

Primena savremenih metoda kompjuterske vizije na robotizovani berač šljiva

Jovan Šumarac, Branko Baljić, David Milovanović, Jelena Kljajić

Apstrakt—Ovaj rad predstavlja analizu savremenih metoda robotske vizije, dubokog učenja i detekcije objekata, za primenu na robotizovanom beraču šljiva. U ovom trenutku detekcija šljiva je nedovoljno istražen domen. Izazove predstavljaju njihova relativno mala veličina, gust rast i pristupstvo okluzija. U Srbiji dodatan problem čine mahom neobrezani i nedovoljno strukturirani voćnjaci. Iz tog razloga, analizirali smo stanje tehnike u domenu detekcije voća i performanse nekoliko najzastupljenijih modela. Prvo je sakupljan originalni skup slika domaće šljive. Zatim su slike preprocesirane, augmentirane i propuštene kroz najnovije modele iz porodice YOLO modela. Sprovedena je detaljna analiza dobijenih rezultata. Korišćene su metrike srednje prosečne preciznosti, kao i veličine modela i brzine detekcije.

Ključne reči— Duboko učenje; Detekcija šljiva; Obrada slike; Robotizovani berač voća; Poljoprivreda 4.0.

I. UVOD

U Srbiji je domaća šljiva (*Prunus domestica*) jedna od najrasprostranjenijih voćnih vrsta i smatra se tradicionalnom vrstom. Srbija se po ukupnoj proizvodnji šljive konstantno svrstava među pet najvećih proizvođača u svetu, što šljivu čini jednom od ekonomski najznačajnijih vrsta. [1] Međutim, nepovoljne okolnosti vezane za berbu (sezonski, nisko plaćen rad, često pod teškim vremenskim uslovima) i nedostatak ljudske radne snage doveli su do potrebe da se proces berbe automatizuje.

Sa druge strane, u svetu su sve više u upotrebi robotizovani i automatizovani berači voća. To su mašine koje se koriste za mehaničku berbu plodova voća, umesto tradicionalne ručne berbe. Ove mašine su obično opremljene sa različitim senzorima, kamerama, robotičkim rukama i drugim tehnološkim rešenjima, kako bi mogle precizno i efikasno beru plodove voća. Robotizovani berači voća imaju nekoliko prednosti u odnosu na ručnu berbu. Pre svega, omogućavaju bržu i efikasniju berbu, što može smanjiti radnu snagu potrebnu za berbu voća, kao i smanjiti vreme berbe,

Jovan Šumarac – Institut “Mihajlo Pupin”, Univerzitet u Beogradu, Volgina 15, 11060, Beograd, Srbija (e-mail: jovan.sumarac@pupin.rs), ORCID ID (<https://orcid.org/0000-0002-2924-4813>)

Branko Baljić – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11020, Beograd, Srbija (e-mail: bb223202m@student.etf.bg.ac.rs), ORCID ID (<https://orcid.org/0009-0002-0382-4817>)

David Milovanović – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11020, Beograd, Srbija (e-mail: md223205m@student.etf.bg.ac.rs), ORCID ID (<https://orcid.org/0009-0009-8096-7484>)

Jelena Kljajić – Institut “Mihajlo Pupin”, Univerzitet u Beogradu, Volgina 15, 11060, Beograd, Srbija (e-mail: jelena.kljajic@pupin.rs), ORCID ID (<https://orcid.org/0000-0002-1277-3276>)

odnosno rezultovati povećanom produktivnošću. Ipak ovakvi berači podrazumevaju kompleksne sisteme od kojih se zahteva rad u realnom vremenu i realnim uslovima te je ova tehnologija i dalje u fazi razvoja. Poseban izazov ovakvim sistemima u Srbiji predstavljaju neodovoljno strukturirani voćnjaci i često neobrezane ili nedovoljno obrezane stabljike šljiva.

Sistem robotske vizije je ključan deo robotizovanih berača voća. Sistem vizije koristi kamere, senzore i druge tehnologije za automatsko prepoznavanje i identifikaciju plodova voća na drvetu. Na osnovu tih informacija, robotska ruka mašine može precizno odabrati i brati plodove, bez oštećenja i gubitaka.

Poslednjih godina, sistemi vizije robotizovanih berača intenzivno su prisutni u naučnim istraživanjima, što je prikazano u nedavnim preglednim radovima [2][3]. Sve studije ističu metode dubokog učenja kao dominantan pristup poslednjih godina. Međutim, ove metode takođe imaju tehničke poteškoće koje sprečavaju široku upotrebu robotskih berača u voćnjacima. U [3] autori kao posebne izazove ističu promenljive uslove osvetljenja, prisustvo različitih prepreka, kao i rad u okruženjima ispunjenim okluzijama. U oblasti dubokog učenja korišćeni su različiti modeli za detekciju plodova. Među njima su najzastupljeniji bili modeli detekcije objekata zasnovani na tzv. regionskim konvolucionim mrežama (eng. *Region-based Convolutional Neural Networks* - RCNN) [4], jednostupanjskim detektorima (eng. *Single Shot Detectors* - SSD) [5] i porodici modela YOLO (eng. *You Only Look Once*) [6].

Počev od objavljivanja prve verzije YOLO modela 2015. godine, sistemi vizije i detekcije objekata zasnovani na različitim verzijama YOLO modela primenjeni su u novijim radovima, i bili su posebno prisutni u studijama [2] i [3]. Poboljšane četvrte verzije ovog modela (YOLOv4) su korišćene za detekciju citrusa u voćnjacima [7] i za robusnu detekciju krušaka u realnom vremenu na mobilnim aplikacijama [8]. U [9] autori predlažu poboljšani, lagani model YOLOv5 za robota za berbu koji automatski prepoznaje jabuke koje se mogu uhvatiti i koje se ne mogu uhvatiti na slici drveta jabuke. U [10] autori koriste poboljšanu sedmu verziju YOLO modela uz augmentaciju podataka za detekciju voćaka kivijsa.

Cilj ovog rada bio je da se prikupe podaci za detekciju šljiva vrste *Prunus domestica* u prirodnim okruženjima jer ova vrsta šljive nije zastupljena u postojećim istraživanjima. Zatim su primenjene metode proširivanja skupa podataka. Potom je novonapravljeni skup podataka propušten kroz najnovije modele iz porodice YOLO modela (YOLOv7 i YOLOv8).

Dobijeni rezultati su analizirani i dati su predlozi za budući rad.

II. METODI I MATERIJALI

A. Akvizicija slika

Skup slika za ovu studiju formiran je od originalnih slika dobijenih u neuređenom (nestrukturiranom) šljiviku koji se nalazi u opštini Sopot (južni deo Beograda), Srbija. Analizirano voće pripada vrsti *Prunus domestica L.*, tačnije sortama "Čačanska lepota" i "Čačanska najbolja". Slike stabala šljive dobijene su po sunčanom vremenu, što su tipični vremenski uslovi za berbu šljiva, sredinom jula 2022. godine, u tri iteracije: kasno ujutro, rano popodne i kasno popodne, kako bi se uzeli u obzir različiti uslovi osvetljenja. Većina šljiva bila je u sredini zrelosti u trenutku dobijanja slike.

Tokom dobijanja slika, cilj je bio da se postignu različiti uglovi i udaljenosti snimanja koje bi simulirale stvarnu situaciju. Slike su snimljene Intel® RealSense™ Depth kamerom D405. S obzirom da je optimalni radni opseg ove kamere između 7 cm i 50 cm, snimanje je pratilo te smernice. Vertikalno, kamera je postavljena između 0,5 m i 2 m od zemlje, što će biti slučaj kada se čitav sistem izgradi. Pošto je snimanje obavljeno u prirodnom okruženju, slike često sadrže preklapajuće plodove, blagu i tešku okluziju u obliku grana i lišća, kao i različite intenzitete i uglove svetlosti (direktni i bočni svetlosni uglovi, kao i pozadinski uglovi).

B. Preprocesiranje slika

Ukupno je dobijeno 1079 slika. Od toga, 757 je činilo skup za obuku, 217 skup za validaciju, a 105 skup za testiranje (odnos između ova tri skupa je, redom, 70:20:10). Raspon slika između ova tri skupa bio je slučajan. Kako bi se sprečilo preobučavanje modela, nije bilo ponavljanja slika u ovim skupovima.

Objekti od interesa uključuju 3 klase: dostupne šljive (voće koje je dovoljno blizu robota da ga dohvati, i nema prepreka ispred njega; posebno prepreka ispred gornjeg dela voćke, gde se nalazi drška), nedostupne šljive (voće koje je blizu, ali nedostupno zbog barem jedne prepreke, u obliku grane, velikog lista/listova i/ili druge šljive) i pozadinske (daleke) šljive. Ova treća, dodatna klasa uključuje šljive koje su predaleko u pozadini da bi se mogle pokupiti sa trenutne pozicije robota, ali pošto je naša opšta ideja da unapred mapiramo poziciju voća u voćnjaku, ova klasa će se koristiti za buduće pozicioniranje robota i korake branja. Obeležavanje svih objekata od interesa zasniva se na najmanjem okružujućem pravougaoniku oko vidljivog dela voćke. U ovim slikama, bilo je 4465 primera dostupnih šljiva, 2255 nedostupnih i 4372 pozadinskih šljiva. Zbog prirode naših podataka, nedostupne šljive su malo manje zastupljene, što je uzeto u obzir tokom analize rezultata.

Smanjenje veličine slika bilo je neophodno zbog potencijalnog računarskog opterećenja (što bi produžilo vreme obuke i procesiranja), ali s druge strane, nismo mogli previše smanjiti veličinu slika, pošto one često sadrže mnogo manjih

objekata interesa i značajnih detalja. Kao kompromis između ova dva uslova, sve slike su smanjene na 640x640.

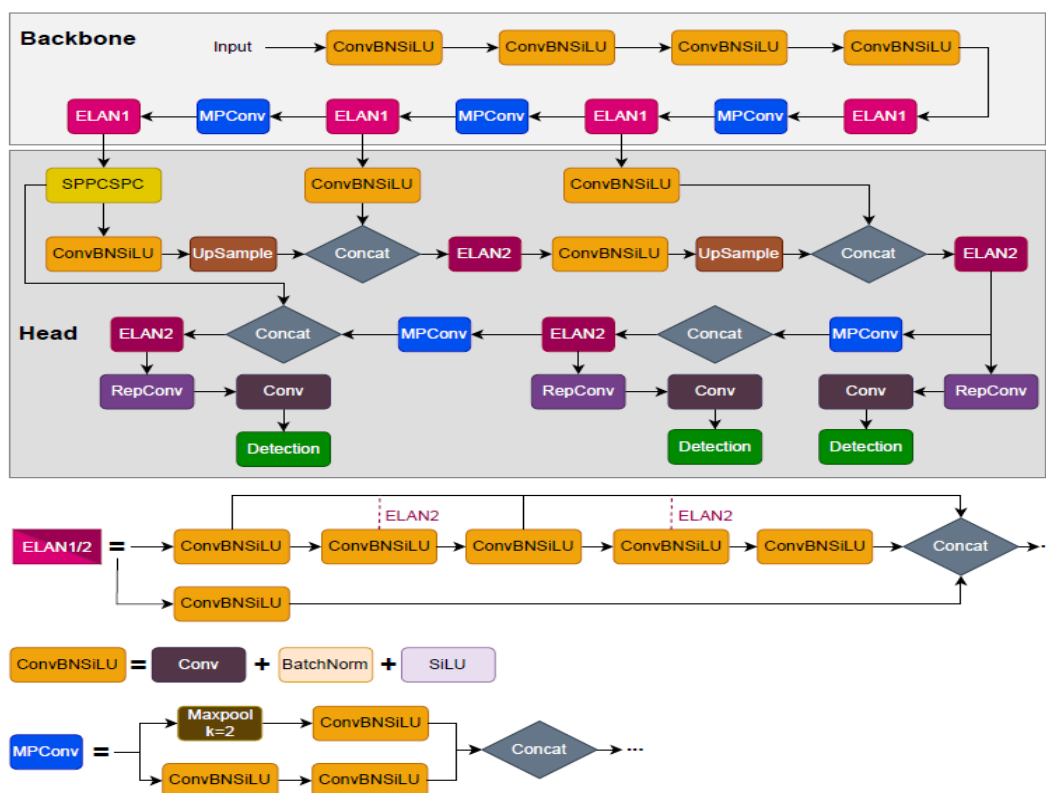
Kako bi se poboljšao kvalitet obuke i tačnost predviđanja, primenjene su sledeće metode uvećanja: horizontalno preklapanje slika, rotacija slike za 90° (u smeru kazaljke na satu i obrnuto) i histogram izjednačavanje. Kao rezultat, ukupno je dobijeno 2685 slika za obuku. Primer anotirane slike dat je u dodatku.

C. Arhitekture modela

Porodica modela YOLO (eng. You Only Look Once) dominira pejzažom detekcije objekata od objavljivanja prvog YOLO modela [6] 2015. godine. YOLO je detektor jednog koraka, što znači da direktno predviđa klasu objekta i koordinate okvira objekata na slici, bez dodatne obrade. Ovo je učinilo YOLO generalno bržim i računarski efikasnijim od mnogih drugih najsavremenijih detektora objekata, ali sa nižom tačnošću. Od tada se YOLO iterativno ažurira, a kasniji modeli poboljšavaju mnoge funkcije, uključujući napredne arhitekture mreže, otkrivanje objekata različitih razmera, kutije sidara (eng. *anchor boxes*) itd. [11]. Kutije sidara su unapred definisani okviri koji se koriste u modelima detekcije objekata i služe kao referenca za mrežu da predvidi lokaciju i veličinu objekata na slici. Pošto se detekcija objekata vrši obeležavanjem granične kutije (eng. *bounding box*) oko pronađenog objekta na slici, proces detekcije zapravo kreće od unapred definisanih kutija sidara koje su određene na osnovu distribucije graničnih kutija na obeleženim slikama iz trening skupa. Lagana priroda i visoka tačnost modela YOLO postavili su standard za najsavremenije metode vizuelne detekcije objekata.

Najnoviji model detekcije objekata iz porodice YOLO u vreme ovog istraživanja - YOLOv7 [12] predstavljen je 2022. godine. U oblasti detekcije objekata u realnom vremenu, on je nadmašio većinu poznatih i korišćenih detektora objekata, kao što su modeli RCNN porodice, kao i drugi modeli iz porodice YOLO. Slika 2 prikazuje arhitekturu YOLOv7 modela. Sastoji se od tri komponente: ulazne slike u željenoj rezoluciji, glavne mreže i glave detekcije. Glavna mreža (tzv. kičma od eng. *backbone network*) izvlači značajne mape obeležja iz slika, dok se sloj glave (eng. *model head*) koristi za predviđanje. Izlaz glavne mreže je mapa obeležja u tri različite veličine. Zatim prolaze kroz finalni sloj (glava detekcije), a na kraju se rezultat predviđanja provlači kroz konvoluciju i vrši se detekcija.

Dizajn arhitekture YOLOv7 zasnovan je na ELAN-u (eng. *efficient layer aggregation network* - efikasna mreža agregacije slojeva). ELAN razmatra dizajn efikasne mreže kontrolisanjem najkraćeg i najdužeg puta gradijenta, tako da dublje mreže mogu konvergirati i učiti efikasno. YOLOv7 modifikuje ELAN arhitekturu, nazvanu E-ELAN (proširena efikasna agregacija slojeva), koja se koristi u dve verzije, ELAN1 i ELAN2 (prikazane na dnu slike 2). Još jedna važna karakteristika arhitekture YOLOv7 je SPPCSPC blok, koji kombinuje modul parcijalne konekcije preko etapa (eng. *Cross Stage Partial Connections* – CSP modul) i modul



Sl. 2. Arhitektura YOLOv7 modela

prostorne piramidalne niskofiltrazne sabirnice (eng. *Spatial Pyramid Pooling* – SPP modul).

U CSP modulu izlazne mape obeležja različitih etapa (ili slojeva) modela se kombinuju na delimičan način. Konkretno, mape obeležja se dele na dva dela, a svaki deo se posebno obrađuje različitim etapama mreže. Izlazi ovih etapa se zatim spajaju i dalje se obrađuju u kasnijim slojevima. Uvođenjem parcijalnih konekcija između različitih etapa mreže, CSPC moduli omogućavaju mreži da razmenjuje informacije na različitim skalama i nivoima apstrakcije. Ovo može pomoći da se poboljša sposobnost mreže da uhvati i niske i visoke nivoe karakteristika ulaznih podataka, što može dovesti do poboljšanja njene ukupne performanse [13].

Osnovna ideja iza SPP modula je da se generišu mape obeležja fiksne veličine, bez obzira na veličinu ulazne slike, čime se omogućava da mreža obrađuje slike različitih veličina i odnosa stranica. SPP modul radi tako što deli ulazne mape obeležja na fiksni broj pravougaonih podregija na više skala. Unutar svake podregije, modul izvršava operaciju agregacije, kao što su maksimalno agregiranje ili prosečno agregiranje, kako bi proizveo vektor fiksne veličine. Izlazni vektori iz svih podregija zatim se spajaju da bi se proizveo konačni SPP izlaz. SPP modul može pomoći u poboljšanju performansi dubokih modela učenja tako što hvata kontekstualne informacije višeskalarno iz ulaznih slika. Agregiranjem obeležja na više skala i lokacija, modul hvata lokalni i globalni kontekst, što je važno za zadatke kao što je detekcija objekata, gde lokacija i veličina objekta mogu varirati unutar ulazne slike. [14]

U januaru 2023. predstavljen je i model YOLOv8. Ono što je karakteristično za ovu verziju jeste da uz nju nije objavljen nikav naučni rad. Ovo je bio slučaj i za često korišćenu verziju YOLOv5. Ovo je jedan od simptoma trenutne situacije u industriji gde komercijalne potrebe diktiraju razvoj novih i boljih modela za detekciju objekata te postaje sve češći slučaj da se modeli objavljuju bez naučnog rada, ili da rad bude objavljen kasnije.

Prema autorima modela YOLOv8 [15] ističe se nekoliko glavnih razlika u odnosu na prethodne modele, posebno v7 i v5. Za početak, YOLOv8 je model bez kutije sidara. To znači da direktno predviđa centar objekta umesto pomeraja od poznate kutije sidara. Kutije sidara su bile problematična stavka u ranijim YOLO modelima, pošto one mogu predstavljati distribuciju kutija ciljane reference, ali ne i distribuciju novog, proizvoljnog skupa podataka. Detekcija bez sidara smanjuje broj predikcija kutija, što ubrzava proces detekcije, i samog obučavanja modela.

Istraživanje dubokog učenja obično se fokusira na arhitekturu modela, ali rutina obuke u YOLOv5 i YOLOv8 modelima je ključni deo njihovog uspeha. YOLOv8 proširuje skup slika tokom obuke u realnom vremenu. U svakoj epohi, model vidi malo drugačiju varijaciju slika koje su mu pružene. Jedna od tih augmentacija zove se mozaik augmentacija. To podrazumeva spajanje četiri slike zajedno, prisiljavajući model da nauči objekte na novim lokacijama, u delimičnoj okluziji i sa različitim okolnim pikselima. Međutim, ova augmentacija je pokazala empirijski dokazan pad performansi ako se izvodi kroz celu rutinu obuke. Prednost je isključiti je

za poslednjih deset epoha obuke. YOLOv8 razlikuje se i u finim razlikama u konvolucionim slojevima u glavnom delu mreže.

III. REZULTATI

U ovom radu predložene metrike za procenu našeg modela su preciznost (P – eng. *precision*), osetljivost (R – eng. *recall*) i srednja prosečna preciznost (mAP – eng. *mean Average Precision*). Preciznost meri tačnost modela u pogledu broja lažno pozitivnih (FP – eng. *False Positive*) predviđanja, odnosno proporciju tačno pozitivnih (TP – eng. *True Positive*) predviđanja (ispravno predviđeni pozitivni primeri) u odnosu na sva pozitivna predviđanja koja je napravio model. Osetljivost meri kompletnost modela u pogledu broja lažno negativnih (FN – eng. *False Negative*) predviđanja, odnosno proporciju tačno pozitivnih predviđanja (ispravno predviđeni pozitivni primeri) u odnosu na sve stvarne pozitivne primere.

Prosečna preciznost (AP – eng. *Average Precision*) se računa kao površina ispod krive preciznost-osetljivost za određenu kategoriju. Srednja prosečna preciznost se potom izračuna kao usrednjena AP vrednost za sve klase objekata. mAP je široko korišćena metrika za detekciju objekata, jer uzima u obzir i preciznost i osetljivost i pruža jedan broj koji sažima ukupne performanse detektora objekata. mAP@0.5 metrika je specifična verzija mAP metrike, gde se prag preseka predviđene granične kutije i stvarne granične kutije objekta (IoU – eng. *Intersection over Union*) postavlja na 0,5. Prag IoU se koristi za odlučivanje da li se detektovana okvirna kutija smatra tačnom ili lažnom pozitivnom. Ako je IoU između predviđene i stvarne okvirne kutije veći od 0,5, detekcija se smatra tačnom pozitivnom. Kompaktan prikaz metrika P, R i AP dat je u (1).

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$AP = \int_0^1 P(R) dR$$
(1)

U ovom radu su eksperimenti sprovedeni na Google Collaboratory platformi. Korišćena verzija Pythona bila je 3.8, Pytorch v 1.13.1, CUDA v 11.6 za obuku modela, a Tesla T4 (16G) je korišćena GPU. Veličina svih slika korišćenih za obuku je 640x640, veličina paketa (eng. *batch-size*) je 16, a model je obučavan 30 epoha. Validacioni skup je korišćen za evaluaciju svih metrika tokom faze obuke. U Tabeli I prikazane su metrike za sve klase šljiva i model YOLOv7, dok je u Tabeli II dato isto za model YOLOv8.

TABELA I
REZULTATI MODELA YOLOv7

Skup	Broj	P	R	mAP@0.5
Sve šljive	1861	0.73	0.69	0.77
Dostupne	822	0.80	0.85	0.91
Nedostupne	623	0.73	0.60	0.72
Pozadinske	416	0.66	0.63	0.67

TABELA II
REZULTATI MODELA YOLOv8

Skup	Broj	P	R	mAP@0.5
Sve šljive	1861	0.69	0.69	0.75
Dostupne	822	0.80	0.85	0.91
Nedostupne	623	0.69	0.59	0.69
Pozadinske	416	0.58	0.63	0.65

Dakle rezultati pokazuju nešto bolje performanse modela YOLOv7. To se ipak najviše vidi u klasama nedostupnih i pozadinskih šljiva, gde su performanse modela YOLOv8 nešto smanjene. Sama mera mAP@0.5 na nivou svih šljiva razlikuje se za 2 procenta u korist YOLOv7 modela.

Međutim pri poređenju modela dubokog učenja moraju se uzeti u obzir još neki parametri. Kako ovo mogu biti poprilično veliki modeli, koji se dugo obučavaju, a ponekad i sporo detektuju, važni parametri su i vreme obučavanja modela, veličina modela (tj. težinskih parametara) kao i brzina detekcije na test skupu. Poređenje ovih parametara dato je u Tabeli III.

TABELA III
POREĐENJE MODELA YOLOv7 I YOLOv8

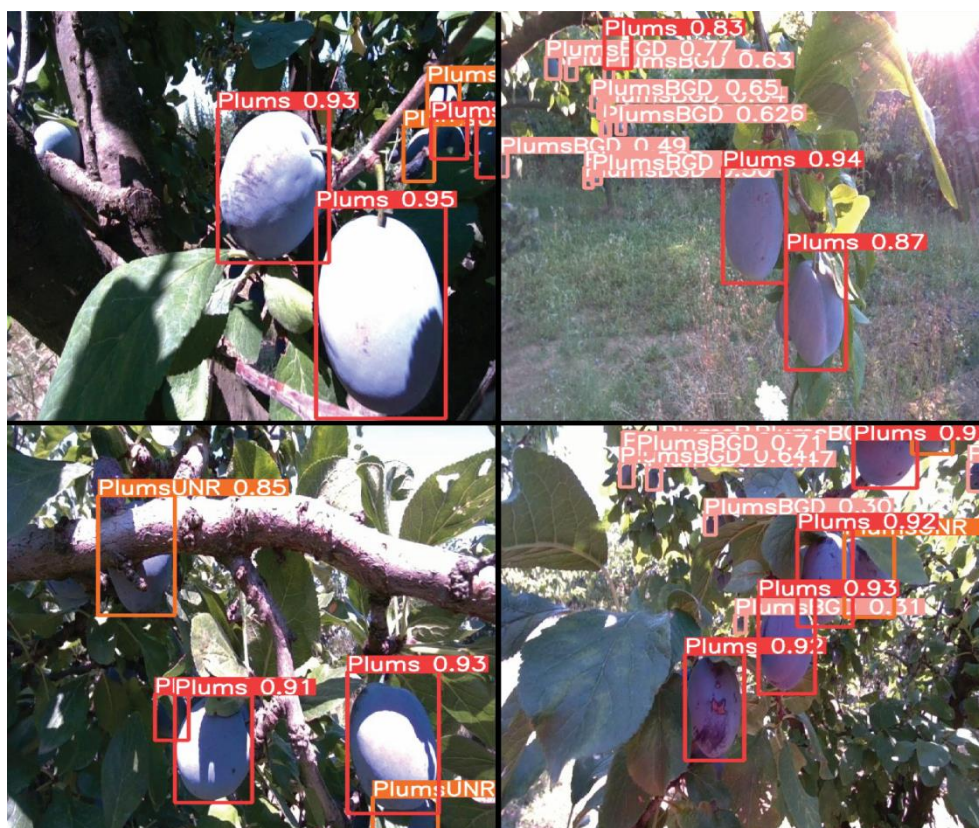
Verzija modela	Vreme obučavanja (min)	Veličina (MB)	Brzina detekcije (ms)	mAP
v7	65.4	75	13.4	0.77
v8	35.3	22	12.8	0.75

Dakle sada se može videti detaljnije poređenje ova dva modela. Iako je YOLOv7 nešto precizniji što se tiče detekcije, model YOLOv8 je приметно manje veličine. Takođe, dok je brzina detekcije na test slikama malo brža kod modela YOLOv8, ono što se značajno razlikuje jeste vreme obučavanja modela. Pri istom broju epoha, na istoj platformi (Google Collaboratory) model YOLOv8 se obučavao 30 minuta manje, odnosno gotovo 85% brže od modela YOLOv7. Ovo je konzistentno sa navedenim prednostima modela YOLOv8.

Na Sl. 3. i 4. mogu se videti primeri detekcije na slikama šljiva. Uzeti su isti ili slični primeri slika kako bi se prikazale suptilne razlike u detekciji modelima YOLOv7 i YOLOv8.



Sl. 3. Primeri detekcije modelom YOLOv7. Crvene kutije su dostupne, zelene nedostupne, a svetlo braon pozadinske šljive



Sl. 4. Primeri detekcije modelom YOLOv8. Crvene kutije su dostupne, narandžaste nedostupne, a svetlo roze pozadinske šljive

Dakle vidi se da YOLOv7 nešto preciznije obeleži šljive, što je posebno jasno na gornje dve slike na Sl. 3. i 4. gde su dostupne šljive jednako označene kod oba modela, ali model YOLOv7 preciznije detektuje ostale klase šljiva.

IV. ZAKLJUČAK

Ciljevi ovog rada bili su: (i) prikupljanje novog skupa podataka za vrstu šljive *Prunus domestica*. Slike su prikupljene u realnom okruženju (neuređeni voćnjaci sa slabo obrezanim stablima) i pod različitim osvetljenjem. (ii) Predstavljena je primena najnovijih modela detekcije objekata iz porodice YOLO modela. Verzije YOLOv7 i YOLOv8 su upoređene i istaknute prednosti svake. (iii) Voće je klasifikovano na način koji odražava potrebe robota za berbu. Iako je to rezultiralo nešto lošijim rezultatima, preciznost, osetljivost i mAP za klasu dostupnih šljiva su 0.8, 0.85 i 0.91 (80%, 85% i 91%), što obećava buduće primene.

Budući rad će se fokusirati na prenošenje i testiranje ovog modela u produkciji, na prototipu robota berača šljiva koji se razvija u Laboratoriji za robotiku Instituta "Mihajlo Pupin". Takođe važan pravac budućeg rada je razvoj preciznijih modela, za detekciju peteljke šljiva, kao i za 3D estimaciju pozicije i orijentacije plodova šljiva.

DODATAK

U nastavku je prikazan primer anotirane slike. Crvene kutije označavaju dostupne, narandžaste nedostupne, a svetlo roze pozadinske šljive.



ZAHVALNICA

Ovaj rad je rezultat istraživanja sprovedenog u okviru projekta "Razvoj demonstracionog prototipa manipulativnog robota za branje jestivih šljiva - dokaz koncepta", ugovor br. 623-1-22, inovacioni vaučer br. 1216 sufinansiran od strane Fonda za inovacionu delatnost R. Srbije, 2022.

LITERATURA

- [1] Tomic J, Štampar F, Glišić I, Jakopić J. Phytochemical assessment of plum (*Prunus domestica* L.) cultivars selected in Serbia. *Food Chem.* 2019 Nov 30;299:125113.
- [2] Anand Koirala, Kerry B. Walsh, Zhenglin Wang, Cheryl McCarthy, Deep learning – Method overview and review of use for fruit detection and yield estimation, *Comp. and El. in Ag.*, Volume 162, 2019, Pages 219-234, ISSN 0168-1699.
- [3] Chiagoziem C. Ukwuoma et al., 2022. "Recent Advancements in Fruit Detection and Classification Using Deep Learning Techniques," *Mat. Prob. in Eng.*, Hindawi, vol. 2022, pages 1-29, January.
- [4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.
- [5] W. Liu et al., "SSD: Single Shot MultiBox Detector," 2015, doi: 10.48550/ARXIV.1512.02325.
- [6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2015. doi: 10.48550/ARXIV.1506.02640
- [7] W. Chen, S. Lu, B. Liu, G. Li, and T. Qian, "Detecting Citrus in Orchard Environment by Using Improved YOLOv4," *Sci. Program.*, vol. 2020, pp. 1–13, Nov. 2020, doi: 10.1155/2020/8859237.
- [8] A. I. B. Parico and T. Ahamed, "Real Time Pear Fruit Detection and Counting Using YOLOv4 Mod - els and Deep SORT," *Sensors*, vol. 21, no. 14, p. 4803, Jul. 2021.
- [9] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, "A Real-Time Apple Targets Detection Method for Pick-ing Robot Based on Improved YOLOv5," *Remote Sens.*, vol. 13, no. 9, p. 1619, Apr. 2021.
- [10] Xia, Y., Nguyen, M., Yan, W.Q. (2023). A Real-Time Kiwifruit Detection Based on Improved YOLOv7. In: Yan, W.Q., Nguyen, M., Stommel, M. (eds) *Image and Vision Computing. IVCNZ 2022. Lecture Notes in Computer Science*, vol 13836. Springer, Cham.
- [11] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," 2018, doi:10.48550/ARXIV.1804.02767.
- [12] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, doi: 10.48550/ARXIV.2207.02696.
- [13] C.-Y. Wang et al., "CSPNet: A New Backbone that can Enhance Learning Capability of CNN," 2019, doi: 10.48550/ARXIV.1911.11929.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," 2014, doi: 10.48550/ARXIV.1406.4729.
- [15] YOLOv8, available at: <https://github.com/ultralytics/ultralytics>

ABSTRACT

Abstract—This paper presents an analysis of contemporary methods of robotic vision, deep learning, and object detection for application on a robotic plum harvester. At this moment, plum detection is an insufficiently explored domain. Challenges are posed by their relatively small size, dense growth, and the presence of occlusions. In Serbia, an additional problem is mostly untrimmed and insufficiently structured orchards. For this reason, we analyzed the state of the art in the domain of fruit detection and the performance of some of the most represented models. First, an original dataset of domestic plum images was collected. Then, the images were preprocessed, augmented, and passed through the latest models from the YOLO family. A detailed analysis of the obtained results was conducted. Mean average precision, as well as model sizes and inference speed, were used as metrics.

Application of contemporary computer vision methods on a robotic plum harvester.

Jovan Šumarac, Branko Baljić, David Milovanović, Jelena Kljajić