciju klasa pšenice u RGB slikama	57
2.6 3D rekonstrukcija primenom NERF metode	63

2.6.1	Metrike i kvantitativna analiza NeRF 3D rekonstrukcije	67
3	Opis skupova podataka	70
3.1	Skup podataka za hiperspektralnu analizu morfoloških karakteristika krunicnih listova paradajza	71
3.2	Skup podataka za detekciju i brojanje klasova pšenice .	77
3.2.1	Skup podataka za semantičku segmentaciju klasova pšenice u termalnim i RGB slikama	77
3.2.2	Skup podataka za segmentaciju instanci klasova pšenice u RGB slikama	80
3.2.3	Procena kvaliteta slika u skupovima podataka za segmentaciju klasova pšenice	89
3.2.4	Preporuke i protokoli za optimalnu akviziciju slika	91
3.2.5	RGB skupovi podataka za 3D NeRF rekonstrukciju mikroparcele pšenice u kontrolisanim uslovima	92
4	Rezultati	96
4.1	Rezultati semantičke segmentacije	96
4.1.1	Studija slučaja I: Paradajz – kvantitativna i vizuelna analiza	96
4.1.2	Studija slučaja II: Pšenica – kvantitativna i vizuelna analiza	99
4.2	Rezultati segmentacije instanci	102
4.2.1	Studija slučaja I: Paradajz – kvantitativna i vizuelna analiza	102
4.2.2	Studija slučaja II: Pšenica – kvantitativna i vizuelna analiza	109
4.2.3	Detekcija klasova pšenice - analiza modela i stabilnost performansi	109
4.2.4	Automatsko brojanje klasova pšenice	117
4.3	Rekonstrukcija NERF metodama i analiza 3D reprezentacije	120
5	Diskusija	124
6	Zaključak	132
	Literatura	137

Spisak slika

2.1	(a) 8-bitna RGB reprezentacija slike; (b) segmentacija nakon primene filtriranja i Find Maxima algoritma. . .	28
2.2	Pseudokolor reprezentacija termalne slike.	28
2.3	Primena praga na osnovu vrednosti u histogramu slike.	30
2.4	Tok predložene metodologije prikazan korak po korak. .	36
2.5	Arhitektura nove razvijene neuronske mreže za semantičku segmentaciju, nazvana <i>Optimized Fully Convolutional Neural Network</i> , <i>oFCN</i> , primenjena na termalne i RGB slike	38
2.6	(a) Analiza potrošnje GPU memorije; (b) performanse modela u zavisnosti od primenjenih različitih aktivacionih funkcija. Rezultati su dobijeni nakon 10 epoha. . .	39
2.7	Izmereno potrebno vreme za proces obuke u zavisnosti od različitih aktivacionih funkcija. Rezultati su dobijeni nakon 10 epoha.	39
2.8	Razvoj gljivičnog oboljenja od 1. do 5. dana u RGB (vidljivom) spektru.	41
2.9	oFCN arhitektura primenjena na dva nivoa segmentacije: a) stabljika i krunični listovi paradajza b) model fokusiran na segmentaciju samo kruničnih listova paradajza.	43
2.10	U-Net arhitektura	44
2.11	Seg-Net arhitektura	45
2.12	Tok obrade hiperspektralnih slika korak po korak u procesu segmentacije na nivou instanci	56
2.13	Primer slika po izvorima podataka	60
2.14	Prikaz 3D rekonstrukcije scene NeRF metodom	64

3.1	a) Orijentacija plodova paradajza tokom hiperspektralnog snimanja. (b) Pseudokolor slika jednog grozda/grane dobijena kombinovanjem tri spektralna kanala (Slika preuzeta iz [1])	72
3.2	RGB vizuelizacija primera sve četiri sorte paradajza iz drugog skupa hiperspektralnih podataka	75
3.3	Hiperspektralni odzivi plodova i kruničnih listova paradajza na primerima iz a) prvog i b) drugog skupa hiperspektralnih podataka	76
3.4	Varijabilnost visine, širine i površine kruničnih listova paradajza u oba skupa hiperspektralnih podataka po sortama paradajza	76
3.5	Levo: RGB reprezentacija slike sa terena; Desno: pseudokolor reprezentacija termalne slike sa terena	78
3.6	Prvi red: desno – RGB originalna reprezentacija slike sa terena; levo – ručno anotirana binarna maska. Donji red: desno – primena rotacije nad RGB originalnom slikom pod uglom od 30 stepeni; levo – primena rotacije nad ručno anotiranom binarnom maskom pod uglom od 30 stepeni.	79
3.7	Dizajn eksperimenta sa parcelama numerisanim prema različitim količinama tretmana azotom, praćen koracima predobrade za izdvajanje regiona od interesa iz sirovih podataka.	84
3.8	Primer kompleksnih slučajeva u procesu anotacije u okviru GWHD skupa podataka, uključujući preklopljene klasove pšenice, pogrešno spojene ili razdvojene labele, delimično isečene objekte na ivicama slike, kao i snimke sa smanjenim osvetljenjem.	86
3.9	Vizuelizacija predložene metoda anotiranja koji uspešno obrađuje gustu raspodelu i preklopljene instance. Ovaj pristup ublažava izazove anotacije efikasnim razdvajanjem grupisanih klasova pšenice.	87
3.10	Procena kvaliteta slika: NIQE	90
3.11	Laboratorijski skupovi podataka: (a) uniformna pozadina, (b) neuniformna pozadina.	93

3.12	Laboratorijski skupovi podataka: (a) Primer RGB slike sa neuniformnom pozadinom, (b) Manuelno kreirane labele u formi poligona c) Binarna maska kreirana kreirana na osnovu poligona d) UNet generisana maska	95
4.1	Prikaz trodimenzionalne reprezentacije hiperspektralnih slika paradajza, ručno anotiranih maski, i rezultata oFCN modela	98
4.2	Prikazan opseg vrednosti Dice koeficijenta po batch-u, kao i njegova ukupna prosečna vrednost.	99
4.3	(a) Originalna termalna slika; (b) ručno anotirana binarna maska; (c) izlaz modela; (d) primenjeni prag od 0,8 na izlaz.	100
4.4	(a) Originalna RGB slika; (b) ručno izrađena binarna maska; (c) izlaz modela; (d) primenjeni prag od 0,8 na izlaz.	101
4.5	Kodovane spektralne informacije jedne slike iz skupa podataka 1, unutar trodimenzionalnog podprostora korišćenjem: a) PCA, b) ICA, c) PPCA, i d) NMF tehnike, prikazano kroz trokanalnu pseudokolor sliku.	104
4.6	Detektovane instance kruničnih listova paradajza pomoću NMFMaskFPN modela koristeći: a) PCA b) ICA c) PPCA d) NMF tehnike za izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama (prvi red), i pripadajuće binarne maske (drugi red).	106
4.7	Analiza performansi Mask R-CNN modela kroz epohe sa različitim backbone-ovima (FPN, DC5, C4) evaluirano na validacionim hiperspektralnim slikama kodovanim NMF tehnikom.	106
4.8	Rezultati detekcije lista paradajza na četiri sorte iz drugog HSI skupa: a) <i>Brioso</i> b) <i>Capricia</i> c) <i>Provine</i> d) <i>Sao Paolo</i> sa NMFMaskFPN	108
4.9	Levo: primeri slika iz različitih izvora sa anotiranim brojem klasova pšenice. Desno: prosečna zastupljenost svakog izvora u skupu kroz 5-fold eksperimente.	109
4.10	Detekcija klasova pšenice na test slikama pomoću RetinaNet R101 iz objedinjene baze podataka; levo primer iz BioS-Wheat skupa, desno iz GWHD skupa	114

4.11	KDE grafici koji prikazuju gustinu rezultata detekcije za svaku sortu u BioS-Wheat skupu podataka, naglašavajući razlike u performansama modela u odnosu na stvaran broj klasova.	118
4.12	Evaluacija performansi zasnovana na poverenju. Gornji red: krive <i>Precision-Confidence</i> (levo) i <i>Recall-Confidence</i> (desno), Donji red: krive <i>Precision-Recall</i> (levo) i <i>F1 score-Confidence</i> (desno)	119
4.13	a) originalni RGB frejm b) originalni RGB frejm sa uklonjenom pozadinom, primenom transformacije u drugi prostor boja HSV i određenog praga za boju c) segmentovani klasovi na originalnom RGB frejmu d) segmentovani klasovi na originalnom frejmu sa uklonjenom pozadinom	121
5.1	Normalizovani koeficijenti dobijeni NMF i PPCA metodom iz prvog HSI skupa sa tri glavne komponente. Dominantni maksimum u komponenti 3, dobijen NMF-om, primećen je na 1398,6 nm, preklapajući se sa regionom 1390–1420 nm (označeno), koji je prethodno [1] identifikovan kao značajan za diferencijaciju vezanu za količinu i posledično manjak vode u kruničnim listovima	126
5.2	Dominantne talasne dužine računatih korišćenjem ground-truth maski i maski dobijenih od ConvMaskFPN modela na test skupovima (a) prvog i (b) drugog HSI skupa podataka	127

Spisak tabela

2.1	Korišćene metode i njihove prosečne performanse na test skupu RGB slika	29
3.1	Pregled prvog hiperspektralnog skupa podataka	72
3.2	Tabelarni prikaz hiperspektralnih podataka	76
3.3	Proširena tabela iz [2] sa opisom BioS-Wheat skupa podataka [3]	82
3.4	BioS-Wheat skup podataka: broj slika i broj klasova pšenice po sorti.	83
3.5	Uporedni pregled skupova podataka GWHD i BioS-Wheat korišćenih u istraživanju	88
3.6	Procena kvaliteta slika — GWHD i BioS-Wheat skupovi podataka	91
4.1	Poređenje postignutih rezultata sve tri arhitekture i broj epoha	97
4.2	Pregled poređenja predložene arhitekture sa dve najčešće korišćene arhitekture za rešavanje problema segmentacije	97
4.3	Rezultati segmentacije - termalne slike	99
4.4	Rezultati segmentacije - RGB slike	100
4.5	Rezultati brojanja klasova na termalnim i RGB slikama	101
4.6	Za svaku tehniku latentnog izdvajanja spektralnih informacija prikazana je suma standardnih devijacija svih koeficijenata, estimirana projekcijom na trodimenzionalni potprostor, kroz sve foldove.	103

4.7	Srednja vrednost i standardna devijacija mAP kroz 6 validacionih foldova za tri varijante Mask R-CNN modela (FPN, C4, i DC5) i korišćene tehnike izdvajanja latentnih hiperspektralnih informacija (PCA, NMF, ICA, PPCA)	105
4.8	Greške predikcije $\bar{\mathbf{e}}/\underline{\mathbf{e}}$ za krunične liste paradajza na test skupu, izražene u procentima	105
4.9	Srednja vrednost i standardna devijacija mAP za NMF-MaskFPN model kroz 8-fold kros-validaciju	107
4.10	mAP dobijen inicijalizacijom ConvMaskFPN sloja koeficijentima iz tehnika: XavierN, KaimingN, Uniform i Normal	108
4.11	mAP dobijen inicijalizacijom ConvMaskFPN sloja koeficijentima iz korišćenih tehnika izdvajanja latentnih hiperspektralnih informacija	108
4.12	GWHD skup — rezultati	111
4.13	BioS-Wheat—rezultati	112
4.14	BioS-Wheat + GWHD skup — rezultati	113
4.15	Relativno poređenje performansi modela treniranih na objedinjenom skupu podataka u odnosu na osnovni GWHD skup.	114
4.16	Prosečne vrednosti mAP50 ostvarene RT-DETR modelom po izvorima podataka	116
4.17	Evaluacija robusnosti modela pod različitim perturbacijama slike. mAP_{50} je istaknuta kao glavna metrika.	116
4.18	Korelacije između označenog i detektovanog broja klasova u BioS-Wheat skupu podataka	117
4.19	Korelacije između stvarnog i detektovanog broja klasova po sortama pšenice iz BioS-Wheat skupa podataka	118
4.20	Kvantitativna evaluacija kvaliteta rekonstruisanog NeRF videa.	122

1 Uvod

Segmentacija slike predstavlja fundamentalni problem u oblasti obrade slike i kompjuterske vizije. Metode analize slike koje omogućavaju preciznu segmentaciju regiona od interesa predstavljaju temelj savremenih istraživanja. Semantička segmentacija i segmentacija na nivou instanci obezbeđuju metodološki okvir za prepoznavanje kompleksnih i heterogenih regiona u različitim modalitetima slika.

Algoritam segmentacije slika ima ključnu ulogu u obradi slika i računarstvu jer omogućava podelu slike na delove ili segmente koji sadrže slične osobine (kao što su boja, intenzitet, tekstura ili oblik). Ovaj proces je od suštinskog značaja za mnoge zadatke u oblasti mašinske vizije jer omogućava identifikaciju objekata, analizu scena i razumevanje sadržaja slike. Algoritmi segmentacije slika imaju ogromnu vrednost i široku primenu u različitim industrijama i naučnim disciplinama. Primena algoritama segmentacije slika je široka: u medicinskoj obradi slika (npr. MRI, CT skeneri, ultrazvuk) segmentacija omogućava automatsko prepoznavanje i izdvajanje struktura kao što su tumori, organi, krvni sudovi ili lezije. Ovo može značajno ubrzati dijagnostičke procese. Kod autonomnih vozila, segmentacija slika je ključna za sisteme koji omogućavaju autonomnim vozilima da prepoznaju objekte na putu, kao što su druga vozila, saobraćajni znakovi ili prepreke, ali i pešaci. U robotici, segmentacija je bitna za navigaciju i manipulaciju objektima. Roboti koriste segmentaciju kako bi identifikovali prepreke u svom okruženju, kao i kako bi pravilno rukovali objektima tokom obavljanja zadataka. U sistemima za video nadzor segmentacija slika pomaže u prepoznavanju aktivnosti i anomalija u okruženju. Identifikovanje neobičnih objekata ili pokreta može omogućiti brzu reakciju na potencijalnu pretnju. Segmentacija satelitskih i vazduhoplovnih snimaka koristi se za prepoznavanje zemljinih pokrivača, kao što su šume, vodene površine, urbane oblasti ili poljoprivredne površine. Ovi podaci su ključni za praćenje promena u životnoj sredini i upravljanje resursima. U ana-

lizi slika segmentacija se koristi za izdvajanje ključnih objekata, koji se zatim mogu dalje analizirati ili povezivati s tekstualnim informacijama u aplikacijama poput prepoznavanja slika ili automatskog generisanja opisa slika. U industriji segmentacija se koristi u automatskoj inspekciji proizvoda. Na primer, u proizvodnoj liniji može se koristiti za detekciju nepravilnosti ili anomalija na proizvodima, poput ogrebotina, nepravilnosti u obliku ili nesavršenosti u kvalitetu. Dakle, od medicine i autonomnih vozila do robotike i video nadzora, segmentacija je osnovni korak za razumevanje i interpretaciju vizuelnih podataka. Tačnost i efikasnost ovog procesa direktno utiču na performanse aplikacija koje zavise od prepoznavanja i analize objekata. Dalji razvoj i unapređenje algoritama segmentacije, kao i primena naprednih tehnika kao što su duboko učenje i veštačka inteligencija, obećavaju dodatna poboljšanja u ovoj oblasti, omogućavajući još precizniju, bržu i pouzdanu analizu slika.

Precizna poljoprivreda je još jedna oblast u kojoj kompjuterska vizija donosi unapređenja. Omogućava automatizovano praćenje stanja useva, detekciju bolesti, plodova i korova, kao i procenu kvaliteta useva.

Precizna poljoprivreda predstavlja savremeni pristup upravljanja poljoprivrednim resursima, zasnovan na primeni naprednih tehnologija, kao što su algoritmi kompjuterske vizije i mašinskog učenja, različiti tipovi senzora i sistemi za obradu podataka. Njena osnovna svrha je povećanje efikasnosti proizvodnje, smanjenje troškova i očuvanje životne sredine. Razvoj algoritama mašinskog učenja omogućava automatsko prepoznavanje regiona i objekata od interesa, kao i donošenje odluka u realnom vremenu, što je od suštinskog značaja za optimizaciju poljoprivrednih procesa i bolje upravljanje resursima. Posebno, oblast oplemenjivanja biljaka (engl. *plant breeding*) i analiza procene kvaliteta proizvoda nakon berbe (engl. *post-harvest*) predstavljaju dva veoma važna segmenta savremene poljoprivredne nauke, koja otvaraju širok spektar istraživačkih mogućnosti u pogledu precizne segmentacije regiona od interesa na slikama različitih biljnih kultura.

U oblasti oplemenjivanja biljaka, jedan od najznačajnijih izazova predstavlja pouzdana detekcija i prebrojavanje klasova pšenice, što je jedan od ključnih parametara za procenu prinosa pšenice, analizu genotipskih razlika i unapređenje procesa za selekciju. S druge strane, nakon berbe, detekcija plodova i listova ima presudnu ulogu u proceni kvaliteta, identifikaciji potencijalnih oštećenja i optimizaciji procesa skladištenja, prerade i transporta. U oba slučaja, algoritmi za segmen-

taciju i detekciju regiona od interesa predstavljaju osnovni tehnološki instrument, omogućavajući automatizovanu, preciznu i efikasnu analizu biljnih struktura u različitim fazama proizvodnog ciklusa.

Kombinacija senzora i kamera sa softverskim rešenjima pruža podatke i informacije koji prevazilaze mogućnosti tradicionalnih metoda, omogućavajući rano otkrivanje pojave stresa biljaka, nutritivnih nedostataka i pojave bolesti ili štetočina, što direktno utiče na povećanje prinosa i efikasnost proizvodnje. Za potrebe precizne detekcije i segmentacije regiona od interesa koriste se različite senzorske tehnologije koje omogućavaju prikupljanje podataka sa spektralnim i prostornim karakteristikama. RGB kamere su najrasprostranjenije zbog jednostavnosti i primenljivosti u realnim uslovima, dok multispektralne i hiperspektralne kamere omogućavaju snimanje u desetinama i stotinama spektralnih opsega, što u oblasti precizne poljoprivrede omogućava detaljnu analizu fiziološkog stanja biljaka i identifikaciju karakteristika nevidljivih u vidljivom spektru. Dodatno, termalne kamere omogućavaju praćenje temperaturnih promena na površini biljaka i plodova, dok RGB + *Depth* sistemi pružaju trodimenzionalne podatke korisne za procenu morfoloških osobina i prostornog rasporeda objekata.

Uprkos tehnološkom napretku, kreiranje potrebnih specifično domenskih i kompleksnih skupova podataka predstavlja značajan izazov. Laboratorijski uslovi omogućavaju kontrolu osvetljaja i položaja objekata, ali ograničavaju prirodnu varijabilnost biljaka i okruženja. Sa druge strane, uslovi na terenu pružaju realističnije i kompleksnije uslove kao što su promenljivo osvetljenje, pokreti biljaka usled vetra i nepristupačan teren. Prikupljanje podataka zahteva interdisciplinarnu saradnju između proizvođača, vlasnika parcela, agronoma i istraživača inženjera iz oblasti obrade podataka, pažljivo osmišljen eksperimentalni dizajn koji obuhvata različite sorte, nivoe đubrenja, faze rasta i način obrade, kao i izbor opreme i postavki za snimanje. Proces snimanja uključuje kalibraciju opreme i višefazno praćenje eksperimenata, zbog čega su kvalitetni i namenski kreirani skupovi podataka u oblasti precizne poljoprivrede retki, skupi i teško dostupni, dok javno dostupne baze podataka često ne zadovoljavaju specifične zahteve za širu primenu i naučna istraživanja u poljoprivredi.

Glavna motivacija ove doktorske disertacije se ogleda u potrebi za razvijanjem efikasnih metoda segmentacije regiona od interesa u složenim, višemodalnim i realnim skupovima podataka u domenu precizne poljoprivrede. Problemi koji se javljaju prilikom transfera znanja pret-

hodno obučениh postojećih modela na specifične poljoprivredne zadatke kao što su segmentacija klasova pšenice u RGB slikama ili kručnih listova plodova paradajza u hiperspektralnim slikama zahtevaju razvoj novih neuronskih mreža koje su prilagođene veličini i kvalitetu dostupnih podataka. Količina slika potrebna za doobučavanje postojećih modela je često nedovoljna, a kompleksnost mreža u smislu broja parametara koji figurišu prilikom obučavanja modela, ograničava njihovu primenu u realnom vremenu i na uređajima smanjenih računarskih sposobnosti, tzv. *edge* uređajima (npr. pametnim telefonima, pametnim satovima i drugim nosivim uređajima, IoT pametnim uređajima i dr.).

Imajući u vidu ove izazove, doprinos ove disertacije se ogleda u tri ključna aspekta:

1. Prikupljanje kompleksnih i namenski kreiranih skupova podataka u laboratorijskim i terenskim uslovima, za više biljnih kultura i modaliteta akvizicije podataka (termalne, RGB, hiperspektralne slike).
2. Razvoj nove arhitekture neuronske mreže za semantičku segmentaciju regiona od interesa po nazivom *Optimized Fully Convolutional Neural Network (oFCN)* koja pokazuje dobre rezultate na slikama različitih modaliteta snimanja i zadacima detekcije regiona od interesa. Mreža je robusna i primenljiva na više skupova podataka i modaliteta, i postiže kompromis u postignutoj tačnosti i nižoj kompleksnosti arhitekture, kako bi se omogućila implementacija na *edge* uređajima i obradu u realnom vremenu. Pokazano je njeno uspešno korišćenje na studijama slučaja za detekciju i segmentaciju klasova pšenice u termalnim i RGB slikama i kručnih listova paradajza u hiperspektralnim slikama. Razvijena arhitektura kombinuje jednostavnost neophodnu za primenu u realnom vremenu sa dovoljno kompleksnosti da registruje složene strukture regiona od interesa u preciznoj poljoprivredi. Dodatno, proširenje istraživanja omogućilo je transfer znanja modela za segmentaciju regiona od interesa na nivou instance i primenu ovakvih metoda na različite zadatke. Ovaj doprinos se ogleda u softverskom rešenju koje je primenljivo i skalabilno, uz visok nivo preciznosti, što predstavlja značajan korak ka automatizaciji i unapređenju procesa u preciznoj poljoprivredi.

3. Integracija predloga *end-to-end* metode koja obuhvata izdvajanje latentnih spektralnih komponenti hiperspektralnog opsega i redukciju dimenzionalnosti hiperspektralnih slika, čuvajući najveći stepen korisnih informacija i primenu postojećih modela za detekciju i segmentaciju regiona od interesa na nivou instance, demonstrirana je na skupu hiperspektralnih podataka.

Jedan od ciljeva istraživanja je i predlog i razvoj metode za automatsko brojanje klasova pšenice na podacima u realnim uslovima, uz što jednostavniji način korišćenja, bez potrebe za montažom kamera na vozila i slične zahtevne opreme. Predviđeno je da se slike pšenice snimaju ručno, korišćenjem kamera ugrađenih u pametne mobilne telefone. Imajući u vidu sve širu primenu mobilnih telefona i konstantan razvoj njihovih funkcionalnosti i dodataka, u budućnosti se očekuje sve više rešenja baziranih upravo na mobilnim uređajima. Ovakav sistem omogućava poljoprivrednicima da jednostavno fotografišu usev i trenutno dobiju informaciju o broju prebrojanih klasova po posmatranoj površini, što je osnov za dalju procenu prinosa pšenice. Veći broj slika koje korisnici naprave omogućava obuhvatanje različitih gustina klasova na parceli i samim tim razvoj modela sa visokom preciznošću na nivou cele parcele. Motivacija za ovu studiju proističe iz činjenice da je broj klasova po jedinici površine jedna od ključnih agronomskih komponenti prinosa. Njegova brža procena doprinosi efikasnijem praćenju agrotehničkih mera, ranoj predikciji prinosa, ali i selekciji u oplemenjivačkim programima. Razvoj ove inovativne metodologije opisan je u podnetoj patentnoj prijavi na nacionalnom nivou. Patentna prijava je objavljena u Glasniku intelektualne svojine [4], a registracija patenta na nacionalnom nivou se očekuje po okončanju propisane procedure. Ovim se potvrđuje inovativni karakter predloženog rešenja i njegov potencijal za praktičnu primenu.

Ciljevi ove doktorske disertacije ističu važnost razvoja i primene algoritama kompjuterske vizije i mašinskog učenja u oblasti precizne poljoprivrede, ističući ključne izazove u segmentaciji regiona od interesa i prikupljanju višemodalnih podataka. Ovaj rad postavlja temelje za dalja istraživanja i praktične primene u oblasti oplemenjivanja biljaka i analize kvaliteta nakon berbe, pokazujući potencijal novih tehnoloških rešenja za povećanje efikasnosti i održivosti poljoprivrede.

U okviru ove doktorske disertacije, istraživanje se usmerava na problem semantičke segmentacije regiona od interesa u slikama različitih modaliteta snimanja, korišćenjem savremenih metoda dubokog učenja.

Semantička segmentacija podrazumeva proces identifikacije i klasifikacije pojedinačnih objekata ili regiona unutar slike prema njihovom značenju i funkcionalnoj ulozi u posmatranom sistemu. Primena dubokih neuronskih mreža omogućava automatsko učenje i prepoznavanje kompleksnih regiona u slikama, čime se ostvaruje detekcija i klasifikacija objekata od interesa, kao i naknadna predikcija relevantnih parametara.

Cilj istraživanja jeste razvoj i optimizacija modela dubokih neuronskih mreža sposobnih da pouzdano analiziraju slike različitih tipova senzora: termalne, RGB i hiperspektralne i da omoguće preciznu segmentaciju objekata kao što su pojedinačne biljke, listovi, plodovi ili klasovi pšenice. Time se omogućava tačna, pravovremena i automatska analiza važnih agronomskih parametara koji su ključni za donošenje odluka u okviru precizne poljoprivrede. Glavni istraživački problem obuhvata više međusobno povezanih izazova: varijabilnost uslova osvetljenja i senki u realnim uslovima na terenu koja otežava konzistentnu segmentaciju; preklapanje i kompleksna geometrija objekata, kao što su listovi, plodovi ili klasovi, što dovodi do gubitka jasnih granica između regiona; visoka dimenzionalnost hiperspektralnih podataka koja zahteva efikasnu redukciju dimenzionalnosti i selekciju reprezentativnih spektralnih karakteristika; razlike u rezoluciji i udaljenosti senzora tokom akvizicije, koje utiču na kvalitet i interpretaciju podataka. Zbog navedenih faktora, segmentacija regiona od interesa u preciznoj poljoprivredi predstavlja složen, multidisciplinarni istraživački problem, koji zahteva integraciju metoda obrade i analize slike, ekstrakcije željenih karakteristika i naprednih algoritama dubokog učenja. Rešenje ovog problema direktno doprinosi povećanju efikasnosti automatizovanih sistema za monitoring i evaluaciju stanja biljaka, što ima značajan uticaj na unapređenje procesa u preciznoj poljoprivredi.

U nastavku doktorske disertacije je detaljno opisana metodologija istraživanja, uključujući razvijene i korišćene algoritme, eksperimentalni dizajn i skupove podataka prikupljene za svaku od sprovedenih studija. Nakon toga, izloženi su dobijeni rezultati i analiza. Posebno poglavlje posvećeno je diskusiji rezultata, u okviru koje se razmatraju njihove implikacije i ograničenja. Na kraju, u zaključnom poglavlju sumirani su glavni doprinosi disertacije i dati pravci za buduća istraživanja.

2 Metodologija

Segmentacija slike predstavlja proces u kome se izvršava podela slike na regione ili objekte od interesa. Suština tradicionalnih algoritma za segmentaciju slike je bazirana na jednom od dva principa: diskontinuitet i sličnost. Princip diskontinuiteta je princip podele slike na osnovu nagle promene intenziteta, odnosno detekcije ivice regiona. Drugi princip se bazira na određivanju sličnosti regiona, po unapred zadatom kriterijumu. U ovu kategoriju spada metoda određivanja praga (engl. *threshold*) rast regiona, razdvajanje i spajanje regiona. Osnovna razlika u pristupima je da, metode koje detektuju ivice regiona ne odvajaju region od pozadine zatvorenom krivom, nego isprekidanom, te zbog toga zahtevaju primenu dodatnih metoda za spajanje isprekidane granice regiona, kao što su: fitovanje krive, Hafova transformacija ili heurističke metode [5]. Pravila u procesu segmentacije slike su: uniformni i homogeni regioni u pogledu neke karakteriske, značajno različiti susedni regioni u pogledu neke karakteristike, kao i jasne, jednostavne, glatke i prostorno tačne granice svakog regiona. Ako sliku označimo kao $\mathbf{I}$, segmentaciju možemo definisati kao proces pri kome razdvajamo sliku na njene regione $R_1, R_2, \dots, R_n$, gde $Q(R_i)$ predstavlja logički uslov koji je zajednički pikselima unutar istog regiona R_i i definisan je nad skupom R_i , [5]:

a) $\bigcup_{i=1}^n R_i = \mathbf{I}$

b) $R_i \cap R_j = \emptyset$ za $\forall i \neq j$

c) $Q(R_i) = 1$ za $i = 0, 1, 2, \dots, n$

d) $Q(R_i \cup R_j) = 0$ za sve susedne regione R_i i R_j

e) R_i je povezan skup za $i = 0, 1, 2, \dots, n^1$

¹Ovaj uslov nije neophodan i ponekad se izostavlja.

Uslovi nalažu da celokupnu sliku definiše skup svih tačaka tj. piksela unutar svih pojedinačnih regiona. Iskaz a) predstavlja matematički iskaz koji govori da je segmentacija kompletna ukoliko svaki piksel pripada određenom posmatranom regionu. Iskaz b) kaže da različiti regioni moraju biti razdvojeni, tj. da nema međusobnog preklapanja. Iskaz c) zahteva predhodno definisan kriterijum povezanosti piksela (sparam njihovih vrednosti, susedstva ili nekog drugog kriterijuma) unutar jednog regiona, odnosno da pikseli istih regiona imaju iste karakteristike po nekom prethodno utvrđenom kriterijumu. Iskaz d) kaže da pikseli koji pripadaju različitim regionima moraju da se razlikuju po unapred zadatom kriterijumu. Iskaz e) govori o tome da pikseli unutar istog regiona mogu biti povezani, ali i ne moraju [5].

Postoje brojne matematičke formulacije segmentacije i izbor odgovarajuće metode koja bi dala zadovoljavajuće rezultate u svim uslovima u realnom okruženju može biti izuzetno zahtevan [6].

U ovoj doktorskoj disertaciji, problem segmentacije je predstavljen kroz detekciju klasova pšenice u slikama različitih modaliteta snimanja i perspektive (RGB slike), kao i detekcija kruničnih listova paradajza (na hiperspektralnim slikama). U problemu koji se analizira u okviru ove doktorske disertacije podaci su slike koje se prikupljaju kamerom na terenu ili u laboratoriji. Samim tim 3D stvarnost predstavljena je u 2D prostoru, pa je okluzija u slici željenih regiona od interesa veoma česta pojava (preklapanje klasova u slici). Iako daleko od trivijalnog problema, ovaj problem ne pripada grupi nerešivih problema [7].

Vrlo je izazovno odrediti koja metoda za segmentaciju je u potpunosti odgovarajuća ili bolja, u odnosu na druge. Za najbolju metodu se proglašava ona koja postiže kompromis u analizi kompleksnosti modela, ali i postignutim rezultatima na podacima i problemu koji se rešava [8]. Neuronske mreže kao izbor grupe algoritama koji se analizira u ovoj doktorskoj disertaciji, izabran je zbog toga što primena drugih algoritama za segmentaciju nailazi na nekoliko problema koje metoda neuronskih mreža rešava ili ih nema. Osim toga, u poređenju sa tradicionalnim tehnikama za segmentaciju slike, neuronske mreže postižu veću preciznost prilikom rešavanja problema klasifikacije slika, semantičke segmentacije ili detekcije objekata [9]. Standardni algoritmi često zahtevaju neku vrstu predobrade slike nad originalnim sirovim podacima: npr. pretvaranje u traženu formu, različite veličine binarnog zapisa (8-bit, 16-bit, 32-bit), korekciju osvetljaja, kontrasta, itd. Dodatnu univerzalniju primenu i repetitivnost otežava i podešavanje adekvatnih

parametara koji zahtevaju različite vrednosti na različitim modalitetima slika. Mogućnost univerzalnog podešavanja da ove metode daju zadovoljavajuće rezultate na više različitih modaliteta snimanja i zadataka je skoro nemoguće postići jer je potrebno dosta vremena da se odaberu pravi pristupi i pronađu adekvatna diskriminatorna obeležja, a vrednosti parametara koji se primenjuju najčešće su određeni heurističkom metodom. To dalje rezultira da tradicionalni algoritmi nisu univerzalno primenjivi na sve modalitete slika koji se sagledavaju u ovoj doktorskoj tezi. Analiza rezultata dobijenih ovim tehnikama data je u poglavlju 4. Teorijski, u ovom slučaju trebalo bi uspostaviti više algoritamskih rešenja za svaki od modaliteta slika ili grupe klasa – npr. klasovi su različite veličine u slikama, različite sorte imaju drugačije morfološke karakteristike. Osim toga, u poslednjih 5 godina aktuelno stanje u oblasti segmentacije slika bazira se na različitim arhitekturama neuronskih mreža, zbog značajno boljih performansi i odsustvu potrebe za zahtevnim inženjeringom obeležja.

Ovo poglavlje predstavlja metodološki okvir istraživanja, usmeren na pregled relevantnih metoda i metodoloških pristupa, razvoj i implementaciju naprednih pristupa za segmentaciju regiona od interesa u slikama različitih senzorskih modaliteta. Polazna osnova obuhvata pregled tradicionalnih metoda kompjuterske vizije, koje se zasnivaju na tehnikama obrade slike i ručno definisanim deskriptorima karakteristika. Ove metode služe kao referentna tačka za razumevanje ograničenja klasičnih pristupa i motivišu prelazak ka savremenim modelima zasnovanim na dubokom učenju.

Centralni deo poglavlja fokusira se na metodološki dizajn arhitekture za semantičku segmentaciju na nivou piksela i segmentaciju na nivou instance, pri čemu se razmatraju različite arhitekture neuronskih mreža, uz detaljan pregled relevantnih metoda. Poseban akcenat stavljen je na integrisani pristup koji povezuje različite tipove segmentacije radi postizanja veće preciznosti u detekciji i prostornoj lokalizaciji objekata. Na taj način, metodologija obezbeđuje sveobuhvatnu osnovu za analizu kompleksnih vizuelnih scena i pouzdanu interpretaciju struktura od interesa.

Razvijeni pristupi biće demonstrirani i validirani kroz dve reprezentativne studije slučaja koje čine empirijsku osnovu disertacije. Prva studija obuhvata detekciju i klasifikaciju klasova pšenice na multisenzorskim skupovima podataka koji uključuju termalne i RGB slike u 2D i 3D prostoru, dok se druga odnosi na segmentaciju kruničnih

listova paradajza u hiperspektralnim snimcima. Ovakva kombinacija omogućava evaluaciju razvijene metodologije u različitim uslovima — od standardnih do spektralno i geometrijski kompleksnih scena — čime se potvrđuje njena robusnost, skalabilnost i opšta primenljivost u kontekstu precizne poljoprivrede i analize biljnih fenotipova.

U ovom poglavlju izložen je metodološki pristup istraživanja, zasnovan na dve studije slučaja koje obuhvataju analizu podataka prikupljenih nad usevima pšenice i paradajza. Segmentacija slike se u savremenoj literaturi najčešće deli na semantičku segmentaciju, segmentaciju instanci i panoptičku segmentaciju, u zavisnosti od nivoa detalja i cilja analize. Semantička segmentacija podrazumeva dodeljivanje svakom pikselu slike pripadnost klasi, pri čemu svi objekti iste klase dele istu oznaku, bez međusobnog razlikovanja pojedinačnih instanci. Nasuprot tome, segmentacija instanci omogućava identifikaciju i razdvajanje pojedinačnih objekata unutar iste klase, kombinujući lokalizaciju i precizno pikselno razgraničenje. Panoptička segmentacija objedinjuje ova dva pristupa, pružajući jedinstven okvir koji istovremeno obezbeđuje klasifikaciju svih piksela i razlikovanje pojedinačnih instanci objekata u sceni. U ovoj doktorskoj disertaciji za oba skupa podataka razmatrani su zadaci semantičke segmentacije i segmentacije instanci, čime je omogućena sveobuhvatna evaluacija predloženih metoda u različitim modalitetima slika.

2.1 Metrike performansi za semantičku segmentaciju

Slično kao u radovima [10, 11, 12], za semantičku segmentaciju se koriste standardne mere za evaluaciju rezultata:

1. *True Positive* (TP) predstavlja piksel koji je ispravno klasifikovan
2. *False Positive* (FP) označava piksel pozadine koji je pogrešno klasifikovan uključujući i višestruke detekcije istog piksela.
3. *False Negative* (FN) predstavlja piksel koji je propušten i pogrešno klasifikovan kao pozadina.
4. Preciznost (engl. *Precision*) predstavlja meru ispravno detekto-

vanih uzoraka u odnosu na sve detektovane uzorke:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.1)$$

5. Recall predstavlja meru detektovanih piksela koji pripadaju klasama klasifikovanim kao klase od interesa (npr. klas pšenice ili krunični list paradajza):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.2)$$

6. Tačnost (engl. *Accuracy*) kao mera ukupne performanse modela:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.3)$$

7. F1 mera (engl. *Sørensen-Dice* indeks, Dice koeficijent sličnosti) predstavlja harmonijsku sredinu preciznosti i odziva:

$$F1_{\text{score}} = 2 \cdot \frac{\text{Preciznost} \cdot \text{Odziv}}{\text{Preciznost} + \text{Odziv}} \quad (2.4)$$

Za prikaz odnosa između modelom predviđene vrednosti i stvarne vrednosti izlazne promenljive koristi se relativna greška. Za sliku i , relativna greška definiše se kao:

$$\text{Error}(i) = \frac{|y_p(i) - y_t(i)|}{y_t(i)} \quad (2.5)$$

Na primer, u gornjem izrazu, za studiju o detekciji i prebrojavanju klasova pšenice, y_p predstavlja automatski izbrojan broj klasova na slici i , dok y_t predstavlja ručno izbrojan (stvarni) broj klasova. Tačnost po slici i računa se kao:

$$\text{Accuracy}(i) = \begin{cases} \frac{y_p(i)}{y_t(i)}, & \text{ako } y_p(i) \leq y_t(i), \\ 2 - \frac{y_p(i)}{y_t(i)}, & \text{inače.} \end{cases} \quad (2.6)$$

Ova definicija pokriva oba slučaja i osigurava da vrednost preciznosti ostane između 0 i 100%. U slučaju da relativna greška ima vrednost

veću od 1, to znači da je predviđena vrednost previše udaljena od referentne vrednosti [13]. Takođe, za procenu linearne povezanosti između stvarnih i predikovanih vrednosti uobičajeno je primeniti Pearson-ov koeficijent korelacije i koeficijent determinacije, koji je jednak kvadratu Pearson-ovog koeficijenta. Ove mere su korišćene u [11]. Pearson-ov koeficijent daje informaciju o regularnosti odnosa između vrednosti podataka, dok je kvadrat Pearson-ovog koeficijenta koeficijent determinacije. Minimalna vrednost ovog koeficijenta koja se u praksi smatra prihvatljivom je 0,25 [14].

2.1.1 Metrike performansi za segmentaciju na nivou instance

Najčešće korišćene metrike za ocenu performansi dubokih neuronskih mreža u zadacima segmentacije instanci su *Intersection over Union (IoU)* i njegove varijante, zajedno sa *Average Precision (AP)*. Za evaluaciju modela za segmentaciju instanci, IoU je korišćen kao međukorak za izračunavanje mAP, koji je definisan kao prosek vrednosti AP izračunatih po klasama i definisan je na sledeći način:

$$\text{mAP} = \frac{1}{n} \sum_i^n \text{AP}_i \quad (2.7)$$

gde n predstavlja broj klasa u posmatranom problemu detekcije objekata. U ovoj studiji, mAP je označen kao mAP₅₀, jer je korišćen kao prag od 50% preklapanja za IoU, slično kao u [2].

U ovoj doktorskoj disertaciji, za procenu performansi modela za segmentaciju kraničnih listova paradajza i klasova pšenice na nivou instanci koristi se metrika prosečne preciznosti (engl. *Mean Average Precision*, (mAP)). Pošto razmatramo problem sa jednom klasom [15, 16], u daljem tekstu mAP označava mAP₅₀, gde je usvojeni prag za izračunavanje IoU postavljen na 50% preklapanja između stvarnog i predviđenog regiona koji odgovara željenoj klasi. Dakle, u nastavku, mAP₅₀ će biti označavan jednostavno kao mAP.

Ono što se razlikuje u odnosu na deo istraživanja koji je vezan za prebrojavanje klasova pšenice je da u cilju procene preciznosti broja detektovanih kraničnih listova, izdvajamo dve greške koje pružaju bolji uvid u performanse modela [17, 18]. Procena tipa grešaka je značajna za buduću operativnu primenu (npr. optimizacija logistike i transpor-

ta, procena trajanja proizvoda na polici), gde bi donošenje odluka bilo zasnovano na agregiranim predikcijama na nivou kruničnih listova, pojedinačnih plodova ili grozdova paradajza. Greška precenjene predikcije se odnosi na broj identifikovanih kruničnih listova u test slici koja nisu prisutna u stvarnim, referentnim podacima, tj. slučajevi kada algoritam pogrešno detektuje krunične liste koji zapravo nisu deo paradajza (engl. *false positive*). Nasuprot tome, greške podcenjene predikcije odnose se na situacije kada algoritam ne detektuje krunične listove koje su potvrđene u referentnim podacima za datu test sliku. Motivacija za korišćenje ove dve metrike leži u činjenici da kod primene jednostavne prosečne apsolutne greške (engl. *Mean Absolute Error*, MAE), vrednosti promašenih ili lažno detektovanih kruničnih listova mogu međusobno da se ponište pri prosečenju.

Greške precenjene $\bar{e}$ i podcenjene $\underline{e}$ predikcije kruničnih listova se izračunavaju prema jednačinama:

$$\bar{e} = \frac{\mathbf{FP}}{\mathbf{P}} \quad , \quad \underline{e} = \frac{\mathbf{FN}}{\mathbf{P}} \quad (2.8)$$

gde je $\mathbf{P}$ stvaran, referentni broj kruničnih listova, $\mathbf{FP}$ broj lažno detektovanih kruničnih listova, a $\mathbf{FN}$ broj nedetekovanih kruničnih listova.

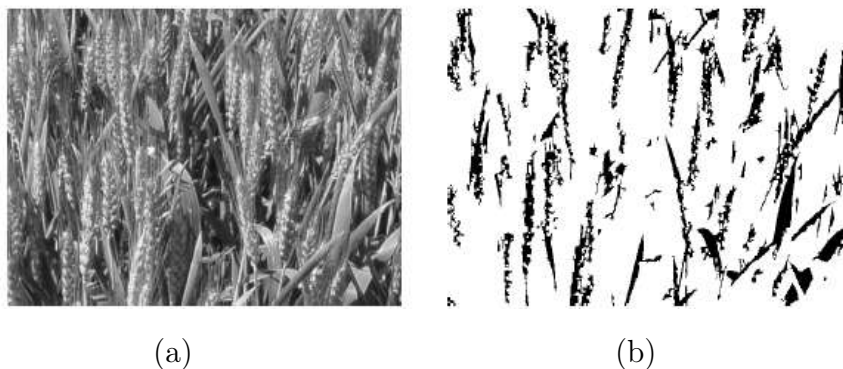
2.2 Hardverska infrastruktura

Za realizaciju eksperimenata u okviru ove doktorske disertacije korišćene su četiri visokoperformansne grafičke procesorske jedinice (GPU). Grafička kartica koja je korišćena za zadatke dubokog učenja i složene računarske operacije je NVIDIA GeForce RTX 2080 Ti, sa 4352 CUDA jezgara i 11 gigabajta memorije velike propusnosti (HBM2). Njihova računarska moć u preciznosti pojedinačnih operacija (FP32) dostiže do 13,4 teraflopsa po jedinici, što ukupno iznosi 53,6 teraflopsa za sve četiri jedinice. Ova značajna računarska snaga omogućila je efikasno treniranje odabranih modela dubokog učenja u praktičnim vremenimskim okvirima.

2.3 Analiza tradicionalnih metoda i motivacija za primenu složenijih modela

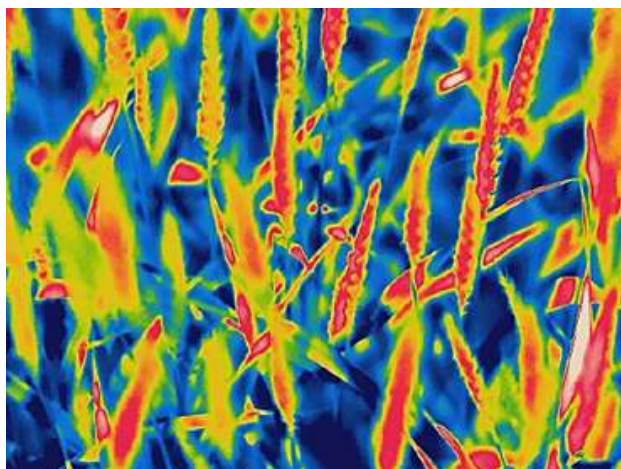
Kako bismo ukazali na potrebu za razvojem i korišćenjem dubokih neuronskih mreža isprobane su standardne metode koje zahtevaju manje heuristički određenih parametara na skupu podataka slika klasova pšenice. U okviru ove doktorske disertacije, za testiranje nekih od jednostavnijih algoritama u kojima figuriše manji broj parametara, koji su heuristički određeni, korišćeno je okruženje *Fiji* za softver *ImageJ* u koji su ove metode ugrađene. Ovaj softver je izabran jer se koristi u nekim od radovima za rešavanje problema detekcije klasa pšenice u RGB i termalnim slikama, gde su slike registrovane iz „ptičije“ perspektive [19, 20]. Proceduralni koraci za problem segmentacije klasova pšenice u RGB slici iz [19]: su primena *Laplasijan* i *Median* filtera, zaključno sa algoritmom *Find Maxima*. Svaki od ovih algoritama nameće određena ograničenja i izbore vrednosti parametara. Za primenu filtera *Laplasijana*, ulazna slika mora biti jednokanalna sa numeričkom preciznošću od 32 bita, za *Median* filter se bira odgovarajuća veličina kernela, dok za *Find Maxima* algoritam treba odrediti optimalan prag, odnosno parametar *prominence* koji direktno utiče na broj detektovanih klasova u slici, nalazeći maksimume, tj. pikove u filtriranoj slici. *Laplasijan*, odnosno filter kojim se realizuje drugi izvod na slici, se koristi sa ciljem identifikacije diskontinuiteta u slici, da naglasi razlike u osvetljaju, i da zemljište i listove odbaci kao regione koje nisu od interesa na osnovu frekvencijskog odziva. Veličina prozora kod upotrebe *Median* filtera je veoma bitna, zbog različite veličine klasova u slici, a samim tim i broja piksela koji im pripadaju, kako bi se uzeli u obzir i klasovi koji zauzimaju manji broj piksela. *Find Maxima* algoritam detektuje lokalne maksimalne vrednosti u slici, tj. pikove, gde nakon prethodnih filtriranja svaki lokalni maksimum predstavlja klas u slučaju da je slika slikana iz ptičije perspektive, kao što i jeste u pomenutom radu [19]. U slučaju slike iz perspektive, tj. pod uglom, kao što je slučaj sa slikama snimanim u ovom istraživanju, pokazaće se da primena predložene procedure nema zadovoljavajući rezultat. Na ovaj način ukazuje se da nedostatak standardne procedure za akviziciju slika nužno diktira dalji razvoj i unapređenje postupka jer tradicionalne tehnike nisu univerzalno primenjive sa istim podešavanjima za različite perspektive slikanja.

Primena odgovarajućeg praga za ovakvu vrstu problema je vrlo izazo-



Slika 2.1: (a) 8-bitna RGB reprezentacija slike; (b) segmentacija nakon primene filtriranja i Find Maxima algoritma.

van zadatak, jer u RGB slikama, klasovi nisu dovoljno čak ni vizuelno razdvojivi, pogotovo u ranijim fenološkim fazama, kada pšenica tek klasa, dok se kod termalnih slika mogu primeniti analize koje uključuje vrednost intenziteta piksela, zbog pseudokolor prikaza u vidu *heat* mape, gde se jasno vidi razlika u temperaturi klasa i ostalih regiona (vidi sliku 2.2).



Slika 2.2: Pseudokolor reprezentacija termalne slike.

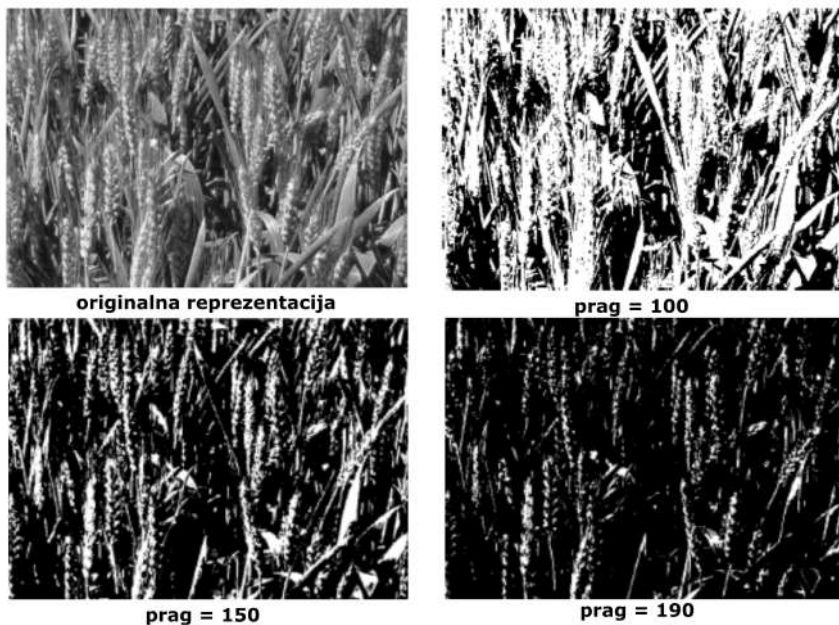
Na *grayscale* jednodimenzionalnoj reprezentaciji RGB slike i odzivu termalne kamere primenjeno je 17 metoda za automatsko određivanje praga, a rezultati su prikazani u Tabeli 2.1.

Tabela 2.1: Korišćene metode i njihove prosečne performanse na test skupu RGB slika

Metoda	Parametri	Preciznost	IoU (Intersection over Union)
Default	125	0,58	0,25
Huang	113	0,52	0,26
Intermodes	172	0,73	0,21
IsoData	125	0,58	0,26
Li	114	0,53	0,26
MaxEntropy	152	0,68	0,23
Mean	120	0,56	0,26
MinError	64	0,27	0,23
Otsu	127	0,59	0,26
Triangle	196	0,77	0,16
Yen	158	0,70	0,23
Manual threshold	100	0,50	0,26
Manual threshold	150	0,70	0,24
Manual threshold	190	0,77	0,17
Random Forest	Gaussian blur, Maximum, Gabor, Entropy, Sobel, Difference of Gaussians, Median, Neighbors, 100 stabala	0,54	0,32
U-Net	–	0,76	0,49
Predložena arhitektura	–	0,82	0,62

Izbor metrike koja će na najbolji način prikazati rezultate i performanse modela u velikoj meri zavisi od vrste podataka. U ovom slučaju, distribucija piksela koji pripadaju i onih koji ne pripadaju klasu pšenice u labelama je veoma nebalansirana. Značajno veći broj piksela u slikama pripadaju pozadini u odnosu na klasove. Stoga, je u tabeli prikazana preciznost ali i IoU (engl. *Intersection Over Union*), metrika koja se najčešće koristi u problemu segmentacije [21, 22, 23].

Iz predstavljenih primera jasno je da primena ovakih algoritama ne rešava problem segmentacije i detekciju klasova pšenice. Osim primene praga i filtriranja slika, dekompozicije i razlaganja slike, u ovom



Slika 2.3: Primena praga na osnovu vrednosti u histogramu slike.

softveru dostupni su još neki ugrađeni algoritmi, kao što je na primer *FastRandomForest*, gde je neophodan odabir velikog broja obeležja (*Gaussian blur*, *Hessian*, *Membrane projections*, *Mean*, *Maximum*, *Anisotropic Diffusion*, *Lipschitz*, *Gabor*, *Entropy*, *Sobel*, *Difference of gaussians*, *Variance*, *Minimum*, *Median*, *Bilateral*, *Kuwahara*, *Neighbours*). Za pravi odabir obeležja potrebno je predznanje za svaku od ovih metoda za izdvajanje istih, a potom i analiza primene baš na rešavanju ovakvog konkretnog problema. Dodatno, treba primeniti ove metode od kojih je svaka određene složenosti i skoro svaka zahteva dodatni izbor parametara. Što se tiče samog algoritma *FastRandomForest*, implemetacija je iz [24]. Prilikom pokretanja ovog algoritma potrebno je definisati i sledeće parametre: broj stabala, minimalan broj uzoraka u svakoj grani, maksimalnu dubinu stabla i broj obeležja koje će uzimati prilikom slučajne selekcije. Dok tradicionalne metode zahtevaju ručno i heurističko podešavanje parametara kako bi se postigla stabilna segmentacija, neuronske mreže ovaj proces automatizuju, učeći kriterijum razdvajanja klasa kroz optimizaciju funkcije cene. Time se segmentacija posmatra kao rezultat procesa učenja, a ne kao skup unapred definisanih pravila. Takođe, uprkos velikom broju nepoznatih

parametara koje neuronske mreže uče tokom obuke modela iz skupa podataka, prevazilaze potrebu za ekspertskim znanjem, izdvajanjem i potpunim razumevanjem velikog broja obeležja. Konačnom obukom modela, neuronske mreže nalaze optimalnu kombinaciju koja je primenljiva u najvećoj meri na sve slike iz tog skupa, sa većom ili manjom preciznošću.

Detaljna razrada predložene arhitekture za semantičku segmentaciju, kao i pristup zasnovan na transferu znanja, biće prikazani kroz dve studije slučaja – na kulturama pšenice i paradajza, uz upotrebu slika različitih modaliteta snimanja, uključujući hiperspektralne i standardne RGB kolor slike. Metodologija segmentacije, koja predstavlja centralnu temu ove doktorske disertacije, biće sistematski predstavljena kroz semantičku segmentaciju i segmentaciju instanci. U tom okviru, pored razvijenih univerzalnih arhitektura za segmentaciju regiona od interesa, akcenat je stavljen i na primenu engl. *transfer learning* pristupa za segmentaciju instanci, čime se omogućava efikasnije iskorišćavanje postojećih modela na novim i domenski specifičnim skupovima podataka. Kao rezultat sprovedenih eksperimenata, obezbeđeno je više prikupljenih skupova podataka, koji predstavljaju značajan doprinos daljim istraživanjima u oblasti. Pored toga, posebna pažnja posvećena je i trodimenzionalnoj reprezentaciji u prostoru, koja omogućava dodatnu interpretabilnost i proširuje primenljivost razvijenih metoda u kontekstu precizne poljoprivrede.

2.4 Semantička segmentacija

Budući da je u okviru ove doktorske disertacije predložena nova arhitektura konvolucione neuronske mreže za semantičku segmentaciju, vredi istaći da je njena glavna prednost, a i konkretno konvolucionih neuronskih mreža, ta da prilikom obučavanja ovog modela figuriše znatno manji broj parametara u poređenju sa arhitekturama sa potpuno povezanim slojevima, tačnije potpuno deljenim težinama u neuronskoj mreži, pa se samim tim sa razvojem ove arhitekture, smanjuje kompleksnost izračunavanja prilikom procesa obuke modela. Neke od konvolucionih arhitektura koje su dale najveći doprinos u oblasti segmentacije slike uključuju: AlexNet [25], VGGNet [26], GoogLeNet [27], ResNet [28], U-Net [29], DenseNet [30] i MobileNet [31]. U novijem periodu, značajan napredak je postignut uvođenjem mehanizama pažnje

(engl. *Attention*) i transformer arhitektura, kao što su: Vision Transformer (ViT) [19], DeiT [32], Swin Transformer [33], SegFormer [34], kao i hibridni modeli poput ConvNeXt [35], koji kombinuju konvolucione i transformer pristupe.

2.4.1 Pregled literature za semantičku segmentaciju klasova pšenice

Ovaj odeljak započinje pregledom relevantne literature u oblasti detekcije klasova pšenice, iz koje je proistekla inicijalna ideja i motivacija za dalji istraživački rad, kao i za razvoj metodološkog pristupa koji je kasnije generalizovan i primenjen na druge zadatke i modalitete.

Počevši od 2008. godine, prva ideja za automatizovano brojanje klasova sa slika snimljenih na polju predložena je korišćenjem analize boja i teksture slike u [36], dok su u [37] iz slika na osnovu morfologije izdvojene fenotipske osobine klasova pšenice kao što su dužina, širina, površina i obim klasa. Autori u [38] predstavili su neuronsku mrežu za klasifikaciju [39] kako bi izdvojili RGB karakteristike boja, uključujući dužinu klasova. U [20] korišćen je Laplasov frekvencijski filtar [40] i median filtar [41] za suzbijanje prisustva šuma oko klasova uklanjanjem visokofrekventnih komponenti u slici. Konačna segmentacija klasova vrši se korišćenjem algoritma *Find Maxima* [42]. Filtriranje korišćenjem Gabor filtar banaka [43] praćena analizom glavnih komponenti (engl. *Principal Component Analysis*, PCA) [44] i K-means algoritmom [45] predložena je u [13] za segmentaciju regiona koji pripadaju klasovima pšenice. Postignuta je prosečna tačnost od 90,7% u poređenju sa ručnim brojanjem, a predloženo rešenje je testirano na realnom skupu podataka koji sadrži slike pšenice u fazama cvetanja i zrenja. U [46] broj klasova na slikama se dobija korišćenjem *Simple Linear Iterative Clustering* metode (SLIC) [47], transformišući jedinicu slike pšenice iz piksela u superpiksel blokove. Tačnost ovog pristupa kreće se od 80,8% do 94,2%. Model segmentacije korišćenjem *Twin Support Vector Machine* (TWSVM-Seg) [48] primenjen je u [49] za segmentaciju klasova pšenice na nivou piksela. Rezultati predloženog pristupa pokazali su visoku korelaciju od 0,79-0,82 sa podacima dobijenim ručnim brojanjem. Skup podataka je prikupljen tokom istraživanja i sadrži 1000 kontura (na nivou piksela). Segmentacija klasova pšenice sa slike omogućava automatsku procenu gustine klasova i predviđanje stanje pšenice, veličinu klasa i fazu zrelosti. Većina podataka prikupljenih u

ovim studijama prilagođena je za specifične svrhe i ciljeve, što otežava njihovu upotrebu za šire praktične primene.

Oslanjajući se na sve veće performanse grafičkih kartica (*Graphics Processing Unit* – GPU) i široku dostupnost senzora za prikupljanje podataka, što omogućava kreiranje velikih skupova podataka [50, 51], duboko učenje postaje najnoviji trend u oblasti kompjuterske vizije za prepoznavanje objekata [52], segmentaciju instanci [53], semantičku segmentaciju [29] i regresiju slika [54, 55], pružajući visoke performanse i nadmašujući druge tehnike obrade slika i klasične metode mašinskog učenja [56, 57]. Primena konvolucionih neuronskih mreža (engl. *Convolution Neural Networks*, *CNN*) ima ključnu ulogu u fenotipizaciji zasnovanoj na slikama [58]. Efikasnost ovog pristupa leži u konvolucionim filtrima, koji rešavaju problem računarske složenosti i potrošnje memorije tako što ograničavaju zavisnost samo na okolinu piksela. Relativna greška u brojanju klasova pšenice procenjena je na oko 10% u [59] za ovu metodu ako je rezolucija slike uniformna. U [60] autori su razvili CNN model zasnovan na enkoder i dekoder arhitekturi za semantičku segmentaciju klasova pšenice koji nadmašuje tradicionalne manuelne metode kompjuterske vizije. U [61] je razvijen model za detekciju klasova pšenice i probabilističko praćenje rasta pšenice na terenu. Kao i u drugim sličnim primenama, u kontekstu brojanja klasova pšenice, uobičajen pristup je treniranje CNN modela koristeći podatke sa jednog dela parcele i evaluacija performansi na podacima sa drugog dela parcele [57, 62, 63, 64, 65].

S obzirom na specifične izazove koji se javljaju na slikama, postojeći, već razvijeni i prethodno trenirani modeli za semantičku segmentaciju klasova pšenice u početku nisu postizali ohrabrujuće rezultate. Problem preklapanja klasova pšenice je težak za tradicionalne metode obrade slika sa visokim semantičkim osobinama, dok postojeće metode brojanja zasnovane na dubokom učenju nisu efikasno rešile problem preklapanja.

Duboko učenje se pokazalo kao obećavajuća metoda koja nadmašuje tradicionalne pristupe koji se koriste u analizi slika i mašinskom učenju [66]. Upotreba konvolucionih neuronskih mreža (CNN) je u porastu i one se sve češće koriste u fenotipizaciji biljaka: za automatsku identifikaciju biljnih vrsta [67], izdvajanje fenotipskih karakteristika kod vrste *Arabidopsis thaliana* [68, 69], ili za detekciju klasova pšenice u kontrolisanim uslovima [70, 71, 72]. Za razliku od potpuno povezanih neuronskih mreža, u kojima je svaki čvor u skrivenim slojevima

povezan sa svim pikselima slike (ova kompleksnost čini ovu arhitekturu neprikladnom za analizu slika), konvolucionni filtri rešavaju problem ograničavanjem zavisnosti samo na okolne piksele. Ovi filtri obično imaju male dimenzije (npr. 3×3 , 5×5 , 7×7), čime se značajno smanjuje broj parametara u mreži. Filtri su skup težina, matrica brojeva, koja se primenjuje na deo slike na ulazu, „prolaze“ preko cele slike i formiraju matricu obeležja – tzv. engl. *feature map* [73]. Primena konvolucionih filtara u skrivenim slojevima omogućava kontekstualno razumevanje piksela slike, čime se simulira način obrade informacija u vizuelnom korteksu. Pored konvolucionih slojeva, izbor aktivacione funkcije ima značajan uticaj na ukupne performanse mreže. Najčešće korišćena aktivaciona funkcija u dubokim mrežama je ReLU (engl. *Rectified Linear Unit*) [74], koja se pokazala efikasnom u procesu treniranja, kao i njene varijacije (engl. *Leaky ReLU*, *Parametric ReLU*), koje su korisne za rešavanje problema „nestajućih gradijenata“ [75]. Ranije razvijene funkcije, poput linearne, stepenaste, sigmoidne, hiperboličkog tangensa i *softsign* funkcije, najčešće se primenjuju u manje kompleksnim mrežama.

U novije vreme i duboko učenje nalazi svoju primenu u fenotipizaciji – tako su u [11, 12] konvolucione neuronske mreže postigle veoma dobre rezultate u rešavanju ovog tipa problema. U [11] autori su koristili *Region-based Convolutional Neural Network* (R-CNN) i RGB slike, dostigavši prosečnu tačnost od 93,30% i F1 meru od 95%. Doprinos rada iz [12] ogleda se ne samo u razvijenoj novoj arhitekturi mreže za ovu specifičnu primenu, već i u formiranju skupa podataka sa jednostavnom crnom pozadinom u kontrolisanom unutrašnjem laboratorijski kreiranom okruženju.

Na tačnost prepoznavanja može da utiče čitav niz faktora: intenzitet osvetljenja, prisustvo senki, šum, različiti uglovi pri snimanju i širina posmatranog vidnog polja. U [10] je detaljno objašnjeno da nije jednostavno odrediti optimalan ugao ili vreme akvizicije slike. U [11, 20] pokazano je da je vetikalna perspektiva dobro rešenje za ugao snimanja. Ipak, naše istraživanje, kao i nekoliko drugih [10, 11, 12, 13, 76], oslanjalo se na kosu perspektivu, koja omogućava dobijanje više dodatnih informacija o boji, obliku i teksturi. Ove informacije mogu biti dragocene ne samo za brojanje klasova, već i za detekciju i praćenje širenja bolesti [11]. U poglavlju 3 pored detaljnog opisa prikupljenih podataka u laboratoriji i na terenu, data je i procena kvaliteta slika, kao i preporuke i protokoli za optimalnu akviziciju slika na terenu.

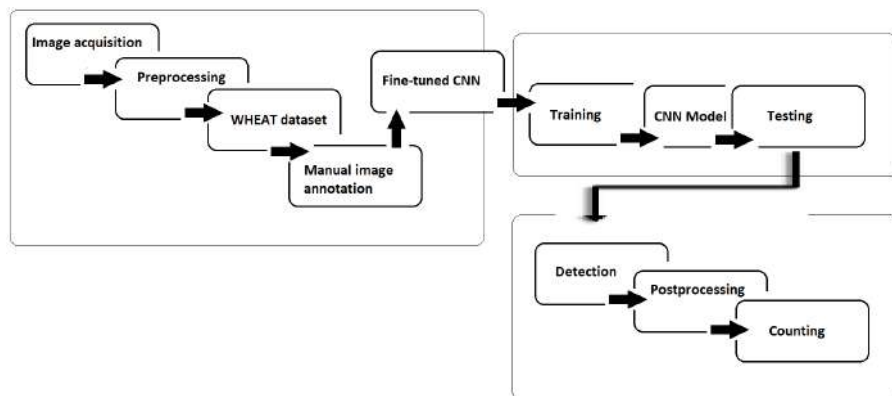
Većina prethodnih studija fokusirala se isključivo na upotrebu RGB slika. Jedino je u [20] pomenuta potencijalna primena termalne kamere, ali je ona kritikovana zbog niske rezolucije i visoke cene. U našem inicijalnom istraživanju odlučili smo se za kombinaciju RGB i termalne kamere za detekciju i brojanje klasova. Zbog svog kompaktnog oblika, klasovi poseduju veći toplotni kapacitet od ostatka biljke. Shodno tome, oni akumuliraju toplotu sporije od listova i stabljike, ali je zadržavaju duže i u većoj meri. Iz tog razloga klasovi na termalnim slikama izgledaju veoma svetlo i lako se razlikuju od pozadine, čak i golim okom [77]. Suprotno tome, na RGB slikama klasovi i pozadina imaju sličnu zelenu ili žutu boju u zavisnosti od posmatrane fenološke faze, pa se njihova segmentacija mora oslanjati na složenije parametre, kao što su oblik i tekstura. Ovo je bio glavni razlog da u našem istraživanju razmatramo oba senzora i modaliteta slika.

2.4.2 Predlog arhitekture za semantičku segmentaciju klasova pšenice iz termalnih i RGB slika

U ovom delu doktorske disertacije razvijen je predlog arhitekture za detekciju klasova pšenice iz termalnih i RGB slika, pod nazivom *optimized Fully Convolutional Neural Network (oFCN)* snimljenih na oglednoj parceli u spoljašnjim uslovima. Sistem se sastoji iz dva glavna modula: modula za segmentaciju delova slike koji pripadaju klasovima pšenice i modula za njihovo brojanje. Pored ovih glavnih modula, celokupan tok rada uključuje korake predobrade slika, validaciju modela, kao i postprocesiranja (vidi Sliku 2.4). U nastavku sledi kratak opis svakog od ovih koraka.

U koraku predobrade sve slike su normalizovane na opseg u intervalu $[0, 1]$. Termalne slike su snimljene kao pseudokolor slike, pri čemu crvena i žuta boja predstavljaju toplije delove (klasove), dok plava označava hladnije delove i pozadinu (najčešće listove i stabljiku). Ovo omogućava bolju distinkciju i interpretaciju objekata od interesa na slikama, što olakšava detekciju klasova pšenice. RGB slike su originalno visoke rezolucije, dimenzija 2048×1536 piksela. Obeležavanje tako velikih slika sa velikim brojem klasova pokazalo se previše vremenski zahtevnim. Stoga, pošto termalne slike pokrivaju centralni deo RGB slika, RGB slike su podeljene na istu veličinu kao termalne slike (640×480 piksela), tako da predstavljaju isto vidno polje i centralni deo vidnog polja.

Ono što čini segmentaciju slike veoma važnom, kao korak u sistemi-



Slika 2.4: Tok predložene metodologije prikazan korak po korak.

ma za detekciju, klasifikaciju ili predikciju, jeste činjenica da rezultati segmentacije utiču na dalju analizu slike i postoji propagacija greške [5]. U ovom delu istraživanja izvršena je segmentacija na nivou piksela koristeći konvolucionu neuronsku mrežu (CNN). Problem segmentacije smo ograničili na binarni problem i dve klase: jedna predstavlja pripadnost posmatranog piksela klasovima pšenice, a druga pozadinu.

Da bi se testirala preciznost segmentacije, referentne oznake gde se klasovi nalaze kreirane su ručno. Rezultat obeležavanja su binarne slike, gde vrednost piksela 1 označava pripadanje regionu klasova, dok vrednost 0 označava pozadinu. Oba tipa slika su podeljena na trening i test skup, gde je 80% slika opredeljeno za trening i 20% za test. Prosečan broj klasova na termalnim slikama bio je 80, dobijen iz ukupnog broja od 11040 ručno prebrojanih klasova na svim odabranim termalnim slikama (8832 za trening i 2208 za test). Prosečan broj klasova na RGB slikama je 36, dobijen iz ukupnog broja od 4968 klasova na svim odabranim RGB slikama (3975 za trening i 993 za test). Razlika u rezoluciji termalnih i RGB senzora rezultirala je manjim brojem klasova na RGB slikama u poređenju sa termalnim slikama. U nastavku sledi detaljno objašnjenje arhitekture CNN-a koju predlažemo.

U okviru ovog dela doktorske disertacije, za problem detekcije klasova, odnosno segmentacije regiona slike koji pripada klasovima, razvijena je arhitektura konvolucione neuronske mreže (CNN) bez korišćenja prethodno treniranih modela iz literature. Prednost korišćenih konvolucionih slojeva u predloženoj arhitekturi ogleda se u značajno manjem broju parametara za učenje zbog retkih međusobnih veza i

deljenih težina, što smanjuje računarsku složenost, štedi memoriju i omogućava brže učenje mreže. Konvolucionni sloj definiše se veličinom konvolucionog filtra, brojem kanala ulazne slike i brojem izlaznih mapa obeležja (filtara), koji zajedno određuju dimenzionalnost i broj parametara sloja [78].

Predložena mreža se sastoji od blokova konvolucionih slojeva, nakon kojih je umetnuta funkcija aktivacije. Najčešće korišćena funkcija aktivacije za duboke neuronske mreže je *Rectified Linear Unit*, RELU [71], definisana kao:

$$\text{ReLU}(x) = \max(0, x) \quad (2.9)$$

Iako ReLU nadmašuje druge aktivacione funkcije po računarskoj efikasnosti i u rešavanju problema nestajanja gradijenata, ona može dovesti do pojave "mrtvih" neurona. Ovi neuroni daju izlaznu vrednost jednaku 0 za svaki uzorak u bazi podataka za treniranje, što se dešava kada je argument funkcije negativan. Prednost ovog ponašanja je što se te nule mogu eliminisati iz mreže, čime se postiže bolja računarska efikasnost, dok se proračuni značajno pojednostavljaju i redukuju na operacije poput poređenja, sabiranja i množenja. Nedostatak ReLU funkcije je što mrtvi neuroni mogu uticati na ukupnu tačnost izlaza mreže [79].

Različita poboljšanja ReLU funkcije su predložena kako bi se ublažili negativni efekti mrtvih neurona. Ovde ćemo pomenuti *Leaky ReLU* (*lReLU*) i *Parametrized ReLU* (*PReLU*) aktivacione funkcije. lReLU je definisana kao:

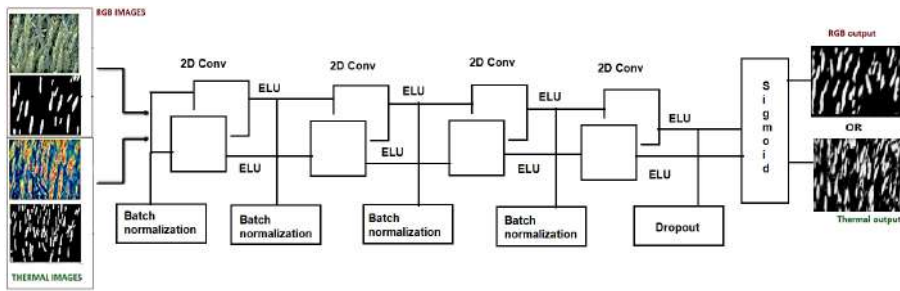
$$\text{lReLU}(x) = \begin{cases} \alpha x, & \text{ako } x < 0 \\ x, & \text{otherwise} \end{cases} \quad (2.10)$$

gde je α pozitivan hiperparametar (obično 0,01), koji omogućava negativnim vrednostima da prolaze kroz mrežu bez nestajanja gradijenta. Vrednost α može biti u opsegu [0, 1], dok je prema literaturi [79] optimalno postavljena na 0,01, mada su za neke tipove podataka veće vrednosti pokazale bolje rezultate [80]. PReLU je modifikacija ReLU funkcije kod koje se hiperparametar α uči iz podataka tokom treniranja mreže, za razliku od lReLU, koja koristi fiksnu vrednost [79]. U ovom delu istraživanja odlučili smo se za *Exponential Linear Units* (*ELU*) funkciju aktivacije [71]. ELU pripada grupi nelinearnih funkcija i definiše se:

$$\text{ELU}(x) = \max(0, x) + \min(0, \alpha * (\exp(x) - 1)) \quad (2.11)$$

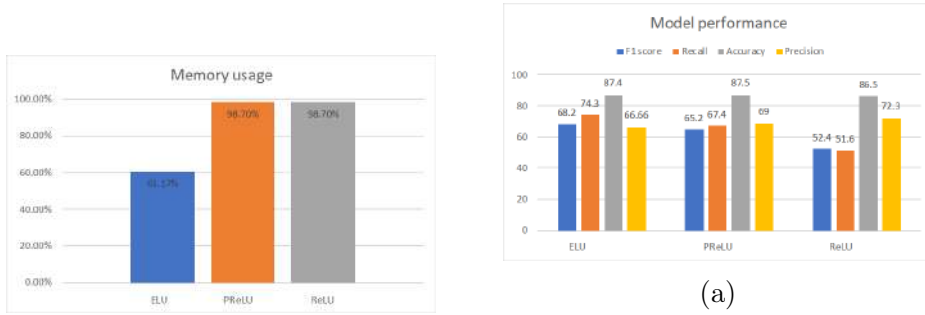
Ova funkcija ubrzava učenje u dubokim mrežama i omogućava postizanje veće tačnosti klasifikacije. Primena ELU aktivacione funkcije pokazala je veću robusnost na šum u podacima i bolju generalizaciju CNN mreža sa više od 5 slojeva u poređenju sa drugim funkcijama aktivacije [80].

Pored blokova konvolucionih slojeva praćenih ELU funkcijom, umetnuta je *batch* normalizacija kao metoda regularizacije. Ova metoda omogućava skraćanje vremena treniranja i sprečava pojavu natprilagođenja modela (engl. *overfitting*). U prvom konvolucionom sloju koristi se kernel veličine 7×7 , dok se u ostalim slojevima koristi kernel 3×3 . Arhitektura predložene mreže je prikazana na Slici 2.5.

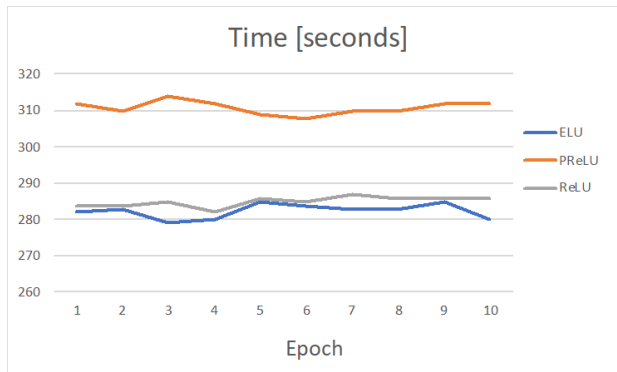


Slika 2.5: Arhitektura nove razvijene neuronske mreže za semantičku segmentaciju, nazvana *Optimized Fully Convolutional Neural Network, oFCN*, primenjena na termalne i RGB slike

U ovom delu istraživanja evaluirali smo performanse aktivacionih funkcija sa aspekta korišćenja memorije na jednoj GPU kartici i vremena potrebnog za 10 epoha treniranja na termalnim slikama. Korišćena GPU kartica je jedna od četiri NVIDIA TITAN X kartice na serveru. Na osnovu upoređivanja dobijenih rezultata (vidi sliku 2.6), izabrali smo ELU aktivacionu funkciju, zbog boljih performansi u odnosu na druge dve aktivacione funkcije. ELU zahteva znatno manju količinu memorije na jednoj GPU kartici, samo 61,17% od ukupne memorije tokom procesa treniranja (7457/12189 MiB), što omogućava korišćenje dublje, odnosno kompleksnije mreže ukoliko je potrebno i znatno brže treniranje, dok ReLU i PReLU zahtevaju 98,70% GPU memorije (12031/12189 MiB) [13, 71].



Slika 2.6: (a) Analiza potrošnje GPU memorije; (b) performanse modela u zavisnosti od primenjenih različitih aktivacionih funkcija. Rezultati su dobijeni nakon 10 epoha.



Slika 2.7: Izmereno potrebno vreme za proces obuke u zavisnosti od različitih aktivacionih funkcija. Rezultati su dobijeni nakon 10 epoha.

ReLU zahteva slično vreme za treniranje tokom 10 epoha kao i ELU, ali pri 1000 epoha proces treniranja postaje sporiji. PReLU zahteva duže vreme za treniranje, s obzirom na to da se hiperparametar takođe uči tokom procesa treniranja. ELU aktivaciona funkcija je izabrana ne samo zbog bržeg treniranja i manje potrošnje memorije, već i zato što ove prednosti ne utiču na performanse tačnosti modela 2.7 [71, 79]).

Cross entropy[81] tj. unakrsna entropija definisana je kao mera razlike dve raspodele verovatnoće za dati posmatrani skup slučajnih uzoraka. Najčešće se koristi u klasifikaciji, ali i u slučaju segmentacije na nivou piksela, kao što je slučaj u ovom istraživanju. U okviru ove teze, rešava se binarni problem u slici: detektovati klas (klasa 1) i sve ostalo (list, stabljika, zemlja – klasa 0). Kao osnova prilikom obučavanja neuronske mreže koristi se binarna funkcija cene (*engl. Binary cross*

entropy). Na ulazu u neuronsku mrežu nalazi se slika X . Svaki piksel u matrici slike se množi na ulazu u prvi sloj sa odgovarajućim težinama. Predikcija izlaza iz duboke neuronske mreže može se zapisati kao:

$$\tilde{y} = f_{\text{dnn}}(X, \mathbf{w}) \quad (2.12)$$

Binarnu unakrsnu entropiju u ovom radu finalno definišemo kao:

$$L_{\text{BCE}}(y, f_{\text{dnn}}(X, \mathbf{w})) = -\left(y \log(f_{\text{dnn}}(X, \mathbf{w})) + (1 - y) \log(1 - f_{\text{dnn}}(X, \mathbf{w}))\right) \quad (2.13)$$

gde y predstavlja pravu vrednost, a $\tilde{y}$ predviđenu vrednost na izlazu iz modela.

Za funkciju gubitka koristi se *Binary Cross Entropy* (BCE), definisana kao:

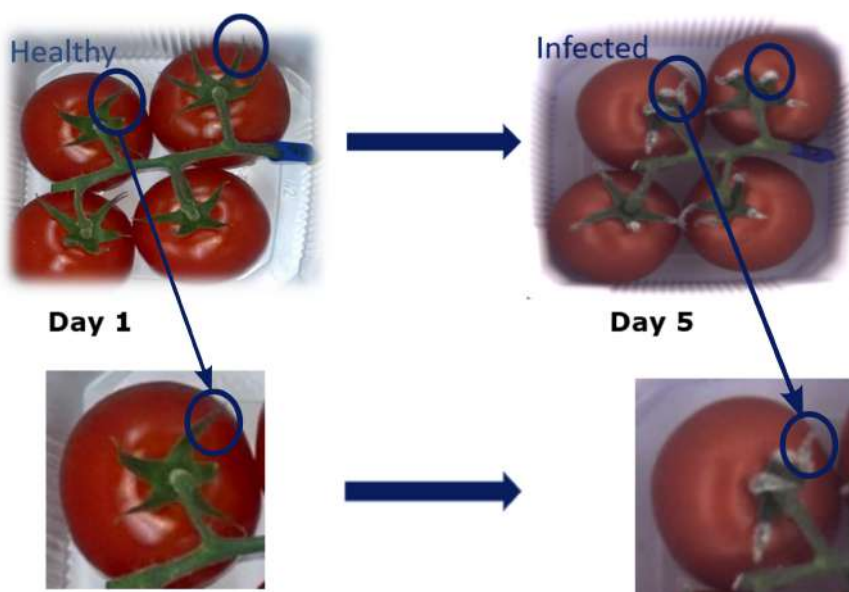
$$\mathcal{H}p(y) = -\frac{1}{N} \sum_{i=1}^N \left[y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i)) \right], \quad (2.14)$$

gde N predstavlja broj piksela u batch-u, y_i labelu (1 za piksele koji pripadaju klasovima, a 0 za pozadinu), dok $p(y_i)$ označava predviđenu verovatnoću da piksel pripada klasi klasova [82, 83]. Ova funkcija gubitka omogućava da se model trenira za binarnu klasifikaciju piksela, gde se za piksele klasova ($y_i = 1$) dodaje logaritam verovatnoće $\log(p(y_i))$, a za piksele pozadine ($y_i = 0$) $\log(1 - p(y_i))$.

2.4.3 Evaluacija predložene arhitekture na hiperspektralnim podacima

Istraživanje u okviru ove studije slučaja je osmišljeno sa namerom da se proceni da li spektralne karakteristike iz bliskog infracrvenog (NIR) opsega mogu pouzdano ukazati na mikroskopska oštećenja čelija i/ili fiziološku deformaciju kraničnih listova paradajza, te da se utvrdi korelacija sa kasnijim intenzitetom infekcije u kontrolisanim uslovima. Kvalitet paradajza zavisi od uslova gajenja i uslova u lancu snabdevanja, poput vlažnosti i temperature, kao i od načina rukovanja tokom berbe i procesa nakon iste (transport, pakovanje, skladištenje itd.). Poput mnogih drugih lako kvarljivih plodova i povrća, paradajz je izrazito podložan gubicima nakon berbe, koji u pojedinim zemljama u razvoju dostižu i do 30%. Poznato je da je paradajz veoma osetljiv

na patogene gljive, poput *Penicillium*, *Aspergillus* i *Mucor*, koje napadaju plodove bogate vlagom i hranljivim materijama. Oštećenja ćelija tkiva paradajza mogu nastati usled promena u spoljnim uslovima, kao i tokom rukovanja proizvodom. Takva oštećenja predstavljaju potencijalna ulazna mesta za spore gljiva koje, u pogodnim uslovima za klijanje, mogu inficirati peteljku, krunične liste i druge delove ploda. Ovaj deo studije je usmeren na ispitivanje osetljivosti kruničnih listova na gljivične infekcije. Pored fizičkih oštećenja, krunični listovi mogu



Slika 2.8: Razvoj gljivičnog oboljenja od 1. do 5. dana u RGB (vidljivom) spektru.

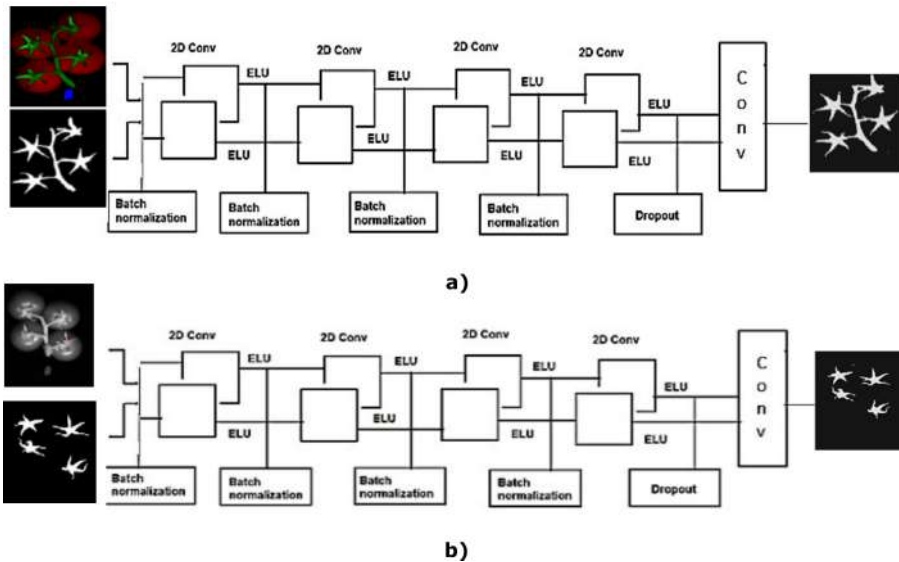
biti i fiziološki jači, tj. otporniji ili slabiji, pri čemu na njihovu otpornost utiču različiti uslovi gajenja, relativne vlažnosti tokom vegetacije, gustine useva i nutritivnog statusa biljke. U prisustvu gljivičnih spora i povoljnih uslova za njihov rast, postavlja se hipoteza da postoji korelacija između predisponirane otpornosti kruničnih listova (pre infekcije) i njihove kasnije infekcije i/ili intenziteta iste. Rano oštećenje ćelija kruničnih listova, odnosno otpornost nije vidljiva golim okom i, koliko je poznato, ne postoji metoda za automatsku detekciju ovih pojava pre nastupanja infekcije. Hiperspektralno snimanje (HSI), naročito u bliskom infracrvenom (NIR) spektru, pokazalo je osetljivost na pojedine tipove oštećenja ćelija, poput nagnječenja, ali do sada nije

primenjeno na oštećenja na vrhovima kruničnih listova paradajza ili za ranu detekciju njihove podložnosti. U inicijalnoj studiji [84] istražena je primena hiperspektralne tehnologije u otkrivanju oštećenja i podložnošću ćelija kruničnih listova paradajza. Radi provere hipoteze, sproveden je eksperiment u kojem su hiperspektralne slike prikupljene iz različitih izvora, za jednu sortu paradajza, pre pojave vidljivih gljivičnih infekcija, s ciljem da se registruje eventualno oštećenje ćelija na kruničnim listovima. Nakon toga, plodovi su izloženi uslovima pogodnim za klijanje gljiva tokom više dana. Na završni dan eksperimenta snimljene su standardne RGB kolor slike kako bi se prikupili vizuelni dokazi o stepenu gljivične infekcije za svaki krunični list. Prikazani su prvi rezultati, u kojima je korišćen pristup zasnovan na mašinskom učenju *Random Forest* radi utvrđivanja korelacije između spektralnih informacija zabeleženih prvog dana eksperimenta i intenziteta gljivične infekcije poslednjeg dana. Svaki krunični list opisan je srednjom vrednošću i standardnom devijacijom svojih hiperspektralnih odziva. Performanse modela procenjavane su metodama 10-ostruke i grupne kros-validacije, pri čemu su grupe odgovarale različitim poreklima paradajza. Dobijeni rezultati pokazuju dobru korelaciju između predikcije i stvarnih vrednosti, sa Pirsonovim koeficijentom od 0,73 i 0,66, kao i značajnim udelom varijanse ($R^2 = 0,52$ i $0,43$) za 10-ostruku, odnosno grupnu kros-validaciju.

Kao pilot ispitivanje [84], ova studija je dizajnirana da pruži rane indikacije izvodljivosti, te stoga uključuje jednu sortu i ograničen skup podataka. Naredna studija obuhvata širi koncept inicijalne studije [1]. Uporedo sa analizom o ranoj indikaciji širenja gljivičnog oboljenja na kruničnim listovima paradajza, sprovedena je inicijalna analiza koja je objavljena u [85, 86], a kasnije i u [87] gde se proširuje hiperspektralni skup podataka i daje detaljan uvid u analizu pristupa za segmentaciju regiona od interesa kao i metoda za efikasno izdvajanje latentnih spektralnih komponenti visokodimenzionalnih hiperspektralnih podataka.

Razvijena arhitektura za semantičku segmentaciju primarno je primenjena na skup podataka namenjen rešavanju problema detekcije klasova pšenice na termalnim i RGB slikama.

Postojeći modeli u ovoj oblasti u najvećoj meri se oslanjaju na pre-trenirane mreže i njihovu primenu na standardnim RGB slikama, dok je mogućnost efikasne adaptacije na druge modalitete slabo istražena. Predložena arhitektura doprinosi upravo u tom smeru — kroz adaptivnu primenu i evaluaciju na više modaliteta, uključujući termalne

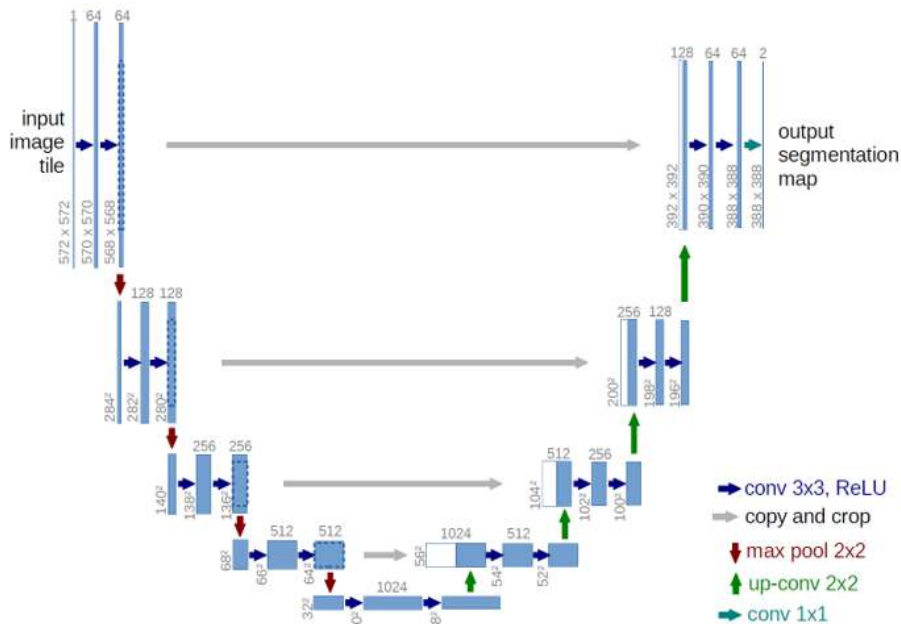


Slika 2.9: oFCN arhitektura primenjena na dva nivoa segmentacije: a) stabljika i krunični listovi paradajza b) model fokusiran na segmentaciju samo kruničnih listova paradajza.

slike u okviru studije detekcije klasova pšenice, kao i hiperspektralne slike korišćene u studiji detekcije kruničnih listova paradajza, sa ciljem predikcije širenja gljivičnog oboljenja. Iako se na prvi pogled ove studije razlikuju po sadržaju i primeni, povezuje ih zajednički tehnički okvir problema vizuelne segmentacije složenih bioloških struktura. U oba slučaja, strukture objekata predstavljaju značajan izazov u domenu kompjuterske vizije — klasovi pšenice karakterišu visoka repetitivnost, međusobno preklapanje i varijabilnost u orijentaciji, dok su hiperspektralni podaci inherentno visoke dimenzionalnosti, a krunični listovi paradajza često uvijeni, delimično zaklonjeni i izloženi intenzivnim refleksijama. Precizna segmentacija u oba scenarija ima suštinski značaj: broj i prostorni raspored klasova direktno utiču na procenu prinosa, dok tačna detekcija piksela na vrhovima kruničnih listova omogućava izdvajanje spektralnih informacija ključnih za rano otkrivanje ili praćenje gljivičnih oboljenja. Time razvijena metodologija ne samo da demonstrira fleksibilnost arhitekture kroz više modaliteta i različite biološke strukture, već i ukazuje na mogućnost njene primene u širem kontekstu analize biljnih fenotipova i drugih oblasti kompjuterske vizije. U daljem nastavku će biti objašnjena adaptacija arhitekture na

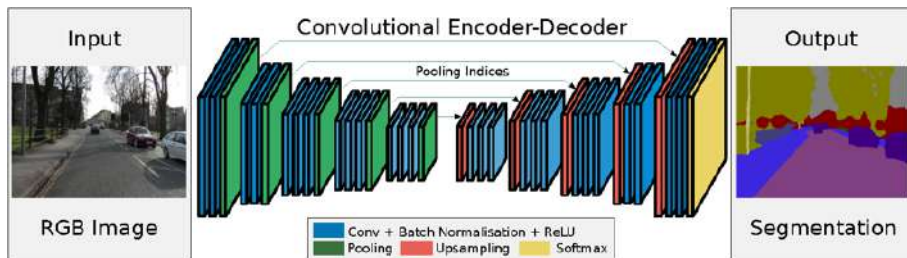
hiperspektralne podatke. Segmentacija kruničnih listova paradajza iz hiperspektralnih slika predstavlja ključni korak u razvoju automatizovanog sistema za procenu osetljivosti paradajza na gljivične infekcije uz minimalnu intervenciju čoveka. Precizna segmentacija vrhova kruničnih listova može značajno unaprediti tačnost prediktivnih modela za ranu procenu rizika od neželjenog razvoja gljivičnih oboljenja. U ovom delu doktorske disertacije analizirane su reprezentativne arhitekture dubokog učenja za segmentaciju slike, koje su široko prihvaćene i često korišćene u literaturi (U-Net 2.10 [88] i Seg-Net 2.11 [89].), kako bismo identifikovali koja od njih omogućava najprecizniju semantičku segmentaciju kruničnih listova paradajza na osnovu hiperspektralnih slika, ali i uporedili preciznost sa arhitekturom koja je razvijena u okviru ove doktorske disertacije.

Za zadatak segmentacije upoređene su performanse tri arhitekture



Slika 2.10: U-Net arhitektura

neuronskih mreža: našeg predloženog, novog optimizovanog konvolucionog modela nazvanog (oFCN) razvijenog u [90], a adaptiranog u [90], 2.9, kao i dve arhitekture koje se ubrajaju među najčešće korišćene u semantičkoj segmentaciji: U-Net 2.10 [88] i Seg-Net 2.11 [89]. Analiza je sprovedena na prvom skupu hiperspektralnih podataka koji obuhva-



Slika 2.11: Seg-Net arhitektura

ta 36 hiperspektralnih slika sa 224 spektralna kanala u opsegu od 900 do 1700 nm (bliski infracrveni deo spektra), sa spektralnom rezolucijom od 3 nm. Podaci su prikupljeni šest dana pre nego što gljivična infekcija postane vidljiva golim okom. Referentno označavanje regiona od interesa izvršeno je korišćenjem pseudokolor snimaka generisanih na osnovu spektra. Za sve mreže primenjen je identičan postupak obuke, uključujući istu podelu podataka na skupove za treniranje, validaciju i testiranje. Performanse modela procenjujane su na osnovu F1 skor metrike, broja epoha potrebnih za postizanje optimuma i broja parametara koji se uče tokom procesa obuke.

2.5 Segmentacija instanci

Proširenje istraživanja na segmentaciju na nivou instanci predstavlja logičan nastavak rada i odgovor na tehnička ograničenja semantičke segmentacije. Dok semantička segmentacija omogućava klasifikaciju svakog piksela prema pripadnosti određenoj klasi, ona ne razlikuje pojedinačne objekte unutar iste klase. U slučajevima kao što su klasovi pšenice sa značajnim preklapanjem ili krupni listovi paradajza sa kompleksnim prostornim rasporedom, ova razlika postaje ključna.

Primena pristupa segmentacije na nivou instance omogućava preciznije razdvajanje međusobno preklapljenih ili delimično zaklonjenih objekata, čime se postiže veća tačnost u kvantitativnim analizama — poput brojanja klasova ili izdvajanja pojedinačnih krupnih listova paradajza. Na tehničkom nivou, ovakav pristup zahteva naprednije arhitekture, koje kombinuju prostornu lokalizaciju i klasifikaciju na nivou piksela, čime se povećava robustnost i upotrebljivost razvijene metodologije u realnim, terenskim uslovima.

2.5.1 Pregled literature za segmentaciju na nivou instance u hiperspektralnim podacima

Kada dolazimo do oblasti analize i procene kvaliteta proizvoda, česte i vremenski zahtevne vizuelne inspekcije mogu pomoći u smanjenju prisustva inficiranih plodova na policama, ali takve metode nisu dovoljno precizne, ni pouzdane, naročito u najranijim fazama infekcije [91, 92, 93]. Stoga, razvoj nedestruktivnih metoda za procenu osetljivosti paradajza pre nego što posledice infekcije postanu vidljive ima ključnu ulogu u lancu snabdevanja. Ovaj deo istraživanja ima značajne praktične implikacije u oceni kvaliteta hrane. Unapređenjem detekcije inficiranih plodova paradajza, dobijeni rezultati mogu doprineti većoj preciznosti i efikasnosti procesa kontrole kvaliteta, obezbeđujući da do maloprodajnih polica stižu samo proizvodi odgovarajućeg kvaliteta. Iako su prethodne studije istraživale različite metode detekcije bolesti paradajza, i dalje postoje značajni nedostaci u sposobnosti efikasne identifikacije infekcije u maloprodajnom lancu. Takođe, većina prethodnih istraživanja fokusirala se na opštu detekciju bolesti, bez razmatranja specifičnih implikacija za ocenu kvaliteta hrane u preciznoj poljoprivredi. Precizna poljoprivreda, zasnovana na tehnologijama poput IoT sistema i mašinskog učenja, omogućava proizvođačima da optimizuju prinos i smanje troškove, obezbeđujući podatke potrebne za donošenje odluka [94, 95, 96, 97]. Nadovezujući se na ove napretke, ovo istraživanje popunjava pomenuti jaz uvođenjem efikasnijeg, automatizovanog pristupa zasnovanog na hiperspektralnom snimanju, koje omogućava preciznu detekciju kruničnih listova paradajza – ključnog koraka u identifikaciji inficiranih plodova u maloprodajnim uslovima, sa potencijalom za primenu u realnom vremenu u procesima kontrole kvaliteta.

Neinvazivno hiperspektralno snimanje se pokazalo kao izuzetno efikasna, nedestruktivna tehnika koja prevazilazi tradicionalne RGB slike, jer obezbeđuje više detaljnih informacija kroz desetine uzastopnih spektralnih kanala u nevidljivom delu elektromagnetnog spektra [98]. U kombinaciji sa metodama mašinskog učenja, nadgledanim i nenadgledanim, HSI se koristi za razvoj preciznijih modela detekcije biljnih bolesti. U [99] autori su pokazali da je blisko-infracrveni (NIR) opseg (700 nm–1300 nm) pružio vrednije informacije za detekciju bolesti paradajza u poređenju sa vidljivim opsegom (350 nm–700 nm), koristeći *Extreme Learning Machine (ELM)* klasifikator. Značaj NIR spektra

potvrđen je i u [100], gde je razvijena metoda koja uključuje *Minimum Noise Fraction (MNF)* transformaciju [101], višedimenzionalnu vizualizaciju [102], izbor spektralnih endmember-a [103] i *Spectral Angle Mapping (SAM)* [104]. Osetljivost hiperspektralnih odziva na određene oblike oštećenja ćelija takođe je potvrđena u različitim studijama – npr. u detekciji modrica kod kivija [105], gde su korišćene metode za redukciju dimenzionalnosti PCA i paralelopiped klasifikacija [106], ili kod borovnica, gde su autori koristili *Support Vector Machine (SVM)* [107, 108]. Dalje, [109] su kombinovali hiperspektralne podatke sa PLS-DA i SVM modelima kroz primenu strategije za fuziju podataka, dok su [110] koristili SAM, multinomijalnu logističku regresiju (MLR) i odluke stabla klasifikacije. Za rano otkrivanje bolesti borovnica, [111] su primenili PLS-DA na dva spektralna opsega (400 nm – 1000 nm i 685 nm – 1000 nm), pokazujući da uži spektar pruža bolje rezultate. U [112] predložen je unapređeni 3D CNN rezidualni model za rano otkrivanje degeneracije ploda borovnica. Pored toga, u [113] razvijena je *Spectral-Spatial Network (SSNet)* za detekciju nečistoća u duvanu, dok je u [114] primenjena rekonstrukcija RGB slika iz hiperspektralnih slika i primena dubokog učenja za klasifikaciju sorti semena kukuruza. Konačno, u [1] korišćene su hiperspektralne slike za ranu predikciju osetljivosti kruničnog lista paradajza na gljivične infekcije, pre pojave vidljivih simptoma, upotrebom *XGBoost* i *Random forest* regresionih modela. Međutim, primena hiperspektralnih podataka nosi određene izazove, poput osetljivosti na atmosferske uslove ili šum senzora, promene osvetljenja i problema spektralne mešavine, što ograničava broj čistih spektralnih odziva zbog niske prostorne rezolucije i heterogenih struktura na ivicama slike [98]. Visoka dimenzionalnost ovakvih slika, iako omogućava finu spektralnu diskriminaciju, zahteva značajne memorijske i računarske resurse [115]. Sličnost i redundansa među susednim kanalima dodatno otežavaju detekciju regiona od interesa, naglašavajući značaj razvoja strategija za efikasno upravljanje hiperspektralnim podacima [116]. Redukcija redundantnih informacija obično se postiže mapiranjem spektralnih kanala u podprostore korišćenjem informativnijih i manje korelisanih kombinacija [116, 117]. Nenadgledane metode, poput PCA i ICA (*Independent Component Analysis*), često se koriste pre primene dubokih neuronskih mreža [98, 116, 118]. Na primer, [116, 119] primenjuju PCA za ekstrakciju reprezentativnih karakteristika, dok su [120] koristili sukcesivni algoritam projekcija [121] za izbor najznačajnijih talasnih dužina i teksturalnih karakteri-

stika. CNN arhitekture su korišćene nakon redukcije dimenzionalnosti [122, 123], ali njihov učinak u detekciji malih objekata može biti ograničen zbog operacija snižavanja rezolucije i gubitka prostorne informacije [123, 124]. Dodatno, složeni modeli za detekciju objekata u hiperspektralnim slikama podložni su fenomenu paradoksa pretreniranja (engl. *peaking paradox*) [115, 125, 126, 127], pri čemu porast broja karakteristika povećava složenost modela i dovodi do većih grešaka u proceni. Nedostatak velikih anotiranih skupova podataka ostaje prepreka u primeni dubokog učenja na hiperspektralnim slikama, ali strategije poput augmentacije podataka nude potencijalna rešenja [128, 129, 130].

Studija prikazana u ovom delu doktorske disertacije predstavlja nastavak istraživanja iz [1], gde je razvijen poluautomatski postupak segmentacije regije kruničnih listova i peteljke paradajza. Predloženi pristup segmentacije pruža generalniju i robusniju metodologiju u odnosu na parametarske metode korišćene u [1], koje su često zavisne od skupa podataka i slabije su prilagodljive na nove podatke iz proširenog skupa podataka. Dok je procena osetljivosti na gljivične infekcije demonstrirana u prethodnoj studiji, razvijena metodologija unapređuje izdvajanje kruničnih listova na fleksibilniji i skalabilniji način, poboljšavajući reproduktivnost rezultata i omogućavajući širu primenu na različitim skupovima podataka.

U okviru ovog dela doktorske disertacije predlaže se upotreba hiperspektralnih snimaka za automatsku segmentaciju na nivou instanci kruničnih listova paradajza, zasnovanu na *Mask R-CNN* modelima sa različitim pridruženim, sekundarnim arhitekturama. U kontekstu hiperspektralnih slika, gde se generišu desetine spektralnih kanala, cilj je izdvojiti ključne informacije iz spektra i ispitati kako različite metode izdvajanja latentnih hiperspektralnih informacija utiču na performanse *Mask R-CNN* modela. Iako pregled literature [131] u određenoj meri pokriva ovu temu, u ovom radu se detaljno istražuje uticaj metoda kodovanja spektralnih informacija na uspešnost detekcije kruničnog lista paradajza na hiperspektralnim slikama.

Istraživači kontinuirano istražuju i predlažu nove arhitekture dubokih neuronskih mreža (DNN) kako bi odgovorili na izazov što preciznije klasifikacije objekata unutar slike kroz pružanje segmentacije na nivou piksela i precizne procene granica objekata [132, 133, 134, 135, 136]. Konkretno, konvolucione neuronske mreže (CNN) pokazale su veliki uspeh u zadacima segmentacije na nivou instanci, mahom na RGB slikama [53, 137, 138]. *Region-based CNN* (R-CNN) [139] uveo je kon-

cept *region-based* detekcije objekata, postavljajući osnovu za kasniji razvoj. *Fast R-CNN* [140] optimizovao je brzinu ovog procesa uvođenjem efikasnijeg okvira detekcije u jednoj fazi. *MultiPath Network* iz [141] koristila je višeskanalne karakteristike za poboljšanje prepoznavanja i segmentacije objekata. *Faster R-CNN* pionirski je primenio *Region Proposal Network* (RPN) za bržu detekciju objekata [52]. Mask R-CNN [142] nadogradio je temelje Faster R-CNN metode uvođenjem segmentacije na nivou instanci pored nivoa piksela. *Non-local* neuronske mreže [143] unapredile su razumevanje slike kroz uvezivanje dugodosežnih zavisnosti karakteristika. PANet [144] predložila je piramidalnu strukturu arhitekture za bolje objedinjavanje karakteristika. *Hybrid Task Cascade* [145] unapredila je performanse kaskadnim izvođenjem više zadataka, poput detekcije i segmentacije na nivou instanci. GCNet [146] uključila je globalni kontekst u konvolucione mreže radi poboljšanja prepoznavanja objekata. Yolact [147] otvorio je put ka segmentaciji instanci u realnom vremenu kroz okvir detekcije u jednoj fazi. Tensor Mask R-CNN model iz [148] uveo je tenzorski pristup za efikasnu segmentaciju na nivou instanci. U radu [149], autori su predstavili arhitekturu zasnovanu na transformerima koja direktno predviđa poligone koristeći Mask R-CNN segmentaciju na nivou instanci kao referentnu anotaciju za računanje funkcije gubitka. Takođe, u [150] predložen je jednostavan i kompaktan ViT model pod nazivom *Universal Vision Transformer* (UViT), koji koristi konstantnu rezoluciju karakteristika i dimenziju skrivenog sloja u svim enkoderskim blokovima. Tokom ovog iterativnog procesa razvoja arhitektura, DNN modeli su pokazali značajna poboljšanja u performansama segmentacije na nivou instanci, unapredili računarsku efikasnost i proširili domen primene od medicinskog snimanja [151] i autonomne vožnje [152] do precizne poljoprivrede [153].

Za hiperspektralne podatke, neuronske mreže mogu se koristiti za zadatke kao što su ekstrakcija spektralno-prostornog obeležja [154, 155, 156] i segmentacija [156, 157, 158], gde se koriste za identifikaciju i segmentaciju pojedinačnih piksela ili regiona koji odgovaraju specifičnim materijalima ili površinama na sceni. Nedavna istraživanja predlažu metode adaptacije modela segmentacije na nivou instance zasnovanih na dubokom učenju za primenu na hiperspektralne podatke, kao što su višeskalne mreže [154, 155, 159], spektralno-prostorna fuzija [129, 154, 160] ili modeli zasnovani na *Attention* mehanizmima [143, 161, 162]. Ove metode imaju za cilj iskorišćavanje spektralne

i prostorne informacije u hiperspektralnim podacima i prevazilaženje ograničenja modela dubokog učenja za visokodimenzionalne podatke. Višeskanalne mreže se koriste za čuvanje prostorne informacije na različitim skalama, dok se metode spektralno-prostorne fuzije primenjuju radi integracije spektralnih i prostornih informacija sa ciljem poboljšanja performansi u zadacima segmentacije. Modeli zasnovani na *Attention* mehanizmima ponderišu značaj različitih karakteristika ili kanala za zadatak segmentacije na nivou instance.

Postojeći modeli za ovakvu segmentaciju zasnovani na dubokom učenju obično su unapred obučeni na velikim RGB skupovima podataka i mogu se dodatno prilagoditi (engl. *fine-tuning*) za specifične zadatke na manjim skupovima podataka. Ovi unapred obučeni modeli mogu se dalje koristiti kao početna tačka za različite zadatke bez potrebe za potpunim obučavanjem dubokih neuronskih mreža od nule, što rezultira uštedom vremena za obuku i računarskih resursa, kao i u poboljšanju performansi finalnog modela. Na ovaj način, unapred obučeni modeli mogu se iskoristiti za primenu znanja stečenog na jednostavnim RGB slikama iz otvorenih skupova podataka i primeniti ga na hiperspektralne slike. Ovaj pristup je naročito koristan u slučajevima kada je broj označenih hiperspektralnih slika ograničen.

Ovaj deo doktorske disertacije istražuje integraciju hiperspektralnih slika sa segmentacijom na nivou instanci zasnovanom na dubokom učenju kroz primenu *Mask R-CNN* arhitektura. Konkretno, tri *backbone* Mask R-CNN mreže: *Faster R-CNN*, *Deformable ConvNet* i *FPN*. *Backbone* predstavlja primarni modul neuronske mreže i ključni deo arhitekture koji izdvaja najvažnije karakteristike iz podataka i obezbeđuje reprezentaciju informacija visokih dimenzija, koja se koristi u posledičnim slojevima za klasifikaciju, detekciju i segmentaciju. U daljem tekstu, radi jednostavnosti se usvaja pojam „backbone“. Ove mreže analizirane su u kombinaciji sa četiri tehnike za izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama — PCA (*Principal Component Analysis*) i njena probabilistička verzija PPCA (*Probabilistic Principal Component Analysis*), ICA (*Independent Component Analysis*) i NMF (*Non-negative Matrix Factorization*) — primenjene na hiperspektralnim skupu podataka *Brioso* sorte paradajza. Ove tehnike su detaljnije objašnjene u narednom poglavlju. Cilj je sistematski proceniti uticaj postojećih metoda za izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama na performanse mreža za segmentaciju na nivou instanci. Radi ocene sposobnosti generalizacije

modela, metod transfer učenja je primenjen na sekundarni skup podataka koji obuhvata tri dodatne sorte paradajza: *Capricia*, *Province* i *Sao Paolo*. Pored toga, u ovom delu doktorske disertacije predlaže se dizajn prilagođenog konvolucionog sloja inicijalizovanog koeficijentima dobijenih metodom kodovanja spektralnih informacija, posebno iz NMF metode, kako bi se omogućio *end-to-end* adaptivni sistem učenja za robusnu i preciznu segmentaciju instanci kruničnih listova paradajza u hiperspektralnim podacima.

U nastavku je dat pregled najčešće korišćenih tehnika redukcije dimenzionalnosti, stavljenih u kontekst efikasnosti za izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama.

2.5.2 Izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama

Hiperspektralna slika može se predstaviti kao tenzor trećeg reda $\mathbf{X} \in \mathbb{R}^{R \times C \times D}$, gde R, C označavaju prostorne dimenzije (tj. širinu i visinu slike), dok D predstavlja spektralnu dimenziju (tj. broj talasnih dužina). Pikseli hiperspektralne slike, povezani sa D -dimenzionalnim spektralnim merenjima, odgovaraju vektorima $\mathbf{x}_n \in \mathbb{R}^D, n = 1, \dots, RC$ unutar tenzora $\mathbf{X}$. Iako veće vrednosti D , a samim tim i veći broj obeležja u spektralnoj oblasti, mogu imati značajan uticaj na analizu posmatranih fenomena (npr. biljne bolesti, gljivične infekcije) [163, 164], one takođe uvode izazove u pogledu računarske složenosti pri obradi hiperspektralnih slika. Stoga je neophodno pronaći balans između potrebnih računarskih resursa i količine informacija koje se zadržavaju u spektru. Pregled metoda za izbor najinformativnijih spektralnih opsega i tehnika za izdvajanje latentnih spektralnih komponenti hiperspektralne informacije u zavisnosti od primene hiperspektralnih slika dat je u [131].

U okviru ove doktorske disertacije predlaže se upotreba četiri tehnike za izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama: PCA (*Principal Component Analysis*) i njena probabilistička verzija PPCA (*Probabilistic Principal Component Analysis*), ICA (*Independent Component Analysis*) i NMF (*Non-negative Matrix Factorization*). Sve četiri tehnike uobičajeno se koriste kao inicijalni koraci predprocesiranja u analizi hiperspektralnih slika, sa ciljevima kao što su redukcija dimenzionalnosti, izdvajanje latentnih spektralnih komponenti hiperspektralne informacije, poboljšanje odnosa signal-šum i

smanjenje složenosti naknadne analize podataka i zadataka mašinskog učenja [131].

PCA, poznata i kao Karhunen-Loève transformacija, široko je korišćena tehnika za redukciju dimenzionalnosti ili ekstrakciju obeležja [165]. Prema [166], PCA se može definisati kao ortogonalna linearna projekcija podataka u nižedimenzionalni, tzv. prostor glavnih komponenti, koja maksimizuje varijansu unutar njega. Posmatrajmo skup hiperspektralnih odziva po pikselu $\{\mathbf{x}_n \in \mathbb{R}^D | n = 1, 2, \dots, N\}$ gde $N = RCL$, (tj. RC je širina i visina slike, a L broj slika). Označimo sa $\bar{\mathbf{x}}$ vektor srednjih vrednosti uzorka. Kovarijaciona matrica $\mathbf{M} \in \mathbb{R}^{D \times D}$ računa se kao:

$$\mathbf{M} = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T. \quad (2.15)$$

Faktorizacijom matrice $\mathbf{M}$ putem karakteristične dekompozicije dobija se:

$$\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \quad (2.16)$$

gde je $\mathbf{U}$ ortogonalna matrica u $\mathbb{R}^{D \times D}$ čije kolone predstavljaju karakteristične vektore matrice $\mathbf{M}$, dok je $\mathbf{\Lambda} \in \mathbb{R}^{D \times D}$ dijagonalna matrica sa karakterističnim vrednostima na dijagonali. Odabirom $K < D$ dominantnih karakterističnih vektora $\mathbf{u}_1, \dots, \mathbf{u}_K$ povezanih sa najvećim karakterističnim vrednostima $\{\lambda_1, \dots, \lambda_K\}$ formira se ortogonalna projekciona matrica $\mathbf{U}_K$. Spektralna informacija u $\mathbf{x}_n$ tada se koduje u K glavnih komponenti:

$$\mathbf{z}_n = \mathbf{U}_K^T (\mathbf{x}_n - \bar{\mathbf{x}}) \quad (2.17)$$

gde $\mathbf{z}_n \in \mathbb{R}^K$ predstavlja reprezentaciju $\mathbf{x}_n$ u K -dimenzionalnom prostoru glavnih komponenti.

Probabilistička reformulacija PCA (PPCA) predstavljena u [167] blisko je povezana sa faktorskom analizom [168, 169]. U okviru PPCA, uvodi se latentna promenljiva $\mathbf{z}_n \in \mathbb{R}^K$ koja odgovara prostoru glavnih komponenti generisanom kolonama matrice $\mathbf{W} \in \mathbb{R}^{D \times K}$. Povezanost sa posmatranom promenljivom $\mathbf{x} \in \mathbb{R}^D$ ostvaruje se linearnom transformacijom:

$$\mathbf{x} = \mathbf{W}\mathbf{z} + \boldsymbol{\epsilon}, \quad (2.18)$$

gde je $\boldsymbol{\epsilon} \in \mathbb{R}^D$ i $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$. Na osnovu Bayes-ovog pravila, posteriorna raspodela dobija se kao:

$$p(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{C}^{-1}\mathbf{W}^T(\mathbf{x} - \boldsymbol{\mu}), \sigma^2 \mathbf{C}^{-1}),$$

gde važi

$$\mathbf{C} = \mathbf{W}^T \mathbf{W} + \sigma^2 \mathbf{I}.$$

Sa procenom maksimalne verodostojnosti parametara $p(\mathbf{z}|\mathbf{x})$, latentna reprezentacija $\mathbf{x}$ može se aproksimirati očekivanjem:

$$\mathbb{E}_{p(\mathbf{z}|\mathbf{x})}[\mathbf{z}] = \mathbf{C}^{-1} \mathbf{W}_{ML}^T (\mathbf{x} - \boldsymbol{\mu}_{ML}). \quad (2.19)$$

Za razliku od PCA i PPCA, koji su zasnovani na maksimizaciji varijanse, ICA ima za cilj pronalaženje statistički nezavisnih komponenti u hiperspektralnom odzivu, što može biti naročito korisno za razdvajanje spektralnih odziva koji pripadaju različitim materijalima [170, 171, 172]. Posmatra se centrirana matrica podataka $\mathbf{Y} \in \mathbb{R}^{D \times N}$ definisana kao:

$$\mathbf{y}_n = \mathbf{x}_n - \bar{\mathbf{x}}. \quad (2.20)$$

ICA modeluje matricu $\mathbf{Y}$ kao:

$$\mathbf{Y} = \mathbf{A}\mathbf{S}, \quad (2.21)$$

gde $\mathbf{A} \in \mathbb{R}^{D \times D}$ predstavlja matricu mešanja, dok $\mathbf{S} \in \mathbb{R}^{D \times N}$ sadrži nezavisne komponente. Cilj je pronalaženje ortogonalne razmešavajuće matrice $\mathbf{Q} \in \mathbb{R}^{D \times D}$ takve da:

$$\mathbf{S} = \mathbf{Q}\mathbf{Y}. \quad (2.22)$$

Poslednja tehnika primenjena u ovoj studiji za izdvajanje latentnih spektralnih komponenti hiperspektralne informacije jeste NMF, čiji je cilj faktORIZACIJA matrice podataka $\mathbf{X}^{D \times N}$ u dve nenegativne matrice $\mathbf{V} \in \mathbb{R}^{D \times K}$ i $\mathbf{H} \in \mathbb{R}^{K \times N}$:

$$\mathbf{X} = \mathbf{V}\mathbf{H}, \quad (2.23)$$

pri čemu K predstavlja željenu redukovanu dimenzionalnost. Nenegativnost $\mathbf{V}$ i $\mathbf{H}$ omogućava interpretaciju u smislu aditivnih komponenti. Nakon optimizacije i minimizacije greške rekonstrukcije, spektar posmatranog piksela $\mathbf{x}_n \in \mathbb{R}^D$ može se aproksimirati u redukovanom K -dimenzionalnom prostoru kao:

$$\mathbf{z}_n \approx \mathbf{H}^T \mathbf{x}_n, \quad (2.24)$$

gde matrica $\mathbf{H} \in \mathbb{R}^{D \times K}$ sadrži latentne NMF komponente, dok $\mathbf{z}_n \in \mathbb{R}^K$ predstavlja redukovanu (latentnu) reprezentaciju piksela $\mathbf{x}_n$ u prostoru NMF komponenti.

U ovom delu istraživanja razmatran je problem adaptacije hiperspektralnih slika za treniranje i testiranje postojećih DNN modela za segmentaciju na nivou instanci u zadatku detekcije kruničnih listova paradajza. Iako je glavna prednost ove vrste podataka bogatstvo spektralnih informacija, što omogućava široku primenu, njihova obrada i dalje zahteva visoke računarske resurse. Napredni DNN modeli za segmentaciju instanci uglavnom su prethodno trenirani na RGB slikama [173], [174], [175]. Primena pristupa transfera učenja na ovakve unapred obučene modele za segmentaciju na nivou instanci, radi detekcije kruničnih listova u hiperspektralnim slikama, zahteva kompresiju spektralnih informacija iz visokodimenzionalnog u latentni prostor od najmanje tri dimenzije. Konkretno, metodologija u ovom istraživanju uključuje upotrebu linearnih tehnika radi efikasnog kodovanja spektralnih informacija u šumom redukovanim hiperspektralnim slikama, uz očuvanje najznačajnijih spektralnih komponenti koje su zatim korišćene za dodatno fino podešavanje unapred obučanih DNN modela na bazi transfera učenja.

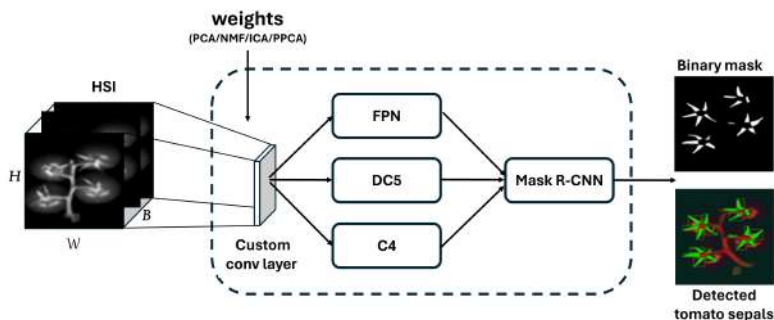
Ovaj pristup predstavlja širi koncept transfera znanja u oblasti mašinskog učenja, koji je stekao popularnost zbog širokog spektra primena [176]. *Transfer learning* podrazumeva unapređivanje procesa učenja u novom zadatku korišćenjem znanja stečenog u srodnom, već rešenom zadatku, radi poboljšanja performansi ciljnog zadatka, što je ujedno i demonstrirano u ovoj studiji. Osnovni princip transfera znanja uključuje mapiranje karakteristika jednog zadatka na drugi, uspostavljaajući korespondencije koje omogućavaju viši inicijalni nivo performansi, strmiju krivu učenja ili bolji konačni rezultat, uz izbegavanje negativnog transfera (situacije kada prethodno naučeno znanje degradira performanse modela u novom zadatku) i automatizaciju procesa mapiranja [177]. Na ovaj način, transfer znanja može prevazići ograničenja vezana za mali broj podataka, smanjiti vreme treniranja i poboljšati performanse modela mašinskog učenja u različitim domenima.

Na osnovu rezultata primenjenih linearnih tehnika za izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama, započeto je fino podešavanje unapred obučanih dubokih modela za detekciju i segmentaciju kruničnih listova paradajza u hiperspektralnim snimcima. Fokus ovog dela istraživanja je na *Mask R-CNN* arhitekturi [142], koja predstavlja dvostepeni i dobro poznati model za zadatke segmentacije na nivou instance. Ova arhitektura je razvijena na osnovu Faster R-CNN arhitekture, uz dodatnu granu koja generiše maske objekata,

paralelno sa predikcijama, tj. *bounding-boxes* [52]. Na taj način, *Mask R-CNN* istovremeno detektuje i segmentuje više objekata unutar slike. Ova arhitektura se sastoji iz dva glavna dela: osnovne mreže (engl. *backbone*) i dva podmrežna modula specifična za određene zadatke. Osnovna mreža, zasnovana na unapred obučenim CNN modelima poput ResNet-a [28], ekstrahuje karakteristike ulazne slike na visokom nivou, koje se dalje prosleđuju kroz dva paralelna podmodula: mrežu za predlaganje regiona (RPN), koja predlaže potencijalne objekte kroz predikciju graničnih ivičnih pravougaonika (engl. *bounding boxes*), i drugu mrežu, koja na osnovu predloženih regiona vrši rangiranje graničnih pravougaonika, istovremeno predviđajući klasne verovatnoće za svaki objekat. Pored toga, generiše se binarna maska za svaki region, čime se omogućava precizna segmentacija objekata na nivou piksela, ali i poređenje sa pristupom semantičke segmentacije. Budući da izlaz *Mask R-CNN* modela sadrži maske objekata, ovaj pristup je opravdano i optimalno rešenje za zadatke koji zahtevaju visoku preciznost lokalizacije i segmentacije, poput segmentacije kruničnih listova paradajza ili klasova pšenice.

Dodatna prednost *Mask R-CNN* arhitekture je njena pogodnost za primenu transfer znanja i prilagodljivost novim problemima [178]. Modeli korišćeni u ovoj doktorskoj disertaciji unapred su trenirani na Microsoft COCO skupu podataka, koji predstavlja veliki korpus sa oko 200k slika [174]. Korišćene su tri varijante osnovnih mreža: FPN (ResNet + FPN), C4 (ResNet) [52] i DC5 ili Dilated-C5 (ResNet sa dilatacijama u conv5 sloju) [179]. Izbor ovih varijanti bio je vođen specifičnostima svake od arhitektura: FPN je izabran zbog multi-skalarne ekstrakcije karakteristika, C4 (ResNet) zbog robusnih performansi, a DC5 (Dilated-C5) zbog pogodnosti za zadatke koji zahtevaju gustu predikciju i visoku prostornu preciznost. Predloženi proces detekcije kruničnih listova paradajza i segmentacije njihovih regiona prikazan je na Slici 2.12, počevši od kodovanja spektralnih informacija, a zatim kroz fino podešavanje *Mask R-CNN* modela sa tri različite osnovne mreže.

Inicijalno istraživanje iz ovog dela doktorske disertacije koje obuhvata segmentaciju kruničnih listova paradajza na nivou instance na hiperspektralnim podacima, delom je objavljeno u [86]. U narednoj studiji, koja je sprovedena na istom skupu podataka kao i u prethodnoj istraživačkoj fazi [1, 84, 85], fokus je bio i na samoj dimenzionalnosti hiperspektralnih slika. Dimenzionalnost spektralnih slika je redukova-



Slika 2.12: Tok obrade hiperspektralnih slika korak po korak u procesu segmentacije na nivou instanci

na na tri kanala korišćenjem metode *Principal Component Analysis* (PCA). Na zadatku segmentacije kruničnih lista testirana su tri osnovna Mask R-CNN modela: ResNet + FPN (*Feature Pyramid Network*), ResNet + C4 (originalni *Faster R-CNN* pristup) i DC5 (*Dilated ConvNet*), uz varijacije u broju epoha, pri čemu su svi modeli inicijalno trenirani na COCO skupu, velikom skupu podataka za detekciju i segmentaciju objekata.

Za svaki model primenjena je ista procedura obuke, pri čemu je korišćeno 15 slika za primenu *transfer learning-a* i podešavanje arhitektura sa unapred obučanim modelima. Performanse segmentacije evaluirane su pomoću standardne metrike mAP (*Mean Average Precision*). Najpreciznija segmentacija postignuta je korišćenjem Mask R-CNN ResNet+C4 modela, sa mAP vrednošću od 92,6% na nezavisnom test skupu.

Proširena studija vezana za segmentaciju na nivou instance kruničnih listova paradajza, zajedno sa analizom metoda za efikasno izdvajanje latentnih spektralnih komponenti visokodimenzionalnih hiperspektralnih slika iz ovog dela doktorske disertacije, detaljno je objašnjena i objavljena u [87].

Osim same segmentacije, kada su u pitanju hiperspektralni podaci, neophodno je adresirati problem vezan za izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama radi povećanja efikasnosti automatizovanih sistema koji koriste visokodimenzionalne hiperspektralne podatke. U okviru ove doktorske disertacije ispitane su četiri različite tehnike: *Principal Component Analysis* (PCA), *Independent Component Analysis* (ICA), *Probabilistic PCA* (PPCA) i *Non-negative*

Matrix Factorization (NMF), sa ciljem primene *transfer learning*-a za segmentaciju na nivou instanci krubičnih listova paradajza korišćenjem modela prethodno obučenih na RGB slikama.

Sprovedena je komparativna analiza tri *Mask Region-based Convolutional Neural Network* (Mask R-CNN) modela sa različitim backboneovima: *Faster R-CNN*, *Deformable ConvNet* i *Feature Pyramid Network* (FPN) na spektralno izdvojenim reprezentacijama hiperspektralnih slika u trodimenzionalnom prostoru sorte paradajza *Brioso*. Model Mask R-CNN sa FPN, integrisan sa NMF tehnikom, postigao je najviši nivo tačnosti, dostižući prosečnu vrednost preciznosti *Mean Average Precision* (mAP) od 94,05%.

Na drugom skupu podataka, koji je uključivao dodatne tri sorte paradajza: *Capricia*, *Provine* i *Sao Paolo*, isti model je ostvario mAP od 86,42% za sve sorte, sa samo jednom lažno pozitivnom detekcijom. Pored toga, implementiran je prilagođeni konvolucioni sloj inicijalizovan procenjenim NMF koeficijentima, čime je postignut mAP rezultat od 87,40%.

Ovi rezultati potvrđuju potencijal integracije izdvajanja latentnih hiperspektralnih informacija sa unapred obučenim modelima dubokog učenja na RGB slikama za segmentaciju na nivou instanci, omogućavajući robusnu i preciznu primenu transfera učenja za potrebe automatizovanih sistema za procenu kvaliteta prehrambenih proizvoda u poljoprivredi.

Pristup segmentacije na nivou instance primenjen je i na drugoj studiji slučaja, detekciji klasova pšenice. U daljem tekstu osvrnućemo se na kratak opis problema, sa fokusom na relevantnu literaturu u oblasti detekcije klasova pšenice.

2.5.3 Pregled literature za segmentaciju na nivou instance za detekciju klasa pšenice u RGB slikama

Prema [49], broj klasova pšenice i masa hiljadu zrna u okviru kvadratnog metra predstavljaju najznačajnije parametre koji doprinose proceni prinosa. Konkretno, broj klasova po jedinici površine jedan je od najvažnijih agronomskih pokazatelja prinosa, budući da ostvaruje najvišu korelaciju sa ukupnim stvarnim prinosom [20]. Konvencionalne metode za procenu prinosa pšenice [180] zasnivaju se na poređenju performansi useva tokom prethodnih godina ili na ručnom brojanju

klasova u regionu od interesa (ROI), zbog čega nisu dovoljno precizne i sklone su subjektivnoj proceni [181, 182]. Ručno brojanje klasova podrazumeva ograničavanje ROI-a pomoću rama površine jednog kvadratnog metra (ili manje), koji se postavlja preko pšenice [20, 181]. Za parcele do jednog hektara, ukupan broj klasova dobija se ekstrapolacijom rezultata iz jednog ROI-a, dok se za veće površine procedura brojanja ROI-a mora ponoviti na više lokacija [20]. Modeli za detekciju klasova pšenice, trenirani na osnovu ovakvih podataka, obično nemaju zadovoljavajući nivo generalizacije, jer ne postižu istu tačnost i preciznost na novim skupovima podataka i akvizicijama [2]. Pristup predložen u ovom delu doktorske disertacije omogućava automatizovanu detekciju i brojanje klasova, što značajno smanjuje potrebu za ručnim radom koji je vremenski zahtevan, fizički naporan i sklon greškama, otežavajući široku primenu u nekontrolisanim uslovima na terenu [90, 183].

Dodatno treba naglasiti da trenutno postoji jako mali broj dostupnih skupova podataka za brojanje klasova pšenice, najčešće prikupljenih od strane pojedinačnih istraživačkih timova, što direktno uvodi pristrasnost i ograničenja za primenu, kao što su određeni genotipovi i specifični uslovi akvizicije. Pored toga, proces anotacije je izuzetno zahtevan i često se razlikuje među institucijama, što zahteva dodatno prilagođavanje za dalje korišćenje na zajedničkim i proširenim javno dostupnim skupovima podataka [2]. Proizvođači i oplemenjivački programi u okviru istraživačkih institucija mogu imati koristi od informacija dobijenih brzim predviđanjem prinosa na osnovu automatskog brojanja klasova [183, 184].

Glavni doprinosi ovog dela teze su: kreiranje novog skupa podataka sa četiri sorte pšenice iz lokaliteta u Srbiji, kao i metodologija transfera znanja zasnovana na dubokom učenju za detekciju klasova na kreiranom skupu podataka.

Usled brzog razvoja i unapređenja tehnologije senzora, kao i naprednih algoritama za obradu slike, pristupi koji se baziraju na kompjuterskoj viziji i veštačkoj inteligenciji na podacima postepeno se uvode i primenjuju u poljoprivredi. Primene su usmerene na rešavanje različitih problema u poljoprivredi [56, 185, 186], kao što su detekcija korova [187, 188], identifikacija biljaka, fenotipizacija biljaka [189, 190] i detekcija bolesti [191, 192, 193, 194, 195, 196].

Na primer, u radu [197] opisana je automatska identifikacija biljnih vrsta [58], dok je u [198, 199] istražena ekstrakcija fenotipskih karakteristika vrste *Arabidopsis Thaliana*, a u [12, 37, 38] detekcija klasova

pšenice u kontrolisanim, laboratorijskim uslovima.

Međutim, identifikacija klasova pšenice predstavlja složen problem u oblasti kompjuterske vizije, prvenstveno zbog varijabilnosti uslova akvizicije slika, uglova posmatranja i različite orijentacije klasova, varijabilnosti izgleda u zavisnosti od fenološke faze, kao i inherentnih genetskih razlika. Pored toga, proces akvizicije slika može biti otežan pojavom zamućenja usled kretanja ili vetra, kao i preklapanjem klasova u gustim populacijama, što dodatno otežava problem automatskog prebrojavanja klasova pšenice [2].

Paralelno sa neuronskim mrežama dizajniranim za semantičku segmentaciju, razvijene su mnoge arhitekture za segmentaciju instanci: Mask R-CNN [142], Mask Lab [200], PANet [144], Hybrid task cascade [145], GCNet [146], Yolact [147], Mask2Former [201], kao i arhitekture za detekciju objekata: RCNN [139], Fast RCNN [140], MultiPath Network [141], Faster RCNN [52], Non-local Neural Networks [143], Tensor Mask [148], Mask Scoring RCNN [202], YOLO verzije i druge.

U poslednje vreme, više autora je predložilo metode fenotipizacije biljaka zasnovane na dubokom učenju [12, 199, 203]. U [183] je prikazana upotreba CNN arhitekture za rešavanje zadatka segmentacije na nivou instanci klasova pšenice u kontrolisanim laboratorijskim uslovima. Najnoviji pristupi detekcije klasova pšenice zasnivaju se na: Faster R-CNN [11, 59, 204], YOLO [205], YOLOv3 [206], YOLOv4 [207], YOLOv5 [208, 209], YOLOv7 [210], YOLOv8 [211] i RIA-SpikeNet [212].

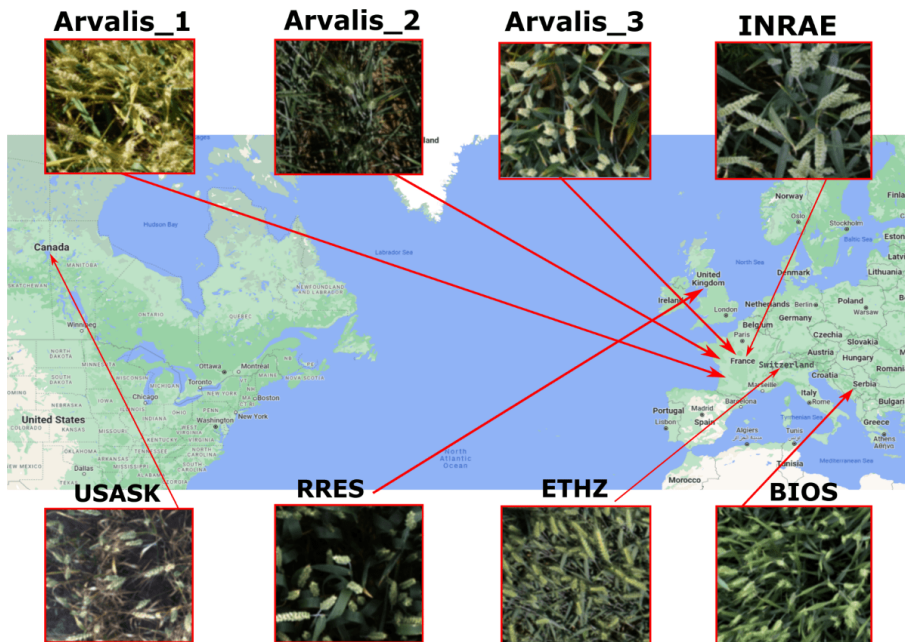
Ovi modeli su korišćeni na javno dostupnom skupu podataka kreiranom kroz *Global Wheat Head Challenge* [2], iz sedam različitih institucija. Trening podaci obuhvataju više od 3,000 slika iz Francuske, Velike Britanije, Švajcarske i Kanade, dok test podaci uključuju oko 1,000 slika iz Australije, Japana i Kine. Kratak pregled najnovijih pristupa dubokog učenja za detekciju klasova pšenice predstavljen je u [184].

U [213], autori razvijaju WheatNet, koji koristi MobileNetV2 [214] kao osnovu i dve dodatne mreže razvijene za svaki modul: za detekciju klasova i brojanje zasnovano na gustini. U [215] primenjen je YOLOv5 [216] kao osnova za detekciju klasova, postignut je prosečni F1 score od 0,93. Performanse DeepMAC modela segmentacije [217] takođe su prikazane u [215], sa prosečnim F1 score-om za detekciju od 0,93 i segmentaciju od 0,86. Metoda iz [218] zasnovana je na RetinaNet [219] i postiže mAP50 (Intersection over Union (IoU) sa primenjenim pragom 0,5) od 0,92.

Pristup u [220] koristi unapređeni EfficientDet-D0 model detekcije objekata [221] uz *transfer learning*, tj. prethodno treniranje *backbone* mreže za izdvajanje visokosloženih semantičkih osobina klasova. U istoj studiji predložena je metoda augmentacije slika *random-cutout*, uklanjajući deo identifikovanih *bounding box-ova* na osnovu broja klasova i njihove veličine, simulirajući preklapanje u realnom okruženju. U finalnoj arhitekturi, nakon što *backbone* suzbije irelevantne informacije, implementiran je dodatni CBAM modul, dostižući tačnost od 94%, što je oko 2% više u odnosu na originalni model.

CTWheatNet model iz [222] koristi različite *attention* module i postiže mAP50 od 93%, dok WH-DETR, jedan od novijih transformer-baziranih pristupa, dostiže 95,7%. RIA-SpikeNet iz [212] dizajniran je za detekciju i brojanje klasova u kompleksnim uslovima na terenu, nadmašujući YOLOv8, sa 81,54% mAP i 90,29% R².

Sve pomenute studije primarno koriste *Global Wheat Head Detection* (GWHD) skup podataka [2] za detekciju klasova pšenice.



Slika 2.13: Primer slika po izvorima podataka

U okviru ove doktorske disertacije, kreiran je skup podataka, na osnovu GWHD skupa, koristeći ga kao bazu. Glavni doprinos se ogleda u proširenju postojećeg skupa podataka povećanjem broja slika, sorti

i lokacija, uz unapređenje kvaliteta podataka.

BioS-Wheat skup podataka sadrži klasove pšenice sa četiri različita genotipa sa jedne evropske lokacije (vidi Sliku 2.13). BioS-Wheat proširuje postojeći GWHD skup podataka u pogledu genotipa, tipa senzora i lokacije. Cilj je da se adresira specifičan problem brojanja klasova pšenice, pružajući novi podskup koji može pomoći budućim istraživanjima u rešavanju ovog problema.

Duboke arhitekture neuronskih mreža koje su korišćene u ovom delu doktorske disertacije poklapaju se sa onima iz studije slučaja i primenu na hiperspektralnim slikama paradajza. To su modifikacije Mask-RCNN arhitekture sa tri različita *backbone*-a i dodatno, arhitektura RetinaNet modela, trenirane i na referentnom GWHD skupu i na proširenom skupu podataka. Dodatno, YOLOv8 i RT-DETR arhitekture korišćene su da demonstriraju doprinos novih podataka korišćenjem modifikacijom poslednjih razvijenih arhitektura za segmentaciju.

U ovom delu doktorske disertacije korišćeni su unapred obučeni modeli iz Detectron2 [223] i Ultralytics [224] platformi. Detectron2 je modularna PyTorch biblioteka razvijena za problem detekcije objekata, koja sadrži obučene modele na različitim javno dostupnim skupovima podataka i pruža osnovne (engl. *baseline*) rezultate za većinu njih.

Ultralytics je javno podelio YOLO (engl. *You Only Look Once*) serija evolucije za detekciju objekata u realnom vremenu i segmentaciju slika. Najnovije verzije YOLO modela imaju za cilj da unaprede prethodne verzije sa novim funkcionalnostima radi poboljšanja performansi, fleksibilnosti i efikasnosti.

Pošto je cilj ove studije detekcija klasova pšenice na RGB slikama, problem se može posmatrati kao problem detekcije objekata, slično početnom izveštaju sa GWHD skupom podataka [2], gde lokalizacija i brojanje klasova u RGB slikama zavise od region-baziranih konvolucionih neuronskih mreža (F-RCNN), koje se sastoje iz dve neuronske mreže: *Region Proposal Network (RPN)* i *Object Detection Network*. Ova arhitektura je postigla najbolji rezultat od 0,77 mAP50 sa prosečnim RMSE od 12,63 klasova po slici i R^2 0,57.

Isto tako, ista arhitektura je korišćena od strane tri najbolje rangiranih timova na GWHD izazovu [225, 226]. Pored ovog osnovnog modela, RetinaNet postiže najveću preciznost, od oko 90% za detekciju klasova, uključujući različite dodatne klasifikacione probleme, kao što su faza zrelosti klasova [227] ili fokus na klasove određene veličine na istom skupu podataka (GWHD) [218]. U radovima [210, 228, 229]

postignuti su vrlo visoki rezultati mAP50, pri čemu YOLOv8 dostiže 93%, dok transformer-bazirane arhitekture [222, 230] i RT-DETR [231] postižu još bolje rezultate, oko 95%. Iako ovi radovi pokazuju izuzetnu preciznost u rešavanju izazova detekcije klasova, oni su koristili isključivo javno dostupni GWHD skup podataka i poboljšavali tačnost kroz augmentaciju podataka i sintetičke podatke.

U ovoj doktorskoj disertaciji, osnovni modeli su tri modela sa istim *backbone-om* kao model u [2], ali i RetinaNet, YOLOv8 i RT-DETR, zbog njihove bolje performanse kao *state-of-the-art* arhitektura: F-RCNN R50 FPN 3x, F-RCNN R101 FPN 3x, F-RCNN X101 32x8d FPN 3x, RetinaNet R101 FPN 3x, YOLOv8 i RT-DETR.

GWHD takmičenje nije rezultirao kreiranjem novih arhitektura koje bi potencijalno mogle poboljšati detekciju klasova pšenice ili rešiti problem preklapanja ili detekciju manjih klasova u kompleksnoj sceni. Većina učesnika koristila je postojeće, javno dostupne arhitekture: EfficientDet [221], Faster-RCNN [52] i Yolo-v3 [232]. Ovo može ukazivati da izbor mrežne arhitekture nije ključni faktor u konačnom učinku modela, budući da nijedan od korišćenih modela nije značajno bolje generalizovao finalni rezultat [226].

Pored toga, svi pobednici koristili su pseudo-oznake (engl. *pseudo-labeling*) [233], koristeći predikcije nad test skupom i fino podešavanje parametara modela kroz treniranje sa pseudo-označenim podacima [226]. Ukupan najbolji rezultat iznosi 0,69 [225], mereno kroz mAP, kao pokazatelj performansi modela prema ponderisanoj tačnosti u zavisnosti od težine svake klase [226].

Cilj ovog istraživanja je razvoj sistema za prepoznavanje i brojanje klasova pšenice, što predstavlja neophodan korak za dalju procenu gustine pokrivenosti, a zatim i za predikciju prinosa pšenice. U inicijalnoj studiji objavljenoj [90] snimci ozime pšenice prikupljeni su u četiri termina tokom vegetacione sezone i ručno označeni radi dobijanja referentnih oznaka, gde se klasovi nalaze na slikama. Semantička segmentacija slika izvršena je na nivou piksela, primenom metoda dubokog učenja. Konkretno, konvolucione neuronske mreže primenjene su na RGB i termalne slike, a rezultati su upoređeni sa ručno označenim referentnim anotacijama kako bi se ocenile performanse razvijene arhitekture. Razvoj sveobuhvatnog sistema za brojanje klasova i predikciju prinosa ima izuzetnu praktičnu vrednost za praćenje stanja useva i donošenje optimalnih odluka u proizvodnji pšenice. U narednoj studiji objavljen je skup podataka koji upotpunjuje javno dostupan GWHD,

preporuka za generisanje skupa podataka na terenu koristeći pametan telefon, kao i predlog metoda za precizniju detekciju klasova pšenice koje se baziraju na principu segmentacije na nivou instance [3].

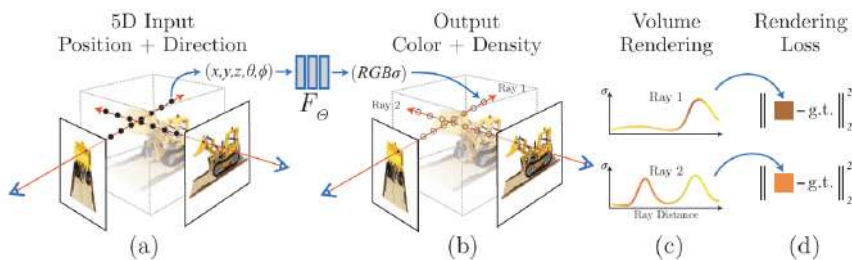
2.6 3D rekonstrukcija primenom NERF metode

U prethodnim poglavljima fokus je bio na razvoju i primeni metoda semantičke i segmentacije na nivou instanci u 2D prostoru, sa ciljem detekcije i analize kompleksnih biljnih struktura na slikama različitih modaliteta snimanja. Iako su ovi pristupi pokazali značajne rezultate, inherentna ograničenja dvodimenzionalne reprezentacije, poput gubitka prostorne informacije, preklapanja objekata i nedovoljno izražene dubinske percepcije, otežavaju pouzdano razdvajanje i kvantifikaciju pojedinačnih elemenata biljke. Ove karakteristike naročito dolaze do izražaja kod struktura poput klasova pšenice, gde su prostorne relacije i orijentacija u prostoru od ključnog značaja za tačnu analizu fenotipskih karakteristika.

Sa ciljem da se prevaziđu navedena ograničenja i obezbedi sveobuhvatnija prostorna reprezentacija biljke, u ovom delu istraživanja sprovedena je validacija upotrebe *Neural Radiance Fields* (NeRF) metoda za trodimenzionalnu rekonstrukciju mikroparcele pšenice u kontrolisanom, laboratorijskim uslovima. Ovaj pristup, zasnovan na dubokoj neuronskoj mreži koja implicitno modeluje geometriju i osvetljenje scene, omogućava generisanje 3D struktura visoke vernosti iz ograničenog broja 2D snimaka. Na taj način, uključivanjem 3D rekonstrukcije kroz NeRF pristup, istraživanje se širi izvan okvira klasičnih 2D tehnika segmentacije, sa ciljem da se ispita potencijal neuronskih mreža u rešavanju problema prostorne kompleksnosti i unapređenju razumevanja morfoloških karakteristika biljaka.

Neural Radiance Fields (NeRF) predstavlja naprednu metodu za 3D rekonstrukciju scene, koristeći duboku neuronsku mrežu za modelovanje volumetrijske gustine i emitovane radijacije na osnovu 5D koordinata (prostorna lokacija i pravac posmatranja), omogućavajući fotorealističnu sintezu novih prikaza sa minimalnim brojem ulaznih slika, (vidi sliku 2.14). [234] Sinteza slike vrši se uzorkovanjem petodimenzionalnih koordinata — prostorne lokacije i smera posmatranja — duž zraka kamere (a). Te koordinate se zatim unose u višeslojnu per-

ceptronsku mrežu (MLP), koja za svaku tačku generiše vrednosti boje i zapreminske gustine (b). Primenom tehnika zapreminskog renderovanja ove vrednosti se kompozituju u konačnu sliku (c). Pošto je funkcija renderovanja diferencijabilna, moguće je optimizovati prikaz scene minimizacijom razlike između sintetizovane i stvarno posmatrane slike (d).



Slika 2.14: Prikaz 3D rekonstrukcije scene NeRF metodom

Dakle, NeRF rešava problem mapiranja pozicije i pravca zraka koji prilaze kroz volumen iz pravca kamere, a kao rezultat daje boju i gustinu posmatranog objekta ili materijala.

U poređenju sa tradicionalnim pristupima kao što su *point clouds* ili mreže (*mesh*), NeRF koristi implicitnu reprezentaciju scene, što omogućava preciznije modelovanje složenih geometrijskih detalja i interakcija svetlosti, bez potrebe za eksplicitnim definisanjem površina ili topologije. Ovaj pristup omogućava glatku interpolaciju između prikaza i tačno renderovanje zaklonjenih ili delimično posmatranih regiona, čime se poboljšava kvalitet 3D rekonstrukcija. Stoga, imajući u vidu da je centralni problem ove doktorske disertacije postizanje visoko precizne segmentacije, a primenjeno na studiji preklapljenih klasova pšenice, završni korak u metodologiji se odnosi na primenu NeRF tehnike za rekonstrukciju 3D scene pšenice na segmentovanim klasovima. Ova metoda omogućava značajno poboljšanje u razumevanju problema preklapanja klasova, koji u 2D prostoru ostaje gotovo nerešiv zbog kompleksnosti scene, velikog broja klasova i njihove prostorne raspodeljenosti. Korišćenjem 3D volumetrijske reprezentacije, NeRF omogućava efikasno razdvajanje pojedinačnih klasova, čime se dobija precizna kvantifikacija broja klasova po jedinici posmatrane površine, što predstavlja ključni ulazni parametar za finalnu, pouzdanu procenu prinosa.

U polju poljoprivrede, NeRF je primenjen za preciznu fenotipizaciju

biljaka, omogućavajući rekonstrukciju 3D struktura biljaka za procenu prinosa i modelovanje biomase pod različitim uslovima osvetljenja za detekciju stresa. Ove primene pokazuju potencijal NeRF-a u unapređenju precizne poljoprivrede kroz detaljno i tačno 3D modelovanje biljaka.

Osnovna verzija NeRFa [235], uvodi neuronski parametarsku reprezentaciju scene kroz petodimenzionalno mapiranje (prostorna pozicija i pravac posmatranja) u gustinu i emitovanu radijaciju, što omogućava diferencijabilno volumetrijsko renderovanje i generisanje novih pogleda (engl. *views*) visoke fotorealističnosti. Na ovim temeljima razvijen je čitav niz varijanti čiji je cilj da unaprede robusnost, brzinu, skalabilnost ili mogućnost primene u složenim i realnim okruženjima [236].

Jedna od prvih ekstenzija je NeRF-W (*NeRF in the Wild*), koji uvodi latentne kodove specifične za svaku sliku, čime se efikasno modeluju neregularni uslovi osvetljenja i privremeni objekti, omogućavajući učenje iz heterogenih i neuređenih kolekcija fotografija [237]. NeRF++ dalje unapređuje parametrizaciju prostora kroz razdvajanje unutrašnje i spoljašnje oblasti scene, što značajno poboljšava rekonstrukciju u neograničenim (engl. *unbounded*) scenama [238]. Na problem *aliasinga* i skaliranja odgovor daju mip-NeRF i mip-NeRF 360, uvodeći poziciono kodiranje na više frekvencijskih skala, dok varijanta „360“ dodatno omogućava rekonstrukciju celovitih panorama [239, 240]. Dalji razvoj ide ka verodostojnijem modelovanju materijala i reflektujućih površina. Tako Ref-NeRF eksplicitno razdvaja difuzne i reflektovane komponente, uključujući i parametre kao što su normale i hrapavost, čime se postiže znatno verodostojnija vizuelizacija sjajnih, visokoreflektujućih materijala [241]. Paralelno, značajan napredak u performansama ostvaren je metodom Instant-NGP, koja uvodi višeskalnu metodu, *hash* kodiranje i omogućava treniranje NeRF modela za nekoliko minuta ili čak sekundi, umesto za sate, što predstavlja praktični pomak ka realnoj upotrebi [242]. Poseban pravac razvoja čine varijante koje uključuju dodatne regularizacije ili geometrijske informacije, kao što su DS-NeRF[243], DietNeRF[244] ili RegNeRF[245], koje omogućavaju rekonstrukciju iz malog broja pogleda (engl. *sparse-view*) uz očuvanje strukturalne konzistentnosti [236, 246]. Konačno, značajan korak ka skalabilnosti predstavlja metode kao što su Mega-NeRF[247] i Block-NeRF[248], koje dekomponuju velike scene (npr. gradske predele) na lokalne blokove povezane globalnim regularizacionim šemama, čime je omogućeno efikasno modelovanje scena na razmeri čitavih urbanih po-

dručja [236].

Dakle, pregled literature pokazuje da evolucija NeRF metoda ide u pravcu rešavanja četiri ključna ograničenja osnovne verzije:

1. ograničena primenljivost u nestrukturisanim i realnim skupovima podataka;
2. nemogućnost efikasnog skaliranja ka većim scenama;
3. visoka računarska složenost i vreme treniranja;
4. ograničeno modelovanje kompleksnih materijala i osvetljenja

Ove inovacije čine osnovu savremenih pristupa u 3D viziji i otvaraju mogućnost njihove primene i u specifičnim oblastima kao što je poljoprivreda, gde precizna rekonstrukcija kompleksnih biljnih struktura ima neposredan značaj za procenu prinosa i fenotipsku analizu.

U poslednjoj deceniji, NeRF-ovi su se afirmisali kao jedan od naj-savremenijih pristupa u oblasti 3D rekonstrukcije biljaka i useva, čime značajno doprinose razvoju precizne poljoprivrede. Model *NeRF-Ag*, osmišljen za trodimenzionalnu rekonstrukciju voćnjaka, integriše višerezolucione latentne enkodere i parametre okruženja, omogućavajući fotorealistično renderovanje uz nivo tačnosti koji značajno prevazilaze tradicionalne COLMAP metode, pri čemu se vreme treniranja skraćuje i do 39 puta [249]. Dodatno, razvijene su NeRF-bazirane metodologije za generisanje tačkastih mreža (engl. *point cloud*) namenjenih visokoprotlačnoj fenotipizaciji biljaka, koje obezbeđuju gotovo savršenu rekonstrukciju koristeći samo jedan stacionarni fotoaparat. Ovaj pristup se pokazao izuzetno pogodnim za automatizovane trake sa rotirajućim postoljem u laboratorijskim uslovima [250]. Na nivou terenskih eksperimenata, performanse NeRF tehnika poređene su sa LiDAR referentnim senzorom, pri čemu je postignut F1 skor od 74,6% nakon svega 30 minuta treniranja. Implementacija strategije ranog zaustavljanja dodatno je smanjila vreme treniranja za gotovo polovinu, uz minimalan pad tačnosti (smanjenje F1 za svega 7,4%) [251]. U kontrolisanim uslovima, u stakleniku, NeRF je primenjen za morfološku analizu paradajza korišćenjem robotizovanih sistema. Generisani gusti i semantički obogaćeni *point cloud-ovi* omogućili su merenja dužine stabljike, površine listova i zapremine plodova, koja su pokazala izuzetno visoku korelaciju sa ručnim merenjima (R^2 do 0,973, MAPE manje od 13,5%) [252]. Najnoviji razvoj u ovoj oblasti predstavlja *OB-NeRF*

platforma, koja omogućava trodimenzionalnu rekonstrukciju kompleksnih biljnih struktura uz drastično skraćeno vreme treniranja (oko 30 sekundi naspram više od 10 sati), visok PSNR i izuzetnu geometrijsku i teksturalnu preciznost ($R^2 > 0,99$, $MAE \approx 2$ mm). Ovakvi rezultati otvaraju perspektivu primene *digital twin* biljaka u okviru naprednih sistema za fenotipizaciju i upravljanje poljoprivrednom proizvodnjom [253]. Za inicijalne eksperimente u okviru ove disertacije korišćen je *Nerfacto* model, savremena implementacija NeRF-a koja je razvijena u okviru okruženja *NeRFStudio*. Za razliku od originalnog NeRF modela [234], koji se oslanja na sinusoidalne funkcije za kodiranje položaja i zahteva višednevno treniranje, *Nerfacto* integriše višerezolucijsku *hash encoding* prema principima Instant-NGP pristupa [254], čime značajno ubrzava proces učenja i omogućava efikasniju reprezentaciju složenih scena. Osim toga, uključuje koncepte iz *Mip-NeRF 360* arhitekture [255], koji obezbeđuju stabilnost pri rekonstrukciji otvorenih i dinamičkih scena, kao i bolju kontrolu distorzije i zamućenja na većim udaljenostima. Zahvaljujući kombinaciji ovih tehnika, *Nerfacto* obezbeđuje veću robusnost u rekonstrukciji vizuelnih detalja, poboljšanu fotometrijsku konzistentnost između frejmova i kraće vreme treniranja u poređenju sa osnovnim NeRF pristupima. Iz tih razloga, *Nerfacto* je odabran kao kompromis između kvaliteta rekonstrukcije i računarske efikasnosti, što ga čini pogodnim za analize koje zahtevaju stabilno i kvantitativno pouzdano renderovanje kompleksnih scena.

2.6.1 Metrike i kvantitativna analiza NeRF 3D rekonstrukcije

Za objektivnu procenu kvaliteta rekonstruisanih scena NeRF metodom i stabilnosti renderovanih frejmova neophodno je koristiti kvantitativne metrike koje obuhvataju kako prostorne karakteristike pojedinačnih frejmova, tako i vremensku konzistentnost između uzastopnih frejmova. Takve metrike omogućavaju detaljnu analizu performansi modela, identifikaciju potencijalnih nedostataka u rekonstrukciji, kao i poređenje različitih pristupa i arhitektura. U ovom poglavlju su predstavljene ključne evaluacione metrike koje uključuju oštrinu slike, odnos signal-šum, *Peak Signal-to-Noise Ratio*, (*PSNR*), *Structural Similarity Index Measure*, (*SSIM*) i vremensku konzistentnost, zajedno sa njihovim matematičkim definicijama i interpretacijom u kontekstu NeRF rekonstrukcija.

1. Oštrina slike (engl. *Sharpness*)

Oštrina slike predstavlja meru lokalnih promena intenziteta piksela, odnosno stepen prisutnosti ivica u slici i kvantifikovana je pomoću varijanse Laplasovog operatora nad sivom skalom slike:

$$S = \text{Var}(\nabla^2 I), \quad (2.25)$$

gde je $\nabla^2 I$ Laplajian slike I , a $\text{Var}(\cdot)$ označava varijansu. Veća vrednost S ukazuje na veću oštrinu slike i prisustvo izraženijih detalja, dok niže vrednosti sugerišu zamućenost ili gubitak fine strukture.

2. Odnos signal-šum (engl. *Signal-to-Noise Ratio, SNR*)

Odnos signal-šum (SNR) meri koliko je signal dominantan u odnosu na prisutan šum i izražava se u decibelima (dB):

$$\text{SNR} = 20 \log_{10} \left(\frac{\mu_I}{\sigma_I} \right), \quad (2.26)$$

gde su μ_I i σ_I srednja vrednost i standardna devijacija intenziteta slike I . Veća vrednost SNR označava „čistiju“ sliku sa manje šuma, dok negativne vrednosti ukazuju na dominaciju šuma u odnosu na signal.

3. Vršni odnos signal-šum, (engl. *Peak Signal-to-Noise Ratio, PSNR*)

Vršni odnos signal-šum (PSNR) predstavlja standardnu metriku za poređenje kvaliteta slike u odnosu na referentnu. U kontekstu vremenskog niza frejmova, PSNR meri razliku između uzastopnih frejmova:

$$\text{PSNR} = 10 \log_{10} \left(\frac{L^2}{\text{MSE}} \right), \quad (2.27)$$

gde je L maksimalna vrednost piksela (255 za 8-bitne slike), a MSE srednja kvadratna greška između dva uzastopna frejma:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (I_i - J_i)^2. \quad (2.28)$$

Veća vrednost PSNR ukazuje na manju grešku i veću konzistentnost između uzastopnih frejmova.

4. Indeks strukturne sličnosti, (engl. *Structural Similarity Index*, SSIM)

Indeks strukturne sličnosti (SSIM) je perceptivna metrika koja ocenjuje sličnost između dve slike na osnovu osvetljaja, kontrasta i strukture. Definiše se kao:

$$\text{SSIM}(I, J) = \frac{(2\mu_I\mu_J + C_1)(2\sigma_{IJ} + C_2)}{(\mu_I^2 + \mu_J^2 + C_1)(\sigma_I^2 + \sigma_J^2 + C_2)}, \quad (2.29)$$

gde su μ_I i μ_J srednje vrednosti, σ_I i σ_J standardne devijacije, σ_{IJ} kovarijansa između slika, dok su C_1 i C_2 konstante koje sprečavaju numeričku nestabilnost. Vrednost SSIM-a se kreće od 0 do 1, pri čemu 1 označava identične slike.

5. Vremenska konzistentnost (engl. *Temporal Consistency*)

Vremenska konzistentnost procenjuje stabilnost i glatkoću promena između susednih frejmova u video sekvenci. U ovom radu, ona je definisana kao prosečna apsolutna razlika sivih nivoa između uzastopnih frejmova:

$$T = \frac{1}{N} \sum_{i=1}^N |I_t(i) - I_{t-1}(i)|, \quad (2.30)$$

gde I_t i I_{t-1} označavaju intenzitete piksela trenutnog i prethodnog frejma. Niže vrednosti T ukazuju na veću vremensku konzistentnost, tj. manju pojavu treperenja i stabilnije renderovanje u vremenskom domenu.

Kombinovanjem ovih metrika moguće je detaljno analizirati kvalitet NeRF rekonstrukcije. Metrike poput PSNR i SSIM daju uvid u perceptivni kvalitet i vernost generisanih frejmova, dok oštrina i SNR ukazuju na kvalitet lokalnih detalja i prisustvo šuma. Temporalna konzistentnost, s druge strane, meri stabilnost modela u vremenskom domenu.

3 Opis skupova podataka

Upotreba modela dubokog učenja je rasprostranjena u mnogim primenama, ali postizanje njihove robusnosti u studijama fenotipizacije biljaka i dalje predstavlja značajan izazov, prvenstveno zbog nedostatka javno dostupnih skupova podataka prikupljenih u uslovima na terenu [2]. Pored toga, budući da metode zasnovane na dubokom učenju zahtevaju velike količine podataka, prisustvo varijacija unutar skupova podataka je ključan faktor za postizanje precizne i objektivne analize.

Neizbežne procedure, uključujući dizajn eksperimentalnog polja, prikupljanje, označavanje i čišćenje podataka kreiranih u terenskim uslovima, dodatno otežavaju proces zbog ograničenog vremenskog okvira i promenljivih vremenskih uslova tokom akvizicije slika. Glavne prepreke u prikupljanju podataka obuhvataju nepovoljne vremenske prilike, potrebu za snimanjem u optimalnoj fenološkoj fazi biljaka, kao i rad na terenu, koji je često intenzivan i vremenski zahtevan.

U okviru ove disertacije, prikupljeni su kako laboratorijski, tako i terenski podaci, kako bi se obuhvatili različiti uslovi eksperimenta i omogućila analiza robusnosti arhitektura i obučениh modela u različitim scenarijima primene. Posebno je važno napomenuti da su hiperspektralni, RGB i skupovi podataka prikupljeni pametnim telefonom retki u literaturi, a njihova upotreba u fenotipizaciji biljaka često nije podržana naučnom evaluacijom kvaliteta prikupljenih podataka, naročito u slučaju mobilnih uređaja. Ovaj izazov je adresiran u okviru ove doktorske disertacije, a rezultati su delimično publikovani u [3].

Jedan od primarnih doprinosa ove teze je kreiranje sveobuhvatnih skupova podataka koji doprinose naučnoj zajednici i omogućavaju napredak u budućim studijama, ne samo u oblasti poljoprivrede, već i u širem domenu kompjuterske vizije. Ovi skupovi podataka osvetljavaju tehničke i inženjerske izazove specifične za fenotipizaciju biljaka, koji se retko pojavljuju u drugim oblastima, čime se dodatno ističe njihov značaj za razvoj robusnih i primenjivih modela dubokog učenja.

3.1 Skup podataka za hiperspektralnu analizu morfoloških karakteristika kruničnih listova paradajza

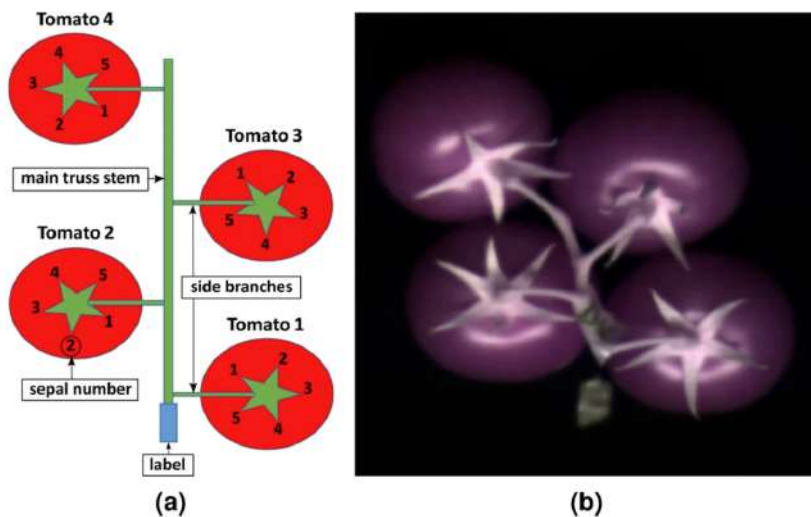
Za potrebe ovog istraživanja, 6 serija paradajza sorte *Brioso RZ* obezbeđeno je od strane 5 komercijalnih proizvođača iz Holandije i Belgije. Paradajz su proizvođači isporučili i dostavili institutu Wageningen Food and Biobased Research (WFBR), koji je deo Univerziteta i istraživačkog centra Wageningen. Predstavnici proizvođača, zajedno sa WFBR institutom, čine konzorcijum u okviru javno-privatnog partnerstva na projektu Humistatus, finansiranog od strane fondacije *TKI – Dutch Horticulture and Starting Materials*. Paradajz je ustupljen isključivo u svrhe istraživanja, a ta saradnja je unapred dogovorena sa predstavnikom proizvođača i ostalim članovima konzorcijuma Humistatus projekta. Takođe, berbu su obavili sami proizvođači, u skladu sa relevantnim institucionalnim, nacionalnim i međunarodnim smernicama i zakonodavstvom. Svi uzorci paradajza ubrani su 12. decembra 2017. godine, nakon čega su 13. decembra prevezeni u Wageningen na sobnoj temperaturi, i skladišteni preko noći na 15 °C, do narednog dana. Serije su označene velikim slovima od A do F, kako bi se razlikovali proizvođači i/ili uslovi gajenja. Serije A–E potiču od različitih proizvođača i svi su gajeni u plastenicima bez dodatnog osvetljenja, a berba je obavljena na kraju životnog ciklusa biljaka. Serije E i F potiču od istog proizvođača, ali iz različitih plastenika – pri čemu je serija F dobijena iz zasada koji je tek započeo proizvodnju, uz primenu dopunskog osvetljenja. Sledećeg dana, plodovi iz svake serije (5–6 komada po seriji) su orezani i raspodeljeni u šest manjih grozdova, od kojih je svaki sadržao tri do četiri paradajza. Tabela 3.1 daje sažet prikaz ovih podataka.

Prvi deo eksperimenta obuhvatao je prikupljanje hiperspektralnih podataka pre nego što su se na paradajzu pojavili vidljivi znaci prisustva gljivičnog oboljenja. Zbog toga su hiperspektralne slike snimljene prvog dana eksperimenta (14. decembar 2017). Snimanje je obavljeno pomoću kamere SPECIM FX17, proizvođača Specim, Finska. Ova kamera registruje spektralne odzive u opsegu od 900 do 1700 nm, sa rezolucionim korakom od 3,46 nm, što daje ukupno 224 spektralna kanala po slici. Kao izvor osvetljenja korišćene su dve halogene lampe. Svaka grana (cvetni grozd) paradajza je postavljena ispod kamere tako da vrh grane bude okrenut ka dnu, i obeležen je plavom trakom (primer

Tabela 3.1: Pregled prvog hiperspektralnog skupa podataka

Oznaka serije	Br. plodova/kr. listova	Uslovi gajenja
A	23/109	Bez dodatnog osvetljenja, kraj proizvodnog ciklusa
B	24/119	Bez dodatnog osvetljenja, kraj proizvodnog ciklusa
C	24/115	Bez dodatnog osvetljenja, kraj proizvodnog ciklusa
D	24/118	Bez dodatnog osvetljenja, kraj proizvodnog ciklusa
E	24/116	Bez dodatnog osvetljenja, kraj proizvodnog ciklusa
F	24/111	Dodatno osvetljenje, prvi rod

je prikazan na slici 3.1 koja je preuzeta iz rada [1]).



Slika 3.1: a) Orijentacija plodova paradajza tokom hiperspektralnog snimanja. (b) Pseudokolor slika jednog grozda/grane dobijena kombinovanjem tri spektralna kanala (Slika preuzeta iz [1])

U drugom delu prvog eksperimenta, odmah nakon snimanja hiperspektralnih podataka, paradajzi su smešteni u kontrolisano okruženje sa relativnom vlažnošću od 100%, što predstavlja idealne uslove za klija-

nje gljivica. Paradajzi su izvađeni iz kontrolisanih uslova 18. decembra 2017. godine, a primećeno je da su gotovo svi krunični listovi bile zaraženi u određenoj meri (osim serije F).

Kako bi se objektivno zabeležio stepen infekcije, napravljen je dodatni referentni skup podataka sa RGB slikama. Snimanje je obavljeno pomoću uređaja *Smart Color Inspector (SCI)*, koji predstavlja zatvoreno i kontrolisano okruženje sa kalibrisanim osvetljenjem i podešavanjima kamere, kako bi se zabeležile verne boje objekata postavljenih unutar uređaja. SCI je dizajniran od strane WFBR i izrađen od strane IPSS Engineering (obe firme sa sedištem u Wageningen, Holandija). Uređaj je opremljen LED panelima (4038 K) sa pet strana, i RGB kamerom (MAKO G-192C POE, Allied Vision Technologies GmbH, Štatroda, Nemačka) koja slike snima odozgo. Za svaku seriju merenja, sistem je kalibrisan pomoću bele pozadine (Forex PVC ploča, bela, 6 mm) i kolor test table sa 24 polja (Color Checker Classic, X-rite Europe GmbH, Regensdorf, Švajcarska).

Kolor slike su korišćene kao osnovni referentni podaci (engl. *ground truth*) za procenu stepena infekcije gljivicama. Slika 3.1 a) prikazuje kolor sliku iz serije C, kao i pravila kodiranja koja su primenjena kako bi se nivo infekcije kruničnog lista povezao sa njihovim hiperspektralnim odzivima (Slika preuzeta iz [1]). U skladu sa pravilima kodiranja, svakom kruničnom listu dodeljen je jedinstveni četvorocifreni identifikator, pri čemu cifre redom označavaju: oznaku serije, broj grozda, broj paradajza i broj kruničnog lista.

Inicijalna studija koja koristi ove podatke, predstavljena je u radu [1], u kom je analizirana osetljivost tek ubranih plodova paradajza i kruničnih listova na kasniju gljivičnu infekciju, oslanjajući se upravo na detaljne karakteristike prikupljene na nivou pojedinačnih piksela, pre nego što su vidljivi golim okom. U toj fazi istraživanja primenjene su metode spektralne analize u kombinaciji sa hiperspektralnim i kolor podacima koji su prethodno opisani, pri čemu je demonstrirana značajna prednost fuzije podataka u odnosu na pojedinačne modalitete. Ova inicijalna analiza motivisala je dalje unapređenje procesa semantičke segmentacije i detekcije regiona od interesa, posebno kruničnog lista paradajza, imajući u vidu da tačnost klasifikacije svakog pojedinačnog piksela direktno utiče na preciznost finalne procene zaraženosti biljke, što je detaljnije razvijeno i metodološki razrađeno u okviru ove disertacije i objavljeno u [87]. U [87] je pored gore opisanog hiperspektralnog skupa podataka korišćen još jedan, dakle, dva skupa podataka koja

sadrže hiperspektralne slike različitih sorti paradajza. Hiperspektralne slike dodatnog skupa podataka, prikupljene su pomoću iste kamere SPECIM FX17, koja meri reflektansu u spektralnom opsegu od 900 nm do 1700 nm sa rezolucijom od 3,46 nm, čime se po slici dobija ukupno 224 spektralna kanala, u oba skupa podataka. Da bi se umanjio uticaj neuniformnih uslova osvetljenja na hiperspektralno snimanje, sve slike snimane su u kontrolisanim uslovima tamne sobe, koristeći halogene lampe kao jedini izvor svetlosti. Ovakva postavka omogućila je ujednačene uslove osvetljenja i eliminisala potencijalne smetnje od spoljnih izvora svetlosti, čime je poboljšana pouzdanost spektralnih podataka. Time je osigurano da hiperspektralni skup podataka ostane neometan spoljnim varijacijama osvetljenja, dok je istovremeno smanjena redundantnost spektralnih informacija, što doprinosi većoj pouzdanosti i ponovljivosti rezultata studije. Proces celokupnog snimanja podataka sproveden je u okviru Wageningen Food and Biobased Research (WFBR), deo Wageningen Univerziteta.

Prvi skup podataka sadrži ukupno 36 hiperspektralnih slika grozdova paradajza, sa prosečno 19 kruničnih listova po slici (u proseku 5 po paradajzu) [1, 87].

Drugi skup podataka prikupljen je u skladu sa drugačijim eksperimentalnim protokolom, koji je podrazumevao manji broj paradajza po slici (jedan paradajz naspram tri ili četiri u prvom skupu podataka). Pored toga, razlika u rastojanju između senzora i objekta tokom snimanja dovela je do smanjene rezolucije (tj. manjeg broja piksela po objektu od interesa tj. kruničnih listova paradajza). Do ove razlike došlo je jer su plodovi paradajza bili pozicionirani na većoj udaljenosti od senzora, što je doprinelo uočenoj razlici u rezoluciji između dva skupa hiperspektralnih podataka. Tokom formiranja drugog skupa podataka korišćene su četiri sorte paradajza: *Brioso*, *Capricia*, *Provine* i *Sao Paolo*.

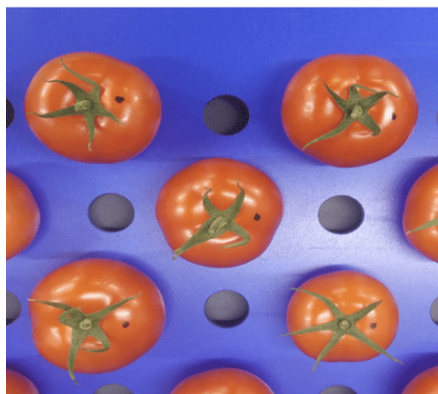
Paradajz je uzgajan u plastenicima u De Lieru, u Holandiji, a obezbedio ga je Growers United u okviru WFBR instituta. Između sorti uočavaju se приметne razlike u veličini – *Brioso* i *Provine* su srednje do krupne sorte, dok su *Capricia* i *Sao Paolo* generalno sitnije i često se nazivaju čeri paradajz. Pored toga, *Sao Paolo* se odlikuje karakterističnom tamnocrvenom bojom. Drugi skup podataka sadrži dve hiperspektralne slike sa po 32 ploda paradajza za sorte *Brioso* i *Sao Paolo*, kao i dve slike sa po 16 plodova paradajza za sorte *Capricia* i *Provine* (videti Tabelu 3.2 i Sliku 3.2). U cilju postizanja sličnih razmera



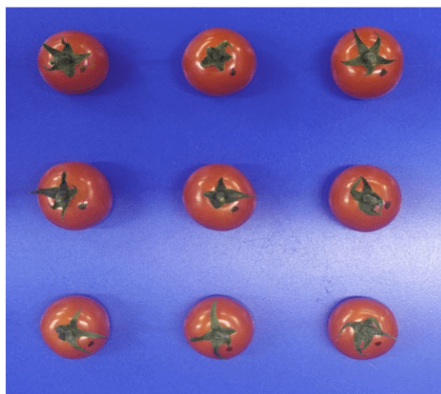
Briosio



Capricia



Provine



Sao Paolo

Slika 3.2: RGB vizuelizacija primera sve četiri sorte paradajza iz drugog skupa hiperspektralnih podataka

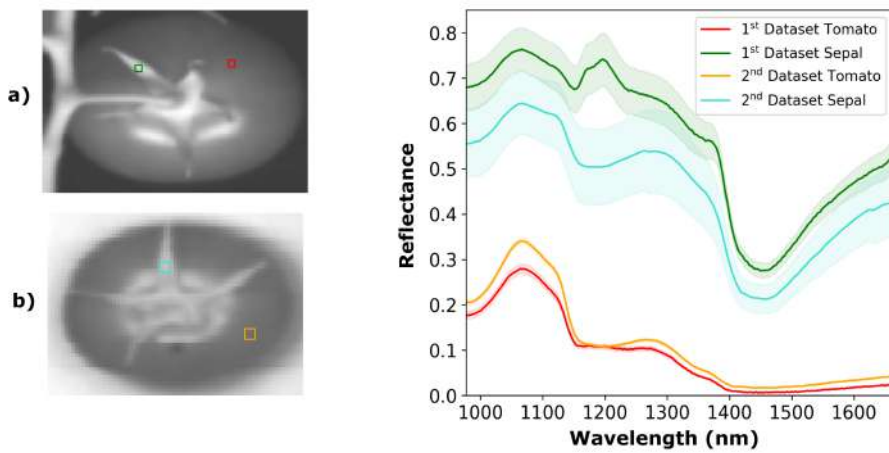
paradajza, odnosno kruničnih listova između skupova, hiperspektralne slike iz drugog skupa podeljene su tako da svaka slika sadrži po jedan paradajz, a zatim su te slike skalirane kako bi odgovarale dimenzijama regije koja pokriva jedan paradajz u slikama iz prvog skupa.

Na slici 3.3 prikazani su primeri spektralnih odziva ploda i kruničnih listova paradajza sorte *Briosio* u oba skupa podataka koji su potvrdili početnu pretpostavku da reflektanse ploda i kruničnih listova paradajza u posmatranom spektralnom opsegu korelišu između dva skupa podataka.

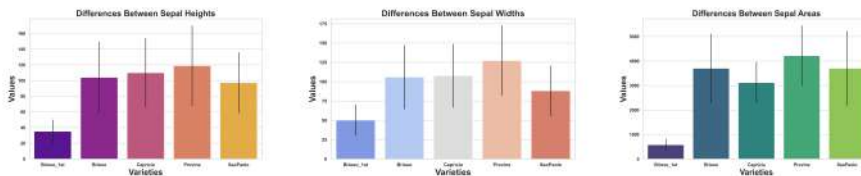
Korelacija između kruničnih listova iznosi 0,8859 sa p -vrednošću $0,87 \times 10^{-35}$, dok je korelacija između plodova 0,8953 sa p -vrednošću $1,75 \times 10^{-35}$. Dodatno, na slici 3.4 je prikazana varijabilnost u širini (50 do

Tabela 3.2: Tabelarni prikaz hiperspektralnih podataka

Skup podataka	Sorta	Broj hipersp. slika	Broj kr. listova
I	<i>Brioso</i>	36	698
II	<i>Brioso</i>	4×32	164
	<i>Capricia</i>		167
	<i>Provine</i>		163
	<i>Sao Paolo</i>		158



Slika 3.3: Hiperspektralni odzivi plodova i kruničnih listova paradajza na primerima iz a) prvog i b) drugog skupa hiperspektralnih podataka



Slika 3.4: Varijabilnost visine, širine i površine kruničnih listova paradajza u oba skupa hiperspektralnih podataka po sortama paradajza

125 piksela), visini (35 do 120 piksela) i površini (500 do 4200 piksela) kruničnih listova paradajza, što pruža ključne informacije o morfološkoj raznolikosti unutar drugog skupa podataka, ali i među skupovima. Hiperspektralne slike drugog skupa podataka dodatno su podeljene ta-

ko da svaka novonastala slika sadrži po jedan paradajz. Sve anotirane maske za detekciju kruničnih listova paradajza kreirane su ručno, korišćenjem LabelMe softvera [256], pri čemu su oznake formirane kao poligoni oko svakog kruničnog lista. Referentni podaci za brojanje kruničnih listova kreirani su numerisanjem označenih listova direktno iz JSON fajlova generisanih u LabelMe softveru.

Kompletna metodologija implementirana je u programskom jeziku *Python* verzije 3,9, korišćenjem sledećih biblioteka: *Spectral Python (SPy)* [257] i *Scikit-Image (skimage)* [258] za predobradu podataka, *Scikit-Learn (scikit)* [259] za implementaciju tehnika kodiranja spektralnih informacija, kao i biblioteke *detectron2* [223] za primenu modela dubokih neuronskih mreža za segmentaciju instanci u rešavanju problema detekcije kruničnih listova paradajza.

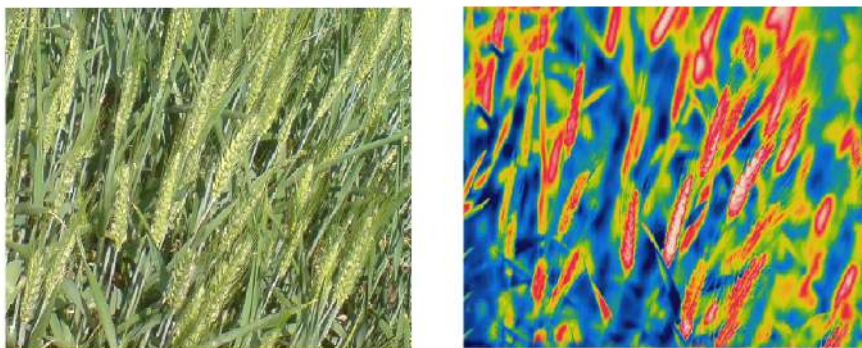
U ovom istraživanju, složene računarske operacije su izvršavane korišćenjem jedinica (GPU) visokih performansi. GPU koji je korišćen za zadatke dubokog učenja su NVIDIA GeForce RTX 2080 Ti. Sa 4352 CUDA jezgara i 11 gigabajta memorije visokog protoka (HBM2). Ovakva konfiguracija GPU obezbeđuje neophodne računarske resurse za ubrzanje faza treniranja i izvođenja svih potrebnih operacija za obuku modela dubokih neuronskih mreža.

3.2 Skup podataka za detekciju i brojanje klasova pšenice

3.2.1 Skup podataka za semantičku segmentaciju klasova pšenice u termalnim i RGB slikama

U ovom delu opisan je skup podataka prikupljen profesionalnom ručnom kamerom, oznake FLIR SC620 kreiran za potrebu detekcije klasova pšenice na termalnim i RGB slikama pomocu semantičke segmentacije.

Za potrebe kreiranja ovog skupa podataka dve parcele su posmatrane na dve lokacije: u Bajmoku (45,97°N, 19,42°E) i u Ravnom Selu (45,45°N, 19,62°E), dve naseljene sredine u Južnobačkom okrugu, u Vojvodini, severnoj pokrajini Srbije. Slike su snimane u četiri datuma (7. i 17. maja u Bajmoku, te 11. i 18. maja u Ravnom Selu), po dva puta na svakoj lokaciji, u dve fenološke faze rasta pšenice: početkom cvetanja, kada je otvoreno približno 10% cvetova, i u ranoj mlečnoj



Slika 3.5: Levo: RGB reprezentacija slike sa terena; Desno: pseudokolor reprezentacija termalne slike sa terena

fazi pšenice.

U Bajmoku je prosečna dnevna maksimalna temperatura u maju iznosila 23°C , a prosečna količina padavina 49 mm [260]. U Ravnom Selu, prosečna dnevna maksimalna temperatura u maju bila je 24°C , sa prosečnim padavinama od 52 mm [261]. Eksperimenti na oglednim parcelama su postavljene na isti način: posejane su dve sorte semena (*NS40S*, razvijena od strane Instituta za ratarstvo i povrtarstvo, Novi Sad, Srbija, i *Ingenio* CCB, Syngenta).

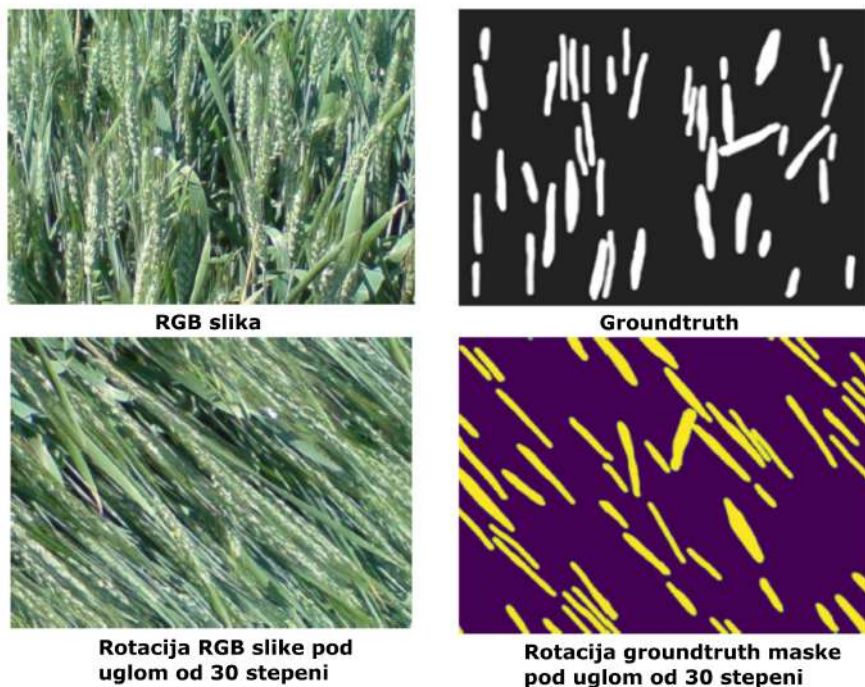
Ogledne parcele su organizovane kao mreža mikro parcela tj. plotova, dimenzija 16×10 , pri čemu je prva pomenuta sorta bila posejana u prvih 5, dok je druga sorta u preostalih 5 redova. To je rezultiralo sa po 80 plotova po sorti na svakoj lokaciji. Svaki plot zauzima površinu od 4 m^2 , sa proredom širine od 0,2 m između svakog reda. Eksperimenti su sprovedeni ujutro, tokom vedrih dana od 9 do 12 časova, kako bi se minimizovao neželjeni šum u merenjima koji može biti izazvan razlikama u osvetljenju.

Kamera korišćena na oglednoj parceli je manuelna termalna FLIR SC620, sa rezolucijom slike od 640×480 piksela i RGB rezolucijom od 2048×1536 piksela. Ova kamera ima preciznost očitavanja od $\pm 2\%$ u temperaturnom opsegu od -40°C do $+500^{\circ}\text{C}$, sa vidnim poljem od $24^{\circ} \times 18^{\circ}/0,3\text{m}$, žižnom daljinom od 40 mm i prostornom rezolucijom od 0,65 mrad [262]. Ova tip kamere detektuje vrlo male temperaturne razlike i prikuplja infracrveno zračenje, što je korisno za razlikovanje i isticanje klasova u odnosu na pozadinu.

Stativ je postavljen na maksimalnu visinu (1,25 m), pri čemu je sen-

zor kamere bio pozicioniran na visinu od 0,55 m iznad useva pšenice (koja je u trenutku snimanja imala visinu od 0,7–0,8 m). Slike su snimljene pod približno uglom od 30° u odnosu na posmatranu pšenicu. Tokom akvizicije slika korišćena je automatska postavka osvetljenja, što je omogućilo posmatranje relativne temperaturne razlike između klasa i stabljike.

Kako je proces akvizicije slika manuelan, nekoliko problema se pojavilo prilikom obrade podataka: pojava zamagljenih slika (engl. *blur*) usled nenamernog pokreta ruke ili pomeranja kamere tokom snimanja. Nakon odbacivanja slika sa problemima zamagljenja, osvetljaja ili senki, odabrano je 138 slika zadovoljavajućeg kvaliteta u oba modaliteta za eksperimentalnu evaluaciju. Primeri termalnih i RGB slika dati su na slici 3.5.



Slika 3.6: Prvi red: desno – RGB originalna reprezentacija slike sa terena; levo – ručno anotirana binarna maska. Donji red: desno – primena rotacije nad RGB originalnom slikom pod uglom od 30 stepeni; levo – primena rotacije nad ručno anotiranom binarnom maskom pod uglom od 30 stepeni.

S obzirom na ograničen broj originalnih slika prikupljenih na terenu, čiji kvalitet i vizuelni prikaz mogu biti pod uticajem različitih vremenskih uslova, nivoa osvetljaja, prisustva zamućenja ili drugih nepredvidivih faktora, neophodno je primeniti metode augmentacije radi sintetičkog proširenja skupa podataka. Korišćenjem rotacionih i *affine* transformacija moguće je ne samo povećati ukupan broj slika, već i obezbediti da model postane robusniji na različite položaje objekata u prostoru, u ovom slučaju klasova pšenice, promene orijentacije, pomeranja izazvana vetrom i druge varijacije koje se prirodno javljaju na terenu. Primena ovih transformacija rešava problem neidealnog ili nekonzistentnog ugla snimanja. [263, 264].

3.2.2 Skup podataka za segmentaciju instanci klasova pšenice u RGB slikama

U ovom delu opisan je skup podataka prikupljen mobilnim telefonom u okviru ove studije, označen kao BioS-Wheat, uz pregled postojećih, javno dostupnih skupova podataka u oblasti poljoprivrede i fenotipizacije, sa posebnim osvrtom na *GHWD - Global Wheat Head Detection* skup kao referentni skup koji se koristi u ovom radu. BioS-Wheat skup podataka proširuje postojeći GWHD u pogledu varijabilnosti genotipa, geografske lokacije i primenjenih tretmana.

Pregled javno dostupnih skupova podataka

Postoje poznati, veliki i javno dostupni skupovi podataka, kao što su *The Microsoft Common Objects in Context (COCO)* [174], Open Images V4 [265] ili ImageNet [266], koji se koriste za treniranje dubokih neuronskih mreža i mogu odgovoriti na mnoge zahteve aplikacija u zadacima detekcije objekata. Neki od specijalizovanih skupova podataka za detekciju objekata sa fokusom na guste scene su PUCPR [267], CARPK [268], koji sadrži slike dobijene pomoću bespilotnih letelica (UAV), i SKU-110K, koji obuhvata slike objekata na policama supermarketa [269].

Međutim, slike iz ovih skupova podataka obično sadrže manji broj objekata koji pripadaju generičkim klasama (automobili, baloni itd.), koji su samo blago preklapljeni ili se uopšte ne preklapaju. Ovo naglašava potrebu za specijalizovanim skupom podataka dizajniranim posebno za naš zadatak. Modeli koje koristimo prethodno su trenirani na

podacima koji ne obuhvataju u potpunosti kompleksnost našeg problema. Na primer, mnogi skupovi podataka sadrže slike sa samo dva objekta koja zauzimaju 80% piksela slike, dok naš dodati skup podataka sadrži slike sa preko 50 klasova pšenice. Skale između objekata na ovim slikama se značajno razlikuju, što otežava mreži da efikasno uči iz podataka. Pored toga, na našim slikama objekti se često mnogo više preklapaju, što dodatno komplikuje proces učenja u poređenju sa jednostavnijim scenarijima.

Među javno dostupnim skupovima podataka kreiranim u studijama fenotipizacije biljaka, velika većina je prikupljena u kontrolisanim laboratorijskim uslovima, dok samo mali broj sadrži podatke generisane na terenu, u nekontrolisanim uslovima, kao što su skupovi podataka za klasifikaciju korova [270, 271], brojanje sorguma [272, 273] i klasova pšenice[274].

Dakle, postojeći skupovi podataka i prethodno obučeni modeli mogu poslužiti kao polazna osnova u poljoprivrednim aplikacijama, ali kada je scena na terenu specifična, kao što je parcela pšenice, problem detekcije objekata od interesa znatno se komplikuje.

GWHD skup podataka je trenutno najveći javno dostupan skup podataka koji sadrži RGB slike pšenice u visokoj rezoluciji prikupljene u uslovima na terenu, zajedno sa anotacijama i geolokacijama klasova pšenice. U okviru ove doktorske disertacije, kao referentni skup koristimo GWHD skup podataka, koji dalje proširujemo novokreiranim BioS-Wheat skupom podataka.

***Global Wheat Head Detection* skup podataka**

GWHD skup podataka se sastoji od RGB slika pšenice prikupljenih različitim vrstama senzora tokom fenoloških faza cvetanja i zrenja, a nastao je kroz saradnju devet institucija. Za potrebe GWHD izazova [225, 226] i ove studije uključeno je sedam skupova podataka (videti Tabelu 3.3).

Pšenica je zasejana u tri stepena gustine (niska, normalna i visoka) i sa razmakom redova od 12,5 cm do 30,5 cm. Tokom prikupljanja podataka korišćene su različite kamere postavljene na fenotipske platforme, pri čemu je visina kamera varirala od 1,8 m do 3 m iznad tla. Žična daljina korišćenih kamera kretala se od 7,7 do 60 mm, dok se prostorna rezolucija (engl. *Ground Sample Distance* – *GSD*) kretala u opsegu od 0,1 do 0,6 mm. Imajući u vidu da je prosečna dužina klasa pšenice oko 1,5

Tabela 3.3: Proširena tabela iz [2] sa opisom BioS-Wheat skupa podataka [3]

Institucija	Država	G.Š.	G.D.	Fenološka faza	Međuredni razmak (cm)	Gustina setve (seeds/m ²)	Br. geno- tipova
Arvalis1	FR	43,7°N	5,8°E	nakon cvetanja i zrenje	17,5	300	20
Arvalis2	FR	43,7°N	5,8°E	nakon cvetanja	17,5	300	20
Arvalis3	FR	49,7°N	3,0°E	nakon cvetanja i zrenje	17,5	300	4
ETHZ	CH	47,4°N	8,6°E	nije pri- menjivo	12,5	400	354
INRAE	FR	43,5°N	1,5°E	nakon cvetanja	16	300	7
RRES	UK	51,8°N	0,36°W	nije pri- menjivo	nije pri- menjivo	350	6
USASK	CA	52,1°N	106°E	nije pri- menjivo	30,5	250	16
BioS- Wheat	RS	45,45°N	19,62°E	nakon cveta- nja	12,5	500- 550	4
BioS- Wheat	RS	45,45°N	19,62°E	srednje mlečna faza	12,5	500- 550	4

cm u prečniku [275], dobijene vrednosti GSD-a dovoljne su za detaljnu reprezentaciju tekstone klasa. Iako visoka raznovrsnost upotrebljenih senzora može doprineti boljoj generalizaciji modela dubokog učenja, ona ujedno uvodi i određene izazove usled potrebe za harmonizacijom podataka [2]. Za GWHD skup podataka autori su izvršili ručnu reviziju slika i uklonili one koje su snimljene u ranijim fenološkim fazama, pre cvetanja. Budući da precizna detekcija klasa pšenice [59] zavisi od rezolucije slike i veličine klasa, što je uslovljeno odabranim genotipom i fazom rasta u trenutku snimanja, autori su dodatno izvršili reskaliranje slika. Konačna verzija GWHD skupa sadrži 3442 RGB slike rezolucije 1024x1024 piksela, od čega 3373 slike sadrže klasove, sa prosečno 40 klasova po slici. Preostale slike obuhvataju površine tla, travu ili krove, ali ne i klasove pšenice. GWHD skup podataka anotiran je od strane više institucija korišćenjem *COCO Annotator-a*, postojeće veb-

platforme za anotaciju [276, 277]. Za neanotirane skupove podataka, autori rada [2] koristili su polunadgledani način anotiranja koji se bazi-
ra na principu dubokog učenja [273], i zasnovan je, u ovom slučaju, na
YoloV3 modelu [232] treniranom na podacima iz dva izvora (Arvalis1
i UTokio1), kako bi dobili oznake za preostale slike. S obzirom na to
da su oznake u GWDH skupu dobijene poluautomatskim pristupom,
javljaju se i potencijalni problemi: zaklanjanje i preklapanje klasova,
senke koje prekrivaju delove klasa, varijacije u osvetljaju i perspektiv-
ne distorzije, što može dovesti do razlika u veličini klasa na slikama.
Ovi faktori doprinose greškama u procesu anotacije, pa je neophodna
pažljiva ljudska verifikacija radi obezbeđivanja tačnosti podataka.

BioS-Wheat skup podataka

Za potrebe ove doktorske disertacije sproveden je eksperimentalni
terenski ogled, čiji je dizajn prikazan na slici 3.7, na parceli u blizini
Ravnog Sela (g. širina 45,45°, g. dužina 19,62°) u Srbiji.
Četiri sorte pšenice — *Ingenio* (In) kompanije Syngenta, *NS40S* (NS),
Futura (Po) i *Futura* (Fu) kompanije NS Seme — posejane su na po
24 plota, sa četiri tretmana đubrenja (količine date na slici 3.7), čime
je postignuta varijabilnost prinosa uslovljena različitim genotipovima i
uslovima prihrane. Sažeti pregled BioS-Wheat skupa podataka, uklju-
čujući broj slika i broj klasova pšenice, dat je u Tabeli 3.4.

Tabela 3.4: BioS-Wheat skup podataka: broj slika i broj klasova pšenice
po sorti.

Sorta	Broj slika	Broj klasova pšenice
<i>Ingenio</i>	1472	14528
<i>NS40S</i>	1344	14080
<i>Futura</i>	1472	13141
<i>Futura</i>	1408	12736
Ukupno	5696	54485

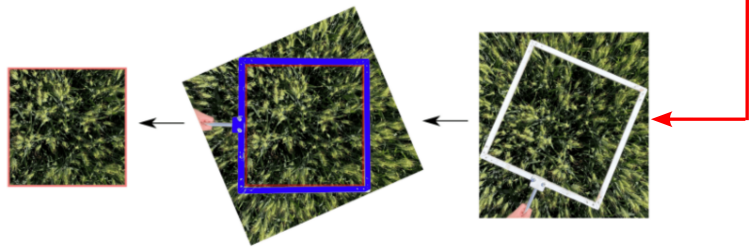
Ovaj skup podataka se odlikuje izuzetno visokom gustinom setve (500–550
semena/m²) i minimalnim međurednim rastojanjem (12,5 cm), što pred-
stavlja znatno intenzivnije uslove u odnosu na druge GWDH podsku-
pove. Ovakvi uslovi dovode do pojačane konkurencije među biljkama
i veoma guste prostorne raspodele klasova, sa čestim zaklanjanjima,

**Basic N treatment
(Autumn)**

In	0	60	180	120	60	180	0	120	180	120	60	0
Ns	120	0	60	180	120	0	60	180	0	60	120	180
Po	60	180	120	0	180	0	120	60	60	120	180	0
Fu	180	120	0	60	60	120	180	0	120	0	180	60

**Top-dress N treatment
(Spring)**

In	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0
Ns	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0
Po	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0
Fu	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0	0	150	150	0



Slika 3.7: Dizajn eksperimenta sa parcelama numerisanim prema različitim količinama tretmana azotom, praćen koracima predobrade za izdvajanje regiona od interesa iz sirovih podataka.

preklapanjima i nepravilnim orijentacijama. Navedeni faktori dodatno komplikuju zadatak detekcije objekata od interesa, tj. klasova pšenice, naročito za modele koji koriste osno poravnate pravougaonike ili su obučeni na redim skupovima podataka, kao što je GWDH. Uključivanje podskupa BioS-Wheat u podatke za obuku povećava raznovrsnost i težinu scenarija učenja, omogućavajući modelima bolje prilagođavanje uslovima visoke gustine useva. Time se unapređuje robustnost i sposobnost generalizacije modela, naročito u primenama u regionima sa intenzivnim agrotehničkim praksama.

Procedura akvizicije slika sprovedena je dva puta, u dve fenološke faze rasta pšenice [278]: vodena zrelost ili *watery ripe* i srednje mlečna zrelost *medium milk*. Snimanje je obavljeno u jutarnjim časovima, u periodu od 9 do 12, tokom vedrih dana, kako bi se ublažile ili izbegle varijacije u uslovima osvetljaja. Za akviziciju je korišćen pamet-

ni mobilni telefon sa trostrukim sistemom kamera: ultra-širokougaona 13 mm ($f/2,4$), širokougaona 26 mm ($f/1,6$) i telefoto 52 mm ($f/2,0$) [279], sa autofokusom i dvostrukom optičkom stabilizacijom slike. Uređaj je postavljen na visini od 1 m iznad useva, tako da se čitava oblast od interesa (ROI), oivičena drvenim ramom dimenzija $0,5\text{ m} \times 0,5\text{ m}$, nalazi u centru slike i obuhvata srednji deo jednog posmatranog plo- ta (videti sliku 3.7). Iako u literaturi nije uspostavljen standardizovan protokol akvizicije, preporučena GSD vrednost od 0,3 mm prema [59] je uzeta kao referentna u ovoj doktorskoj disertaciji. Dobijeni podaci sa terena naknadno su rotirani tako da drveni ram bude paralelan sa osama slike, a ROI je izdvojen ekstrakcijom svih piksela unutar rama (slika 3.7). Segmentacija rama izvedena je konverzijom iz RGB u HSI (engl. *hue, saturation, intensity*) prostor boja [280] i primenom metode postavljenog praga u okviru *hue* kanala, nakon čega sledi morfološka operacija zatvaranja radi redukcije šuma. Konačna segmentacija rama dobijena je algoritmom povezanih komponenti implementiranim u biblioteci *OpenCV* [281, 282], pri čemu je odabrana komponenta najveće površine. Horizontalne ivice rama detektovane su *Canny* detektorom [283, 284], a zatim je primenjena *Hough* transformacija [285]. Ugao rotacije procenjen je kao ugao između detektovanih horizontalnih linija rama i horizontalne ose slike. Nakon rotacije (slika 3.7, izdvojen je ROI unutar rama. Na kraju, ROI slike su podeljene prema postupku iz [2], čime je dobijeno 5696 slika, potrebnih dimenzija 1024×1024 piksela.

Za izdvojene slike manuelno su kreirane anotacije za semantičku segmentaciju, koje su potom prilagođene zadacima segmentacije instanci. Nepreciznosti uzrokovane zaklanjanjem ili preklapanjem klasova pšenice ublažene su primenom metode *Distance Transform* [286] i primenom praga, pri čemu je prag postavljen na polovinu maksimalne distance po povezanoj komponenti. Krajnji obuhvatni pravougaonici (engl. *bounding boxes*) za svaki označeni klas generisani su algoritmom povezanih komponenti [281, 282].

Predloženi postupak zasnovan na metodama *Distance Transform* i povezanim komponentama može biti podložan greškama spajanja ili razdvajanja klasova pšenice, naročito u slučajevima visoke gustine ili njihovog preklapanja. Ispitana je robusnost predloženog pristupa i uočeno je da se takve greške uglavnom javljaju u specifičnim situacijama (npr. kod potpuno preklapljenih klasova pšenice), što predstavlja uobičajen izazov za sve automatske i poluautomatske metode anotacije, a može se pojaviti čak i pri ručnom označavanju.

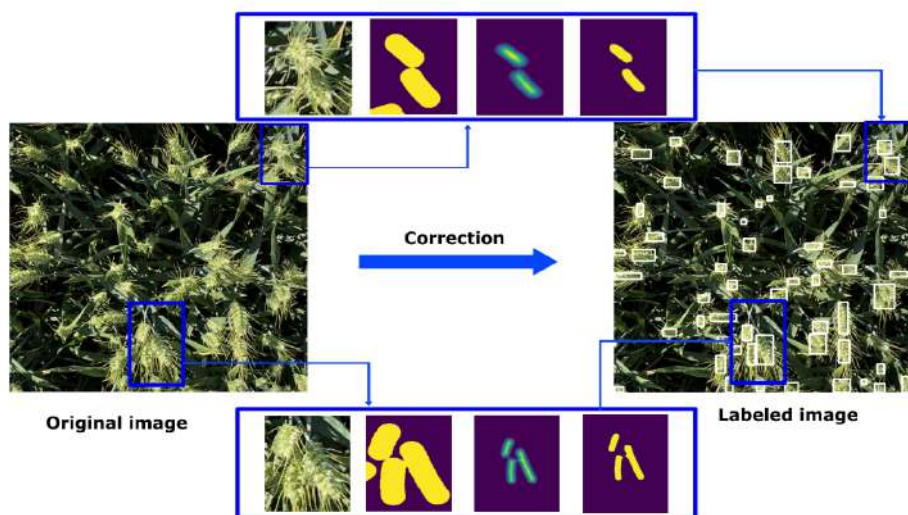


Slika 3.8: Primer kompleksnih slučajeva u procesu anotacije u okviru GWHD skupa podataka, uključujući preklopljene klasove pšenice, pogrešno spojene ili razdvojene labele, delimično isečene objekte na ivicama slike, kao i snimke sa smanjenim osvetljenjem.

Postoje studije koje ističu inherentne poteškoće u preciznom označavanju gusto preklopljenih ili vizuelno složenih slučajeva u okviru GWHD skupa podataka. Kao što je prikazano na slici 3.8 u radu [2], takvi izazovni scenariji obuhvataju preklopljene klasove pšenice, delimično isečene klasove na ivicama slike, kao i snimci sa nedovoljnim osvetljenjem ili drugim suboptimalnim uslovima snimanja. U cilju ublažavanja problema koji se javljaju prilikom primene polunadgledanog pristupa za označavanje objekata od interesa, predloženi metod kreiranja anotacija u okviru ove teze smanjuje uticaj visokog stepena preklapanja i drugih poteškoća u procesu anotacije efikasnim razdvajanjem grupisanih instanci, što je prikazano na slici 3.9.

Ovim pristupom omogućeno je pouzdanije detekcija i kvantifikovanje objekata na nivou pojedinačnih instanci, čak i u uslovima velike gustine i vizuelne složenosti, čime se značajno unapređuje robusnost i konzistentnost anotacija unutar skupa podataka. BioS-Wheat skup podataka koristi sledeću konvenciju: prva dva slova naziva sorte služe kao oznake za razlikovanje sorti, dok broj u opsegu 0–23 označava plot koja pripada datoj sorti (uključujući različite tretmane đubrenja). Iako skup sadrži i ove dodatne informacije, u okviru ovog istraživanja sve sorte sa različitim tretmanima posmatrane su kao jedan izvor, u skladu sa praksom GWHD. Traženi format anotacija je rečnik koji sadrži: ID slike, širinu i visinu slike, koordinate pravougaonika svakog klasa pšenice i izvor slike. Dodatno, u tabeli 3.5 prikazano je višedimenzionalno poređenje skupova podataka GWHD i BioS-Wheat, ističući razlike u veličini skupa podataka, granularnosti anotacija, obuhvatu domena i uslovima snimanja.

Dok GWHD obezbeđuje raznovrsnost u gustini setve i korišćenim



Slika 3.9: Vizuelizacija predložene metoda anotiranja koji uspešno obrađuje gustu raspodelu i preklapljene instance. Ovaj pristup ublažava izazove anotacije efikasnim razdvajanjem grupisanih klasova pšenice.

platformama za snimanje, uključivanje skupa BioS-Wheat donosi sledeće prednosti: predstavlja izuzetno gusto sejane parcele sa čestim okluzijama, preklapanjem klasova i nepravilnim orijentacijama klasa, čime odražava složenije i realnije scenarije iz prakse. Uključivanje BioS-Wheat skupa u istraživanje time povećava robusnost i generalizaciju modela za fenotipizaciju, omogućavajući im da uspešno obrade uslove visoke gustine i složenosti u polju, koji nisu u potpunosti obuhvaćeni postojećim skupovima podataka. Ova kombinacija skupova podataka obezbeđuje sveobuhvatniju osnovu za obuku i evaluaciju modela, podržavajući kako metodološke inovacije, tako i praktičnu primenljivost u detekciji klasova pšenice.

Tabela 3.5: Uporedni pregled skupova podataka GWHD i BioS-Wheat korišćenih u istraživanju

Karakteristika	Skup GWHD	Skup BioS-Wheat
Broj slika	3442	5696
Broj anotiranih objekata	197606	54485
Klase	1 (klas pšenice)	1 (klas pšenice)
Tip anotacije	Pravougaone (<i>axis-aligned</i>) granice objekata	Pravougaone (<i>axis-aligned</i>) granice objekata
Preciznost anotacije	Visoka; poluautomatsko označavanje naknadno ručno korigovano; manje greške usled okluzije, osvetljenja i perspektive	Visoka; ručna verifikacija; veoma guste parcele sa čestim okluzijama i preklapanjima
Uslovi na terenu	Spoljašnje parcele pšenice; gustine setve: niska, normalna, visoka; razmak između redova 12,5 – 30,5 cm	Spoljašnje parcele pšenice; izuzetno visoka gustina setve (500 – 550 semena po m ²); minimalni razmak između redova 12,5 cm
Faza rasta	Nakon cvetanja (potpuno izrasli klasovi)	Vodena, mlečna i srednje mlečna faza
Ugao snimanja / perspektiva	Nadirni ili blizu nadirnog ugla; platforme za fenotipizaciju ili ručne kamere	Nadirni ugao; pametni telefon postavljen 1 m iznad useva sa okvirom ROI od 0,5 m × 0,5 m
Uslovi osvetljenja	Prirodno dnevno svetlo, uglavnom jutarnji sati; umerene varijacije osvetljenja	Jutarnji sati, vedri dani bez oblaka; minimalne varijacije osvetljenja
Pozadina / složenost scene	Zemlja, trava, korov; umerena okluzija i vizuelna složenost	Guste parcele pšenice; česte okluzije, preklapanje klasova, nepravilne orijentacije klasova
Rezolucija slike / GSD	1024 × 1024 px; GSD ≈ 0,1 – 0,6 mm u zavisnosti od platforme	4032 × 3024 px; GSD 0,3 mm po pikselu
Tip kamere / senzora	Različite RGB kamere na platformama za fenotipizaciju	Pametni telefon sa trostrukim kamerama (ultraširoka 13 mm, široka 26 mm, telefoto 52 mm)

3.2.3 Procena kvaliteta slika u skupovima podataka za segmentaciju klasova pšenice

Svakodnevno se na različitim platformama deli i razmenjuju milioni digitalnih slika, zbog čega je precizna procena kvaliteta slika (engl. *Image Quality Assessment*, IQA) od ključnog značaja za unapređenje vizuelnog prikaza [287]. Pored svakodnevne primene na društvenim mrežama, IQA ima važnu ulogu i u naučne i tehničke svrhe [288, 289], naročito prilikom poređenja slika dobijenih različitim senzorima. Pouzdana predviđanja kvaliteta slike obezbeđuju da se slike pravilno obrađuju, skladište i prikazuju, što direktno utiče na njihovu upotrebljivost i interpretaciju [287]. Kako bi se opravdala upotreba pametnog telefona uz naučne i profesionalne kamere, u ovoj studiji analiziran je kvalitet snimljenih slika unutar skupova podataka BioS-Wheat i GWHD, primenom objektivnih i često korišćenih metrika za procenu kvaliteta slike. Fokus je stavljen na metode bez referentne slike, jer se ne može izdvojiti referentna slika, a cilj je bio dobijanje opšte procene kvaliteta slika u oba skupa [290]. Ove metode izračunavaju obeležja koja opisuju strukturu slike, relevantnu za ljudsku percepciju njenog kvaliteta.

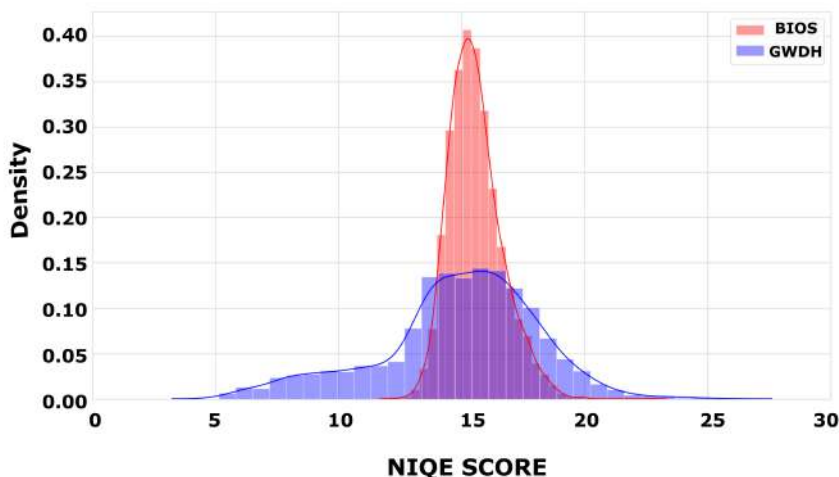
Za procenu su primenjene tri metode: *Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE)* [291], *Natural Image Quality Evaluator (NIQE)* [292] i *Perception-based Image Quality Evaluator (PIQE)* [293] koje predstavljaju najčešće korišćene bezreferentne metrike kvaliteta slike.

BRISQUE [291] procenjuje kvalitet slike direktnom analizom njenih statističkih karakteristika u prostornom domenu, bez potrebe za referentnom slikom. Ova metrika polazi od pretpostavke da prirodne slike poseduju dobro definisane statističke obrasce, te degradacije identifikuje kao odstupanja od tih očekivanih regularnosti.

NIQE [292] predstavlja potpuno nereferentnu metriku koja konstruše generički statistički model iz velike kolekcije visokokvalitetnih prirodnih slika i procenjuje kvalitet posmatrane slike kroz stepen njenog odstupanja od tog modela. Prednost NIQE metrike je što joj nisu potrebni referentni podaci sa degradacijama za treniranje modela.

PIQE [293] procenjuje kvalitet slike segmentacijom na blokove i identifikovanjem regiona koji najviše doprinose degradaciji. Ovakav pristup omogućava robusnost na različite tipove šuma i zamućenja, kao i lokalizovanu detekciju vizuelno relevantnih artefakata, na osnovu čega se formira ukupna ocena kvaliteta slike.

Upotreba više metrika je od suštinskog značaja za pouzdanu validaciju prikladnosti korišćenja slika dobijenih pametnim telefonom, čime se obezbeđuje robusna procena kvaliteta u prisustvu različitih tipova distorzija. Prednost metode BRISQUE ogleda se u tome što je „*opinion-aware*“, jer je obučena na osnovu subjektivnih ljudskih ocena kvaliteta. PIQE obezbeđuje lokalne i globalne ocene kvaliteta, ali može biti manje računarski efikasna u poređenju sa preostale dve metode. BRISQUE daje dobre rezultate kod određenih distorzija, dok ga dopunjuju NIQE i PIQE koje obuhvataju širi spektar proizvoljnih izobličenja [294]. Sve tri metode ocenjuju kvalitet slike na skali od 0 do 100, pri čemu 100 označava najlošiji kvalitet. Rezultati pokazuju da slike u BioS-Wheat skupu podataka odlikuje visok kvalitet. U Tabeli 3.6 su sumirani svi dobijeni rezultati procene kvaliteta slika (prosečne vrednosti sa standardnom devijacijom po skupovima podataka), dok su na Slici 3.10 prikazane distribucije rezultata kvaliteta dobijenih NIQE metodom primenjenom na oba skupa podataka. Iako NIQE skor,



Slika 3.10: Procena kvaliteta slika: NIQE

koji daje bolje rezultate za GWHD, ukazuje na povoljnije rezultate, distribucija pokazuje veću stabilnost za BioS-Wheat skup u poređenju sa GWHD. Ukupni dobijeni rezultati ukazuju da BioS-Wheat skup podataka poseduje slike veoma visokog kvaliteta prema izabranim metrikama, i sugerišu da se korišćenje pametnog mobilnog telefona može uključiti u praksu akvizicije slika *in situ* u uslovima na terenu.

Tabela 3.6: Procena kvaliteta slika — GWHD i BioS-Wheat skupovi podataka

Skup	BRISQUE	NIQE	PIQE
GWHD	$37,08 \pm 13,24$	$14,87 \pm 3,2$	$23,83 \pm 12,48$
BioS-Wheat	$10,82 \pm 5,18$	$15,55 \pm 1,07$	$8,05 \pm 3,17$

U narednom pasusu biće data preporuka za generisanje visoko-kvalitetnih podataka sa terena, u ovom slučaju, klasova pšenice pametnim mobilnim telefonom, radi olakšane akvizicije podataka i daljeg obogaćivanja i proširivanja postojećih skupova podataka.

3.2.4 Preporuke i protokoli za optimalnu akviziciju slika

Standardizovani protokol za snimanje klasova pšenice na terenu, u različitim spoljašnjim uslovima, još uvek ne postoji, jer prethodne studije nisu usklađene [21]. Tokom akvizicije slika potrebno je uzeti u obzir različite faktore, kao što su doba dana, koje utiče na varijacije u osvetljaju, fenološke faze pšenice, perspektiva kamere i ugao prikaza, jer oni direktno utiču na tačnost prepoznavanja klasova u slikama. U radovima [20] i [183], nadir perspektiva se navodi kao relevantan izbor prilikom snimanja, posebno kada su sorte posejane velikom gustinom. Pošto visina pšenice može varirati oko 20%, postavljanje visine kamere prilikom snimanja ostaje otvoreno pitanje, koje direktno utiče na vidljivost klasova u slikama [2]. Obično se u studijama koriste RGB kamere, dok je potencijal korišćenja termalne kamere spomenut u [20], uz opreznu procenu zbog njene visoke cene i niže rezolucije. Slični razlozi obeshrabruju upotrebu profesionalnih kamera na terenu, što dodatno opravdava istraživanje mogućnosti detekcije klasova pšenice pomoću standardnih i široko dostupnih kamera na pametnim mobilnim telefonima.

Naša studija podržava zaključke iz studije [2] da ne postoji jedinstvena fenološka faza koja je idealna za snimanje svih genotipova. Prema rezultatima prepoznavanja postignutim u ovoj studiji, najbolji period za akviziciju slika je nakon cvetanja, kada su klasovi u potpunosti vidljivi, još uvek su lagani, zadržavaju uspravan položaj i time se izbegava ili minimalizuje potencijalno preklapanje klasova i listova. Na osnovu ovih preporuka i iskustava stečenih na terenu, slike BioS-Wheat

skupa podataka su snimljene prema sledećem protokolu:

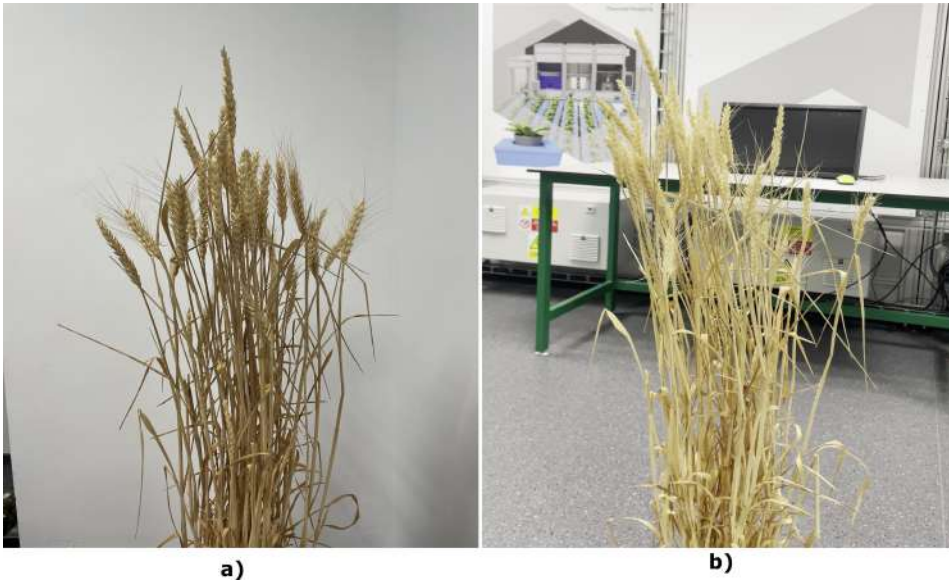
1. nadir perspektiva;
2. visina kamere postavljena na 1 m iznad biljaka kako bi se obuhvatio drveni okvir, u skladu sa zahtevima iz [2];
3. korišćena kamera sa rezolucijom od 4032×3024 piksela, što rezultira udaljenošću uzorka sa tla (GSD) od 0,3 mm po pikselu.

3.2.5 RGB skupovi podataka za 3D NeRF rekonstrukciju mikroparcele pšenice u kontrolisanim uslovima

Za evaluaciju primene NeRF metoda inicijalno su kreirana dva eksperimentalna skupa podataka u kontrolisanim laboratorijskim uslovima: prvi sa pšenicom u saksiji postavljenom na homogenu, uniformno belu pozadinu, i drugi sa pšenicom u okruženju bogatom strukturnim i teksturalnim detaljima. Analiza prvog skupa pokazala je da uniformna pozadina dodatno otežava problem segmentacije klasova, jer se proces registracije segmenata suočava sa nedostatkom kontrastnih referentnih tačaka. Budući da je sama segmentacija klasova inherentno složen zadatak usled njihove slične morfologije, visoke homogenosti i međusobnog preklapanja, nedostatak kompleksnije pozadine može značajno smanjiti kvalitet 3D rekonstrukcije. Nasuprot tome, prisustvo raznolike i bogate pozadine doprinosi stabilnijoj registraciji i pouzdanijem odvajanju klasova u rekonstruisanim trodimenzionalnim scenama. Na slici 3.11 dat je primer slike sa uniformnom i neuniformnom pozadinom.

Za potrebe ove disertacije kreiran je prilagođen skup podataka za rekonstrukciju pšenice u 3D koristeći NeRF metodu. Proces je započet prikupljanjem sirovih podataka iz video snimaka, izvedenih u kontrolisanim laboratorijskim uslovima. Ukupno je generisano šest nezavisnih skupova podataka, obuhvatajući pet video zapisa i oko 70 pojedinačnih slika. Iz ovih video zapisa ekstraktovano je približno 1852 frejmova za skupove sa uniformnom belom pozadinom i oko 1200 frejmova za varijante sa kompleksnom, dinamičkom pozadinom.

Prvi korak uključivao je filtriranje dobijenih frejmova kako bi se eliminisale redundantne ili neupotrebljive slike, čime se smanjila količina podataka bez gubitka relevantnih vizuelnih informacija. Nakon



Slika 3.11: Laboratorijski skupovi podataka: (a) uniformna pozadina, (b) neuniformna pozadina.

filtriranja, primenjena je procedura pripreme podataka kroz instrukcije COLMAP-a i *NeRF Studio* biblioteke [295], što je uključivalo generisanje transformacione matrice i parametara kamere, uključujući angularnu i linearnu brzinu snimanja. Ova priprema je rezultirala JSON fajlovima koji sadrže sve potrebne informacije o poziciji kamere, orijentaciji i parametarskim transformacijama, uz adekvatnu strukturu foldera za kompatibilnost sa NeRF treninzima.

Za ekstrakciju frejmova i pripremu ovakvog skupa podataka korišćen je alat *nerfstudio* [295], koji je omogućio automatsko usklađivanje frejmova i pripremu ulaznih podataka za NeRF implementaciju. Za povećanje varijabilnosti i robustnosti ovog skupa podataka, pored laboratorijskih snimaka, korišćena je aplikacija *Spectacular Rec*, koja je omogućila snimanje tri različite varijante video sekvenci: dve u kojima se biljka rotirala, i jedna u kojoj je kamera kružila oko statične biljke.

Za pripremu skupa podataka za segmentaciju u 3D prostoru, namenjenog NeRF rekonstrukciji, neophodno je generisati binarne maske za svaki frejm. RGB slike su obrađene pre treniranja modela, u HUE prostoru, maskiranjem samo delova koji pripadaju pšenici, kako bi se olakšala segmentacija i što preciznije uhvatio željeni region od interesa. Binarne maske su kreirane korišćenjem gorepomenutih modela za

semantičku segmentaciju iz [90] i U-Net arhitekture. Postignute performanse pokazuju da su oba pristupa efikasna, pri čemu je model iz [90] ostvario *Dice Score* od 82,25% i IoU od 75,21%, dok je U-Net model postigao nešto bolje rezultate, sa *Dice Score-om* od 84,48% i IoU od 81,2%. Trening oba modela sproveden je uz primenu *Early Stopping* metode kako bi se sprečila nadprilagođenje modela. Oba modela su trenirana na po jednoj RGB slici/frejmu iz sekvence sa neuniformnom pozadinom, sa odgovarajućom ručno anotiranom maskom. Model iz [90] dostigao je optimalnu tačku obuke nakon 306 epoha sa greškom na trening skupu od 0,05, i 0,0495 na validacionom skupu, dok je U-Net model dostigao optimum nakon 406 epoha sa greškom koja iznosi 0,051 na trening, i 0,0497 na validacionom skupu. Ovi rezultati pokazuju da oba modela pružaju pouzdanu osnovu za generisanje preciznih binarnih maski za 3D rekonstrukciju pomoću NeRF metode.

Originalne RGB slike i odgovarajuće binarne maske organizovane su u zasebne direktorijume sa sekvencijalnim indeksima radi automatskog uparivanja. Prostor kamere i orijentacije procenjeni su pomoću COLMAP alata [296, 297], koji kombinuje *Structure-from-Motion (SfM)* i *Multi-View Stereo (MVS)* metode za rekonstrukciju 3D scene iz 2D slika. Proces uključuje ekstrakciju i podudaranje karakterističnih tačaka, te izgradnju retke reprezentacije oblaka tačaka, tj. *sparse point cloud*-a sa parametrima kamera (cameras.bin, images.bin, points3D.bin). Dobijeni rezultati konvertovani su u Nerfstudio format komandom `ns-process-data colmap`, čime je generisan direktorijum sa slikama i datotekom `transforms.json` neophodnom za treniranje NeRF modela. Za RGB+binary NeRF korišćen je isti `transforms.json`, čime su identične transformacije primenjene i na RGB slike i na pripadajuće maske. Maske omogućavaju eksplicitno odvajanje objekta od pozadine tokom treniga, fokusirajući model na relevantne volumetrijske informacije. Trening je sproveden pomoću Nerfacto procedure sa definisanim parametrima za maske, boju pozadine i 30000 iteracija. Tokom procesa generisane su intermedijarne reprezentacije (RGB frejmovi, depth i normal mape) koje služe za formiranje finalnog point cloud-a.



a)



b)



c)



d)

Slika 3.12: Laboratorijski skupovi podataka: (a) Primer RGB slike sa neuniformnom pozadinom, (b) Manuelno kreirane labele u formi poligona c) Binarna maska kreirana kreirana na osnovu poligona d) UNet generisana maska

4 Rezultati

4.1 Rezultati semantičke segmentacije

4.1.1 Studija slučaja I: Paradajz – kvantitativna i vizuelna analiza

Za zadatak semantičke segmentacije krunchnih listova plodova paradajza upoređene su performanse tri arhitekture neuronskih mreža: našeg predloženog, novog optimizovanog konvolucionog modela nazvanog (oFCN) razvijenog u [90], a adaptiranog u [90], 2.9, kao i dve arhitekture koje se ubrajaju među najčešće korišćene u semantičkoj segmentaciji: U-Net 2.10 [88] i Seg-Net 2.11 [89]. Analiza je sprovedena na prvom skupu hiperspektralnih podataka 3.1 koji obuhvata 36 hiperspektralnih slika sa 224 spektralna kanala u opsegu od 900 do 1700 nm (bliski infracrveni deo spektra), sa spektralnom rezolucijom od 3 nm (podaci dati u 3.1 i 3.2. Podaci su prikupljeni šest dana pre nego što gljivična infekcija postane vidljiva golim okom. Referentno označavanje regiona od interesa izvršeno je korišćenjem pseudokolor snimaka generisanih na osnovu spektra. Za sve mreže primenjen je uniforman postupak obuke, uključujući istu podelu podataka na skupove za treniranje, validaciju i testiranje (36 hiperspektralnih slika su raspoređene u trening, validaciju i test skup u odnosu: 84%, 8%, i 8% (30-3-3). Projekcija za svaku tehniku kodovanja spektralnih informacija je naučena izvođenjem 6-fold kros-validacije sa usvojenim odnosom podela skupa podataka.). Performanse modela procenjujane su na osnovu F1 skor metrike, broja epoha potrebnih za postizanje optimuma i broja parametara koji se uče tokom procesa obuke.

Najpreciznija segmentacija, uz najmanje zahteve za memorijskim resursima, postignuta je korišćenjem oFCN arhitekture, sa F1 skorom od 91,59% nakon 78 epoha. U-net je ostvario F1 skor od 75,23% nakon

150 epoha, dok je Seg-net dostigao skor od 43,39% nakon 46 epoha. Dobijeni rezultati potvrđuju da predložena oFCN arhitektura predstavlja najefikasnije rešenje za preciznu segmentaciju kraničnih listova paradajza na osnovu hiperspektralnih snimaka u bliskom infracrvenom spektru. Tabelarni prikaz rezultata dat je u tabeli 4.1.

Tabela 4.1: Poređenje postignutih rezultata sve tri arhitekture i broj epoha

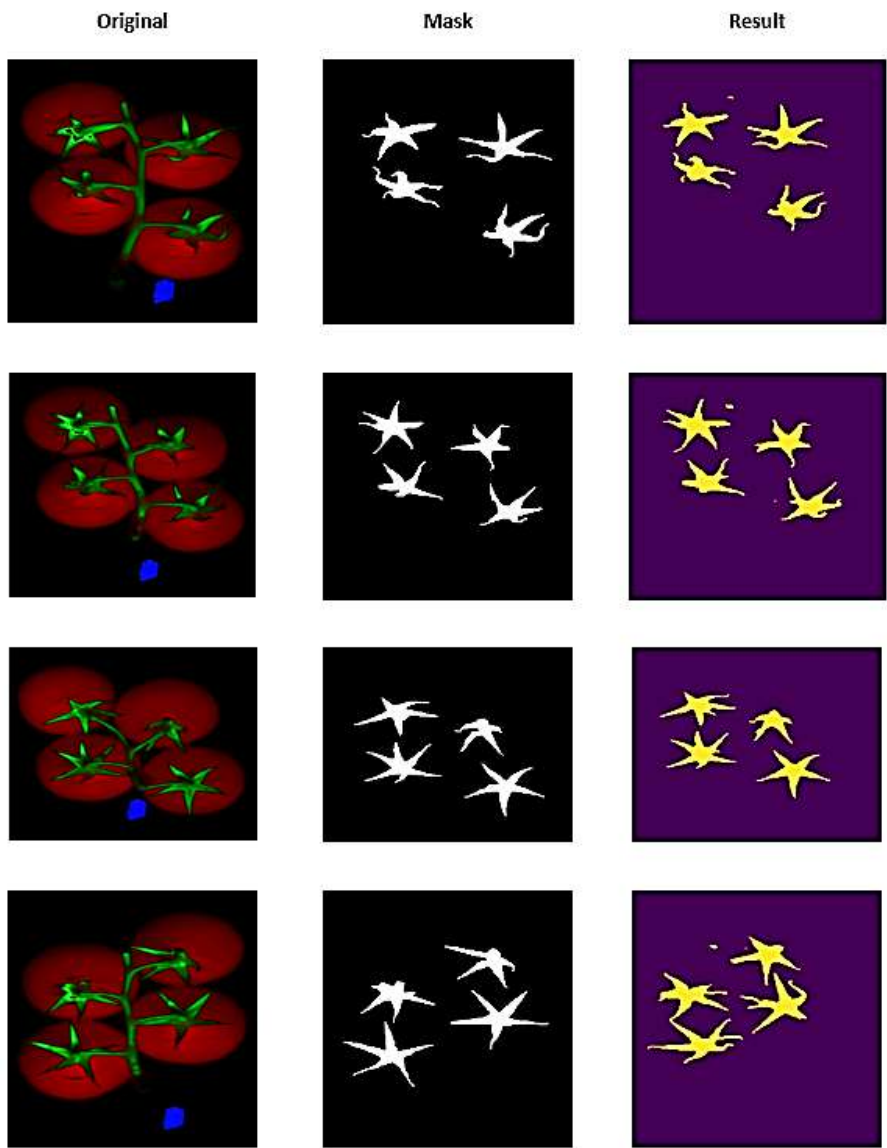
Arhitektura	oFCN	U-Net	Seg-Net
Broj epoha	78	150	46
F1 skor	91,59%	75,23%	43,39%

Analiza kompleksnosti u pogledu broja parametara koji figurišu u procesu obuke modela neuronske mreže koja je razvijena u ovoj doktorskoj disertaciji je sprovedena na hiperspektralnim slikama krune paradajza. Pored broja parametara, kompleksnost je sagledana i u pogledu potrebnog broja epoha, tj. iteracija kako bi se dostigao optimum, tj. minimizovala funkcija cene. Poređenje arhitekture neuronske mreže iz ove doktorske disertacije je izvršeno sa dva najpoznatija modela za segmentaciju slika, U-Net [88] i Seg-Net [34].

Tabela 4.2: Pregled poređenja predložene arhitekture sa dve najčešće korišćene arhitekture za rešavanje problema segmentacije

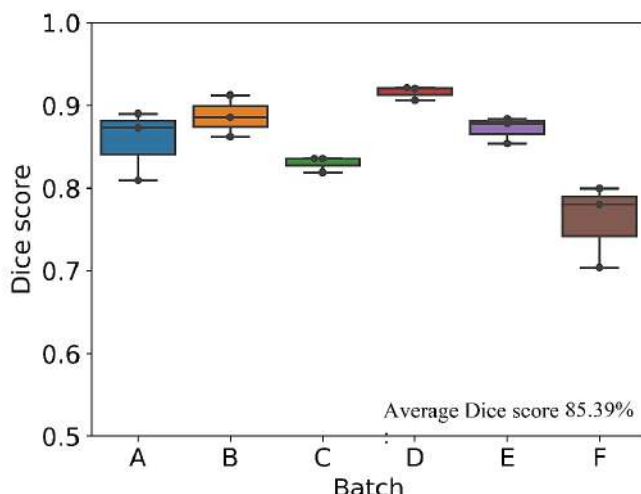
Arhitektura	Br. epoha	Br. parametara ($\times 10^6$)	<i>Dice score</i>
U-Net	150	8,69	75,23%
Seg-Net	46	2,99	43,39%
Predložena arhitektura	78	0,96	91,56%

Na osnovu rezultata iz tabele 4.2, može se zaključiti da predložena arhitektura nadmašuje druge dve arhitekture u pogledu broja parametara koji figurišu u obuci mreže, kao i preciznosti izraženoj preko *Dice Score-a*, iako se radi o velikoj dimenzionalnosti slika na ulazu (224 kanala) i očekuje se da takva kompleksnost podataka utiče na rast parametara



Slika 4.1: Prikaz trodimenzionalne reprezentacije hiperspektralnih slika paradajza, ručno anotiranih maski, i rezultata oFCN modela

koji figurišu u procesu obuke modela u poređenju sa modelom koji je obučen na slikama pšenice.



Slika 4.2: Prikazan opseg vrednosti Dice koeficijenta po batch-u, kao i njegova ukupna prosečna vrednost.

4.1.2 Studija slučaja II: Pšenica – kvantitativna i vizuelna analiza

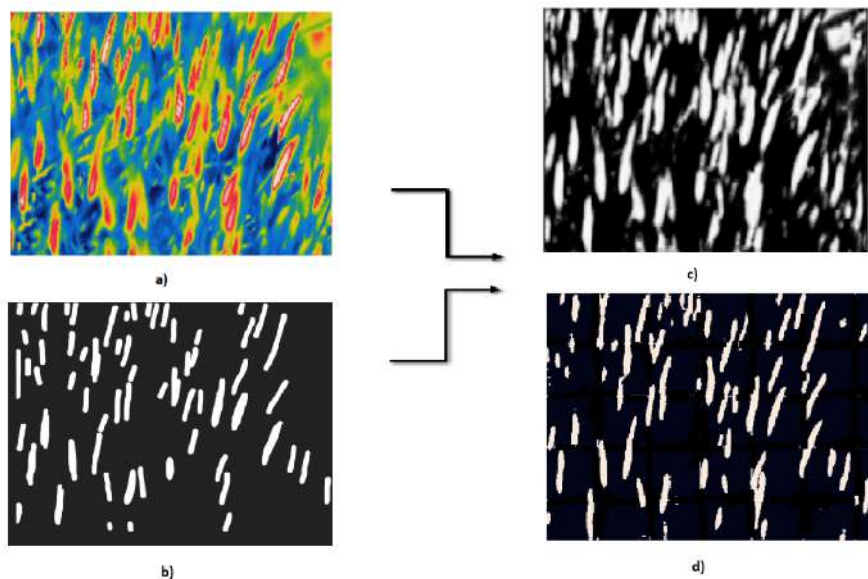
Predložena arhitektura duboke neuronske konvolucione mreže primenjena na termalnim slikama postigla je rezultate od 75,63% F1 koeficijenta. Ova mera je najznačajnija za prvi deo sistema, tj. semantičku segmentaciju i detekciju klasova pšenice. Informacije o podacima, kao i podela i količina termalnih slika opisana je u 2.4.2. Ukupna tačnost modela iznosi 92,59%, preciznost 85,36%, dok je *Recall* dostigao 94,73%. U tabeli 4.3 prikazani su ovi rezultati.

Tabela 4.3: Rezultati segmentacije - termalne slike

F1 koeficijent	Tačnost	Recall	Preciznost
0,75	0,92	0,94	0,85

Kako bi se postigla što preciznija segmentacija, prag je izabran na osnovu najveće postignute vrednosti F1 skora. Izabrani prag ima vrednost 0,8 i primenjuje se na izlaz modela, koji se koristi u narednom koraku za automatsko brojanje klasova (vidi sliku 4.4)[71, 79].

Za RGB slike rezultati su nešto drugačiji (vidi Tabelu 4.4). Model je postigao F1 skor od 68,46%, tačnost 67,74%, preciznost 70,83%, dok



Slika 4.3: (a) Originalna termalna slika; (b) ručno anotirana binarna maska; (c) izlaz modela; (d) primenjeni prag od 0,8 na izlaz.

je odziv (recall) dostigao 84,23%. Na izlazu iz RGB modela primenjuje se isti prag od 0,8.

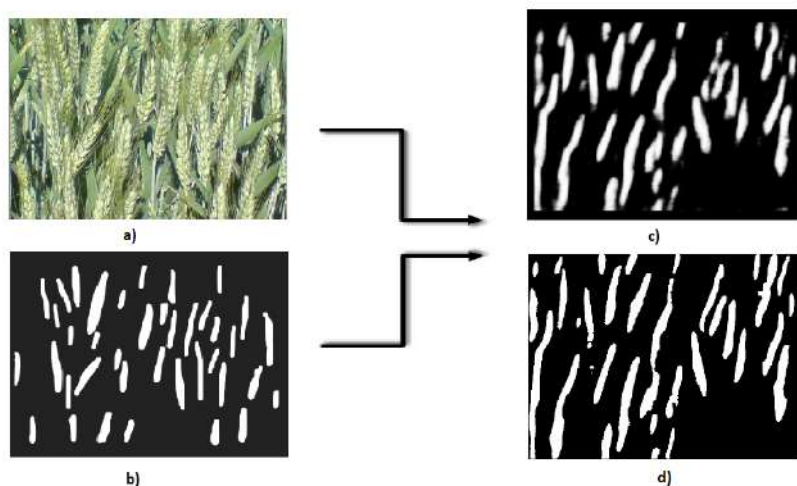
Tabela 4.4: Rezultati segmentacije - RGB slike

F1 koeficijent	Tačnost	Recall	Preciznost
0,68	0,67	0,84	0,70

Brojanje klasova pšenice

Za referentne podatke, klasovi su prebrojani na originalnim, ručno kreiranim maskama od strane dve osobe. Prosečna vrednost se uzima kao tačan broj klasova na slikama [10, 20]. Klasovi su takođe prebrojani ručno od strane dve osobe, na maskama dobijenim iz modela za obe vrste slika. Propagacija greške usled nedetektovanih klasova tokom segmentacije dovela je do relativne greške od 9,5% za termalne slike i 14,33% za RGB slike.

Automatska metoda brojanja klasova primenjuje se na izlazu iz modela, nakon primenjenog praga od 0,8 i osnovne morfološke operacije,



Slika 4.4: (a) Originalna RGB slika; (b) ručno izrađena binarna maska; (c) izlaz modela; (d) primenjeni prag od 0,8 na izlaz.

erozije u jednoj iteraciji, sa veličinom kernela 5×5 za termalne slike i 7×7 za RGB slike. Za termalne slike, automatska metoda brojanja klasova ima grešku od 11,74% u odnosu na ručno prebrojane klasove na maskama dobijenim iz modela. Takođe, tačnost automatski prebrojanih klasova je 89,22% u poređenju sa ručno prebrojanim klasovima, na ukupno 27 test slika. Za RGB slike, greška iznosi 18,46% u odnosu na ručno prebrojane klasove na maskama dobijenim iz modela, dok je tačnost automatski prebrojanih klasova 82,27% u poređenju sa ručno prebrojanim klasovima na svim test slikama. Rezultati su prikazani u Tabeli 4.5.

Tabela 4.5: Rezultati brojanja klasova na termalnim i RGB slikama

Modalitet	Termalne slike	RGB
GT	80,26	36,58
Prebrojano na slici	75,96	36,63
Automatski pristup	72,66	33,95
Tačnost	89,22	82,27
Greška	11,74	18,46
Pirsonov koeficijent	0,73	0,64
Koeficijent determinacije	0,53	0,40

4.2 Rezultati segmentacije instanci

4.2.1 Studija slučaja I: Paradajz – kvantitativna i vizuelna analiza

Detaljna analiza uticaja izbora tehnike za izdvajanje latentnih spektralnih komponenti u hiperspektralnih slikama u kombinaciji sa jednim od obučanih modela za segmentaciju instanci na konačnu preciznost identifikacije kruničnih listova paradajza prikazana je u nastavku. PCA se često fokusira na detekciju linearnih korelacija među varijablama, što može biti nedovoljno za detekciju složenih relacija prisutnih među hiperspektralnim kanalima, a takođe može imati ograničenja u scenarijima gde srednje i kovarijanske statistike same po sebi ne predstavljaju dovoljno kompleksnu strukturu podataka. PPCA, probabilistička ekstenzija PCA, može uhvatiti unutrašnju probabilističku strukturu podataka, što je posebno relevantno pri radu sa neizvesnošću hiperspektralnih informacija. ICA, s druge strane, ima za cilj otkrivanje statistički nezavisnih odziva u podacima, što može otkriti osnovna fizička ili hemijska svojstva u hiperspektralnim slikama [298]. NMF, sa svojim ne-negativnim ograničenjem, može otkriti aditivne komponente što je korisno za interpretaciju složenih spektralnih informacija, posebno u domenu segmentacije, jer su objekti na slikama često sastavljeni od aditivnih delova, a negativne vrednosti mogu biti besmislene u mnogim fizičkim kontekstima (npr. spektralni refleks ili intenzitet piksela). Ova karakteristika čini NMF pogodnim za zadatke segmentacije u hiperspektralnim slikama, gde intenzitet piksela (i izvedene karakteristike) moraju biti nenegativni. Ovo ograničenje osigurava da izračunate komponente budu interpretabilne, predstavljajući aditivne kombinacije originalnih spektralnih kanala, što je od suštinskog značaja pri analizi bioloških struktura poput kruničnih listova paradajza. Pored toga, NMF se fokusira na izdvajanje značajnih delova u spektru, što je u skladu sa ciljem segmentacije kruničnih listova paradajza. Nasuprot tome, PCA i ICA se fokusiraju na maksimizaciju varijanse i statističke nezavisnosti, ali ne garantuju uvek delove koji su direktno interpretabilni ili pogodni za segmentacione zadatke. Sposobnost NMF-a da otkrije komponente koje predstavljaju aditivne delove čini ga efikasnijim za izolaciju različitih regiona ili karakteristika u hiperspektralnim slikama. Dok PCA i ICA proizvode komponente sa nižom varijabilnošću (što je indikativno kroz niže standardne devijacije), te su komponente često apstraktnije

i manje interpretabilne. NMF komponente, iako pokazuju nešto veću varijabilnost od PCA i ICA, i dalje su relativno stabilne, što sugerise da NMF ostvaruje dobar balans između izračunatih relevantnih komponenti i stabilnosti. Literatura dodatno podržava robusnost NMF-a u hiperspektralnom "*unmixingu*" zbog nenegativnog ograničenja i interpretabilnosti komponenti [299]. Segmentacija hiperspektralnih slika je uspešno postignuta nakon NMF-a u [300], kao i u ovoj doktorskoj disertaciji. Dakle, izbor tehnike za analizu hiperspektralnih podataka treba biti vođen specifičnim ciljevima, karakteristikama i kompleksnostima skupa podataka, uz uvažavanje prednosti i ograničenja svake metode.

Za ovu svrhu, korišćen je prvi skup hiperspektralnih podataka (3.1). Projekcija za svaku tehniku kodovanja spektralnih informacija je naučena izvođenjem 6-fold kros-validacije sa usvojenim odnosom podela skupa podataka.

Sumirani prikaz standardnih devijacija svih koeficijenata, procenjenih projekcijom na trodimenzionalni podprostor kroz sve foldove, prikazan je u Tabeli 4.6.

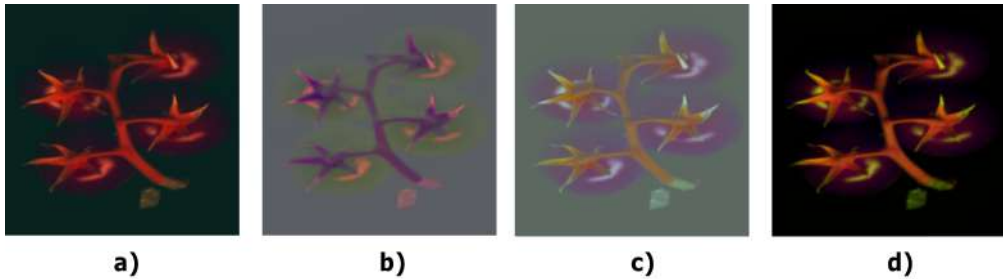
Tabela 4.6: Za svaku tehniku latentnog izdvajanja spektralnih informacija prikazana je suma standardnih devijacija svih koeficijenata, estimirana projekcijom na trodimenzionalni potprostor, kroz sve foldove.

PCA	PPCA	ICA	NMF
0,38	2,87	10,21	13,52

Vrednosti prikazane u tabeli predstavljaju stvarne vrednosti izražene u jedinicama $\times 10^{-4}$.

Ovo omogućava kvantitativno poređenje varijabilnosti između različitih tehnika za izdvajanje latentnih hiperspektralnih informacija (PCA, PPCA, ICA i NMF) i dodatno podržava izbor NMF-a za segmentaciju krničnih listova paradajza. Iako NMF pokazuje najveću sumu standardnih devijacija kroz sve foldove, ovo nije nužno mana. Povećana varijabilnost reflektuje sposobnost metode da detektuje složenije karakteristike koje su pogodne za segmentaciju. NMF-ovo nenegativno ograničenje osigurava interpretabilnost komponenti i njihovu korepondenciju sa realnim regionima slike, što je ključno za identifikaciju različitih delova krničnih listova paradajza. U poređenju sa drugim tehnikama, PCA i PPCA imaju manju sumu standardnih devijacija, što ukazuje na stabilnije, ali manje informativne i složene komponente. ICA se fokusira i detektuje nezavisne komponente, ali sa većom

varijabilnošću, što može smanjiti pouzdanost segmentacije. Razlike u naučenim projekcijama među foldovima nisu bile statistički značajne. Pošto su varijacije među foldovima bile prisutne samo u četvrtom decimalnom mestu, nasumično je odabrana jedna od šest mogućih transformacija po tehnici za buduće eksperimente. Primeri transformisanih hiperspektralnih slika iz prvog HSI skupa prikazani su na Slici 4.5.



Slika 4.5: Kodovane spektralne informacije jedne slike iz skupa podataka 1, unutar trodimenzionalnog podprostora korišćenjem: a) PCA, b) ICA, c) PPCA, i d) NMF tehnike, prikazano kroz trokanalnu pseudokolor sliku.

Predloženi pre-trenirani modeli segmentacije instanci [52, 142, 179] su dodatno prilagođeni na hiperspektralne slike, gde je 196-dimenzionalna spektralna informacija sa smanjenim šumom predstavljena u trodimenzionalnom podprostoru koristeći pomenute tehnike za izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama. Augmentacija tokom inferencije uključuje promenu veličine ulazne slike uz očuvanje relativnog odnosa širine i visine slike. Specifična veličina slike se nasumično bira unutar definisanog opsega, što osigurava minimalnu veličinu kraće ivice dok se održava proporcija stranica. Ova augmentacija omogućava raznovrsne reprezentacije, poboljšavajući robusnost modela. Promene uključuju transformacije: *resize* i *flip*, adresirajući varijacije u položaju kruničnih listova. *Flip* simulira različite orijentacije kruničnih listova, dok *resize* standardizuje veličinu, olakšavajući analizu karaktera.

Tokom 6-fold kros-validacije, korišćena je po jedna slika po batch-u (približno 20 instanci) radi optimizacije memorije na GPU-u. Korišćena je *warm-up cosine learning rate* [301] metoda, koja postepeno povećava LR do ciljane vrednosti, zatim smanjuje po kosinusnom obrascu kroz 5000 epoha, čime se osigurava stabilna konvergencija [301, 302, 303].

Tabela 4.7 prikazuje performanse predloženog Mask R-CNN modela sa različitim backbone-ovima u smislu mAP, za četiri tehnike izdvajanja latentnih hiperspektralnih informacija NMF sa Mask R-CNN + FPN backbone-om (dalje NMFMaskFPN) postiže najbolji rezultat mAP $94,05 \pm 3,25\%$.

Tabela 4.7: Srednja vrednost i standardna devijacija mAP kroz 6 validacionih foldova za tri varijante Mask R-CNN modela (FPN, C4, i DC5) i korišćene tehnike izdvajanja latentnih hiperspektralnih informacija (PCA, NMF, ICA, PPCA)

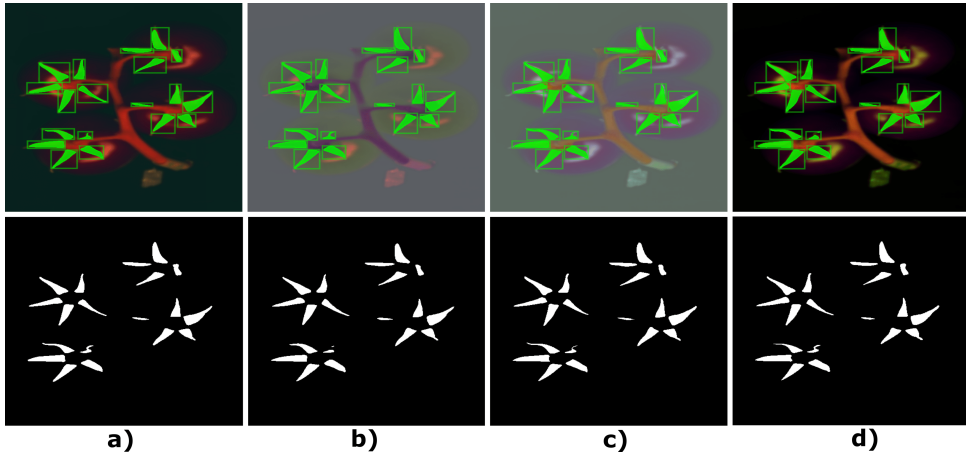
Model	PCA	PPCA	ICA	NMF
C4	$92,92 \pm 3,43$	$93,50 \pm 4,12$	$93,33 \pm 3,03$	$92,19 \pm 3,84$
DC5	$91,45 \pm 3,17$	$93,54 \pm 3,23$	$93,21 \pm 2,71$	$92,09 \pm 3,39$
FPN	$92,39 \pm 3,33$	$92,93 \pm 3,39$	$92,85 \pm 4,22$	$94,05 \pm 3,25$

Slične rezultate ilustruje slika 4.6. Dodatni metrički uvid uključuje tačnost, FNR i FPR kroz epohe (Slika 4.7). Tabela 4.8 prikazuje preceanjene i podceanjene greške predikcije izražene u procentima za test skup kruničnih listova paradajza, kao dopunu performansi prikazanim u tabeli sa mAP metrikama. Vrednosti predstavljaju procentualna odstupanja modela u odnosu na referentne instance. Prikazani rezultati omogućavaju poređenje efekata različitih metoda redukcije dimenzionalnosti (PCA, PPCA, ICA, NMF) u kombinaciji sa arhitekturama C4, DC5 i FPN. Najmanje greške (**0/0**) ukazuju na najbolju usklađenost predikcije sa stvarnim oblicima lista, i potvrđuje performanse najboljeg modela iz prethodne tabele 4.7.

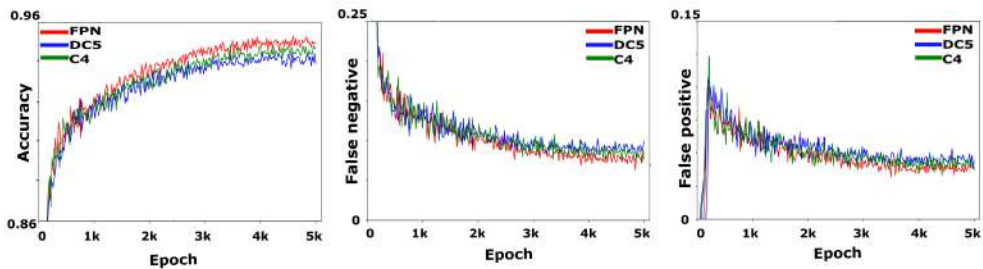
Tabela 4.8: Greške predikcije $\bar{e}/\underline{e}$ za krunične liste paradajza na test skupu, izražene u procentima

Model	PCA	PPCA	ICA	NMF
C4	1,69/0	1,69/0	1,69/1,69	3,39/0
DC5	0/1,69	0/1,69	1,69/0	0/1,69
FPN	0/1,69	1,75/1,69	1,75/3,39	0/0

Sledeći korak predstavlja evaluaciju NMFMaskFPN modela treniranog na hiperspektralnim slikama sorte *Brioso* (skup podataka br. 1)



Slika 4.6: Detektovane instance kruničnih listova paradajza pomoću NMFMaskFPN modela koristeći: a) PCA b) ICA c) PPCA d) NMF tehnike za izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama (prvi red), i pripadajuće binarne maske (drugi red).



Slika 4.7: Analiza performansi Mask R-CNN modela kroz epohe sa različitim backbone-ovima (FPN, DC5, C4) evaluirano na validacionim hiperspektralnim slikama kodovanim NMF tehnikom.

na hiperspektralnim slikama iste sorte iz skupa podataka br. 2. Hiperspektralne slike iz drugog po redu skupa podataka pripremljene su istom procedurom kao i za prvi skup. Iako je dobijeni mAP na drugom skupu podataka nizak, model je detektovao 37% ukupnog broja lista (61/164 na 32 slike), od kojih je 77% tačno pozitivno, a 0,29% netačno pozitivno, uključujući 2 lista sa dvostrukom detekcijom. Loši rezultati mogu biti posledica razlike u distribuciji između skupova. U prvom skupu podataka, model je treniran na grozdovima sa četiri ploda, učeći da prepozna kompleksne oblike i odnose između plodova.

Nasuprot tome, drugi skup podataka sadrži slike sa samo jednim paradjazom i nekoliko listova, što dovodi do lošijih performansi. Ostali faktori uključuju razlike u kontekstu, skali (pojedinačni list je veći u odnosu na grozd) i fokus modela je na detekciji grozdova. Prekomerno učenje specifičnih karakteristika može takođe otežati generalizaciju.

U narednom koraku treniran je NMFMaskFPN model kroz 8-fold kros-validaciju, koristeći 86% podataka iz drugog po redu skupa podataka za treniranje (28/32 slike po sorti, u proseku 130 krunična lista po sorti, ili sve sorte objedinjene: 112/128 slika, približno 570 lista), i 7% za validaciju i testiranje (2/23 slike po sorti, što je 16 lista, ili 8/112 slika spojeno, 41 list). Ista procedura treniranja za prvi skup HSI slika primenjena je na četiri pojedinačne sorte i spojeni, tj. objedinjeni skup podataka. Rezultati su prikazani u Tabeli 4.9.

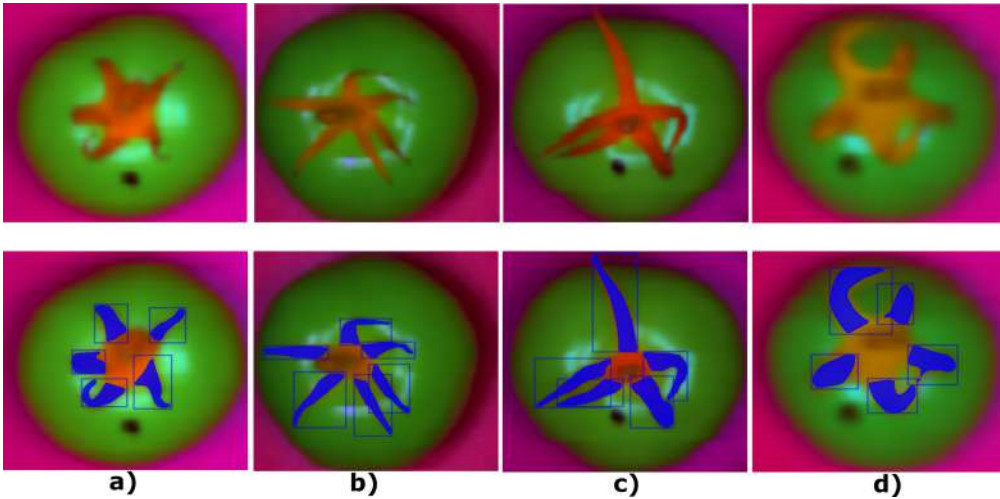
Tabela 4.9: Srednja vrednost i standardna devijacija mAP za NMF-MaskFPN model kroz 8-fold kros-validaciju

<i>Brioso</i>	<i>Capricia</i>	<i>Provine</i>	<i>Sao Paolo</i>	Merged
94,0 ± 8,4	97,4 ± 7,4	98,5 ± 4,3	82,2 ± 12,3	86,4 ± 7,2

Rezultati pokazuju da predloženi pristup transfer učenja postiže bolje performanse i robusnost pri promenjenim eksperimentalnim dizajnom i različitom broju instanci po slici, poboljšavajući generalizaciju među sortama (vidi Slika 4.8).

Predloženi pristup je dao najbolje rezultate za sortu *Provine*, a najslabije za *Sao Paolo*, zbog male veličine ploda i lista i niže rezolucije. Uprkos nižoj preciznosti segmentacije kod nekih sorti, ukupna tačnost detekcije kruničnog lista ostaje visoka. Model iz Tabele 4.9 postigao je 98,52% preciznosti na test skupu za sorte *Brioso*, *Provine* i *Sao Paolo*, dok je za *Capricia* zabeležena samo jedna netačna pozitivna detekcija.

U narednom koraku istraživanja, analizirane su performanse Mask R-CNN modela dodavanjem konvolucionog sloja na početku arhitekture (ConvMaskFPN) za izdvajanje latentnih spektralnih komponenti hiperspektralne informacije. Ispitali smo inicijalizaciju težina korišćenjem XavierNormal, KaimingNormal, Uniform i Normal [304]. Rezultati (Tabela 4.10) na objedinjenom skupu podataka dobijeni su prilagođavanjem osnovnog Mask R-CNN + FPN modela kroz 5000 epoha. Performanse sa XavierN i KaimingN bile su blago bolje od Uniform i Normal inicijalizacija.



Slika 4.8: Rezultati detekcije lista paradajza na četiri sorte iz drugog HSI skupa: a) *Briosio* b) *Capricia* c) *Provine* d) *Sao Paolo* sa NMF-MaskFPN

Tabela 4.10: mAP dobijen inicijalizacijom ConvMaskFPN sloja koeficijentima iz tehnika: XavierN, KaimingN, Uniform i Normal

XavierN	KaimingN	Uniform	Normal
82, 24 ± 4, 89	82, 18 ± 4, 49	81, 12 ± 4, 52	82, 07 ± 3, 13

Tabela 4.11: mAP dobijen inicijalizacijom ConvMaskFPN sloja koeficijentima iz korišćenih tehnika izdvajanja latentnih hiperspektralnih informacija

PCA	PPCA	ICA	NMF
86, 95 ± 3, 63	87, 35 ± 2, 44	86, 96 ± 2, 01	87, 40 ± 3, 81

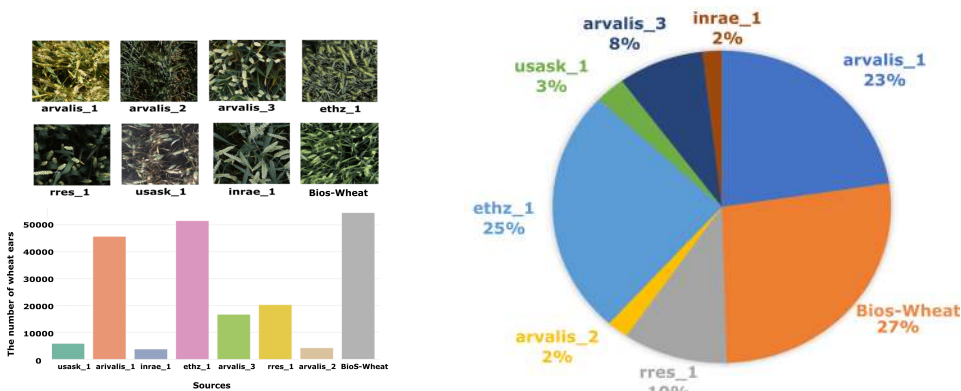
Težine dodatnog sloja su inicijalizovane naučenim koeficijentima iz tehnika za izdvajanje latentnih hiperspektralnih informacija. Nakon prilagođenja, ConvMaskFPN, performanse na validacionom i test skupu pokazale su blaga poboljšanja preciznosti, posebno kod inicijalizacije sa NMF koeficijentima, postižući poboljšanje od 1% (vidi Tabele 4.10 i 4.11). Ovi rezultati pokazuju potencijal primene pred-inicijalizacije arhitektura neuronskih mreža NMF koeficijentima za poboljšanje detekcije lista paradajza.

4.2.2 Studija slučaja II: Pšenica – kvantitativna i vizuelna analiza

U ovom odeljku fokus je na rezultatima dobijenim analizom kvaliteta slika i performanse modela za detekciju klasova pšenice posebno za GWHD, Bios-Wheat, kao i za objedinjeni skup podataka, koji uključuje oba skupa podataka, GWHD i BioS-Wheat, u daljem tekstu označen kao BioS-Wheat&GWHD (vidi tabele 3.3, 3.4 i 3.6).

4.2.3 Detekcija klasova pšenice - analiza modela i stabilnost performansi

U ovom delu istraživanja primenjena je 5-fold kros validacija kako bi se procenila sposobnost generalizacije razmatranih modela. Iako skup podataka obuhvata osam institucija, tj. izvora podataka, podskupovi nemaju istu količinu podataka; stoga smo odabrali 5-fold krosvalidaciju umesto pristupa *leave-one-source-out (LOSO)* kako bismo osigurali da svaki fold sadrži dovoljnu zastupljenost svih klasa. Foldovi su generisani heuristički, a kroz višestruka ponavljanja, slučajne podele dosledno su davali najrobustnije i rezultate. Za svaki fold, skup podataka je pažljivo određen uz održavanje ukupne raspodele klasa kako bi podskupovi bili reprezentativni, što je ilustrovano na slici desno (Slika 4.9).



Slika 4.9: Levo: primeri slika iz različitih izvora sa anotiranim brojem klasova pšenice. Desno: prosečna zastupljenost svakog izvora u skupu kroz 5-fold eksperimente.

Odabrani pristup je i u skladu sa principima teorije distribucije informacija: nasumično uzorkovanje maksimizuje entropiju trening i

test skupa, dok minimizuje potencijalnu pristrasnost uvedenu struktuiranim ili izvorno zasnovanim podelama, čime se dobija objektivnija i generalnija procena performansi modela kroz čitavu distribuciju podataka [305, 306]. Dodatno, pristup LOSO unakrsne validacije nije pogodan za ovaj skup podataka zbog ekstremne neuravnoteženosti među izvorima, koji čine od 2% do 27% ukupnog skupa podataka. Takva podela bi rezultirala veoma varijabilnim i potencijalno obmanjujućim rezultatima na test skupu: izvori koji čine samo 2–3% skupa generisali bi nepouzidane procene i vrednosti mAP50, dok bi izvori od 23–27% nesrazmerno dominirali evaluacijom i davali pristrasne rezultate. Kao posledica, trening bi mogao biti nestabilan pri izostavljanju velikih izvora, a ukupne procene generalizacije modela bile bi pogrešne. Nasuprot tome, stratifikovana nasumična podela koja poštuje i ukupnu raspodelu klasa i proporcije originalne zastupljenosti svakog od izvora pruža robusniju, reproduktivnu i interpretabilnu evaluaciju performansi modela. Idealan scenario podrazumeva da model treniran na više skupova podataka, kao u našem slučaju, može biti fino podešen, dodatno treniran na razumnom broju novih primera iz dodatog skupa podataka i i dalje postići solidne performanse. Međutim, ne može se očekivati da model održi isti nivo performansi pod strogim kriterijumom LOSO, zbog velikog broja faktora koji doprinose varijabilnosti između izvora, uključujući osvetljenje, fazu rasta useva i uslove snimanja. Ovakav pristup predstavlja praktičan kompromis između robusnog treniranja modela i njegove sposobnosti za generalizaciju. Performanse su prikazane kako za validacioni skup, tako i za izdvojeni test skup, čime je obezbeđena robusna evaluacija.

U prvom eksperimentu izabrani modeli (F-RCNN R 50, F-RCNN R 101, F-RCNN X 101, RetinaNet R 101, YOLOv8, RT-DETR) su parametrizovani (engl. *fine-tuned*) isključivo na GWHD skupu, koji sadrži 3373 slike i 147793 anotirana klasa pšenice. Od toga, 2699 slika je korišćeno za treniranje, 337 za validaciju i 337 za testiranje. Kao dodatni uvid u performanse i efikasnost modela, prikazano je i potrebno vremena za obuku i *inference*, kako bi se dobila celovita slika efikasnosti modela. Vreme obuke ukazuje na potrebne resurse, dok vreme inferencije odražava potencijalnu brzinu izvršavanja pri praktičnoj primeni.

Tabela 4.12 prikazuje poređenje šest odabranih modela evaluiranih na GWHD skupu, sa prosečnim vrednostima mAP50, vremenom treniranja i vremenom inferencije po slici. Svi rezultati su u skladu sa SOTA (*state-of-the-art*) rezultatima iz novijih studija [307, 308, 309].

Tabela 4.12: GWHD skup — rezultati

Model	mAP50	Vreme treniranja	Vreme inferencije
F-RCNN R 50	85,37 $\pm$ 0,87	1h 40min 56s	65 ms
F-RCNN R 101	85,07 $\pm$ 0,38	2h 10min 52s	97 ms
F-RCNN X 101 32x8d	84,84 $\pm$ 0,37	4h 6min 17s	116 ms
RetinaNet R 101	84,79 $\pm$ 0,46	2h 7min 43s	85 ms
YOLOv8	85,91 $\pm$ 0,46	5h 12min 18s	18 ms
RT-DETR	87,47 $\pm$ 0,21	8h 48min 2s	137 ms

Rezultati na GWHD skupu ukazuju na kompromis između tačnosti, vremena treniranja i brzine inferencije kod evaluiranih modela. RT-DETR ostvaruje najveću tačnost detekcije sa mAP50 od 87,47%, zahvaljujući svojoj transformer arhitekturi, ali uz cenu najdužeg potrebnog vremena obuke (više od 8 sati) i najsporije inferencije (137 ms po slici). Među CNN detektorima, YOLOv8 pruža optimalan balans, dostižući konkurentnu tačnost (85,91%) uz najbržu inferenciju (18 ms), što ga čini pogodnim za primenu u realnom vremenu, naročito u kontekstu proširenog skupa sa podacima sa pametnih mobilnih telefona. Varijante Faster R-CNN pokazuju stabilne rezultate (85% mAP), pri čemu ResNet-50 backbone daje najkraće vreme treniranja (oko 1h 40min) i umerenu brzinu inferencije (65 ms). RetinaNet sa ResNet-101 backbone-om postiže nešto slabije rezultate u poređenju sa najboljim modelima, ali se pozicionira kao srednje dobro rešenje u ovom poređenju.

U drugom eksperimentu, rezultati dobijeni korišćenjem isključivo BioS-Wheat skupa podataka u skladu su sa karakteristikama samog skupa. Sa 5696 slika koje sadrže visoku gustinu klasova pšenice po slici, u poređenju sa osnovnim GWHD skupom, koji obuhvata 3442 slike sa manjim brojem klasova po slici, arhitekture poput RetinaNet R101, YOLOv8 i RT-DETR postižu relativno visoke vrednosti mAP50 (86–91%). Niže performanse postižu arhitekture bazirane na F-RCNN, posebno R50 i R101, i odražavaju njihovu limitiranu sposobnost precizne detekcije velikog broja malih klasova pšenice. Vreme treniranja proporcionalno raste sa složenošću modela i gustinom, a samim tim i brojem klasova pšenice po slici, dok vreme inferencije ostaje u skladu sa očekivanjima arhitekture modela, pri čemu YOLOv8 postiže najbržu obradu, dok je RT-DETR sporiji zbog *transformer-based* pristupa

koji generiše veliki broj predloga za detekciju klasova pšenice. Ovi rezultati ukazuju da BioS-Wheat skup podataka pruža dovoljnu pokrivenost i složenost, omogućavajući smisleno unapređenje kao dodatak postojećem GWHD osnovnom skupu podataka. Dobijeni rezultati na BioS-Wheat skupu, prikazani su u Tabeli 4.13.

Tabela 4.13: BioS-Wheat—rezultati

Model	mAP50	Vreme treniranja	Vreme inferencije
F-RCNN R 50	79,58 $\pm$ 1,24	1h 45min 12s	65 ms
F-RCNN R 101	71,21 $\pm$ 1,75	2h 15min 07s	27 ms
F-RCNN X 101 32x8d	83,58 $\pm$ 2,34	4h 25min 11s	116 ms
RetinaNet R 101	86,4 $\pm$ 1,96	2h 10min 22s	85 ms
YOLOv8	87,2 $\pm$ 1,41	5h 20min 10s	18 ms
RT-DETR	90,1 $\pm$ 0,12	9h 01min 36s	137 ms

U trećem eksperimentu fokus je na performansama istih modela treniranih na referentnom GWHD skupu podataka i fino podešenih na objedinjeni skup (BioS-Wheat + GWHD). Ceo skup sadrži 9059 slika klasova pšenice sa 202278 anotacija, od čega je 7247 slika korišćeno za treniranje, a po 453 za validaciju i 453 za testiranje. Ovaj pristup je u skladu sa najboljim praksama transfer učenja, gde se prioritet daje maksimizaciji količine podataka za dodatnu obuku modela, dok je manji validacioni skup dovoljan za praćenje performansi [310, 311]. Hiperparametri za proces treniranja modela su delimično modifikovani u odnosu na podrazumevanu *default* Detectron2 konfiguraciju kako bi se bolje prilagodili skupu podataka. Konkretno, *learning rate* je smanjen na 0,0001, veličina batch-a je postavljena na 4 slike po iteraciji, za raspored iterativnog smanjivanja *learning rate-a* korišćena je metoda *WarmupCosineLR*, što omogućava postepeno uspostavljanje krive učenja praćeno kosinusnom dekompozicijom, čime se poboljšava stabilnost konvergencije. Svi ostali parametri treniranja, uključujući optimizator, ostali su na podrazumevanim, *default* vrednostima; posebno, korišćen je stohastički gradijentni spust (SGD) sa momentom, kako je definisano u originalnoj konfiguraciji. Empirijski, testirali smo LR u opsegu od 0,0001 do 0,01 i veličine batch-a od 2 do 8; niže vrednosti LR-a i batch-a omogućile su stabilniju konvergenciju i bolje performanse na validacionom skupu, dok veći LR ili veći batch-evi često uzrokuju osci-

lacije u treningu ili sporije konvergencije.

Modeli sa najstabilnijim i SOTA rezultatima u različitim studijama — YOLOv8 [312] i RT-DETR [313] — pokazali su značajna poboljšanja u tačnosti u poređenju sa Faster R-CNN varijantama, iako su zahtevali znatno više vremena i resursa za treniranje (YOLOv8 oko 5,9h, RT-DETR preko 9h za 100 epoha, vidi Tabelu 4.14).

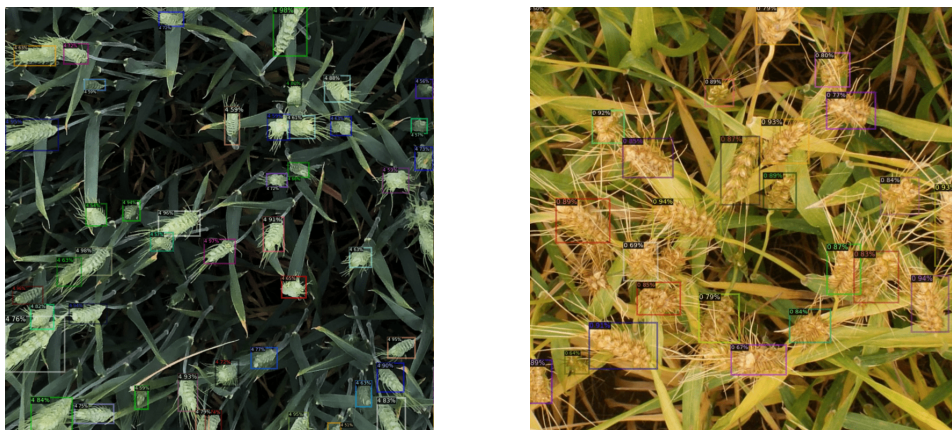
Tabela 4.14: BioS-Wheat + GWHD skup — rezultati

Model	mAP50	Vreme treniranja	Vreme inferencije
F-RCNN R 50	81,48 $\pm$ 2,31	1h 49min 34s	65 ms
F-RCNN R 101	72,36 $\pm$ 0,69	2h 22min 12s	27 ms
F-RCNN X 101 32x8d	84,33 $\pm$ 0,37	4h 36min 21s	116 ms
RetinaNet R 101	87,79 $\pm$ 0,76	2h 17min 41s	85 ms
YOLOv8	88,00 $\pm$ 0,56	5h 34min 11s	18 ms
RT-DETR	91,00 $\pm$ 0,41	9h 14min 42s	137 ms

Prilikom evaluacije na objedinjenom skupu podataka, koji uključuje i BioS-Wheat, pojedini modeli pokazali su smanjenu preciznost detekcije u poređenju sa treniranjem isključivo na podskupovima sa redom gustinom. Povećana vizuelna kompleksnost koju donosi BioS-Wheat, karakterisana gušćim sklopom i većom gustinom useva, pokazala se naročito izazovnom za modele poput Faster R-CNN R101 i RetinaNet, koji su zabeležili blagi pad performansi. Ipak, modeli sa snažnijim sposobnostima generalizacije, poput YOLOv8 i RT-DETR, zadržali su relativno stabilne mAP vrednosti, što ukazuje na njihovu otpornost na zaklanjanje i međusobno preklapanje objekata. Primeri slika sa detektovanim klasovima pšenice iz objedinjene baze prikazani su na slici 4.10.

Integracija BioS-Wheat skupa podataka otkrila je različita ponašanja evaluiranih arhitektura. Relativno poređenje performansi modela treniranih na objedinjenom skupu podataka u odnosu na osnovni GWHD skup dati su u Tabeli 4.15

F-RCNN modeli, posebno sa R101 backbone-om, pokazali su pad performansi, sa naglim smanjenjem od 12,71 mAP50, što ukazuje na limitiranu robusnost prema heterogenosti podataka i moguće natprilagođenje (*overfitting*) na distribuciju originalnog GWHD skupa podataka. Suprotno tome, RetinaNet, YOLOv8 i RT-DETR su imali koristi od proširenog skupa podataka, što ukazuje na jače sposobnosti generali-



Slika 4.10: Detekcija klasova pšenice na test slikama pomoću RetinaNet R101 iz objedinjene baze podataka; levo primer iz BioS-Wheat skupa, desno iz GWHD skupa

Tabela 4.15: Relativno poređenje performansi modela treniranih na objedinjenom skupu podataka u odnosu na osnovni GWHD skup.

Model	$\Delta mAP50$	Δ vreme treniranja
F-RCNN R 50	-3,89	+08min 38s
F-RCNN R 101	-12,71	+11min 20s
F-RCNN X 101 32x8d	-0,51	+30min 04s
RetinaNet R 101	+3,00	+09min 58s
YOLOv8	+2,09	+21min 53s
RT-DETR	+3,53	+26min 40s

zacije; među njima, RT-DETR je pokazao najveće poboljšanje (3,53 mAP50), naglašavajući sposobnost ovog modela da iskoristi dodatnu varijabilnost u trening podacima. Vreme treniranja je dosledno povećano kod svih modela, što je očekivano s obzirom na veći volumen ulaza, dok su vremena inferencije generalno ostala stabilna. Posebno je važno istaći da BioS-Wheat unosi agronomski gušće sorte pšenice i složenije uslove na terenu karakterisane visokom gustinom setve, izraženim zaklanjanjem, preklapanjem klasova i povećanom vizuelnom složenošću koji su u velikoj meri bili nedovoljno zastupljeni u originalnom GWHD skupu. Ovo obogaćivanje povećava sposobnost modela da generalizuju u različitim i izazovnijim scenarijima.

Umerena kvantitativna poboljšanja zabeležena u metričkim vred-

nostima modela stoga ne predstavljaju znak redundanse ili šuma, već ukazuju na činjenicu da skup podataka testira granice postojećih arhitektura. Takvi podaci otkrivaju slabe tačke konvencionalnih detektora i time postaju dragoceni za proveru robusnosti i sposobnosti generalizacije. Sa praktične strane, BioS-Wheat odražava realne uslove na terenu intenzivne poljoprivredne proizvodnje, gde je vizuelna složenost neizbežna. Iako složenije arhitekture detekcije (npr. korišćenje orijentisanih granica, mehanizama pažnje zasnovanih na transformer arhitekturama ili segmentacije instanci) mogu bolje da odgovore na ovakve izazove, one često dolaze uz cenu smanjenja brzine inferencije i povećanja računarskih zahteva. Imajući u vidu dugoročan cilj postizanja detekcije klasova pšenice u realnom vremenu u praktičnim uslovima, ovaj kompromis nije zanemarljiv. Stoga, uključivanje BioS-Wheat skupa u razvoj modela predstavlja strateški korak: unosi kompleksnost koja povećava robusnost bez preteranog prilagođavanja idealizovanim uslovima, a istovremeno zadržava mogućnost brze inferencije.

Sveukupno, problem višestruke detekcije i prebrojavanja klasova dolazi posebno do izražaja u ovako zahtevnom zadatku. Mnogi autori, poput [218], postižu višu preciznost uklanjanjem određenih oblasti graničnih okvira pomoću prilagođenog RetinaNet modela. Iako ovakav pristup povećava preciznost, smanjuje sposobnost detekcije veoma malih ili veoma velikih klasova (nametnuti prag na površini graničnog okvira isključuje one manje od 20 i veće od 2×10^5 piksela). Još jedan način rešavanja ovog problema jeste dovođenje svih slika u horizontalnu orijentaciju, kako predlažu autori iz [226], ili korišćenje orijentisanih graničnih okvira koji bolje obuhvataju neosno poravnate objekte, kao u radu [314]. Dodatno, analiziran je RT-DETR model, koji je ostvario najbolje rezultate mAP50 na objedinjenom skupu (videti Tabelu 4.15), i to u performansama detekcije po izvorima (videti Tabelu 4.16). Visoka varijabilnost u standardnoj devijaciji mAP50 metrike ukazuje na osetljivost na razlike unutar izvora. Uključivanje ovako raznovrsnih izvora, iako izazovno, doprinosi robusnosti objedinjene baze podataka.

U celini, namera uključivanja BioS-Wheat skupa nije primarno unapređenje performansi detekcije klasova pšenice, već osiguravanje da modeli budu izloženi punom spektru realne varijabilnosti na terenu, čime se razvoj usklađuje sa agronomskom relevantnošću i operativnom izvodljivošću. Stoga, dodatni eksperimenti su posvećeni ispitivanju robusnosti modela i dizajnirani su tako da procene kako se najbolji model ponaša pod realnim perturbacijama koje se mogu javiti prilikom mo-

Tabela 4.16: Prosečne vrednosti mAP50 ostvarene RT-DETR modelom po izvorima podataka

Arvalis1	BioS-Wheat	RRES1	Arvalis2	ETHZ1	USASK1	Arvalis3	INRAE1
35,33 ± 0,76	53,17 ± 0,40	53,83 ± 1,93	32,31 ± 3,81	41,68 ± 1,52	42,43 ± 3,31	41,30 ± 8,26	52,82 ± 2,37

bilne ili terenske primene, kao što su varijacije osvetljenja, Gausovo i zamućenje usled pokreta (engl. *motion blur*), kao i JPEG kompresija. Dobijeni rezultati, prikazani u Tabeli 4.17 ukazuju da RT-DETR model održava visok mAP50 u većini uslova, što predstavlja primarni kriterijum uspešnosti modela u ovoj studiji. Na primer, pri varijacijama osvetljenja od $\pm 30\%$, mAP50 ostaje dosledno oko 0,85, demonstrirajući snažnu robusnost na promene osvetljenja. Slično tome, Gaussovo zamućenje sa malim do umerenim kernelima (3×3 , 5×5 , 7×7) dovodi do minimalnog smanjenja mAP50, što ukazuje da model efikasno održava performanse detekcije uprkos redukciji kvaliteta slike. JPEG kompresija pokazuje očekivanu osetljivost pri ekstremnoj kompresiji ($q=10$, mAP50 opada na 0,81), ali pri umerenim nivoima ($q=30$ i 50) model i dalje postiže mAP50 blizu 0,86, potvrđujući praktičnu otpornost.

Tabela 4.17: Evaluacija robusnosti modela pod različitim perturbacijama slike. mAP_{50} je istaknuta kao glavna metrika.

Perturbacija	Preciznost	Recall	F1	mAP_{50}
Brightness -30	0,91	0,87	0,89	0,90
Brightness 0	0,92	0,89	0,90	0,91
Brightness +30	0,91	0,87	0,88	0,90
Gaussian 3	0,90	0,87	0,88	0,89
Gaussian 5	0,89	0,85	0,87	0,88
Gaussian 7	0,88	0,83	0,86	0,87
JPEG 10	0,82	0,77	0,80	0,81
JPEG 30	0,87	0,84	0,85	0,86
JPEG 50	0,90	0,87	0,88	0,89
Motion 3	0,91	0,88	0,89	0,90
Motion 5	0,90	0,87	0,88	0,89
Motion 7	0,88	0,85	0,90	0,87

Sveukupno, rezultati naglašavaju da, iako ekstremne perturbacije utiču na performanse, model ostaje robusan u većini realnih distorzija slike, sa visokim mAP50, što podržava njegovu pogodnost za primenu u promenljivim uslovima snimanja u realnom okruženju.

4.2.4 Automatsko brojanje klasova pšenice

Preciznost brojanja klasova pšenice je procenjena na osnovu pomenutih metrika: RCE i AC. Najbolji model postigao je RCE vrednost od 0,11 i AC od 0,885 na BioS-Wheat skupu podataka. Pored ovih metrika, za procenu preciznosti korišćeni su i R^2 skor [315] i standardni koeficijenti korelacije: Pearson [316], Spearman [317], Kendall [318]. Takođe, da bismo stekli uvid u broj klasova koji su višestruko detektovani ili propušteni, izračunati su srednja kvadratna greška (MSE) i koren srednje kvadratne greške (RMSE) između detektovanog i stvarnog broja klasova. Rezultati su prikazani u Tabeli 4.18.

Tabela 4.18: Korelacije između označenog i detektovanog broja klasova u BioS-Wheat skupu podataka

Pearson	Spearman	Kendall	R^2	MSE	RMSE
0,849	0,724	0,846	0,693	2,670	1,634

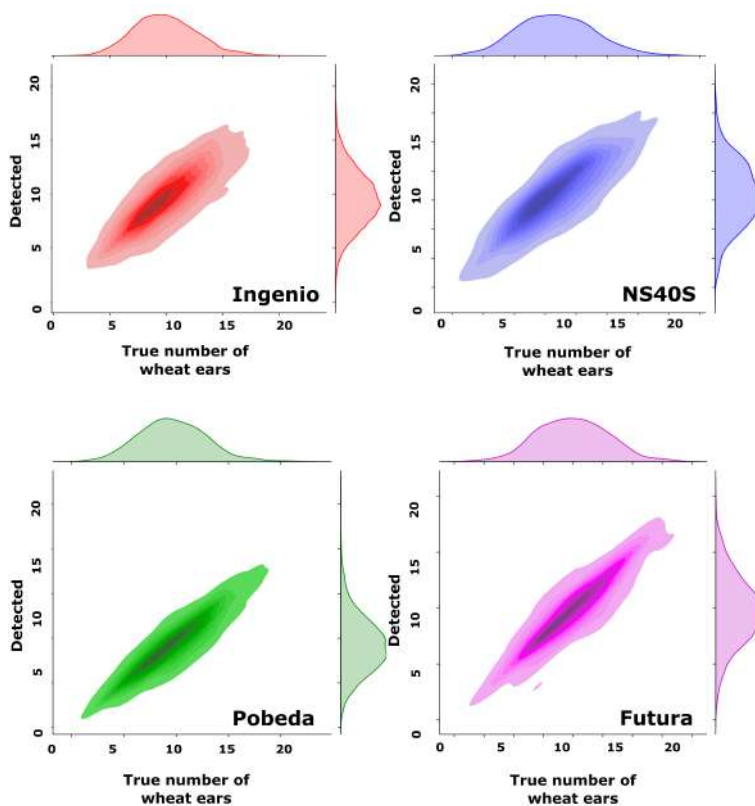
Dobijene metrike ukazuju da model ima dobro, ali ne savršeno slaganje na BioS-Wheat skupu podataka. Pearson, Spearman i Kendall koeficijenti pokazuju snažnu povezanost između stvarnog i detektovanog broja klasova, što znači da model dobro prepoznaje opšti trend i rangiranje. Međutim, R^2 , MSE i RMSE vrednosti pokazuju da predviđanja nisu u potpunosti usklađena sa stvarnim brojem klasova, što otkriva prostor za unapređenja. Da bismo detaljnije razumeli performanse modela i identifikovali konkretne slabosti, sprovedena je analiza po sortama pšenice (Tabela 4.19).

Analiza pokazuje da je sorta *Futura* imala najjače korelacije i najviši R^2 skor, što ukazuje da model najbolje detektuje ovu sortu. Reč je o visokoprinosnoj sorti hlebne pšenice sa gustim, jasno oblikovanim i ujednačenim klasovima, što doprinosi njihovoj lakšoj detekciji. Kod sorte *Futura*, model takođe postiže visoku preciznost i dobre korelacije, čime se svrstava na drugo mesto po tačnosti detekcije. Za sorte *NS40S* i posebno *Ingenio*, model daje niže rezultate, sa slabijim korelacijama i manjim R^2 skorovima.

Tabela 4.19: Korelacije između stvarnog i detektovanog broja klasova po sortama pšenice iz BioS-Wheat skupa podataka

Sorta	Pearson	Spearman	Kendall	R^2
<i>Ingenio</i>	0,809	0,684	0,812	0,616
<i>NS40S</i>	0,846	0,707	0,839	0,694
<i>Futura</i>	0,889	0,778	0,886	0,767
Futura	0,905	0,896	0,795	0,805

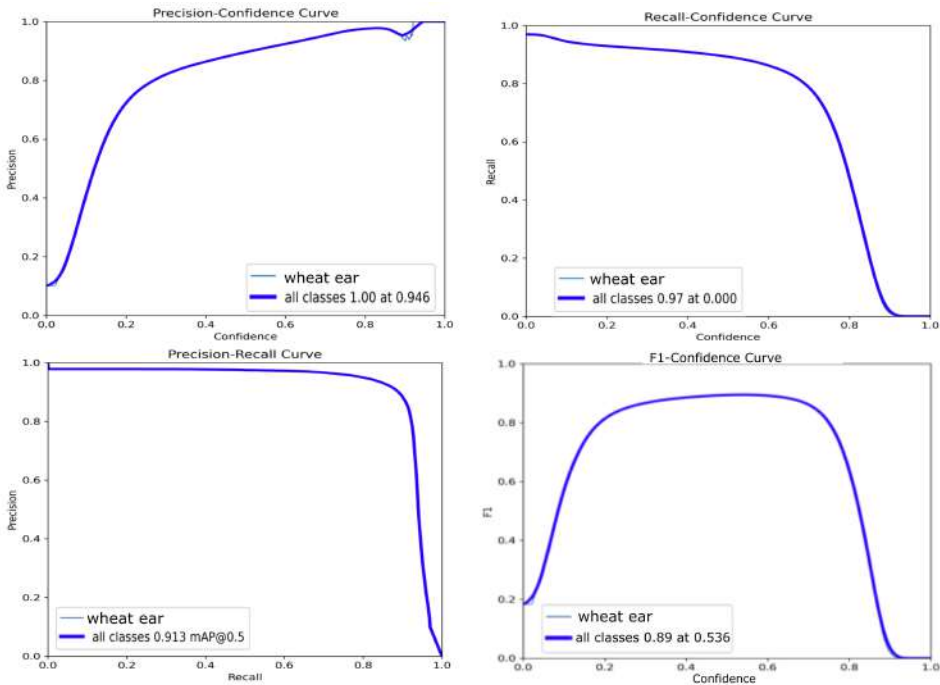
Pored toga, KDE (engl. *Kernel Density Estimation*) prikazi na Slici 4.11 ilustruju slaganje između detektovanog i stvarnog broja klasova po sortama.



Slika 4.11: KDE grafici koji prikazuju gustinu rezultata detekcije za svaku sortu u BioS-Wheat skupu podataka, naglašavajući razlike u performansama modela u odnosu na stvaran broj klasova.

Kod sorti *Futura* i *Pobeda*, tačke su gusto raspoređene duž dijagonale, što ukazuje na visoku preciznost. Sorta *Ingenio* pokazuje veću varijabilnost, ali i dalje zadržava osnovni trend slaganja. Najveću neusaglašenost ima sorta *NS40S*, kod koje se uočava široko rasipanje tačaka i najslabija tačnost detekcije.

Pored toga, uključujemo krive poverenja-performansi koje ilustruju preciznost, odziv i F1 skor u zavisnosti od pragova poverenja (videti Slika 4.12), koje pokazuju stabilnu kalibraciju i dosledne trendove performansi kroz sve foldove, potvrđujući sistemske prednosti RT-DETR arhitekture.



Slika 4.12: Evaluacija performansi zasnovana na poverenju. Gornji red: krive *Precision-Confidence* (levo) i *Recall-Confidence* (desno), Donji red: krive *Precision-Recall* (levo) i *F1 score-Confidence* (desno)

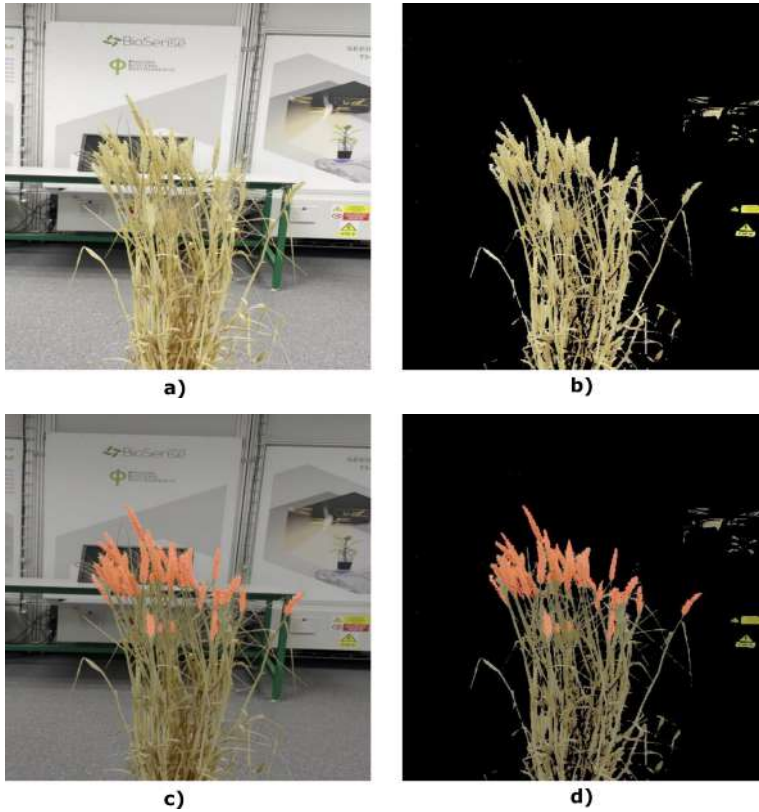
4.3 Rekonstrukcija NeRF metodama i analiza 3D reprezentacije

Nakon pripreme podataka, treniranje NeRF modela je izvršeno na grafičkim karticama, pri čemu je za skup sa uniformnom belom pozadinom trening trajao približno 4,5 sata, dok su varijante sa kompleksnom pozadinom zahtevale dodatno vreme zbog veće vizuelne kompleksnosti scene (8h). Uniformna pozadina značajno otežava rekonstrukciju pomoću NeRF-a, jer ovaj model zasniva učenje na geometriji i boji scene i razlikama između više uglova (engl. *viewpoints*). U odsustvu jasnih teksturalnih ili strukturnih referentnih tačaka, NeRF nema stabilne tačke za optimizaciju projekcija između kamera, što vodi do zamućenja i nestabilnih geometrijskih rekonstrukcija. Dodatno, i sama pšenica predstavlja uniforman i ponavljajući patern, što dodatno smanjuje lokalni kontrast i otežava modelu da razlikuje pozadinu od objekta od interesa. Stoga, evaluacija rezultata je ukazala na značajno poboljšanje kvaliteta rekonstrukcije pri korišćenju kompleksne pozadine u poređenju sa uniformnom, što se manifestovalo kroz stabilnije i detaljnije 3D rekonstrukcije klasova pšenice. Za finalnu vizualizaciju generisani su renderi sa vremenskom rezolucijom od 4 i 8 sati kako bi se dobili kontinuirani video zapisi rekonstruisanih scena.

Naredni korak obuhvatao je primenu metoda dubokog učenja za semantičku segmentaciju, kao što je UNet radi generisanja maski koje precizno obuhvataju detektovane klasove pšenice. Ove maske su zatim poslužile kao ulaz za ponovnu primenu NeRF modela (RGB + binarne maske za svaki frejm), čime je omogućena rekonstrukcija segmentovanih klasova pšenice u trodimenzionalnom prostoru, što značajno poboljšava sposobnost kvantifikacije i analize preklapajućih struktura koje je u 2D gotovo nemoguće razdvojiti.

Za procenu kvaliteta rekonstruisanog NeRF videa i njegove vremenske stabilnosti, izračunate su sledeće metrike: oštrina slike tj. frejma (engl. *Sharpness*), odnos signal–šum (engl. *Signal-to-Noise Ratio*, SNR), vremenska konzistentnost (engl. *Temporal Consistency*), (engl. *Peak Signal-to-Noise Ratio*, PSNR) i indeks strukturne sličnosti (engl. *Structural Similarity Index*, SSIM). Ove metrike omogućavaju sveobuhvatnu procenu prostorne i vremenske stabilnosti generisanih rendera. U Tabeli 4.20 prikazana je kvantitativna evaluacija kvaliteta NeRF rekonstrukcije.

Metrika oštine, izražena varijansom Laplasijanovog operatora, iznosi



Slika 4.13: a) originalni RGB frejm b) originalni RGB frejm sa uklonjenom pozadinom, primenom transformacije u drugi prostor boja HSV i određenog praga za boju c) segmentovani klasovi na originalnom RGB frejmu d) segmentovani klasovi na originalnom frejmu sa uklonjenom pozadinom

1334,64, što ukazuje na adekvatnu reprodukciju ivica i finih detalja u rekonstruisanim frejmovima. Tipične vrednosti u intervalu 1000–2000 označavaju zadovoljavajući nivo prostorne definisanosti, dok niže vrednosti (<500) ukazuju na zamućenje slike, a više (>5000) na prisustvo šuma ili preterano izoštravanje. Vrednost SNR-a od $-2,63$ dB ukazuje na prisustvo izraženog šuma u intenzitetima piksela, što sugerise da je signalna komponenta slabije izražena u odnosu na varijabilnost osvetljenja ili rekonstrukuisani šum modela. Negativna SNR vrednost je uobičajena kod rendera koji trpe od blagog oscilovanja osvetljenja između frejmova ili nedovoljno stabilne radiometrijske kalibracije. Prosečna vrednost vremenske konzistentnosti od 5,22 ukazuje na umeren

Tabela 4.20: Kvantitativna evaluacija kvaliteta rekonstruisanog NeRF videa.

Metrika	Vrednost	Tumačenje
Laplacian variance	1334,64	Dobro očuvani detalji i jasno definisane ivice
SNR (dB)	-2,63	Prisustvo šuma i oscilacija u intenzitetu piksela
Frame difference	5,22	Umereno glatke promene između uzastopnih frejmova
PSNR (dB)	24,44	Umeren nivo sličnosti između susednih frejmova
SSIM	0,863	Visoka strukturna konzistentnost i stabilnost

nivo glatkoće u promenama između sukcesivnih kadrova. Niže vrednosti (1–3) ukazuju na visok stepen stabilnosti i odsustvo treperenja, dok vrednosti iznad 5 označavaju manje nestabilnosti u teksturi ili svetlosti, što se može manifestovati kroz blago treperenje površina u renderu. Vrednost PSNR-a od 24,44 dB ukazuje na umeren stepen sličnosti između uzastopnih frejmova. Vrednosti iznad 30 dB tipično označavaju visok kvalitet i stabilnost prikaza, dok vrednosti između 20 i 25 dB, kao u ovom slučaju, predstavljaju zadovoljavajući kompromis između detaljnosti i stabilnosti prikaza za realne NeRF rekonstrukcije. Indeks strukturne sličnosti SSIM od 0,8630 ukazuje na visok nivo koherentnosti strukture između frejmova. Vrednosti bliže jedinici označavaju očuvanje teksturalnih i strukturnih odnosa, dok vrednosti ispod 0,8 ukazuju na značajne promene u prostornim karakteristikama slike. Dobijeni rezultat sugerise da je model uspešno očuvao prostornu organizaciju scene, uz manja odstupanja u složenim i reflektujućim regionima.

U celini, kombinovani rezultati metrika potvrđuju da rekonstruisani video poseduje dobru prostornu oštrinu, visoku strukturnu stabilnost, uz umeren nivo vremenske konzistentnosti i prisustvo izvesnog šuma, što je karakteristično za NeRF rekonstrukcije izvedene na osnovu stvarnih snimaka u promenljivim svetlosnim uslovima.

Na ovaj način, dobijeni inicijalni rezultati služe kao pouzdana referentna tačka za poređenje sa unapređenim verzijama i omogućavaju

kvantitativnu analizu realnog doprinosa različitih modifikacija u kontekstu složenih scena kao što je pšenica i problem preklapanja klasova. Renderovani snimci se mogu pogledati na datom linku ¹.

Buduća istraživanja prirodno se mogu usmeriti ka integraciji metodologija NeRF sa novim paradigmama generalizovanog segmentacionog modelovanja, pre svega kroz okvir SAM-3D (*Segment Anything Model* u 3D domenu [319]). Dok vanilla NeRF i njegove varijante omogućavaju visoko-fotorealističnu rekonstrukciju složenih scena, njihov glavni izazov ostaje nedovoljna sposobnost direktne semantičke interpretacije rekonstruisanih volumena. Suprotno tome, SAM-3D modeli obećavaju mogućnost univerzalne i automatizovane segmentacije objekata u trodimenzionalnom prostoru, bez potrebe za ručno pripremljenim anotacijama.

Integracija NeRF-a i SAM-3D mogla bi da omogući višeslojnu analizu biljnih struktura u preciznoj poljoprivredi — od vizuelno verne rekonstrukcije do automatskog razdvajanja klasova pšenice, listova ili plodova u sklopu iste scene. Ovakav pristup bi značajno unapredio tačnost procene prinosa, omogućio detaljnu fenotipizaciju i otvorio put ka stvaranju *digital twins* biljnih kultura i okruženja u realnom vremenu. Time se postiže objedinjavanje vizuelne i semantičke dimenzije 3D rekonstrukcije, što predstavlja ključni korak u razvoju naredne generacije sistema za segmentaciju i detekciju željenih objekata u prostoru.

¹https://drive.google.com/drive/folders/1vN0kzfJxr3uXZrbaUXjI_BzHI0Kd3gMq?usp=sharing

5 Diskusija

U okviru predstavljene studije o segmentaciji kruničnih listova paradajza, upoređena je performansa Mask R-CNN modela sa različitim backbone arhitekturama (C4, DC5 i FPN) i različitim tehnikama za izdvajanje latentnih hiperspektralnih informacija. Mask R-CNN model sa FPN backbone-om, integrisan sa NMF tehnikom za izdvajanje latentnih spektralnih informacija za transformaciju hiperspektralnih podataka, nazvan NMFMaskFPN, postigao je najviši mAP i pokazao superiornu stabilnost, čime se ističe kao pouzdan izbor za segmentaciju instanci kruničnih listova paradajza.

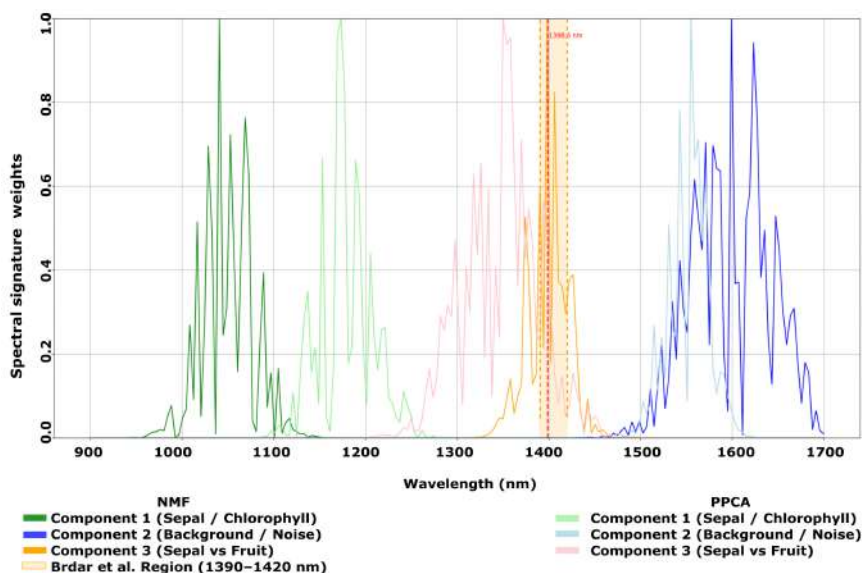
Dodatno, varijacije u precenjenoj i podcenjenoj predikciji kruničnih listova za različite modele i tehnike za izdvajanje latentnih hiperspektralnih informacija sugerišu da oba izbora značajno utiču na tačnost rezultata, u zavisnosti od specifične strukture paradajza. Ovo naglašava značaj adaptacije prethodno naučenih modela na nove skupove HSI podataka, posebno kada uslovi pri akviziciji slika variraju. Dodatna podešavanja odabranog NMFMaskFPN modela na ciljanom HSI skupu podataka optimizovalo je njegove performanse i poboljšalo generalizaciju, demonstrirajući robusnost predloženog pristupa kroz različite sorte paradajza. Uvođenje dodatnog prilagođenog sloja kroz ConvMaskFPN model i inicijalizacija sa NMF koeficijentima rezultovala je skromnim, ali uočljivim poboljšanjem ukupne preciznosti od 1%, dostižući najviši mAP od 87,4%, čime se omogućavaju pouzdanije predikcije.

Razlike u prostim morfometrijskim karakteristikama kruničnih listova među sortama paradajza ukazuju na složenost problema precizne segmentacije kruničnih listova. Procenat piksela na ivicama kruničnih listova direktno utiče na srednju vrednost i standardnu devijaciju hiperspektralnog odziva po kruničnom listu, a nepravilna segmentacija ovih piksela može drastično promeniti ovaj odziv i potencijalno dovesti do pogrešne klasifikacije zaraženih kruničnih listova [1]. Segmentacione maske generisane u ovoj doktorskoj disertaciji imaju za cilj olakšavanje

budućih istraživanja, osiguravajući da je svaki piksel precizno segmentiran i čuvajući kritične strukturne detalje kruničnih listova. Ovo je posebno važno za vrhove kruničnih listova, koji su izuzetno tanki i podložni greškama segmentacije, ali zadržavaju ključne informacije [320]. Čak i mala odstupanja u tačnosti u ovim regionima mogu dovesti do gubitka važnih morfoloških informacija, potencijalno utičući na robustnost procene osetljivosti.

U ovoj analizi procenjeni su i upoređeni dominantni doprinosi predstavljeni kroz tri glavne komponente dobijene NMF tehnikom, koja je identifikovana kao najbolje rangirana metoda za izdvajanje latentnih spektralnih komponenti hiperspektralne informacije. Ove komponente su izračunate iz sirovih hiperspektralnih podataka (prvi skup podataka opisan u 3.1, u 3.1) kako bi se identifikovale najznačajnije spektralne karakteristike i uporedile sa rezultatima iz [1]. Komponente su prikazane preko celokupnog spektralnog opsega 900–1700 nm, pri čemu je svaka kriva normalizovana radi lakše uporedbe (Slika 5.1). Jasno se uočava maksimum u komponenti 3, dobijen NMF-om, oko 1398,6 nm, u skladu sa opsegom identifikovanim u [1] kao kritičnim za razlikovanje sadržaja vode u tkivima paradajza. Ova korelacija sugerise da ova *data-driven* NMF spektralna reprezentacija podržava prethodna zapažanja i dodatno potvrđuje korisnost ovog spektralnog prozora za razlikovanje kruničnih listova i plodova. Komponenta 1 (NMF) pokazuje izražen maksimum na oko 1050 nm, često povezan sa procenom sadržaja hlorofila [321, 322], dok komponenta 2 (NMF) pokazuje maksimum na 1592 nm, obično pripisan šumu ili pozadinskim signalima. Nasuprot tome, PPCA komponente su blago pomerene: komponenta 1 sa maksimumom na 1180 nm sugerise širi spektralni odgovor kruničnih listova ili preklapanje biohemijskih signala; komponenta 2 sa maksimumom na 1550 nm povezana je sa sadržajem vode i rasipanjem tkiva; dok komponenta 3 sa maksimumom 1350 nm reflektuje strukturne razlike na biljci. Ovi rezultati podržavaju hipotezu da je region 1398,6 nm posebno dominantan i važan za probleme segmentacije regiona od interesa u HSI, naročito za razlikovanje plodova bogatih vlagom od relativno suvih kruničnih listova, u skladu sa [1].

Za evaluaciju preciznosti segmentacije najboljeg ConvMaskFPN modela (inicijalizovanog NMF koeficijentima) izračunate su srednje vrednosti i standardne devijacije za svaki talasni opseg originalnog 196-dimenzionalnog hiperspektralnog odziva po kruničnom listu paradajza, koristeći piksele označene referentno anotiranom maskom i



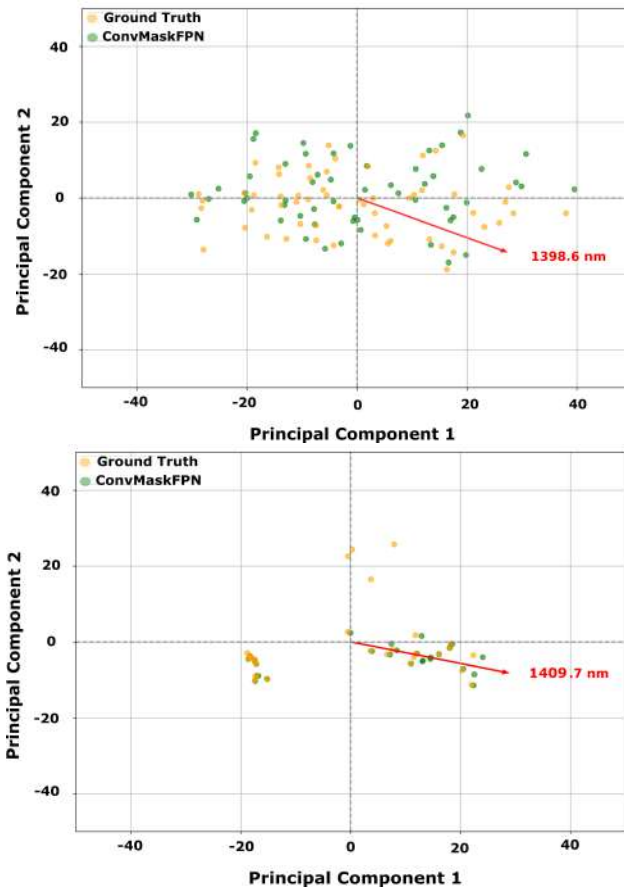
Slika 5.1: Normalizovani koeficijenti dobijeni NMF i PPCA metodama iz prvog HSI skupa sa tri glavne komponente. Dominantni maksimum u komponenti 3, dobijen NMF-om, primećen je na 1398,6 nm, preklapajući se sa regionom 1390–1420 nm (označeno), koji je prethodno [1] identifikovan kao značajan za diferencijaciju vezanu za količinu i posledično manjak vode u kruničnim listovima

predikcijom modela na test podacima. Koeficijenti Pearsonove korelacije između vektorizovanih vrednosti karakteristika iz referentnih vrednosti i predikcionih maski su: za prvi skup HSI podataka $r = 0,8100$, $p = 1,1940 \times 10^{-9}$ i za drugi HSI skup podataka $r = 0,7523$, $p = 6,2944 \times 10^{-9}$, što potvrđuje visok nivo ukupne preciznosti segmentacije kruničnih listova. Precizna segmentacija na nivou piksela ostaje izazov, naročito za ključne regione poput vrhova kruničnih listova.

Daljom PCA analizom vektora karakteristika predikovanih i referentnih maski, prikazan je doprinos svakog talasnog opsega u kreiranju glavnih komponenti (Slika 5.2).

Najznačajnije karakteristike su slične između skupova podataka, što ukazuje na ključnu ulogu ovih spektralnih regiona u odluci modela i predikciji osetljivosti ploda paradajza i kruničnih listova na buduće gljivične infekcije.

U budućem radu planira se detaljnija analiza uzroka nesigurnosti u detekciji kompleksnih struktura kruničnih listova, sa ciljem poboljšanja



Slika 5.2: Dominantne talasne dužine računatih korišćenjem ground-truth maski i maski dobijenih od ConvMaskFPN modela na test skupovima (a) prvog i (b) drugog HSI skupa podataka

u treniranju modela ili strategiji u efikasnijoj predobradi HSI podataka. Iako su predložene tehnike za izdvajanje latentnih hiperspektralnih informacija računarski efikasne i jednostavne za implementaciju [323], one imaju ograničenja. Napredne, nadgledane ili nenadgledane tehnike, kao što su autoenkoderi [324, 325] i varijacioni autoenkoderi (VAE) [326, 327], mogu poboljšati korak efikasnijeg izdvajanja latentnih spektralnih informacija za hiperspektralne slike kroz učenje složenijih, nelinearnih reprezentacija ovako kompleksnih podataka. Takođe, planira se istraživanje nenadgledanih pristupa segmentacije, koji obećavaju u domenima gde je generisanje visokokvalitetnih labela izazovno, kao što je hiperspektralno snimanje sa specifičnim klasama.

Osim rezultata u studiji slučaja vezanoj za paradajz, počev od inicijalnog pristupa semantičke segmentacije, važno je razumeti razloge zbog kojih se dešavaju greške prilikom detekcije klasova pšenice. U fazi detekcije klasova postoji nekoliko izazova: zaklanjanje (engl. *occlusion*) i preklapanje (engl. *overlap*) predstavljaju ključne probleme koji izazivaju greške u segmentaciji. Različite veličine klasova unutar slike, usled perspektive snimanja, kao i različite rezolucije termalnih i RGB slika, predstavljaju dodatne izazove. Ugao akvizicije slike, osvetljenje na polju, vreme snimanja i drugi uslovi na eksperimentalnom polju takođe značajno utiču na konačne rezultate [11, 12, 13].

Objavljene studije u oblasti detekcije klasova pšenice uglavnom se oslanjaju na ograničene skupove podataka vezane za specifične geografske lokacije, što predstavlja izazov prilikom prenosa razvijenih i obučениh modela na druge regione. Prema našem saznanju, nijedan drugi skup podataka sa označenim klasovima pšenice nije sistematski prikupljen pomoću pametnog mobilnog telefona. Uključivanje podskupa BioS-Wheat u GWHD uvodi značajno povećanje složenosti podataka, naglašavajući uticaj agronomske raznolikosti na performanse modela. Visoka gustina setve i minimalni razmak između redova dovode do povećane okluzije objekata i gustog prostornog rasporeda, što predstavlja izazov za standardne detektore objekata, posebno one sa fiksnim veličinama (engl. *anchor box*) ili ograničenim kapacitetom za detekciju nepravilnih oblika. Ipak, treniranje na ovakvim podacima ima prednost jer izlaže modele stvarnoj varijabilnosti uslova, čime se podstiče robusnost, naročito u scenarijima primene na terenu gde su uslovi manje kontrolisani, a usevi gušći.

Ova doktorska disertacija obuhvata korake prikupljanja i harmonizacije podataka, validirane metodama za procenu kvaliteta slika. Takođe, primenjene metode dubokog učenja omogućile su sagledavanje uticaja dodavanja različitih geografskih lokacija na prepoznavanje klasova pšenice u različitim prostornim uslovima.

Metode dubokog učenja koje su ovde upoređene obuhvataju Faster R-CNN i RetinaNet, prepoznate u literaturi kao one koje postižu najvišu preciznost [59, 227], zatim YOLOv8 i RT-DETR. Superiorne performanse RT-DETR i YOLOv8 modela mogu se u velikoj meri objasniti njihovom sposobnošću da efikasno modeluju gusto raspoređene i preklapljene objekte od interesa, što je ilustrativno u BioS-Wheat skupu podataka. U slučaju RT-DETR, arhitektura zasnovana na transformerima koristi globalnu *self-attention* mehaniku, što omogućava modelu

da uhvati kompleksnije zavisnosti i kontekstualne relacije unutar slike. Ovo olakšava preciznije razlikovanje pojedinačnih klasova pšenice, čak i kada su preklapljeni u velikoj meri ili delimično međusobno zaklonjeni, kroz učenje relacijskih indikatora koji prevazilaze lokalna receptivna polja detektora. Takvo globalno rezonovanje omogućava RT-DETR modelu da preciznije mapira pojedinačne instance klasa pšenice i smanjuje greške detekcije ili potiskivanje susednih objekata. S druge strane, YOLOv8 postiže robusnost kroz *anchor-free* detekcioni pristup i optimizovanu *multi-scale backbone* mrežu. Direktnim predviđanjem centara objekata, umesto oslanjanja na fiksne anchor box-eve, YOLOv8 prevazilazi ograničenja usklađivanja anchora u regionima sa gustom raspodelom klasova pšenice. Ovaj dizajn je naročito koristan u parcelama pšenice, gde se klasovi pojavljuju u različitim veličinama i često se preklapaju. Poboljšana ekstrakcija multi-scale karakteristika dodatno unapređuje detekciju manjih i delimično vidljivih klasova, što rezultira stabilnijom lokalizacijom i smanjenjem grešaka preklapanja. Za poređenje, tradicionalni detektori koji su takođe zasnovani na anchorima, kao što su Faster R-CNN i RetinaNet, pokazuju degradaciju performansi u ovim uslovima, jer guste raspodele klasova pšenice dovode do redundantnosti anchora, nepravilno poravnatih predloga predikcije pozicije i prekomerne non-maximum supresije. Sveukupno, ovi rezultati pokazuju kako razlike u arhitekturama RT-DETR i YOLOv8 modela omogućavaju superiornu adaptaciju na kompleksne slike visoke gustine u Bios-Wheat skupu podataka, koji oslikava realno okruženje u poljoprivredi, potvrđujući njihovu robusnost u odnosu na starije detekcione modele. Dobijeni rezultati potvrđuju prethodne nalaze da modeli dubokog učenja i pristupi transfera znanja za detekciju objekata mogu uspešno da se primene na skupove podataka sa slikama iz različitih zemalja, sa različitim gustinama klasova i u različitim uslovima gajenja. Iako oblik i zrna klasova pšenice imaju tendenciju stabilnosti tokom faze rasta, kontrast između klasova, pozadine i listova u slikama postaje izraženiji, što dodatno otežava njihovu detekciju [11, 59, 328].

Da bi se odgovorilo na uobičajen izazov prisutan kod svih automatskih i poluautomatskih metoda označavanja klasova pšenice koji se može javiti i pri ručnoj anotaciji, u budućem radu razmotriće se alternativne strategije za razdvajanje instanci, kao što su metode zasnovane na *Watershed* algoritmu, učljive metode za separaciju instanci ili direktni pristupi detekciji objekata koristeći modele poput najnovijih verzija YOLO, čime se otvaraju potencijalni pravci za unapređenje

robusnosti u gustim i preklapljenim scenarijima.

Integracija BioS-Wheat skupa podataka efikasno funkcioniše kao „stres test“ za generalizaciju modela. Dok su klasični modeli zasnovani na anchorima, poput F-RCNN sa ResNet-101 backbone-om, pokazali značajan pad performansi pod ovim uslovima, savremene arhitekture kao što su RT-DETR i YOLOv8 ostale su stabilne ili su ostvarile poboljšanje. Ova razlika proističe iz činjenice da BioS-Wheat uvodi veću gustinu klasova pšenice i veću verovatnoću preklapanja, što izaziva ograničenja kod modela koji zavise od PRN arhitekture i prethodno definisanih anchora. Suprotno tome, arhitekture sa globalnim mehanizmima pažnje ili *anchor-free*, multi-scale detekcionim sposobnostima bolje se adaptiraju na ovako složene scenarije, ističući poboljšanu robusnost savremenih detektora objekata u gustim i zaklonjenim okruženjima.

Analizom sorti u okviru BioS-Wheat skupa podataka pokazalo se da *Futura* i *Pobeda* imaju jače slaganje i konzistentnost između detektovanog i stvarnog broja klasova, što ukazuje na višu tačnost detekcije kod ovih sorti. Sorte *Ingenio* i *NS40S* pokazuju veću varijabilnost, što može značiti nižu preciznost detekcije. Ova analiza sugerise da model najbolje funkcioniše za sorte *Futura* i *Pobeda*, dok su za ostale sorte potrebna dodatna poboljšanja. U okviru ove studije analizirani su različiti modeli kako bi se identifikovali oni koji postižu dovoljnu robusnost u performansama na podacima različitog geografskog porekla i u različitim fazama razvoja useva, čime se izbegava ograničena primenljivost samo na slike prikupljene u određenoj fenološkoj fazi. To znači da potencijalni korisnici neće biti ograničeni protokolima za vreme snimanja slika, budući da takvi protokoli trenutno ne postoje.

Postizanje dovoljne robusnosti na različitim sortama, fazama rasta i lokacijama može predstavljati dobru osnovu za uspostavljanje procedure za *crowdsourcing* slika, što bi dodatno olakšalo proširenje skupova podataka i unapređenje modela za detekciju i prebrojavanje klasova pšenice. Iako je RetinaNet postigla nešto nižu vremensku zahtevnost u radu sa malim brojem klasa, Faster R-CNN je pokazao bolje ukupne rezultate među osnovnim DNN modelima, dok je RT-DETR ostvario najbolju prosečnu preciznost, ali i najveću složenost, zahtevnost u resursima i najduže trajanje treniranja.

Dobijene metrike ukazuju da rekonstruisani NeRF video pokazuje dobar balans između prostorne oštine i strukturne konzistentnosti, uz umerenu vremensku stabilnost. Vrednost varijanse Laplasijana potvr-

đuje očuvanje detalja i ivica, što ukazuje da model uspešno rekonstruiše fine teksturalne strukture scene. Negativna vrednost SNR-a, međutim, sugeriše prisustvo šuma i varijacija u intenzitetima piksela, što može biti posledica neujednačenog osvetljenja u ulaznim snimcima ili nedovoljno stabilnog renderovanja tokom treniranja. Umerene vrednosti PSNR-a (24,44 dB) i SSIM-a (0,863) pokazuju da model generalno održava sličnost između uzastopnih frejmova, ali se i dalje javljaju lokalne nestabilnosti u regionima složene geometrije i refleksije. Za unapređenje kvaliteta rekonstrukcije mogu se primeniti različite strategije. Pre svega, poboljšanje tačnosti i konzistentnosti segmentacionih maski i preciznije ručne anotacije doprinele bi boljem razdvajanju klasove pšernice i pozadine, čime bi se smanjio šum u 3D reprezentaciji. Dodatno, korišćenje varijanti NeRF modela sa eksplicitnim vremenskim regularizacijama (npr. D-NeRF, T-NeRF ili RegNeRF) moglo bi značajno da poveća vremensku konzistentnost i smanji treperenje površina. Uvođenje fotometrijskih i geometrijskih regularizacionih termina tokom treniranja takođe bi stabilizovalo osvetljenje i refleksije između frejmova. Pored toga, povećanje raznovrsnosti i gustine ulaznih uglova kamere, kao i korišćenje posebnih tehnika za *denoising* ili normalizaciju i ujednačavanje osvetljaja, može dodatno unaprediti kvalitet rekonstrukcije i povećati SNR. Sveukupno, iako trenutni rezultati ukazuju na zadovoljavajući nivo strukturne i prostorne vernosti, dalja optimizacija kroz poboljšane maske, stabilizovane varijante NeRF arhitekture i bolju radiometrijsku kontrolu ulaznih podataka može doprineti postizanju veće vremenske stabilnosti i realističnosti rekonstruisanih scena.

6 Zaključak

Ovo istraživanje demonstrira efikasnost segmentacije kruničnih listova paradajza na nivou instance unutar hiperspektralnih slika kroz integraciju dobro utemeljenih tehnika, uključujući PCA, PPCA, ICA i NMF, za izdvajanje latentnih spektralnih komponenti hiperspektralnih informacija, u kombinaciji sa unapred treniranim modelima dubokog učenja. *Transfer learning* se pokazao kao značajan pristup, budući da su modeli unapred trenirani na COCO skupu podataka i dodatno prilagođeni hiperspektralnim slikama sa izdvojenim spektralnim informacijama ostvarili izuzetne performanse.

Među testiranim modelima, Mask R-CNN sa FPN arhitekturom i NMF tehnikom za izdvajanje latentnih spektralnih komponenti hiperspektralnih informacija (NMFMaskFPN) pokazao je najstabilnije i najpouzdanije performanse, obezbeđujući precizne segmentacije sa niskom standardnom devijacijom mAP vrednosti. Dodatno, inicijalizacija težina dodatnog sloja unapred treniranim koeficijentima, dobijenim izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama tehnikom NMF, dovela je do blagog poboljšanja ukupne preciznosti. Ovi rezultati predstavljaju značajan korak ka razvoju *end-to-end* okvira za izdvajanje latentnih spektralnih komponenti u hiperspektralnim slikama i dalju segmentaciju relevantnih regija unutar istih.

Varijacije u procenjenoj i podcenjenoj predikciji kruničnih listova među različitim modelima i tehnikama za izdvajanje latentnih hiperspektralnih informacija dodatno naglašavaju potrebu za pažljivim izborom i optimizacijom modela za specifične strukture paradajza. Rezultati ove studije ukazuju na mogućnosti naprednih praktičnih primena hiperspektralnog snimanja u domenima upravljanja lancem snabdevanja nakon berbe, uključujući procenu kvaliteta, detekciju defekata i praćenje posledica skladištenja osetljivih plodova.

Buduća istraživanja mogla bi detaljnije da se bave optimizacijom parametara *transfer learninga* i da istraže naprednije metode za izdva-

janje latentnih hiperspektralnih informacija, prilagođene jedinstvenim karakteristikama hiperspektralnih podataka. Pored toga, takva istraživanja mogla bi da se fokusiraju na dalje prilagođavanje i optimizaciju predloženog modela za specifične sorte paradajza ili na prevazilaženje izazova koje nameću ekološke promene i uslovi na terenu. Predstavljeno interdisciplinarno istraživanje ne samo da pruža značajne uvide u trenutni status primene hiperspektralnog snimanja za procenu kvaliteta nakon berbe, već sugeriše i potencijalne pravce budućih istraživanja i praktičnih primena u poljoprivrednoj praksi.

Osim hiperspektralnih slika, ovo istraživanje je pokazalo da je moguće koristiti različite modalitete slika, termalne i RGB, za detekciju klasova pšenice. Predloženi pristup ne zahteva montažu kamera na vozilima niti skupe i složene postavke. Sistem za akviziciju slika zahteva ručnu kameru i stativ, što omogućava jednostavno kretanje kroz parcele na polju ili pametan mobilni telefon.

Kompletan sistem je značajan za fenotipizaciju biljaka, imajući u vidu da je broj klasova značajan za procenu prinosa pšenice. Cilj ovakvog istraživanja u parksi predstavlja rešenje koje omogućava poljoprivrednicima da brzo i jednostavno dobiju informaciju o broju klasova, koji bi bio ulazni podatak za što precizniju procenu prinosa pšenice. Ovo je važno ne samo za planiranje buduće žetve, već i za praćenje biljaka u određenoj fazi rasta i prevenciju smanjenih prinosa ili širenja bolesti.

U nastavku istraživanja, nakon detaljne analize postignutih rezultata u okviru semantičke segmentacije, istraživački fokus je proširen ka metodi segmentacije instanci. Ovaj pristup omogućio je preciznije izdvajanje i razlikovanje pojedinačnih klasova pšenice unutar slike, čime se prevazilaze ograničenja klasične semantičke segmentacije koja sve objekte iste klase tretira kao jedinstvenu celinu.

Prednost ove metode segmentacije ogleda se i u mogućnosti primene transfera prethodno obučениh modela na velikim, javno dostupnim skupovima podataka (engl. *transfer learning*), što značajno ubrzava proces treniranja modela i povećava preciznost, naročito u uslovima kada je raspoloživa količina anotiranih podataka ograničena. Dodatno, u okviru ove istraživačke teme, proširen je multimodalni skup podataka, iako se svi kanali odnose na RGB prikaz slika, oni potiču sa različitih senzora, čime se obezbeđuje raznovrsnost u pogledu rezolucije, osvetljenja i perspektive snimanja. Ovakva heterogenost omogućava bolju generalizaciju modela i povećava robustnost segmentacionih rezultata

u realnim uslovima primene.

Na taj način, prelazak sa semantičke na segmentaciju instanci ne predstavlja samo tehnički iskorak, već i strateški pravac razvoja istraživanja, sa ciljem da se postignu precizniji, skalabilniji i praktično primenljiviji rezultati na terenu koristeći pristupačnije kamere koje su integrisane u mobilne pametne telefone.

Uključivanje podskupa *BioS-Wheat* u skup podataka *GWDH* značajno unapređuje njegovu reprezentativnost uvođenjem slika visoke gustine sa izraženim preklapanjima. Iako ovakva složenost može dovesti do blagog pada performansi modela, ona u krajnjem ishodu podržava razvoj robusnijih i prilagodljivijih detektora klasova pšenice. Dobijeni rezultati naglašavaju značaj uključivanja agronomski raznovrsnih i vizuelno zahtevnih podataka u proces obuke modela, kako bi se obezbedila pouzdanost performansi u širem spektru uslova na terenu.

Pored toga, ovo istraživanje pokazuje da upotreba pristupačnih senzora i tehnologija koje su „dovoljno dobre“ može doneti brojne benefite. Takav pristup olakšava prikupljanje podataka, što direktno doprinosi unapređenju algoritama zasnovanih na podacima i njihovim performansama, ali i omogućava proširenje baza podataka među različitim zemljama, genotipovima i vrstama.

Takođe, može se zaključiti da NeRF metoda za rekonstrukciju scene pokazuje visok stepen strukturne i prostorne vernosti, uz umerenu vremensku stabilnost i prisustvo blagog šuma u intenzitetima piksela. Dobijene vrednosti metrika ukazuju na to da model uspešno očuva ključne detalje i prostorne odnose u sceni, što potvrđuje konzistentnost 3D reprezentacije kroz različite perspektive kamere. Međutim, uočene oscilacije u intenzitetu piksela i manja variranja između uzastopnih frejmova sugerišu potrebu za dodatnim optimizacijama u fazi treniranja modela, posebno u pogledu regularizacije i stabilizacije renderovanja u vremenskom domenu. Ukupno posmatrano, ostvareni rezultati ukazuju na zadovoljavajući nivo vizuelne stabilnosti i realističnosti NeRF rekonstrukcije, što potvrđuje potencijal primenjene metode za analizu i 3D rekonstrukciju složene i kompleksne scene u poljoprivredi.

Poljoprivredna industrija je transformisana pojavom pristupačnih tehnologija, posebno pametnih telefona, koji su poljoprivrednicima omogućili nove načine efikasnijeg upravljanja proizvodnjom. Pametni telefoni mogu se koristiti za praćenje stanja zemljišta i useva, pristup vremenskim i meteorološkim informacijama u realnom vremenu, daljinsko upravljanje operacijama, primenu precizne poljoprivrede i efikasno

prikupljanje podataka. Ovi tehnološki napredci doprineli su povećanju produktivnosti i efikasnosti u poljoprivredi, osnažujući poljoprivrednike da donose informisanije odluke i povećaju svoje prinose.

Izvor finansiranja

Ovo istraživanje je podržano od strane programa istraživanja i inovacija Evropske unije Horizon 2020 kroz projekat *ANTARES* (SGA-CSA. br. 739570 u okviru FPA br. 664387, <https://doi.org/10.3030/739570>), Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije (Grant br. 451-03-136/2025-03/200358), kao i Pokrajinskog sekretarijata za visoko obrazovanje i naučnoistraživačku delatnost Autonomne Pokrajine Vojvodine kroz projekat „Razvoj sistema za podršku odlučivanju u poljoprivrednoj proizvodnji primenom fuzije podataka i veštačke inteligencije“ (Grant br. 003074633 2024 09418 003 000 000 001 04 002). Osim toga deo ovog istraživanja podržan je kroz dva programa Fonda za Inovacionu Delatnost Republike Srbije, Dokaz Koncepta i Transfer Tehnologije pod nazivom "Alat za ranu predikciju prinosa baziran na veštačkoj inteligenciji".

Bibliografija

- [1] Brdar, S., Panić, M., Hogeveen-van Echtelt, E., Mensink, M., Grbović, Ž., Woltering, E., Chauhan, A., “Predicting sensitivity of recently harvested tomatoes and tomato sepals to future fungal infections”, *Scientific Reports*, Vol. 11, No. 1, 2021, str. 23109.
- [2] David, E., Madec, S., Sadeghi-Tehran, P., Aasen, H., Zheng, B., Liu, S., Kirchgessner, N., Ishikawa, G., Nagasawa, K., Badhon, M. A. *et al.*, “Global wheat head detection (gwhd) dataset: a large and diverse dataset of high-resolution rgb-labelled images to develop and benchmark wheat head detection methods”, *Plant Phenomics*, Vol. 2020, 2020.
- [3] Grbović, Ž., Panić, M., Stefanović, D., Brdar, S., Crnojević, V., Lončar-Tutrukalo, T., “Transfer learning models for wheat ear detection on multi-source dataset”, *Scientific Reports*, Vol. X, No. X, 2025, str. XX.
- [4] Zavod za intelektualnu svojinu Republike Srbije, “Glasnik zavoda za intelektualnu svojinu, broj 04/2022”, <https://www.zis.gov.rs/wp-content/uploads/Glasnik-04-2022.pdf>, pristupljeno: 2025-12-19. 2022.
- [5] Gonzalez, R. C., *Digital image processing*. Pearson education india, 2009.
- [6] Szeliski, R., *Computer Vision: Algorithms and Applications*. London: Springer, 2010.
- [7] Cooper, M. C., “The tractability of segmentation and scene analysis”, *International Journal of Computer Vision*, Vol. 30, No. 1, 1998, str. 27–42.

- [8] Duda, R. O., Hart, P. E., “Pattern classification and scene analysis”, A Wiley-interscience publication, 1973.
- [9] O’Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., Riordan, D., Walsh, J., “Deep learning vs. traditional computer vision”, in Science and information conference. Springer, 2019, str. 128–144.
- [10] Zhou, C., Liang, D., Yang, X., Yang, H., Yue, J., Yang, G., “Wheat ears counting in field conditions based on multi-feature optimization and twsvm”, *Frontiers in Plant Science*, Vol. 9, 2018, str. 1024.
- [11] Hasan, M. M., Heidler, J., et al., “Detection and analysis of wheat spikes using convolutional neural networks”, *Plant Methods*, Vol. 14, No. 1, 2018, str. 100.
- [12] Pound, M., Atkinson, J., Wells, D., Pridmore, T., French, A., “Deep learning for multi-task plant phenotyping”, in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, str. 2055–2063.
- [13] Alharbi, N., Zhou, J., Wang, W., “Automatic counting of wheat spikes from wheat growth images”, in *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*. Venice, Italy: IEEE, 2018, str. 346–355.
- [14] Hastie, T., Tibshirani, R., Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, 2017.
- [15] Padilla, R., Passos, W. L., Dias, T. L., Netto, S. L., Da Silva, E. A., “A comparative analysis of object detection metrics with a companion open-source toolkit”, *Electronics*, Vol. 10, No. 3, 2021, str. 279.
- [16] Dai, J., Li, Y., He, K., Sun, J., “R-fcn: Object detection via region-based fully convolutional networks”, *Advances in neural information processing systems*, Vol. 29, 2016.

- [17] Banerjee, A., Chitnis, U., Jadhav, S., Bhawalkar, J., Chaudhury, S., “Hypothesis testing, type i and type ii errors”, *Industrial psychiatry journal*, Vol. 18, No. 2, 2009, str. 127.
- [18] Benjamini, Y., “Discovering the false discovery rate”, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Vol. 72, No. 4, 2010, str. 405–416.
- [19] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale”, *arXiv preprint arXiv:2010.11929*, 2020.
- [20] Fernandez-Gallego, J. *et al.*, “Wheat ear counting in-field conditions: high throughput and low cost approach using rgb images”, *Plant Methods*, 2018.
- [21] Fernandez-Gallego, J. A., Buchaillot, M. L., Aparicio Gutiérrez, N., Nieto-Taladriz, M. T., Araus, J. L., Kefauver, S. C., “Automatic wheat ear counting using thermal imagery”, *Remote Sensing*, Vol. 11, No. 7, 2019, str. 751.
- [22] Rahman, M. A., Wang, Y., “Optimizing intersection-over-union in deep neural networks for image segmentation”, in *International Symposium on Visual Computing*. Springer, 2016, str. 234–244.
- [23] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Saverese, S., “Generalized intersection over union: A metric and a loss for bounding box regression”, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, str. 658–666.
- [24] van Beers, F., Lindström, A., Okafor, E., Wiering, M. A., “Deep neural networks with intersection over union loss for binary image segmentation”, in *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, 2019, str. 438–445.
- [25] Krizhevsky, A., Sutskever, I., Hinton, G. E., “Imagenet classification with deep convolutional neural networks”, in *Advances in neural information processing systems*, 2012.

- [26] Simonyan, K., Zisserman, A., “Very deep convolutional networks for large-scale image recognition”, arXiv preprint arXiv:1409.1556, 2014.
- [27] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., “Going deeper with convolutions”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, str. 1–9.
- [28] He, K., Zhang, X., Ren, S., Sun, J., “Deep residual learning for image recognition”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, str. 770–778.
- [29] Ronneberger, O., Fischer, P., Brox, T., “U-net: Convolutional networks for biomedical image segmentation”, in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, str. 234–241.
- [30] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., “Densely connected convolutional networks”, Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, str. 4700–4708.
- [31] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., “Mobilenets: Efficient convolutional neural networks for mobile vision applications”, in arXiv preprint arXiv:1704.04861, 2017.
- [32] Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jégou, H., “Training data-efficient image transformers & distillation through attention”, in International conference on machine learning. PMLR, 2021, str. 10 347–10 357.
- [33] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., “Swin transformer: Hierarchical vision transformer using shifted windows”, Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, str. 10 012–10 022.
- [34] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., Luo, P., “Segformer: Simple and efficient design for semantic segmentation with transformers”, Advances in Neural Information Processing Systems, Vol. 34, 2021, str. 12 077–12 090.

- [35] Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S., “A convnet for the 2020s”, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, str. 11 976–11 986.
- [36] Cointault, F., Guerin, D., Guillemin, J.-P., Chopinet, B., “In-field triticum aestivum ear counting using colour-texture image analysis”, *New Zealand Journal of Crop and Horticultural Science*, Vol. 36, No. 2, 2008, str. 117–130.
- [37] Bi, K., Jiang, P., Li, L., Shi, B., Wang, C., “Non-destructive measurement of wheat spike characteristics based on morphological image processing”, *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 26, No. 12, 2010, str. 212–216.
- [38] Kun, B., Pan, J., Chongwei, T., Feifei, H., Cheng, W., “The design of wheat variety bp classifier based on wheat ear feature”, *Chinese Agricultural Science Bulletin*, Vol. 6, 2011, str. 465–467.
- [39] Hecht-Nielsen, R., “Theory of the backpropagation neural network”, in *Neural networks for perception*. Elsevier, 1992, str. 65–93.
- [40] Woods, J. W., “Chapter 7 - image enhancement and analysis”, in *Multidimensional Signal, Image, and Video Processing and Coding (2nd Edition)*, 2nd ed., Woods, J. W., (ur.). Boston: Academic Press, 2012, str. 223-256, dostupno na: <https://www.sciencedirect.com/science/ARTICLE/pii/B9780123814203000072>
- [41] Brownrigg, D. R., “The weighted median filter”, *Communications of the ACM*, Vol. 27, No. 8, 1984, str. 807–818.
- [42] Thorsten Wagner, J.-Y. T., “Find maxima detection algorithm, image j”, [https://imagej.net/imagej-wiki-static/Find_maxima_\(Trackmate_module\)](https://imagej.net/imagej-wiki-static/Find_maxima_(Trackmate_module)), 2016.
- [43] Matz, G., Hlawatsch, F., “Linear time–frequency filters: On-line algorithms and applications”, in *Applications in Time-Frequency Signal Processing*. CRC Press, 2018, str. 205–272.
- [44] Shlens, J., “A tutorial on principal component analysis”, *arXiv preprint arXiv:1404.1100*, 2014.

- [45] Fix, E., Hodges, J. L., “Discriminatory analysis. nonparametric discrimination: Consistency properties”, *International Statistical Review/Revue Internationale de Statistique*, Vol. 57, No. 3, 1989, str. 238–247.
- [46] Du, Y., Cai, Y., Tan, C., Li, Z., Yang, G., Feng, H., Dong, H., “Field wheat ears counting based on superpixel segmentation method”, *Scientia Agricultura Sinica*, Vol. 52, No. 1, 2019, str. 21–33.
- [47] Kim, K.-S., Zhang, D., Kang, M.-C., Ko, S.-J., “Improved simple linear iterative clustering superpixels”, in *2013 IEEE International Symposium on Consumer Electronics (ISCE)*, 2013, str. 259-260.
- [48] Huang, H., Wei, X., Zhou, Y., “Twin support vector machines: A survey”, *Neurocomputing*, Vol. 300, 2018, str. 34-43, dostupno na: <https://www.sciencedirect.com/science/ARTICLE/pii/S0925231218302923>
- [49] Zhou, C., Liang, D., Yang, X., Yang, H., Yue, J., Yang, G., “Wheat ears counting in field conditions based on multi-feature optimization and twsvm”, *Frontiers in Plant Science*, Vol. 9, 2018, str. 1024, dostupno na: <https://www.frontiersin.org/ARTICLE/10.3389/fpls.2018.01024>
- [50] Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. A., Asari, V. K., “A state-of-the-art survey on deep learning theory and architectures”, *Electronics*, Vol. 8, No. 3, 2019, str. 292.
- [51] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. *et al.*, “Imagenet large scale visual recognition challenge”, *International journal of computer vision*, Vol. 115, No. 3, 2015, str. 211–252.
- [52] Ren, S., He, K., Girshick, R., Sun, J., “Faster r-cnn: Towards real-time object detection with region proposal networks”, *Advances in neural information processing systems*, Vol. 28, 2015.
- [53] He, K., Gkioxari, G., Dollár, P., Girshick, R., “Mask r-cnn”, in *Proceedings of the IEEE international conference on computer vision*, 2017, str. 2961–2969.

- [54] Aich, S., Stavness, I., “Global sum pooling: A generalization trick for object counting with small datasets of large images”, arXiv preprint arXiv:1805.11123, 2018.
- [55] Xiong, H., Lu, H., Liu, C., Liu, L., Cao, Z., Shen, C., “From open set to closed set: Counting objects by spatial divide-and-conquer”, in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, str. 8362–8371.
- [56] Kamilaris, A., Prenafeta-Boldú, F. X., “Deep learning in agriculture: A survey”, Computers and Electronics in Agriculture, Vol. 147, 2018, str. 70 - 90, dostupno na: <http://www.sciencedirect.com/science/ARTICLE/pii/S0168169917308803>
- [57] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., “Image segmentation using deep learning: A survey”, arXiv preprint arXiv:2001.05566, 2020.
- [58] Barré, P., Stöver, B. C., Müller, K. F., Steinhage, V., “Leafnet: A computer vision system for automatic plant species identification”, Ecological Informatics, Vol. 40, 2017, str. 50 - 56, dostupno na: <http://www.sciencedirect.com/science/ARTICLE/pii/S1574954116302515>
- [59] Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., Heritier, E., Baret, F., “Ear density estimation from high resolution rgb imagery using deep learning technique”, Agricultural and forest meteorology, Vol. 264, 2019, str. 225–234.
- [60] Sadeghi-Tehran, P., Virlet, N., Ampe, E. M., Reyns, P., Hawkesford, M. J., “Deepcount: in-field automatic quantification of wheat spikes using simple linear iterative clustering and deep convolutional neural networks”, Frontiers in plant science, Vol. 10, 2019, str. 1176.
- [61] Gibbs, J. A., Burgess, A. J., Pound, M. P., Pridmore, T. P., Murchie, E. H., “Recovering wind-induced plant motion in dense field environments via deep learning and multiple object tracking”, Plant physiology, Vol. 181, No. 1, 2019, str. 28–42.

- [62] Fukushima, K., Miyake, S., Ito, T., “Neocognitron: A neural network model for a mechanism of visual pattern recognition”, *IEEE transactions on systems, man, and cybernetics*, No. 5, 1983, str. 826–834.
- [63] Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., Lang, K. J., “Phoneme recognition using time-delay neural networks”, *IEEE transactions on acoustics, speech, and signal processing*, Vol. 37, No. 3, 1989, str. 328–339.
- [64] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., “Gradient-based learning applied to document recognition”, *Proceedings of the IEEE*, Vol. 86, No. 11, 1998, str. 2278–2324.
- [65] Jin, X., Madec, S., Dutartre, D., de Solan, B., Comar, A., Barret, F., “High-throughput measurements of stem characteristics to estimate ear density and above-ground biomass”, *Plant Phenomics*, Vol. 2019, 2019.
- [66] Ubbens, J. R., Stavness, I., “Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks”, *Frontiers in Plant Science*, Vol. 8, 2017, str. 1190.
- [67] Bi, K., Jiang, P., Li, L., Shi, B., Wang, C., “Non-destructive measurement of wheat spike characteristics based on morphological image processing”, *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 26, No. 12, 2010, str. 212–216.
- [68] Bi, K., Jiang, P., Wei, T., Huang, F., Wang, C., “The design of wheat variety bp classifier based on wheat ear feature”, *Chinese Agricultural Science Bulletin*, Vol. 26, No. 6, 2011, str. 464–468.
- [69] Pound, M. P., Atkinson, J. A., Wells, D. M., Pridmore, T. P., French, A. P., “Deep learning for multi-task plant phenotyping”, in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, str. 2055–2063.
- [70] Aghdam, H. H., Heravi, E. J., *Guide to Convolutional Neural Networks*. Springer International Publishing AG, 2017.
- [71] Clevert, D., Unithiner, T., Hochreiter, S., “Fast and accurate deep network learning by exponential linear units (elus)”, in *International Conference on Learning Representations (ICLR)*, 2016.

- [72] Heaton, J., Artificial Intelligence for Humans, Volume 3: Deep Learning and Neural Networks. Heaton Research Inc, 2015.
- [73] Nielsen, M. A., Neural Networks and Deep Learning. Determination Press, 2015.
- [74] Cointault, F., Guerin, D., Guillemin, J.-P., Chopinet, B., "In-field triticum aestivum ear counting using colour texture image analysis", New Zealand Journal of Crop and Horticultural Science, Vol. 36, No. 2, 2008, str. 117–130.
- [75] Alharbi, N., Zhou, J., Wang, W., "Automatic counting of wheat spikes from wheat growth images", in Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM). Venice, Italy: IEEE, 2018, str. 346–355.
- [76] Lu, H., Cao, Z., Xiao, Y., Zhuang, B., Shen, C., "Tasselnet: counting maize tassels in the wild via local counts regression network", Plant Methods, 2017.
- [77] "Thermal imaging definition", <http://whatis.techtarget.com/definition/thermal-imaging>, accessed: 2018-12-11.
- [78] Heaton, J., Artificial Intelligence for Humans, Volume 3: Deep Learning and Neural Networks. Heaton Research Inc., 2015.
- [79] Aghdam, H., Heravi, E., Guide to Convolutional Neural Networks. Springer International Publishing, 2017.
- [80] Xu, B., Wang, N., Chen, T., "Empirical evaluation of rectified activations in convolutional network", ICML Deep Learning Workshop, arXiv, 2015.
- [81] Breiman, L., "Random forests", Machine Learning, Vol. 45, No. 1, 2001, str. 5–32.
- [82] "Understanding binary cross-entropy / log loss: A visual explanation", <https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181>, accessed: 2018-12-17.

- [83] “Pytorch documentation: Loss functions”, <https://pytorch.org/docs/stable/modules/torch/nn/modules/loss.html>, accessed: 2018-12-17.
- [84] Brdar, S., Hogeveen, E., Panic, M., Mensink, M., Grbovic, Z., Harchioui, N. E., Chauhan, A., “A case study on prediction of sensitivity of tomato sepals to fungal infection using hyperspectral imaging”, dostupno na: <https://doi.org/10.5281/zenodo.4005770> Aug. 2020.
- [85] Željana Grbović, Brkić, M., Panić, M., Brdar, S., van Echtelt, E. H., Chauhan, A., “Semantic segmentation of tomato sepals on hyperspectral images using deep learning”, dostupno na: <https://doi.org/10.5281/zenodo.4896828> May 2021.
- [86] Željana Grbović, Panić, M., Brdar, S., van Echtelt, E. H., Mensink, M., Woltering, E. W., Chauhan, A., “Deep learning instance level segmentation of tomato sepals on hyperspectral images”, dostupno na: <https://doi.org/10.5281/zenodo.7257774> Sep. 2022.
- [87] Grbović, Ž., Panić, M., Filipović, V., Brdar, S., De Villiers, H., Mensink, M., Chauhan, A., “An efficient encoding spectral information in hyperspectral images for transfer learning of mask r-cnn for instance segmentation of tomato sepals”, IEEE Access, 2025.
- [88] Ronneberger, O., Fischer, P., Brox, T., “U-net: Convolutional networks for biomedical image segmentation”, in International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, Cham, 2015, str. 234–241.
- [89] Badrinarayanan, V., Kendall, A., Cipolla, R., “Segnet: A deep convolutional encoder-decoder architecture for image segmentation”, IEEE transactions on pattern analysis and machine intelligence, Vol. 39, No. 12, 2017, str. 2481–2495.
- [90] Grbovic, Z., Panic, M., Marko, O., Brdar, S., Crnojevic, V., “Wheat ear detection in rgb and thermal images using deep neural networks”, in Proceedings of the 15th International Conference on Machine Learning and Data Mining (MLDM), New York, USA, 2019.

- [91] Ibrahim, F., Al-Ebady, N. *et al.*, “Evaluation of antifungal activity of some plant extracts and their applicability in extending the shelf life of stored tomato fruits”, *J. Food Process. Technol.*, Vol. 5, No. 6, 2014, str. 340.
- [92] Zhang, Y., De Stefano, R., Robine, M., Butelli, E., Bulling, K., Hill, L., Rejzek, M., Martin, C., Schoonbeek, H.-j., “Different reactive oxygen species scavenging properties of flavonoids determine their abilities to extend the shelf life of tomato”, *Plant Physiology*, Vol. 169, No. 3, 2015, str. 1568–1583.
- [93] Pobiega, K., Przybył, J. L., Żubernik, J., Gniewosz, M., “Prolonging the shelf life of cherry tomatoes by pullulan coating with ethanol extract of propolis during refrigerated storage”, *Food and Bioprocess Technology*, Vol. 13, 2020, str. 1447–1461.
- [94] Durai, S. K. S., Shamili, M. D., “Smart farming using machine learning and deep learning techniques”, *Decision Analytics Journal*, Vol. 3, 2022, str. 100041.
- [95] Shaikh, T. A., Rasool, T., Lone, F. R., “Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming”, *Computers and Electronics in Agriculture*, Vol. 198, 2022, str. 107119.
- [96] Prabakaran, S., T, D., H, G., K, A. P., K, K., “Smart farm management using machine learning”, in *2024 International Conference on IoT, Communication and Automation Technology (ICI-CAT)*, 2024, str. 225-229.
- [97] Prashanth, J. S., Krishna, G. B., Prasad, A., Rao, P. R., “Smart farming revolution: A cutting-edge review of deep learning and iot innovations in agriculture”, in *Operations Research Forum*, Vol. 6, No. 1. Springer, 2025, str. 1–39.
- [98] Lone, Z. A., Pais, A. R., “Object detection in hyperspectral images”, *Digital Signal Processing*, Vol. 131, 2022, str. 103752, dostupno na: <https://www.sciencedirect.com/science/article/pii/S1051200422003694>
- [99] Ding, S., Zhao, H., Zhang, Y., Xu, X., Nie, R., “Extreme learning machine: algorithm, theory and applications”, *Artificial Intelligence Review*, Vol. 44, 2015, str. 103–115.

- [100] Zhang, M., Qin, Z., Liu, X., Ustin, S. L., “Detection of stress in tomatoes induced by late blight disease in california, usa, using hyperspectral remote sensing”, *International Journal of Applied Earth Observation and Geoinformation*, Vol. 4, No. 4, 2003, str. 295–310.
- [101] Nielsen, A. A., “Kernel maximum autocorrelation factor and minimum noise fraction transformations”, *IEEE Transactions on Image Processing*, Vol. 20, No. 3, 2010, str. 612–624.
- [102] Sachinopoulou, A., “Multidimensional visualization”, VTT Technical Research Centre of Finland, Espoo, Finland, Research Notes VTT Research Notes 2114, 2001, dostupno na: <https://publications.vtt.fi/pdf/tiedotteet/2001/T2114.pdf>
- [103] Martínez, P., Pérez, R. M., Plaza, A. J., Aguilar, P. L., Cantero, M. C., Plaza, J., “Endmember extraction algorithms from hyperspectral images”, *Annals of Geophysics*, Vol. 49, 2006, dostupno na: <https://api.semanticscholar.org/CorpusID:16229574>
- [104] Kruse, F., Lefkoff, A., Dietz, J., “Expert system-based mineral mapping in northern death valley, california/nevada, using the airborne visible/infrared imaging spectrometer (aviris)”, *Remote Sensing of Environment*, Vol. 44, No. 2-3, 1993, str. 309–336.
- [105] Lü, Q., Tang, M., “Detection of hidden bruise on kiwi fruit using hyperspectral imaging and parallelepiped classification”, *Procedia Environmental Sciences*, Vol. 12, 2012, str. 1172–1179.
- [106] Richards, J. A., Jia, X., Richards, J. A., Jia, X., “The interpretation of digital image data”, *Remote Sensing Digital Image Analysis: An Introduction*, 1999, str. 75–88.
- [107] Jiang, Y., Li, C., Takeda, F., “Nondestructive detection and quantification of blueberry bruising using near-infrared (nir) hyperspectral reflectance imaging”, *Scientific Reports*, Vol. 6, No. 1, 2016, str. 1–14.
- [108] Noble, W. S., “What is a support vector machine?”, *Nature biotechnology*, Vol. 24, No. 12, 2006, str. 1565–1567.

- [109] Fan, S., Li, C., Huang, W., Chen, L., “Data fusion of two hyperspectral imaging systems with complementary spectral sensing ranges for blueberry bruising detection”, *Sensors*, Vol. 18, No. 12, 2018, str. 4463.
- [110] Ma, H., Zhao, K., Jin, X., Ji, J., Qiu, Z., Gao, S., “Spectral difference analysis and identification of different maturity blueberry fruit based on hyperspectral imaging using spectral index”, *International Journal of Agricultural and Biological Engineering*, Vol. 12, No. 3, 2019, str. 134–140.
- [111] Huang, Y., Wang, D., Liu, Y., Zhou, H., Sun, Y., “Measurement of early disease blueberries based on vis/nir hyperspectral imaging system”, *Sensors*, Vol. 20, No. 20, 2020, str. 5783.
- [112] Qiao, S., Wang, Q., Zhang, J., Pei, Z., “Detection and classification of early decay on blueberry based on improved deep residual 3d convolutional neural network in hyperspectral images”, *Scientific Programming*, Vol. 2020, 2020.
- [113] Zhou, C., Li, Z., Wang, D., Xue, S., Zhu, T., Ni, C., “Ssnet: Exploiting spatial information for tobacco stem impurity detection with hyperspectral imaging”, *IEEE Access*, Vol. 12, 2024, str. 55 134-55 145.
- [114] Li, J., Xu, F., Song, S., Ji, Q., Liu, J., “Hyperspectral rgb imaging combined with deep learning for maize seed variety identification”, *IEEE Access*, Vol. 12, 2024, str. 184 477-184 486.
- [115] Paoletti, M., Haut, J., Plaza, J., Plaza, A., “Deep learning classifiers for hyperspectral imaging: A review”, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 158, 2019, str. 279-317, dostupno na: <https://www.sciencedirect.com/science/article/pii/S0924271619302187>
- [116] He, X., Tang, C., Liu, X., Zhang, W., Sun, K., Xu, J., “Object detection in hyperspectral image via unified spectral-spatial feature aggregation”, *arXiv preprint arXiv:2306.08370*, 2023.
- [117] Wang, Q., Zhang, F., Li, X., “Hyperspectral band selection via optimal neighborhood reconstruction”, *IEEE Transactions on*

- Geoscience and Remote Sensing, Vol. 58, No. 12, 2020, str. 8465–8476.
- [118] Sneha, Kaul, A., “Hyperspectral imaging and target detection algorithms: a review”, *Multimedia Tools and Applications*, Vol. 81, No. 30, 2022, str. 44 141–44 206.
- [119] Chen, H., Miao, F., Chen, Y., Xiong, Y., Chen, T., “A hyperspectral image classification method using multifeature vectors and optimized kelm”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14, 2021, str. 2781–2795.
- [120] Xie, C., Shao, Y., Li, X., He, Y., “Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging”, *Scientific reports*, Vol. 5, No. 1, 2015, str. 1–11.
- [121] Soares, S. F. C., Gomes, A. A., Araujo, M. C. U., Galvão Filho, A. R., Galvão, R. K. H., “The successive projections algorithm”, *TrAC Trends in Analytical Chemistry*, Vol. 42, 2013, str. 84–98.
- [122] Makantasis, K., Karantzalos, K., Doulamis, A., Doulamis, N., “Deep supervised learning for hyperspectral data classification through convolutional neural networks”, in *2015 IEEE international geoscience and remote sensing symposium (IGARSS)*. IEEE, 2015, str. 4959–4962.
- [123] Yan, L., Yamaguchi, M., Noro, N., Takara, Y., Ando, F., “A novel two-stage deep learning-based small-object detection using hyperspectral images”, *Optical Review*, Vol. 26, 2019, str. 597–606.
- [124] Ruder, S., “An overview of gradient descent optimization algorithms”, *arXiv preprint arXiv:1609.04747*, 2016.
- [125] Theodoridis, S., Koutroumbas, K., *Pattern recognition*. Elsevier, 2006.
- [126] Kallepalli, A., Kumar, A., Khoshelham, K., “Entropy based determination of optimal principal components of airborne prism experiment (apex) imaging spectrometer data for improved land cover classification”, *The International Archives*

- of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XL-8, 2014, str. 781–786, dostupno na: <https://isprs-archives.copernicus.org/articles/XL-8/781/2014/>
- [127] Sima, C., Dougherty, E. R., “The peaking phenomenon in the presence of feature-selection”, *Pattern Recognition Letters*, Vol. 29, No. 11, 2008, str. 1667–1674.
 - [128] Signoroni, A., Savardi, M., Baronio, A., Benini, S., “Deep learning meets hyperspectral image analysis: A multidisciplinary review”, *Journal of Imaging*, Vol. 5, No. 5, 2019, str. 52.
 - [129] Li, W., Wu, G., Zhang, F., Du, Q., “Hyperspectral image classification using deep pixel-pair features”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 55, No. 2, 2016, str. 844–853.
 - [130] Ran, L., Zhang, Y., Wei, W., Zhang, Q., “A hyperspectral image classification framework with spatial pixel pair features”, *Sensors*, Vol. 17, No. 10, 2017, str. 2421.
 - [131] Sun, W., Du, Q., “Hyperspectral band selection: A review”, *IEEE Geoscience and Remote Sensing Magazine*, Vol. 7, No. 2, 2019, str. 118–139.
 - [132] LeCun, Y., Bengio, Y., Hinton, G., “Deep learning”, *nature*, Vol. 521, No. 7553, 2015, str. 436–444.
 - [133] Khamparia, A., Singh, K. M., “A systematic review on deep learning architectures and applications”, *Expert Systems*, Vol. 36, No. 3, 2019, str. e12400.
 - [134] Shrestha, A., Mahmood, A., “Review of deep learning algorithms and architectures”, *IEEE access*, Vol. 7, 2019, str. 53 040–53 065.
 - [135] Mathew, A., Amudha, P., Sivakumari, S., “Deep learning techniques: an overview”, *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020, 2021*, str. 599–608.
 - [136] Gheisari, M., Ebrahimzadeh, F., Rahimi, M., Moazzamigodarsi, M., Liu, Y., Dutta Pramanik, P. K., Heravi, M. A., Mehbodniya, A., Ghaderzadeh, M., Feylizadeh, M. R. *et al.*, “Deep

- p learning: Applications, architectures, models, tools, and frameworks: A comprehensive survey”, CAAI Transactions on Intelligence Technology, 2023.
- [137] Long, J., Shelhamer, E., Darrell, T., “Fully convolutional networks for semantic segmentation”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, str. 3431–3440.
 - [138] Pinheiro, P. O., Lin, T.-Y., Collobert, R., Dollár, P., “Learning to refine object segments”, in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer, 2016, str. 75–91.
 - [139] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M., “Deep learning for generic object detection: A survey. arxiv 2018”, arXiv preprint arXiv:1809.02165, 2019.
 - [140] Girshick, R., “Fast r-cnn”, in Proceedings of the IEEE international conference on computer vision, 2015, str. 1440–1448.
 - [141] Zagoruyko, S., Lerer, A., Lin, T.-Y., Pinheiro, P. O., Gross, S., Chintala, S., Dollár, P., “A multipath network for object detection”, arXiv preprint arXiv:1604.02135, 2016.
 - [142] He, K., Gkioxari, G., Dollár, P., Girshick, R., “Mask r-cnn”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 42, No. 2, 2020, str. 386–397.
 - [143] Wang, X., Girshick, R., Gupta, A., He, K., “Non-local neural networks”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, str. 7794–7803.
 - [144] Liu, S., Qi, L., Qin, H., Shi, J., Jia, J., “Path aggregation network for instance segmentation”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, str. 8759–8768.
 - [145] Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Shi, J., Ouyang, W. *et al.*, “Hybrid task cascade for instance segmentation”, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, str. 4974–4983.

- [146] Cao, Y., Xu, J., Lin, S., Wei, F., Hu, H., “Gcnet: Non-local networks meet squeeze-excitation networks and beyond”, in Proceedings of the IEEE/CVF international conference on computer vision workshops, 2019, str. 0–0.
- [147] Bolya, D., Zhou, C., Xiao, F., Lee, Y. J., “Yolact: Real-time instance segmentation”, in Proceedings of the IEEE/CVF international conference on computer vision, 2019, str. 9157–9166.
- [148] Chen, X., Girshick, R., He, K., Dollár, P., “Tensormask: A foundation for dense object segmentation”, in Proceedings of the IEEE/CVF international conference on computer vision, 2019, str. 2061–2069.
- [149] Lazarow, J., Xu, W., Tu, Z., “Instance segmentation with mask-supervised polygonal boundary transformers”, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, str. 4382–4391.
- [150] Chen, W., Du, X., Yang, F., Beyer, L., Zhai, X., Lin, T.-Y., Chen, H., Li, J., Song, X., Wang, Z. *et al.*, “A simple single-scale vision transformer for object detection and instance segmentation”, in European Conference on Computer Vision. Springer, 2022, str. 711–727.
- [151] Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., Prince, J. L., Rueckert, D., Summers, R. M., “A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises”, Proceedings of the IEEE, Vol. 109, No. 5, 2021, str. 820–838.
- [152] Mozaffari, S., Al-Jarrah, O. Y., Dianati, M., Jennings, P., Mouzakitis, A., “Deep learning-based vehicle behavior prediction for autonomous driving applications: A review”, IEEE Transactions on Intelligent Transportation Systems, Vol. 23, No. 1, 2020, str. 33–47.
- [153] Coulibaly, S., Kamsu-Foguem, B., Kamissoko, D., Traore, D., “Deep learning for precision agriculture: A bibliometric analysis”, Intelligent Systems with Applications, Vol. 16, 2022, str. 200102.

- [154] Chen, Y., Zhao, X., Jia, X., “Spectral–spatial classification of hyperspectral data based on deep belief network”, *IEEE journal of selected topics in applied earth observations and remote sensing*, Vol. 8, No. 6, 2015, str. 2381–2392.
- [155] Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P., “Deep feature extraction and classification of hyperspectral images based on convolutional neural networks”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 54, No. 10, 2016, str. 6232–6251.
- [156] Cao, X., Yao, J., Xu, Z., Meng, D., “Hyperspectral image classification with convolutional neural network and active learning”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 58, No. 7, 2020, str. 4604–4616.
- [157] Luo, Y., Zou, J., Yao, C., Zhao, X., Li, T., Bai, G., “Hsi-cnn: a novel convolution neural network for hyperspectral image”, in *2018 International Conference on Audio, Language and Image Processing (ICALIP)*. IEEE, 2018, str. 464–469.
- [158] Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., Benediktsson, J. A., “Deep learning for hyperspectral image classification: An overview”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, No. 9, 2019, str. 6690–6709.
- [159] Zhong, Z., Li, J., Ma, L., Jiang, H., Zhao, H., “Deep residual networks for hyperspectral image classification”, in *2017 IEEE international geoscience and remote sensing symposium (IGARSS)*. IEEE, 2017, str. 1824–1827.
- [160] Debes, C., Merentitis, A., Heremans, R., Hahn, J., Frangiadakis, N., van Kasteren, T., Liao, W., Bellens, R., Pižurica, A., Gautama, S. *et al.*, “Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 7, No. 6, 2014, str. 2405–2418.
- [161] Zhao, Z., Zhang, Q., Cheng, H., Wei, Y., Li, Y., “Hyperspectral image classification with markov random fields and a convolutional neural network”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 54, No. 10, 2016, str. 5873–5885.

- [162] Chen, Y., Lin, L., Chen, L., “Attention-based spectral-spatial classification of hyperspectral images”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 55, No. 2, 2017, str. 844–853.
- [163] Joseph, D. S., Pawar, P. M., Pramanik, R., “Intelligent plant disease diagnosis using convolutional neural network: A review”, *Multimedia Tools and Applications*, Vol. 82, No. 14, 2023, str. 21 415–21 481.
- [164] Azmi, A. N. N., Khairunniza-Bejo, S., Jahari, M., Yule, I., “Early detection of plant disease infection using hyperspectral data and machine learning”, *IoT and AI in Agriculture: Self-sufficiency in Food Production to Achieve Society 5.0 and SDG’s Globally*, 2023, str. 423–446.
- [165] Jolliffe, I. T., *Principal component analysis for special types of data*. Springer, 2002.
- [166] Hotelling, H., “Analysis of a complex of statistical variables into principal components.”, *Journal of educational psychology*, Vol. 24, No. 6, 1933, str. 417.
- [167] Tipping, M. E., Bishop, C. M., “Probabilistic principal component analysis”, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Vol. 61, No. 3, 1999, str. 611–622.
- [168] Basilevsky, A. T., *Statistical factor analysis and related methods: theory and applications*. John Wiley & Sons, 2009.
- [169] Barber, D., *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.
- [170] Munipalle, V. K., Nelakuditi, U. R., Nidamanuri, R. R., “Impact of dimensionality reduction techniques on classification of hyperspectral images”, in *2023 3rd International Conference on Intelligent Technologies (CONIT)*. IEEE, 2023, str. 1–6.
- [171] Lennon, M., Mercier, G., Mouchot, M., Hubert-Moy, L., “Independent component analysis as a tool for the dimensionality reduction and the representation of hyperspectral images”, in *IGARSS 2001. Scanning the Present and Resolving the Future*.

- Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217), Vol. 6. IEEE, 2001, str. 2893–2895.
- [172] Wang, J., Chang, C.-I., “Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis”, IEEE transactions on geoscience and remote sensing, Vol. 44, No. 6, 2006, str. 1586–1600.
- [173] Pena, S. B., Abreu, M. M., Magalhães, M. R., “Planning landscape with water infiltration. empirical model to assess maximum infiltration areas in mediterranean landscapes”, Water resources management, Vol. 30, No. 7, 2016, str. 2343–2360.
- [174] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., “Microsoft coco: Common objects in context”, in European conference on computer vision. Springer, 2014, str. 740–755.
- [175] Everingham, M., Van Gool, L., Williams, C. K., Winn, J., Zisserman, A., “The pascal visual object classes (voc) challenge”, International journal of computer vision, Vol. 88, No. 2, 2010, str. 303–338.
- [176] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q., “A comprehensive survey on transfer learning”, Proceedings of the IEEE, Vol. 109, No. 1, 2021, str. 43-76.
- [177] Torrey, L., Shavlik, J., “Transfer learning”, in Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI global, 2010, str. 242–264.
- [178] Rehman, Z. U., Khan, M. A., Ahmed, F., Damaševičius, R., Naqvi, S. R., Nisar, W., Javed, K., “Recognizing apple leaf diseases using a novel parallel real-time processing framework based on mask rcnn and transfer learning: An application for smart agriculture”, IET Image Processing, Vol. 15, No. 10, 2021, str. 2157–2168.
- [179] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y., “Deformable convolutional networks”, dostupno na: <https://arxiv.org/abs/1703.06211> 2017.

- [180] Mac Key, J., Qualset, C. O., “Conventional methods of wheat breeding”, Genetic Improvement in Yield of Wheat, Vol. 13, 1986, str. 7–24.
- [181] Crop, N. D., Report, P., “5 steps for estimating wheat yields”, <https://www.farmprogress.com/wheat/5-steps-estimating-wheat-yields>, 2018.
- [182] Cai, Y., Guan, K., Lobell, D., Potgieter, A. B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L. *et al.*, “Integrating satellite and climate data to predict wheat yield in australia using machine learning approaches”, Agricultural and forest meteorology, Vol. 274, 2019, str. 144–159.
- [183] Hasan, M., Chopin, J., Laga, H., Miklavcic, S., “Detection and analysis of wheat spikes using convolutional neural networks”, Plant Methods, Vol. 14, 11 2018.
- [184] Thakur, A. K., Singh, S., Goyal, N., Gupta, K., “A comparative analysis on the existing techniques of wheat spike detection”, in 2021 2nd International Conference for Emerging Technology (INCET). IEEE, 2021, str. 1–6.
- [185] Zhu, N., Liu, X., Liu, Z., Hu, K., Wang, Y., Tan, J., Huang, M., Zhu, Q., Ji, X., Jiang, Y. *et al.*, “Deep learning for smart agriculture: Concepts, tools, applications, and opportunities”, International Journal of Agricultural and Biological Engineering, Vol. 11, No. 4, 2018, str. 32–44.
- [186] Kondaveeti, H. K., Brahma, G. P., Sahithi, D. V., “Deep learning applications in agriculture: The role of deep learning in smart agriculture”, in Artificial Intelligence and IoT-Based Technologies for Sustainable Farming and Smart Agriculture. IGI Global, 2021, str. 325–345.
- [187] Dyrmann, M., Jørgensen, R. N., Midtiby, H. S., “Roboweedsupport-detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network”, Advances in Animal Biosciences, Vol. 8, No. 2, 2017, str. 842–847.

- [188] McCool, C., Perez, T., Upcroft, B., “Mixtures of lightweight deep convolutional neural networks: Applied to agricultural robotics”, *IEEE Robotics and Automation Letters*, Vol. 2, No. 3, 2017, str. 1344–1351.
- [189] Pound, M. P., Atkinson, J. A., Wells, D. M., Pridmore, T. P., French, A. P., “Deep learning for multi-task plant phenotyping”, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2017.
- [190] Taghavi Namin, S., Esmailzadeh, M., Najafi, M., Brown, T. B., Borevitz, J. O., “Deep phenotyping: deep learning for temporal phenotype/genotype classification”, *Plant methods*, Vol. 14, No. 1, 2018, str. 1–14.
- [191] Lu, J., Hu, J., Zhao, G., Mei, F., Zhang, C., “An in-field automatic wheat disease diagnosis system”, *Computers and electronics in agriculture*, Vol. 142, 2017, str. 369–379.
- [192] Saleem, M. H., Potgieter, J., Arif, K. M., “Plant disease detection and classification by deep learning”, *Plants*, Vol. 8, No. 11, 2019, dostupno na: <https://www.mdpi.com/2223-7747/8/11/468>
- [193] Genaev, M. A., Skolotneva, E. S., Gulyaeva, E. I., Orlova, E. A., Bechtold, N. P., Afonnikov, D. A., “Image-based wheat fungi diseases identification by deep learning”, *Plants*, Vol. 10, No. 8, 2021, dostupno na: <https://www.mdpi.com/2223-7747/10/8/1500>
- [194] Goyal, L., Sharma, C. M., Singh, A., Singh, P. K., “Leaf and spike wheat disease detection & classification using an improved deep convolutional architecture”, *Informatics in Medicine Unlocked*, Vol. 25, 2021, str. 100642, dostupno na: <https://www.sciencedirect.com/science/ARTICLE/pii/S2352914821001313>
- [195] Li, L., Zhang, S., Wang, B., “Plant disease detection and classification by deep learning—a review”, *IEEE Access*, Vol. 9, 2021, str. 56 683–56 698.
- [196] Kumar, D., Kukreja, V., “Deep learning in wheat diseases classification: A systematic review”, *Multimedia Tools and Applications*, 2022, str. 1–45.

- [197] Lad, A., Raval, M. S., “Improving wheat head detection: a data-centric approach by domain variance reduction”, in Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing, 2022, str. 1140–1143.
- [198] Namin, S. T., Esmailzadeh, M., Najafi, M., Brown, T. B., Borevitz, J. O., “Deep phenotyping: deep learning for temporal phenotype/genotype classification”, *Plant methods*, Vol. 14, No. 1, 2018, str. 66.
- [199] Ubbens, J., Stavness, I., “Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks”, *Frontiers in Plant Science*, Vol. 8, 2017, str. 1190.
- [200] Chen, L.-C., Hermans, A., Papandreou, G., Schroff, F., Wang, P., Adam, H., “Masklab: Instance segmentation by refining object detection with semantic and direction features”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, str. 4013–4022.
- [201] Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., Girdhar, R., “Masked-attention mask transformer for universal image segmentation”, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, str. 1290–1299.
- [202] Huang, Z., Huang, L., Gong, Y., Huang, C., Wang, X., “Mask scoring r-cnn”, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, str. 6409–6418.
- [203] Singh, A. K., Ganapathysubramanian, B., Sarkar, S., Singh, A., “Deep learning for plant stress phenotyping: trends and future perspectives”, *Trends in plant science*, Vol. 23, No. 10, 2018, str. 883–898.
- [204] Narisetti, N., Neumann, K., Röder, M. S., Gladilin, E., “Automated spike detection in diverse european wheat plants using textural features and the frangi filter in 2d greenhouse images”, *Frontiers in Plant Science*, 2020, str. 666.
- [205] Gong, B., Ergu, D., Cai, Y., Ma, B., “A method for wheat head detection based on yolov4”, *Research Square*, 2020.

- [206] Zhan, J., “Wheat ear detection based on one-stage object detector”, SPIE, str. 1396–1401, 2023.
- [207] Gong, B., Ergu, D., Cai, Y., Ma, B., “Real-time detection for wheat head applying deep neural network”, *Sensors*, Vol. 21, No. 1, 2020, str. 191.
- [208] Xu, X., Qiao, H., Ma, X., Yin, G., Wang, Y., Zhao, J., Li, H., “An automatic wheat ear counting model based on the minimum area intersection ratio algorithm and transfer learning”, *Measurement*, Vol. 216, 2023, str. 112849.
- [209] Qian, G., “Wheatnet-cs: A wheat ear detection algorithm for complex background”, Springer, str. 357–368, 2023.
- [210] Li, Z., Zhu, Y., Sui, S., Zhao, Y., Liu, P., Li, X., “Real-time detection and counting of wheat ears based on improved yolov7”, *Computers and Electronics in Agriculture*, Vol. 218, 2024, str. 108670.
- [211] Li, R., Meng, J., Wu, Y., Zhang, D., He, Y., “Wheat ear detection based on fastercanet-yolov8s algorithm”, *New Zealand Journal of Crop and Horticultural Science*, 2024, str. 1–21.
- [212] Wen, C., Ma, Z., Ren, J., Zhang, T., Zhang, L., Chen, H., Su, H., Yang, C., Chen, H., Guo, W., “A generalized model for accurate wheat spike detection and counting in complex scenarios”, *Scientific reports*, Vol. 14, No. 1, 2024, str. 24189.
- [213] Khaki, S., Safaei, N., Pham, H., Wang, L., “Wheatnet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting”, arXiv preprint arXiv:2103.09408, 2021.
- [214] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., “Mobilenetv2: Inverted residuals and linear bottlenecks”, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, str. 4510–4520.
- [215] Dandrifosse, S., Ennadifi, E., Carlier, A., Gosselin, B., Dumont, B., Mercatoris, B., “Deep learning for wheat ear segmentation and ear density measurement: From heading to maturity”, *Computers and Electronics in Agriculture*, Vol. 199, 2022, str. 107161.

- [216] Solovyev, R., Wang, W., Gabruseva, T., “Weighted boxes fusion: Ensembling boxes from different object detection models”, *Image and Vision Computing*, Vol. 107, 2021, str. 104117.
- [217] Birodkar, V., Lu, Z., Li, S., Rathod, V., Huang, J., “The surprising impact of mask-head architecture on novel class segmentation”, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, str. 7015–7025.
- [218] Wen, C., Wu, J., Chen, H., Su, H., Chen, X., Li, Z., Yang, C., “Wheat spike detection and counting in the field based on spike-retinanet”, *Frontiers in Plant Science*, Vol. 13, 2022.
- [219] Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., “Focal loss for dense object detection”, in *Proceedings of the IEEE international conference on computer vision*, 2017, str. 2980–2988.
- [220] Wang, Y., Qin, Y., Cui, J., “Occlusion robust wheat ear counting algorithm based on deep learning”, *Frontiers in Plant Science*, Vol. 12, 2021, str. 1139.
- [221] Tan, M., Pang, R., Le, Q. V., “Efficientdet: Scalable and efficient object detection”, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, str. 10 781–10 790.
- [222] Guan, Y., Pan, J., Fan, Q., Yang, L., Yin, X., Jia, W., “Ctwheat-net: Accurate detection model of wheat ears in field”, *Computers and Electronics in Agriculture*, Vol. 225, 2024, str. 109272.
- [223] Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., Girshick, R., “Detectron2”, <https://github.com/facebookresearch/detectron2>, 2019.
- [224] Ultralytics, “Yolo documentation”, dostupno na: <https://docs.ultralytics.com/> Accessed: 2024-10-31. 2023.
- [225] “Global wheat detection competition leaderboard”, <https://www.kaggle.com/competitions/global-wheat-detection/leaderboard>, Kaggle, accessed: December 14, 2022. 2022.
- [226] David, E., Ogidi, F., Guo, W., Baret, F., Stavness, I., “Global wheat challenge 2020: Analysis of the competition design and winning models”, *arXiv preprint arXiv:2105.06182*, 2021.

- [227] Li, J., Li, C., Fei, S., Ma, C., Chen, W., Ding, F., Wang, Y., Li, Y., Shi, J., Xiao, Z., “Wheat ear recognition based on retinanet and transfer learning”, *Sensors*, Vol. 21, No. 14, 2021, str. 4845.
- [228] Ban, X., Liu, P., Xu, L., Zhao, J., “A lightweight model based on yolov8n in wheat spike detection”, in *2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*. IEEE, 2023, str. 1–6.
- [229] Mas, M. S., Saidah, S., Ibrahim, N., “Detection and counting of wheat ear using yolov8.”, *International Journal of Electrical & Computer Engineering (2088-8708)*, Vol. 14, No. 5, 2024.
- [230] Zhao, Z., Feng, W., Zou, X., Li, Q., “Method for wheat ear detection based on csr-detr”, in *2024 6th International Conference on Electronic Engineering and Informatics (EEI)*. IEEE, 2024, str. 464–467.
- [231] Yang, Z., Yang, W., Yi, J., Liu, R., “Wh-detr: An efficient network architecture for wheat spike detection in complex backgrounds”, *Agriculture*, Vol. 14, No. 6, 2024, str. 961.
- [232] Redmon, J., Farhadi, A., “Yolov3: An incremental improvement”, *arXiv preprint arXiv:1804.02767*, 2018.
- [233] Lee, D.-H. *et al.*, “Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks”, in *Workshop on challenges in representation learning, ICML*, Vol. 3, 2013, str. 896.
- [234] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., “NeRF: Representing scenes as neural radiance fields for view synthesis”, *arXiv preprint arXiv:2003.08934*, 2020, dostupno na: <https://arxiv.org/abs/2003.08934>
- [235] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., “Nerf: Representing scenes as neural radiance fields for view synthesis”, *Communications of the ACM*, Vol. 65, No. 1, 2021, str. 99–106.

- [236] Gao, K., Gao, Y., He, H., Lu, D., Xu, L., Li, J., “Nerf: Neural radiance field in 3d vision, a comprehensive review”, arXiv preprint arXiv:2210.00379, 2022.
- [237] Martin-Brualla, R., Radwan, N., Aggarwal, V., Srinivasan, P. P., Barron, J. T., Mildenhall, B., “Nerf in the wild: Neural radiance fields for unconstrained photo collections”, arXiv preprint arXiv:2008.02268, 2020, dostupno na: <https://arxiv.org/abs/2008.02268>
- [238] Zhang, K., Jacobson, A., Poole, B., Tancik, M., Mildenhall, B., Barron, J. T., Srinivasan, P. P., Ng, R., “Nerf++: Analyzing and improving neural radiance fields”, arXiv preprint arXiv:2010.07492, 2020, dostupno na: <https://arxiv.org/abs/2010.07492>
- [239] Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P. P., “Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields”, in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, str. 5855–5864, dostupno na: <https://arxiv.org/abs/2103.13415>
- [240] Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., Hedman, P., “Mip-nerf 360: Unbounded anti-aliased neural radiance fields”, arXiv preprint arXiv:2111.12077, 2021, dostupno na: <https://arxiv.org/abs/2111.12077>
- [241] Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J. T., Srinivasan, P. P., “Ref-nerf: Structured view-dependent appearance for neural radiance fields”, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, dostupno na: <https://arxiv.org/abs/2112.03907>
- [242] Müller, T., Evans, A., Schied, C., Keller, A., “Instant neural graphics primitives with a multiresolution hash encoding”, ACM Transactions on Graphics (Proceedings of SIGGRAPH 2022), 2022, dostupno na: <https://arxiv.org/abs/2201.05989>
- [243] Deng, K., Liu, A., Zhu, J.-Y., Ramanan, D., “Depth-supervised nerf: Fewer views and faster training for free”,

- arXiv preprint arXiv:2107.02791, 2021, dostupno na: <http://arxiv.org/abs/2107.02791v1>
- [244] Jain, A., Tancik, M., Abbeel, P., “Putting nerf on a diet: Semantically consistent few-shot view synthesis”, in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2021, str. 5885–5894.
 - [245] Niemeyer, M., Barron, J. T., Mildenhall, B., Sajjadi, M. S. M., Geiger, A., Radwan, N., “Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs”, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
 - [246] Wang, G., Chen, Z., Loy, C. C., Liu, Z., “Sparsenerf: Distilling depth ranking for few-shot novel view synthesis”, in Proceedings of the IEEE/CVF international conference on computer vision, 2023, str. 9065–9076.
 - [247] Turki, H., Nießner, M., Srivastava, D. R., “Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs”, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, str. 12 922–12 931.
 - [248] Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B., Srinivasan, P. P., Barron, J. T., Kretzschmar, H., “Block-nerf: Scalable large scene neural view synthesis”, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, str. 8248–8258.
 - [249] Zhang, J., Wang, X., Ni, X., Dong, F., Tang, L., Sun, J., Wang, Y., “Neural radiance fields for multi-scale constraint-free 3d reconstruction and rendering in orchard scenes”, *Computers and Electronics in Agriculture*, 2024, neRF-Ag model, up to $39\times$ faster, photorealistic orchard scenes.
 - [250] Ku, K., Jubery, T. Z., Rodriguez, E., Balu, A., Sarkar, S., Krishnamurthy, A., Ganapathysubramanian, B., “Nerf-based point cloud reconstruction using a stationary camera for agricultural applications”, arXiv preprint arXiv:2503.21958, 2025, stationary camera high-throughput phenotyping with near-100

- [251] Arshad, M. A., Jubery, T., Afful, J., Jignasu, A., Balu, A., Ganapathysubramanian, B., Sarkar, S., Krishnamurthy, A., “Evaluating neural radiance fields for 3d plant geometry reconstruction in field conditions”, *Plant Phenomics*, 2024, f1 = 74.6
- [252] Choi, H.-B., Park, J.-K., Park, S. H., Lee, T. S., “Nerf-based 3d reconstruction pipeline for acquisition and analysis of tomato crop morphology”, *Frontiers in Plant Science*, 2024, high correlation (R^2 up to 0.973) between NeRF 3D phenotyping and manual measurement.
- [253] Wu, S., Hu, C., Tian, B., Huang, Y., Yang, S., Li, S., Xu, S., “A 3d reconstruction platform for complex plants using ob-nerf”, *Frontiers in Plant Science*, 2025, oB-NeRF: 30 s train, $R^2 > 0.99$, MAE = 2 mm, for digital twin phenotyping.
- [254] Müller, T., Evans, A., Schied, C., Keller, A., “Instant neural graphics primitives with a multiresolution hash encoding”, *ACM transactions on graphics (TOG)*, Vol. 41, No. 4, 2022, str. 1–15.
- [255] Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., Hedman, P., “Mip-nerf 360: Unbounded anti-aliased neural radiance fields”, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, str. 5470–5479.
- [256] Russell, B. C., Torralba, A., Murphy, K. P., Freeman, W. T., “Labelme: a database and web-based tool for image annotation”, *International journal of computer vision*, Vol. 77, 2008, str. 157–173.
- [257] Boggs, T., “Spectral python (spy)”, <http://www.spectralpython.net>, 2014.
- [258] Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T., “scikit-image: image processing in python”, *PeerJ*, Vol. 2, 2014, str. e453.
- [259] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. *et al.*, “Scikit-learn: Machine learning in python”, *the Journal of machine Learning research*, Vol. 12, 2011, str. 2825–2830.

- [260] “Meteoblue climate data for bajmok, serbia”, https://www.meteoblue.com/en/weather/forecast/modelclimate/bajmok_serbia_3204619, accessed: 2018-12-12.
- [261] “Meteoblue climate data for ravno selo, serbia”, https://www.meteoblue.com/en/weather/forecast/modelclimate/ravno_selo_serbia_3191802, accessed: 2018-12-12.
- [262] “Sc6x0 series datasheet”, <http://www.videol.ch/files/documents/sc6x0-e.pdf>, accessed: 2018-12-12.
- [263] University of Edinburgh, “Affine transformation – hipr2”, <https://homepages.inf.ed.ac.uk/rbf/HIPR2/affine.htm>, accessed: 2021-02-21.
- [264] Wikipedia contributors, “Affine transformation — wikipedia, the free encyclopedia”, https://en.wikipedia.org/wiki/Affine_transformation, accessed: 2021-02-21.
- [265] Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Mallocci, M., Kolesnikov, A. *et al.*, “The open images dataset v4”, *International Journal of Computer Vision*, Vol. 128, No. 7, 2020, str. 1956–1981.
- [266] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., “Imagenet: A large-scale hierarchical image database”, in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, str. 248–255.
- [267] De Almeida, P. R., Oliveira, L. S., Britto Jr, A. S., Silva Jr, E. J., Koerich, A. L., “Pklot—a robust dataset for parking lot classification”, *Expert Systems with Applications*, Vol. 42, No. 11, 2015, str. 4937–4949.
- [268] Hsieh, M.-R., Lin, Y.-L., Hsu, W. H., “Drone-based object counting by spatially regularized regional proposal network”, in *Proceedings of the IEEE international conference on computer vision*, 2017, str. 4145–4153.
- [269] Goldman, E., Herzig, R., Eisenschtat, A., Goldberger, J., Hasner, T., “Precise detection in densely packed scenes”, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, str. 5227–5236.

- [270] Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., Siegwart, R., “Weedmap: A large-scale semantic weed mapping framework using aerial multi-spectral imaging and deep neural network for precision farming”, *Remote Sensing*, Vol. 10, No. 9, 2018, str. 1423.
- [271] Teimouri, N., Dyrmann, M., Nielsen, P. R., Mathiassen, S. K., Somerville, G. J., Jørgensen, R. N., “Weed growth stage estimator using deep convolutional neural networks”, *Sensors*, Vol. 18, No. 5, 2018, str. 1580.
- [272] Guo, W., Zheng, B., Potgieter, A. B., Diot, J., Watanabe, K., Noshita, K., Jordan, D. R., Wang, X., Watson, J., Ninomiya, S. *et al.*, “Aerial imagery analysis—quantifying appearance and number of sorghum heads for applications in breeding and agronomy”, *Frontiers in plant science*, 2018, str. 1544.
- [273] Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., Singh, A. K., Singh, A., Hirafuji, M., Ninomiya, S. *et al.*, “A weakly supervised deep learning framework for sorghum head detection and counting”, *Plant Phenomics*, Vol. 2019, 2019.
- [274] Xiong, H., Cao, Z., Lu, H., Madec, S., Liu, L., Shen, C., “Tassel-netv2: in-field counting of wheat spikes with context-augmented local regression networks”, *Plant methods*, Vol. 15, No. 1, 2019, str. 1–14.
- [275] Genaev, M. A., Komyshev, E. G., Smirnov, N. V., Kruchinina, Y. V., Goncharov, N. P., Afonnikov, D. A., “Morphometry of the wheat spike by analyzing 2d images”, *Agronomy*, Vol. 9, No. 7, 2019, str. 390.
- [276] Brooks, J., “Coco annotator”, URL <https://github.com/jsbrooks/coco-annotator>, 2019.
- [277] “COCO Annotator”, dostupno na: <https://github.com/jsbrooks/coco-annotator> Accessed: October 28, 2022. 2022.
- [278] Monograph, B., “Growth stages of mono-and dicotyledonous plants”, Federal Biological Research Centre for Agriculture and Forestry, 2001.

- [279] “Apple”, dostupno na: <https://support.apple.com/en-us/111875> 2020 (accessed February 27, 2025).
- [280] Joblove, G. H., Greenberg, D., “Color spaces for computer graphics”, in Proceedings of the 5th annual conference on Computer graphics and interactive techniques, 1978, str. 20–25.
- [281] Di Stefano, L., Bulgarelli, A., “A simple and efficient connected components labeling algorithm”, in Proceedings 10th international conference on image analysis and processing. IEEE, 1999, str. 322–327.
- [282] “OpenCV, Connected Components”, dostupno na: https://docs.opencv.org/3.4/d3/dc0/group__imgproc__shape.html Accessed: October 28, 2022. 2022.
- [283] Canny, J., “A computational approach to edge detection”, IEEE Transactions on pattern analysis and machine intelligence, No. 6, 1986, str. 679–698.
- [284] Xu, Z., Baojie, X., Guoxin, W., “Canny edge detection based on open cv”, in 2017 13th IEEE international conference on electronic measurement & instruments (ICEMI). IEEE, 2017, str. 53–56.
- [285] Hart, P. E., Duda, R., “Use of the hough transformation to detect lines and curves in pictures”, Communications of the ACM, Vol. 15, No. 1, 1972, str. 11–15.
- [286] Felzenszwalb, P., Huttenlocher, D., “Distance transforms of sampled functions. technical report tr2004–1963”, Cornell Computing and Information Science, 2004.
- [287] Saha, A., Mishra, S., Bovik, A. C., “Re-iqa: Unsupervised learning for image quality assessment in the wild”, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, str. 5846–5855.
- [288] Chen, H., Wang, Z., Yang, Y., Sun, Q., Ma, K., “Learning a deep color difference metric for photographic images”, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, str. 22 242–22 251.

- [289] Erlenbusch, F., Merkt, C., de Oliveira, B., Gatter, A., Schwenker, F., Klauck, U., Teutsch, M., “Thermal infrared single image dehazing and blind image quality assessment”, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, str. 459–469.
- [290] Mittal, A., Moorthy, A. K., Bovik, A. C., “No-reference image quality assessment in the spatial domain”, IEEE Transactions on image processing, Vol. 21, No. 12, 2012, str. 4695–4708.
- [291] Mittal, A., Moorthy, A. K., Bovik, A. C., “Blind/referenceless image spatial quality evaluator”, in 2011 conference record of the forty fifth asilomar conference on signals, systems and computers (ASILOMAR). IEEE, 2011, str. 723–727.
- [292] Mittal, A., Soundararajan, R., Bovik, A. C., “Making a “completely blind” image quality analyzer”, IEEE Signal processing letters, Vol. 20, No. 3, 2012, str. 209–212.
- [293] Venkatanath, N., Praneeth, D., Bh, M. C., Channappayya, S. S., Medasani, S. S., “Blind image quality evaluation using perception based features”, in 2015 Twenty First National Conference on Communications (NCC). IEEE, 2015, str. 1–6.
- [294] “Image Quality Metrics”, dostupno na: <https://www.mathworks.com/help/images/image-quality-metrics.html> Accessed: November 10, 2022. 2022.
- [295] SharkWipf, “nerf_dataset_preprocessing_helper: Intelligent(?) image sharpness filtering for optimized nerf training datasets”, https://github.com/SharkWipf/nerf_dataset_preprocessing_helper, 2024.
- [296] Schönberger, J. L., Frahm, J.-M., “Structure-from-motion revisited”, in Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [297] Schönberger, J. L., Zheng, E., Pollefeys, M., Frahm, J.-M., “Pixelwise view selection for unstructured multi-view stereo”, in European Conference on Computer Vision (ECCV), 2016.

- [298] Galindo-Prieto, B., Westad, F. O., “Classification in hyperspectral images by independent component analysis, segmented cross-validation and uncertainty estimates”, *Journal of Spectral Imaging*, 2018.
- [299] Feng, X.-R., Li, H.-C., Wang, R., Du, Q., Jia, X., Plaza, A., “Hyperspectral unmixing based on nonnegative matrix factorization: A comprehensive review”, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 15, 2022, str. 4414-4436.
- [300] Wei, H., Smith, J. P., “Machine learning for deconvolution and segmentation of hyperspectral imaging data from biopharmaceutical resins”, *Molecular Pharmaceutics*, Vol. 21, No. 11, 2024, str. 5565–5576.
- [301] Liu, Z., “Super convergence cosine annealing with warm-up learning rate”, in *CAIBDA 2022; 2nd International Conference on Artificial Intelligence, Big Data and Algorithms*, 2022, str. 1-7.
- [302] Xu, Z., Dai, A. M., Kemp, J., Metz, L., “Learning an adaptive learning rate schedule”, *arXiv preprint arXiv:1909.09712*, 2019.
- [303] Defazio, A., Cutkosky, A., Mehta, H., Mishchenko, K., “When, why and how much? adaptive learning rate scheduling by refinement”, *arXiv preprint arXiv:2310.07831*, 2023.
- [304] He, K., Zhang, X., Ren, S., Sun, J., “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification”, in *Proceedings of the IEEE international conference on computer vision*, 2015, str. 1026–1034.
- [305] Reza, F. M., *An introduction to information theory*. Courier Corporation, 1994.
- [306] Hajar, M., El Badaoui, M., Raad, A., Bonnardot, F., “Discrete random sampling: Theory and practice in machine monitoring”, *Mechanical Systems and Signal Processing*, Vol. 123, 2019, str. 386–402.

- [307] Ye, J., Yu, Z., Wang, Y., Wei, L., Liu, X., Tian, X., “Wheatlfanet: in-field detection and counting of wheat heads with high-real-time global regression network”, *Plant Methods*, Vol. 19, No. 1, 2023, str. 103.
- [308] Shi, L., Sun, J., Dang, Y., Zhang, S., Sun, X., Xi, L., Wang, J., “Yolov5s-t: A lightweight small object detection method for wheat spikelet counting”, *Agriculture*, Vol. 13, No. 4, 2023, str. 872.
- [309] Shen, X., Zhang, C., Liu, K., Mao, W., Zhou, C., Yao, L., “A lightweight network for improving wheat ears detection and counting based on yolov5s”, *Frontiers in Plant Science*, Vol. 14, 2023, str. 1289726.
- [310] Yosinski, J., Clune, J., Bengio, Y., Lipson, H., “How transferable are features in deep neural networks?”, *Advances in neural information processing systems*, Vol. 27, 2014.
- [311] Shen, M., “Rethinking data selection for supervised fine-tuning”, *arXiv preprint arXiv:2402.06094*, 2024.
- [312] Jocher, G., Chaurasia, A., Qiu, J., “Ultralytics YOLOv8”, *dostupno na: <https://github.com/ultralytics/ultralytics> Version 8.0.0, License: AGPL-3.0*. 2023.
- [313] Lv, W., Xu, S., Zhao, Y., Wang, G., Wei, J., Cui, C., Du, Y., Dang, Q., Liu, Y., “Detrs beat yolos on real-time object detection”, 2023.
- [314] Yu, J., Chen, W., Liu, N., Fan, C., “Oriented feature pyramid network for small and dense wheat heads detection and counting”, *Scientific Reports*, Vol. 14, No. 1, 2024, str. 8106.
- [315] Gogtay, N. J., Thatte, U. M., “Principles of correlation analysis”, *Journal of the Association of Physicians of India*, Vol. 65, No. 3, 2017, str. 78–81.
- [316] Cohen, I., Huang, Y., Chen, J., Benesty, J., Benesty, J., Chen, J., Huang, Y., Cohen, I., “Pearson correlation coefficient”, *Noise reduction in speech processing*, 2009, str. 1–4.

- [317] Sedgwick, P., "Spearman's rank correlation coefficient", *Bmj*, Vol. 349, 2014.
- [318] Schaeffer, M. S., Levitt, E. E., "Concerning kendall's tau, a nonparametric correlation coefficient.", *Psychological Bulletin*, Vol. 53, No. 4, 1956, str. 338.
- [319] Cen, J., Zhou, Z., Fang, J., Shen, W., Xie, L., Jiang, D., Zhang, X., Tian, Q. *et al.*, "Segment anything in 3d with nerfs", *Advances in Neural Information Processing Systems*, Vol. 36, 2023, str. 25 971–25 990.
- [320] Filipovic, V., Grbović, Ž., Chauhan, A., de Villiers, H., Panić, M., Brdar, S., "Influence of spatial region selection on infection sensitivity prediction of tomato calyx using hyperspectral imaging", in *The 14th International Conference on Information Society and Technology*, 2024.
- [321] Wang, Z., Tan, X., Ma, Y., Liu, T., He, L., Yang, F., Shu, C., Li, L., Fu, H., Li, B. *et al.*, "Combining canopy spectral reflectance and rgb images to estimate leaf chlorophyll content and grain yield in rice", *Computers and Electronics in Agriculture*, Vol. 221, 2024, str. 108975.
- [322] Huang, P., Yang, P., Yang, L., Xiao, F., Feng, Y., Wang, Y., "Non-destructive detection and visualization of chlorophyll content in cherry tomatoes based on hyperspectral technology and machine learning", *Agriculture*, Vol. 14, No. 12, 2024, str. 2247.
- [323] Ghojogh, B., Ghodsi, A., Karray, F., Crowley, M., "Factor analysis, probabilistic principal component analysis, variational inference, and variational autoencoder: Tutorial and survey", *arXiv preprint arXiv:2101.00734*, 2021.
- [324] Baldi, P., "Autoencoders, unsupervised learning, and deep architectures", in *Proceedings of ICML workshop on unsupervised and transfer learning. JMLR Workshop and Conference Proceedings*, 2012, str. 37–49.
- [325] Jaiswal, G., Rani, R., Mangotra, H., Sharma, A., "Integration of hyperspectral imaging and autoencoders: Benefits, applications, hyperparameter tuning and challenges", *Computer Science Review*, Vol. 50, 2023, str. 100584.

- [326] Kingma, D. P., Welling, M. *et al.*, “An introduction to variational autoencoders”, *Foundations and Trends® in Machine Learning*, Vol. 12, No. 4, 2019, str. 307–392.
- [327] Loughlin, C., Manolakis, D., Pieper, M., Ingle, V., Bostick, R., Weisner, A., “Spectral variability modeling with variational autoencoders for hyperspectral target analysis”, in *Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXIX*, Vol. 12519. SPIE, 2023, str. 214–226.
- [328] Zhang, J., Min, A., Steffenson, B. J., Su, W., Hirsch, C. D., Anderson, J., Yang, C., Wei, J., “Wheat-net: An automatic dense wheat spike segmentation method based on an optimized hybrid task cascade model”, in *Unmanned Aerial Systems in Precision Agriculture*. Springer, 2022, str. 87–106.

План третмана података

Назив пројекта/истраживања
Семантичка сегментација дубоким неуронским мрежама са применом на сликама различитих модалитета у прецизној пољопривреди Semantic segmentation using deep neural networks across different image modalities in precision agriculture
Назив институције/институција у оквиру којих се спроводи истраживање
а) Факултет техничких наука, Универзитет у Новом Саду
Назив програма у оквиру ког се реализује истраживање
Истраживање се реализује у оквиру израде докторске дисертације на студијском програму: Енергетика, електроника и телекомуникације.
1. Опис података
1.1 Врста студије <i>Укратко описати тип студије у оквиру које се подаци прикупљају</i> <u>Докторска дисертација</u>
1.2 Врсте података а) квантитативни б) квалитативни
1.3. Начин прикупљања података а) анкете, упитници, тестови б) клиничке процене, медицински записи, електронски здравствени записи в) генотипови: навести врсту _____ г) административни подаци: навести врсту _____

- д) узорци ткива: навести врсту _____
- ђ) снимци, фотографије: навести врсту _____
- е) текст, навести врсту Актуелна литература у области истраживања
- ж) мапа, навести врсту _____
- з) остало: описати _____

1.3 Формат података, употребљене скале, количина података

1.3.1 Употребљени софтвер и формат датотеке:

- а) Excel фајл, датотека _____
- б) SPSS фајл, датотека _____
- в) **PDF фајл, датотека** _____
- г) **Текст фајл, датотека** _____
- д) **JPG фајл, датотека** _____
- е) **Остало, датотека** Пајтон _____

1.3.2. Број записа (код квантитативних података)

- а) број варијабли _____
- б) број мерења (испитаника, процена, снимака и сл.) _____

1.3.3. Поновљена мерења

- а) да
- б) **не**

Уколико је одговор да, одговорити на следећа питања:

- а) временски размак измедју поновљених мера је _____

- б) варијабле које се више пута мере односе се на _____
- в) нове верзије фајлова који садрже поновљена мерења су именоване као _____

Напомене: _____

Да ли формати и софтвер омогућавају дељење и дугорочну валидност података?

а) *Да*

б) *Не*

Ако је одговор не, образложити _____

2. Прикупљање података

2.1 Методологија за прикупљање/генерисање података

2.1.1. У оквиру ког истраживачког нацрта су подаци прикупљени?

- а) експеримент, навести тип _____
- б) корелационо истраживање, навести тип _____
- ц) **анализа текста, навести тип** Анализа доступне литературе и база података
- д) остало, навести шта _____

2.1.2 Навести врсте мерних инструмената или стандарде података специфичних за одређену научну дисциплину (ако постоје).

Подаци су прикупљани претрагом база података међународних институција које се баве анализама и прикупљањем истих.

2.2 Квалитет података и стандарди

2.2.1. Третман недостајућих података

а) Да ли матрица садржи недостајуће податке? Да **Не**

Ако је одговор да, одговорити на следећа питања:

а) Колики је број недостајућих података? _____

б) Да ли се кориснику матрице препоручује замена недостајућих података? Да Не

в) Ако је одговор да, навести сугестије за третман замене недостајућих података

2.2.2. На који начин је контролисан квалитет података? Описати

2.2.3. На који начин је извршена контрола уноса података у матрицу?

Податке у табеле је уносио аутор дисертације.

3. Третман података и пратећа документација

3.1. Третман и чување података

3.1.1. Подаци ће бити депоновани у Репозиторијум докторских дисертација

Универзитета у Новом Саду.

3.1.2. URL адреса _____

3.1.3. DOI _____

3.1.4. Да ли ће подаци бити у отвореном приступу?

- а) *Да*
- б) *Да, али после ембарга који ће трајати до _____*
- в) *Не*

Ако је одговор не, навести разлог _____

3.1.5. Подаци неће бити депоновани у репозиторијум, али ће бити чувани.

Образложење

3.2 Метаподаци и документација података

3.2.1. Који стандард за метаподатке ће бити примењен? Стандард који примењује Репозиторијум Универзитета у Новом Саду.

3.2.1. Навести метаподатке на основу којих су подаци депоновани у репозиторијум.

Жељана Грбовић (2026), Семантичка сегментација дубоким неуронским мрежама са применом на сликама различитих модалитета у прецизној пољопривреди (необјављена докторска дисертација, Универзитет у Новом Саду)

Ако је потребно, навести методе које се користе за преузимање података, аналитичке и процедуралне информације, њихово кодирање, детаљне описе варијабли, записа итд.

3.3 Стратегија и стандарди за чување података

3.3.1. До ког периода ће подаци бити чувани у репозиторијуму? Неограничено

3.3.2. Да ли ће подаци бити депоновани под шифром? Да **Не**

3.3.3. Да ли ће шифра бити доступна одређеном кругу истраживача? Да **Не**

3.3.4. Да ли се подаци морају уклонити из отвореног приступа после извесног времена?

Да **Не**

Образложити

4. Безбедност података и заштита поверљивих информација

Овај одељак МОРА бити попуњен ако ваши подаци укључују личне податке који се односе на учеснике у истраживању. За друга истраживања треба такође размотрити заштиту и сигурност података.

4.1 Формални стандарди за сигурност информација/података

Истраживачи који спроводе испитивања с људима морају да се придржавају Закона о заштити података о личности (https://www.paragraf.rs/propisi/zakon_o_zastiti_podataka_o_licnosti.html) и одговарајућег институционалног кодекса о академском интегритету.

4.1.2. Да ли је истраживање одобрено од стране етичке комисије? Да **Не**

Ако је одговор Да, навести датум и назив етичке комисије која је одобрила истраживање

4.1.2. Да ли подаци укључују личне податке учесника у истраживању? Да **Не**

Ако је одговор да, наведите на који начин сте осигурали поверљивост и сигурност информација везаних за испитанике:

- а) Подаци нису у отвореном приступу
 - б) Подаци су анонимизирани
 - ц) Остало, навести шта
-

5. Доступност података

5.1. Подаци ће бити

а) **јавно доступни**

б) *доступни само уском кругу истраживача у одређеној научној области*

ц) затворени

Ако су подаци доступни само уском кругу истраживача, навести под којим условима могу да их користе:

Ако су подаци доступни само уском кругу истраживача, навести на који начин могу приступити подацима:

5.4. Навести лиценцу под којом ће прикупљени подаци бити архивирани.

Ауторство – некомерцијално – без прераде.

6. Улоге и одговорност

6.1. Навести име и презиме и мејл адресу власника (аутора) података

Жељана Грбовић, zeljanagrbovic@gmail.com

6.2. Навести име и презиме и мејл адресу особе која одржава матрицу с подацима

6.3. Навести име и презиме и мејл адресу особе која омогућује приступ подацима другим истраживачима
