UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Luka Kovačič

Zaznavanje in klasifikacija poškodb na cestišču iz slik z globokimi nevronskimi mrežami

MAGISTRSKO DELO

MAGISTRSKI ŠTUDIJSKI PROGRAM DRUGE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA
SMER: RAČUNALNIŠTVO IN INFORMATIKA

Mentor: doc. dr. Andraž Podlesnik

Ljubljana, 2026

To delo je ponujeno pod licenco Creative Commons Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna (CC BY-SA 4.0). To pomeni, da se lahko besedilo, slike, grafi in druge sestavine tega dela prosto delijo, reproducirajo, priobčujejo javnosti in predelujejo za vsak namen, tudi komercialno, če se jasno navede avtor (po možnosti tudi naslov in povezava do izvirnika), priloži povezava do licence, označijo morebitne spremembe, pri predelavah pa se uporabi enaka licenca (CC BY-SA 4.0). Licenca ne dovoljuje dodajanja dodatnih pravnih ali tehničnih omejitev in ne velja za dele, za katere nosilec pravic ni avtor. Podrobnosti licence so dostopne na spletni strani https://creativecommons.org.

© 2026 LUKA KOVAČIČ

Zahvala

Zahvaljujem se mentorju za strokovne nasvete in potrpežljivost pri vodenju dela ter družini za podporo med študijem.

Kazalo vsebine

1 Uvod **1**
1.1 Opredelitev problema 1
1.2 Cilji in raziskovalna vprašanja 2
1.3 Prispevki dela 2
1.4 Metode dela 2
1.5 Struktura naloge 3

2 Globoke nevronske mreže za detekcijo objektov **5**
2.1 Konvolucijske nevronske mreže 5
2.2 Postopki učenja 5
2.3 Dvostopenjski detektorji 6
2.4 Enostopenjski detektorji in družina YOLO 7
2.5 Sidra, izguba in izločanje podvojenih zaznav 8
2.6 Merila uspešnosti 9
2.7 Sinteza 9

3 Pregled sorodnih del pri zaznavanju poškodb cestišča **11**
3.1 Klasične metode in prvi koraki globokega učenja 11
3.2 Segmentacija razpok 11
3.3 Zbirke in detektorji za poškodbe cest 12
3.4 Izziv CRDDC 2022 in njegove rešitve 12
3.5 Umestitev dela 13

4 Zasnova sistema za zaznavanje poškodb **15**
4.1 Zbirka RDD2022 15
4.2 Delitev in priprava podatkov 16
4.3 Izbira modelov in začetnih uteži 17
4.4 Razširjanje podatkov 17
4.5 Protokol učenja 18
4.6 Protokol vrednotenja 18

5 Implementacija in vrednotenje **21**
5.1 Razvojno okolje in potek učenja 21
5.2 Točnost po razredih poškodb (RV1) 22
5.3 Točnost in hitrost (RV2) 24
5.4 Prispevek začetnih uteži in razširjanja podatkov (RV3) 25
5.5 Analiza napak (RV4) 26
5.6 Pomen rezultatov za uporabo 29
5.7 Primerjava z literaturo 30
5.8 Omejitve veljavnosti 30

6 Sklep **33**

6.1 Odgovori na raziskovalna vprašanja....33
6.2 Prispevki in omejitve....33
6.3 Smernice za nadaljnje delo....34

Literatura....35

Priloge....39
Priloga A: Razredi poškodb zbirke RDD2022....39
Priloga B: Hiperparametri učenja....39
Priloga C: Postopek vrednotenja v obliki psevdokode....40
Priloga D: Matrika zmede modela YOLOv8m v številkah....40
Priloga E: Natančnost, priklic in posamezne ponovitve....41
Priloga F: Uporaba orodij umetne inteligence....41

Seznam uporabljenih kratic

AIU	povprečni presek glede na unijo (angl. average intersection over union)
AP	povprečna natančnost (angl. average precision)
BiFPN	uteženo dvosmerno piramido značilk (angl. bidirectional feature pyramid network)
BY-SA	del imena licence CC BY-SA (Creative Commons Priznanje avtorstva-Deljenje pod enakimi pogoji)
CC BY-SA	licenca Creative Commons Priznanje avtorstva-Deljenje pod enakimi pogoji
CNN	konvolucijska nevronska mreža (angl. convolutional neural network)
COCO	nabor Common Objects in Context
CRDDC	izziv Crowdsensing-based Road Damage Detection Challenge
E-ELAN	razširjena učinkovita agregacija plasti (angl. extended efficient layer aggregation network)
FN	spregledani označeni primerek (angl. false negative)
FP	lažni pozitiv (angl. false positive)
FPHBN	mreža Feature Pyramid and Hierarchical Boosting Network
FPN	piramida značilk (angl. feature pyramid network)
GRDDC	izziv Global Road Damage Detection Challenge
ILSVRC	tekmovanje ImageNet Large Scale Visual Recognition Challenge
IoU	delež preseka glede na unijo (angl. intersection over union)
mAP	povprečje AP po razredih (angl. mean average precision)
MDPT	ime ekipe Pham idr. na lestvici izziva CRDDC 2022
NMS	izločanje podvojenih zaznav (angl. non-maximum suppression)
PASCAL	ime izziva PASCAL Visual Object Classes
RDD-YOLO	izboljšana različica modela YOLOv8 za poškodbe vozišča
RDD2022	zbirka Road Damage Dataset 2022
RPN	omrežje za predlaganje območij (angl. region proposal network)
SD	standardni odklon

SGD	stohastični gradientni spust (angl. stochastic gradient descent)
SGG-RS	ime ekipe SGG-RS-Group na lestvici izziva CRDDC 2022
SSD	enostopenjski detektor Single Shot MultiBox Detector
TP	pravilna zaznava (angl. true positive)
VGG	omrežje skupine Visual Geometry Group
VGG-16	različica omrežja VGG s 16 plastmi z utežmi
VOC	izziv PASCAL Visual Object Classes
YOLO	družina detektorjev You Only Look Once
ZDA	Združene države Amerike
ZF	mreža Zeilerja in Fergusa

Povzetek

Cestišča je treba redno pregledovati, slike z vozil pa omogočajo cenejši pregled, če jih zna računalnik razvrstiti. Na javno dostopnem delu zbirke RDD2022 (38 385 slik, štirje razredi poškodb) smo primerjali dvostopenjski detektor Faster R-CNN ter enostopenjska detektorja YOLOv7 in YOLOv8m. Ker označb testnih slik zbirke ni, smo podatke razdelili sami po slikah. Vsak model smo učili trikrat, rezultate pa podali z intervali zaupanja. YOLOv8m doseže mAP@0,5 0,62 in je s 10,9 ms na sliko tudi najhitrejši, YOLOv7 doseže 0,56, Faster R-CNN pa 0,49. Začetne uteži COCO in razširjanje podatkov skupaj prispevata 13,4 odstotne točke. Največ napak so spregledane poškodbe; odvisne so predvsem od velikosti poškodbe in države, manj od razreda.

Ključne besede: zaznavanje poškodb cestišča, globoke nevronske mreže, detekcija objektov, YOLO, RDD2022

Abstract

Road surfaces need regular inspection, and images taken from vehicles make it cheaper if a computer can classify them. On the publicly available part of the RDD2022 dataset (38,385 images, four damage classes) we compared the two-stage detector Faster R-CNN with the one-stage detectors YOLOv7 and YOLOv8m. Because the labels of the dataset's test images are not published, we split the data by image ourselves. Each model was trained three times and the results are given with confidence intervals. YOLOv8m reaches an mAP@0.5 of 0.62 and is also the fastest at 10.9 ms per image, compared with 0.56 for YOLOv7 and 0.49 for Faster R-CNN. COCO initial weights and data augmentation together add 13.4 percentage points. Most errors are missed damages, and they depend mainly on damage size and country, less on the class.

Keywords: road damage detection, deep neural networks, object detection, YOLO, RDD2022

1 Uvod

Slabo stanje cest je po navedbah organizatorjev izziva CRDDC 2022 eden glavnih dejavnikov prometnih nesreč [1, str. 1], zato je pregledovanje vozišč nujno. Mnoge lokalne oblasti zaradi pomanjkanja sredstev pregledov ne morejo opraviti pravočasno, nekatere občine pa stanje vozišč ugotavljajo z dragimi specializiranimi senzorji, ki jih za pregled celotne države ni mogoče uporabiti [1, str. 1]. Cenejšo pot je pokazala skupina Maeda idr. [2]: nabor 9053 slik vozišč, posnetih s pametnim telefonom v vozilu v sedmih japonskih občinah, je omogočil učenje detektorjev poškodb z globokimi nevronskimi mrežami. Nabor so pozneje razširili z drugimi državami in ga leta 2022 objavili kot RDD2022 (angl. *Road Damage Dataset 2022*) [3, str. 1–3].

V tem delu na zbirki RDD2022 primerjamo tri detektorje objektov in preučimo, kaj jim pomaga pri zaznavanju štirih razredov poškodb. Zbirka obsega 47 420 slik iz šestih držav [3, str. 1], označbe pa so objavljene samo za učni del, to je 38 385 slik s 55 007 označenimi primerki [3, str. 3]. Označb testnih slik organizatorji ne objavijo; rešitve se na njih vrednotijo prek spletnega strežnika izziva [3, str. 12]. Za nalogo to pomeni dvoje. Rezultatov ne moremo neposredno primerjati z lestvico izziva CRDDC 2022 (angl. *Crowdsensing-based Road Damage Detection Challenge*), ki meri F1 na nedostopnih testnih slikah [1, str. 4]. Poleg tega moramo učni del zbirke sami razdeliti na učno, validacijsko in testno množico, kar podrobno opišemo v razdelku 4.2.

1.1 Opredelitev problema

Detektor poškodb mora na sliki najti vsako poškodbo, jo zamejiti s pravokotnikom in ji določiti vrsto. Zbirka RDD2022 razlikuje vzdolžne razpoke (D00), prečne razpoke (D10), mrežaste razpoke (D20) in udarne jame (D40) [3, str. 2]. Težavo povzroča raznolikost zbirke: slike so posnete s pametnimi telefoni, kamerami na posebnem vozilu, z ulične panorame ponudnika Google Street View in z brezpilotnika, v ločljivostih od 512 × 512 do 3650 × 2044 točk [3, str. 7]. Pham idr. [4, str. 2] so ugotovili, da se modeli, naučeni na podatkih ene države, v drugih državah ne obnesejo dobro, zato je pri vrednotenju pomembno ločiti rezultate po podmnožicah zbirke.

Odprta ostajajo tri vprašanja, ki jih delo obravnava. Prvo je, katera družina detektorjev pri enakih podatkih in enakem protokolu da najboljše razmerje med točnostjo in hitrostjo. Drugo je, koliko k točnosti prispevata predhodno učenje na splošnem naboru slik in razširjanje učnih podatkov. Tretje je, kje sistem dejansko dela napake: pri katerih razredih, katerih podmnožicah in velikostih poškodb ter kakšne vrste so te napake.

1.2 Cilji in raziskovalna vprašanja

Cilj dela je ovrednotiti, kako dobro in kako hitro tri tipične rešitve zaznavajo in razvrščajo poškodbe vozišča na zbirki RDD2022 ter kaj o rezultatu odloča. Odgovoriti želimo na štiri raziskovalna vprašanja:

- **RV1:** Kakšno točnost, merjeno s povprečno natančnostjo pri pragu prekrivanja 0,5 (mAP@0,5) in z mero F1 po razredih, dosežejo detektorji Faster R-CNN, YOLOv7 in YOLOv8m pri razredih D00, D10, D20 in D40 na testni množici?
- **RV2:** Kakšno je razmerje med točnostjo in hitrostjo teh detektorjev na isti strojni opremi in kateri ponuja najboljše razmerje za obdelavo videa v vozilu?
- **RV3:** Koliko k točnosti detektorja YOLOv8m prispevata začetne uteži, naučene na naboru COCO, in razširjanje učnih podatkov?
- **RV4:** Pri katerih razredih, podmnožicah zbirke in velikostih poškodb je napak največ in katere vrste napak prevladujejo?

1.3 Prispevki dela

Delo prinaša štiri prispevke. Najprej opisuje ponovljiv protokol delitve, učenja in vrednotenja na zbirki RDD2022, ki ga je mogoče izvesti brez dostopa do skritih testnih označb. Nato primerja tri detektorje pri enakih podatkih in istem postopku izbire praga zaupanja, rezultate pa podaja kot povprečje treh ponovitev z različnimi semeni, skupaj z intervali zaupanja razlik, izračunanimi z vzorčenjem po slikah. Tretji prispevek je ablacijska študija, ki loči vpliv začetnih uteži od vpliva razširjanja podatkov. Četrti je analiza napak po razredih, državah in velikostih poškodb z razčlenitvijo na spregledane primerke, napačno lokalizirane zaznave, zamenjave razredov in lažne alarme na ozadju.

1.4 Metode dela

Teoretični del (poglavji 2 in 3) temelji na pregledu literature o konvolucijskih mrežah, detektorjih objektov in zaznavanju poškodb vozišča. Eksperimentalni del (poglavji 4 in 5) uporablja kontrolirano primerjavo modelov: vsi modeli dobijo iste podatke, isto delitev in isti postopek izbire praga zaupanja. Točnost merimo s povprečno natančnostjo (angl. *average precision*, AP) pri pragu prekrivanja 0,5, hitrost s časom obdelave ene slike, razlike med modeli pa s parjenim vzorčenjem po testnih slikah (angl. *bootstrap*).

1.5 Struktura naloge

V poglavju 2 predstavimo teoretične osnove: konvolucijske mreže, postopke učenja, dvostopenjske in enostopenjske detektorje ter merila uspešnosti. Poglavje 3 pregleda zaznavanje poškodb vozišča od klasičnih metod do rešitev izziva CRDDC 2022 in opredeli vrzel, ki jo delo zapolnjuje. Poglavje 4 opisuje zbirko, delitev podatkov, modele in protokol vrednotenja. V poglavju 5 poročamo o rezultatih in odgovorimo na raziskovalna vprašanja, v poglavju 6 pa povzamemo ugotovitve, omejitve in smernice za nadaljnje delo. Priloge vsebujejo definicije razredov, hiperparametre, postopek vrednotenja v obliki psevdokode, matriko zmede v številkah in navedbo uporabe orodij umetne inteligence.

2 Globoke nevronske mreže za detekcijo objektov

To poglavje obravnava tiste osnove, na katerih temeljijo odločitve v poglavjih 4 in 5. Najprej predstavimo konvolucijske mreže in postopke, s katerimi jih učimo, nato dvostopenjske in enostopenjske detektorje objektov, na koncu pa merila, s katerimi uspešnost detektorjev primerjamo. Poglavje se konča s sintezo, ki iz teorije izpelje izbiro modelov in poskusov.

2.1 Konvolucijske nevronske mreže

Globoko učenje (angl. *deep learning*) uporablja računske modele z več plastmi obdelave, ki se naučijo predstavitev podatkov na več ravneh abstrakcije [5]. Na slikah se je kot najuspešnejša izkazala konvolucijska nevronska mreža (angl. *convolutional neural network*, CNN), katere plasti z drsečimi filtri obdelujejo lokalna območja slike. Prelomni rezultat so objavili Krizhevsky idr. [6]: mreža s petimi konvolucijskimi in tremi polno povezanimi plastmi ter 60 milijoni parametrov je na tekmovanju ILSVRC 2012 (angl. *ImageNet Large Scale Visual Recognition Challenge*) dosegla delež napačnih petih najverjetnejših razredov 15,3 %, drugouvrščena rešitev pa 26,2 %. Avtorji so učenje pospešili z izvedbo konvolucije na grafičnem procesorju in nenasičenimi aktivacijskimi funkcijami, prekomerno prilagajanje polno povezanih plasti pa so omilili z metodo izpuščanja nevronov (angl. *dropout*).

Naslednji korak naprej je bila globina mreže. Simonyan in Zisserman [7] sta z zaporedji majhnih filtrov velikosti 3×3 pokazala, da se točnost pri razvrščanju slik bistveno izboljša, če globino povečamo na 16 do 19 plasti z utežmi. Zelo globoke mreže pa je težko učiti, zato so He idr. [8] plasti zasnovali kot učenje ostankov glede na vhod plasti; tako so preizkusili mreže do 152 plasti, osemkrat globlje od mrež VGG, ansambel takih mrež pa je s 3,57 % napake zmagal na tekmovanju ILSVRC 2015. Za detekcijo poškodb je tak razvoj pomemben, ker detektorji kot hrbtenico (angl. *backbone*) pogosto uporabljajo mrežo, predhodno naučeno na razvrščanju slik; Faster R-CNN iz knjižnice torchvision, ki ga uporabimo v poglavju 5, ima na primer hrbtenico ResNet-50 [9].

Pred globokim učenjem so v detekciji prevladovali ročno zasnovani modeli. Felzenszwalb idr. [10] so predstavili mešanice večskalnih deformabilnih modelov delov, naučenih z latentnim strojem podpornih vektorjev, ki so v času objave dosegali vodilne rezultate na tekmovanjih PASCAL.

2.2 Postopki učenja

Učenje detektorja poteka z optimizacijo funkcije izgube nad velikim številom parametrov. Kingma in Ba [11] sta predstavila optimizator Adam, ki za vsak parameter posebej prilagaja

korak učenja s pomočjo ocen prvega in drugega momenta gradientov; kot privzete vrednosti navajata začetni korak 0,001 ter razpadni stopnji 0,9 in 0,999. Ioffe in Szegedy [12] sta predlagala normalizacijo vhodov plasti po paketih (angl. *batch normalization*), ki omogoča višje stopnje učenja in je pri razvrščanju slik dosegla enako točnost s štirinajstkrat manj koraki učenja. Avtorja obeh postopkov poročata o razvrščanju slik in drugih splošnih nalogah, ne o detekciji poškodb vozišča, zato njuno učinkovitost za našo nalogo preverjamo sami.

Dve skupini postopkov nadomeščata pomanjkanje označenih podatkov. Prenos učenja (angl. *transfer learning*) pomeni, da model ali njegov del, naučen v eni domeni, uporabimo kot izhodišče v drugi, kjer je označenih podatkov premalo [13]. Pri detekciji objektov se kot izhodišče običajno uporabijo uteži, naučene na naboru Microsoft COCO (angl. *Common Objects in Context*), ki vsebuje 2,5 milijona označenih primerkov na 328 tisoč slikah [14]. Razširjanje podatkov (angl. *data augmentation*) pa učne podatke povečuje z geometrijskimi in barvnimi preoblikovanji, mešanjem slik ali naključnim brisanjem delov slike in je po pregledu Shorten in Khoshgoftaar [15] rešitev na ravni podatkov, kadar je označenih primerov malo. Pregleda opisujeta oba postopka kot splošna; kako močno vplivata na detekcijo poškodb vozišča, pa je empirično vprašanje, ki ga v razdelku 5.4 obravnavamo z ablacijo.

2.3 Dvostopenjski detektorji

Detekcija objektov mora poleg razreda določiti tudi lego predmeta. Dvostopenjski detektorji to rešijo v dveh korakih: najprej predlagajo območja, kjer bi lahko bil objekt, nato vsako območje razvrstijo in popravijo njegove meje. Girshick [16] je v mreži Fast R-CNN območja zanimanja z združevanjem (angl. *RoI pooling*) preslikal v značilke stalne velikosti ter razvrščanje in popravek mej učil z eno skupno funkcijo izgube; učenje mreže VGG16 je bilo devetkrat hitrejše, testiranje pa 213-krat hitrejše kot pri izvirni metodi R-CNN. Predlaganje območij je ostalo ozko grlo, zato so Ren idr. [17] dodali omrežje za predlaganje območij (angl. *region proposal network*, RPN), ki z detekcijsko mrežo deli značilke celotne slike; z mrežo VGG-16 so dosegli hitrost 5 slik na sekundo, pri čemer je vključen celoten postopek.

Da bi detektor zaznal tudi majhne objekte, so Lin idr. [18] oblikovali piramido značilk (angl. *feature pyramid network*, FPN), ki z naknadno potjo od zgoraj navzdol in stranskimi povezavami gradi semantično bogate zemljevide značilk na vseh merilih. Takšen vrat (angl. *neck*) ima tudi različica modela Faster R-CNN z imenom ResNet-50-FPN v knjižnici torchvision [9], pri modelu EfficientDet pa ga nadomešča dvosmerna oblika [19].

2.4 Enostopenjski detektorji in družina YOLO

Enostopenjski detektorji območja ne predlagajo, ampak iz slike neposredno napovedujejo pravokotnike in razrede. Redmon idr. [20] so z mrežo YOLO (angl. *you only look once*) sliko razdelili na mrežo celic $S \times S$; celica, v kateri je središče objekta, napove pravokotnike in zaupanje v njih. Mreža ima 24 konvolucijskih in dve polno povezani plasti, osnovna različica pa obdeluje 45 slik na sekundo. Liu idr. [21] so v mreži SSD (angl. *single shot multibox detector*) uvedli privzete pravokotnike različnih razmerij in meril na več zemljevidih značilk, kar zmanjša težave z objekti različnih velikosti; na zbirki PASCAL VOC 2007 so z vhodom 300×300 dosegli 74,3 % mAP pri 59 slikah na sekundo na grafični kartici Nvidia Titan X.

Razvoj družine YOLO je nato prinesel sidrne pravokotnike in učenje na več velikostih slik [22]. Tan idr. [19] so za detektor EfficientDet predlagali uteženo dvosmerno piramido značilk (BiFPN) in sestavljeno skaliranje ločljivosti, globine in širine celotnega detektorja, kot hrbtenico pa uporabili mrežo EfficientNet. Wang idr. [23] so z modelom YOLOv7 predstavili učljive trike, ki točnost povečajo brez dodatnega stroška pri sklepanju (angl. *trainable bag-of-freebies*), med njimi prilagajanje strukture modela po učenju (angl. *re-parameterization*) in razširjeno učinkovito agregacijo plasti E-ELAN; model YOLOv7 s 36,9 milijona parametrov in 104,7 milijarde operacij na sliko pri vhodu 640×640 doseže 51,2 % AP na naboru COCO. Avtorji trdijo, da njihov model v območju od 5 do 160 slik na sekundo presega vse takrat znane detektorje po hitrosti in točnosti [23]; to trditev avtorjev v tem delu preverjamo samo za poškodbe vozišča.

Za YOLOv8 nismo našli znanstvenega članka, zato se opiramo na dokumentacijo izdelovalca: model uporablja glavo brez sidrnih pravokotnikov, različica YOLOv8m ima 25,9 milijona parametrov in 78,9 milijarde operacij na sliko ter na naboru COCO doseže 50,2 mAP pri pragovih prekrivanja od 0,5 do 0,95 [24]. Izdelovalec je medtem objavil tudi novejše različice (YOLO11, YOLO12 in YOLO26) [24]; za našo primerjavo smo izbrali YOLOv8m, ker je izhodišče nedavne izboljšane različice RDD-YOLO za poškodbe vozišča (razdelek 3.4), in jo implementirali s knjižnico Ultralytics [25].

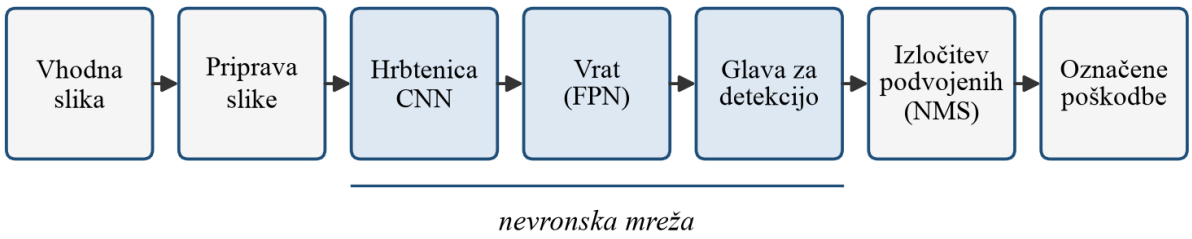
V tabeli 1 povzemamo značilnosti obravnavanih detektorjev. Podatki o točnosti in hitrosti iz različnih člankov niso primerljivi, ker so izmerjeni na različnih naborih podatkov, ločljivostih in grafičnih procesorjih; zato jih v tabelo ne vključujemo, hitrost in točnost pa v poglavju 5 sami izmerimo pod enakimi pogoji.

Detektor	Vrsta	Hrbtenica v izvirnem delu	Posebnost	Leto
Faster R-CNN	dvostopenjski	VGG-16 ali ZF	omrežje RPN deli značilke z detekcijsko mrežo	2017
SSD	enostopenjski	VGG-16	privzeti pravokotniki na več zemljevidih značilk	2016
YOLO	enostopenjski	24 konvolucijskih plasti	mreža celic, napoved pravokotnikov po celicah	2016
EfficientDet	enostopenjski	EfficientNet	BiFPN in sestavljeno skaliranje	2020
YOLOv7	enostopenjski	mreža z E-ELAN	učljivi triki, prilagoditev strukture po učenju	2023
YOLOv8	enostopenjski	po dokumentaciji sodobna hrbtenica in vrat	glava brez sidrnih pravokotnikov	2023

Tabela 1: Značilnosti izbranih detektorjev objektov

Vir: povzeto po [17, 19–21, 23, 24]

Postopek, ki ga ponazarja slika 1, je skupen vsem zgoraj opisanim detektorjem, čeprav se razlikujejo v podrobnostih posameznih korakov.



Slika 1: Zgradba tipičnega detektorja objektov

Vir: lastna ponazoritev

2.5 Sidra, izguba in izločanje podvojenih zaznav

Detektorji pravokotnike napovedujejo glede na referenčne pravokotnike. Ren idr. Ren idr. [17] so uvedli sidrne pravokotnike (angl. *anchor boxes*), ki so reference na več merilih in razmerjih stranic in se izognejo naštevanju slik ali filtrov različnih meril; v modelu SSD isti namen opravljajo privzeti pravokotniki [21]. Model za vsak referenčni pravokotnik napove popravek meja in zaupanje v razred. Pri učenju mreža Fast R-CNN optimizira eno skupno izgubo za razvrščanje in popravek mej [16], SSD pa utežno vsoto izgube lokalizacije in izgube zaupanja [21]. Redmon idr. Redmon idr. [20] v mreži YOLO optimizirajo vsoto kvadratov napak, za katero sami zapišejo, da ni povsem usklajena s ciljem čim višje povprečne natančnosti. Model YOLOv8, ki ga uporabimo v poglavju 5, sider ne uporablja [24].

Ker več referenčnih pravokotnikov lahko pokrije isti objekt, se zaznave pred izhodom očistijo z izločanjem podvojenih zaznav (angl. *non-maximum suppression*, NMS). Ren idr. Ren idr. [17] ga uporabijo za predloge območij s pragom IoU 0,7, SSD ga navaja kot zadnji korak

postopka [21], pri YOLO pa po poročilu avtorjev doda 2 do 3 odstotke mAP [20]. V knjižnici Ultralytics je prag IoU pri NMS nastavljiv parameter [26]. Posebej za zbirko RDD2022 Pham idr. Pham idr. [4, str. 6] opozarjajo, da označbe niso povsem točne, saj so lahko prečne razpoke zaradi perspektive označene kot vzdolžne, in zato uporabijo glajenje oznak (angl. *label smoothing*), ki zaupanje v pravilen razred omeji pod vrednost 1. V našem delu glajenja oznak ne uporabimo.

2.6 Merila uspešnosti

Zaznava je pravilna, če se njen pravokotnik dovolj prekriva z označenim in ima isti razred. Prekrivanje merimo z deležem preseka glede na unijo (angl. *intersection over union*, IoU): $IoU = |A \cap B| / |A \cup B|$, kjer sta A in B ploščini primerjanih pravokotnikov. Izziv CRDDC 2022 šteje zaznavo za pravilno, če je prekrivanje večje od 0,5 in razred ustreza označenemu [1, str. 4]; enako merilo uporabimo mi. Izziv ga je prevzel po vzoru izziva PASCAL VOC [1, str. 2].

Iz števila pravilnih zaznav (TP), napačnih zaznav (FP) in spregledanih označenih primerkov (FN) sledita natančnost $P = TP / (TP + FP)$ in priklic $R = TP / (TP + FN)$, njuno harmonično sredino pa imenujemo $F1 = 2PR / (P + R)$ [1, str. 4]. F1 je odvisna od izbranega praga zaupanja, s katerim zavržemo negotove zaznave. Povprečna natančnost (AP) je zato mera, ki prag zaobide: določimo jo kot ploščino pod krivuljo natančnosti glede na priklic, ko prag zaupanja spreminjamo, mAP pa je povprečje AP po razredih. Pham idr. [4, str. 5] navajajo običajno prakso, da se za izbiro najboljšega modela uporabi mAP pri pragu prekrivanja 0,5 (mAP@0,5) na validacijski množici, za poročanje na testni množici pa F1; v tem delu poročamo o obeh.

2.7 Sinteza

Iz pregleda sledi, da se obravnavani detektorji razlikujejo v treh značilnostih, ki so pomembne za poškodbe vozišča. Prva je število korakov: dvostopenjski detektorji po Tanu idr. pogosto veljajo za prilagodljivejše in natančnejše, enostopenjski pa za preprostejše in učinkovitejše [19]. Druga je način napovedovanja pravokotnikov, ki je pri YOLOv7 sidrni, pri YOLOv8 pa po navedbah izdelovalca brez sider [24]. Tretja so postopki učenja, ki jih lahko uporabimo ali izpustimo: začetne uteži z naborom COCO in razširjanje podatkov [14, 15].

Iz tega izhaja izbira poskusov: dvostopenjski detektor z vratom FPN (Faster R-CNN) primerjamo z enostopenjskima YOLOv7 in YOLOv8m (RV1, RV2), samo za YOLOv8m izvedemo ablacijo z začetnimi utežmi in razširjanjem (RV3), napake pa razčlenimo po razredih, državah in velikostih (RV4).

3 Pregled sorodnih del pri zaznavanju poškodb cestišča

Poglavje sledi razvoju področja od ročno zasnovanih postopkov do detektorjev, naučenih na zbirki RDD2022, in na koncu opredeli, kaj sorodna dela pustijo odprto. Pri vsakem delu navajamo, na katerih podatkih je preizkušeno, saj se razpoke na betonu, zidovih in cestah v slikah močno razlikujejo.

3.1 Klasične metode in prvi koraki globokega učenja

Shi idr. [27] so za zaznavanje razpok na cestah predlagali okvir CrackForest, ki temelji na naključnih strukturiranih gozdovih in celostnih kanalih značilk; po navedbah avtorjev je presegel konkurenčne metode in obravnava težave, kot sta neenakomerna intenziteta vzdolž razpoke in šum s podobno teksturo. Cha idr. [28] navajajo, da spremenljive razmere v resničnem svetu, na primer svetloba in sence, otežujejo širšo uporabo postopkov obdelave slik pri iskanju poškodb gradbenih konstrukcij. Zato so razpoke v betonu zaznavali s CNN, naučeno na 40 tisoč slikah velikosti 256 × 256 točk, in s tehniko drsečega okna obdelali tudi veliko večje slike; po njihovem poročilu je mreža dosegla okoli 98 % točnost, na 55 slikah druge konstrukcije v različnih razmerah pa je bila boljša od zaznavanja robov po Cannyju in Sobelu. Delo se nanaša na beton in ne na cestišče, kaže pa, da se mreža lahko nauči značilk razpok brez ročnega načrtovanja.

3.2 Segmentacija razpok

Vzporedno se je razvila segmentacija, ki vsaki točki slike določi razred. Zou idr. [29] so mrežo DeepCrack zgradili na kodirniško-dekodirniški arhitekturi SegNet in združili značilke več meril; mrežo so naučili na eni zbirki razpok, ovrednotili pa na treh drugih, kjer je povprečna mera F presegla 0,87. Liu idr. [30] so z istim imenom opisali drugačno hierarhično arhitekturo za segmentacijo razpok. Yang idr. [31] so z mrežo FPHBN združili piramido značilk in hierarhično uteženе vzorce, prav tako za razpoke na vozišču, in predlagali merilo AIU, rezultate pa preverili na petih zbirkah.

Segmentacija je smiselna, kadar sta pomembni oblika in debelina razpoke, a zahteva označbe po točkah. Dais idr. [32] so na zidanih površinah ob uporabi prenosa učenja dosegli 95,3 % točnost pri klasifikaciji izrezov in 79,6 % F1 na ravni točk. Tabernik idr. [33] so za površinske napake v industriji predlagali segmentacijsko arhitekturo, ki se je naučila iz le približno 25 do 30 slik z napakami, Božič idr. [34] pa so v istem področju pokazali, da so lahko označbe manj podrobne, če jih kombiniramo s polnimi. Ti deli se nanašajo na industrijske površine in ne na

ceste, zato jih obravnavamo kot dokaz, da je ceno označevanja mogoče znižati, ne pa kot primerjavo za našo nalogo. Zbirka RDD2022 ima označbe v obliki pravokotnikov [3, str. 10], zato je naravna oblika naloge detekcija.

3.3 Zbirke in detektorji za poškodbe cest

Maeda idr. [2] so za cestne poškodbe pripravili prvo večjo zbirko s 15 435 označenimi primerki na 9053 slikah, posnetih s pametnim telefonom v vozilu, in primerjali točnost in hitrost detektorjev na strežniku z grafičnim procesorjem in na telefonu. Zbirka se je razvijala v več korakih: RDD2018 z osmimi razredi poškodb, nato RDD2019, razširjena z umetno ustvarjenimi slikami, in RDD2020 z dodanima državama Indijo in Češko [3, str. 1–2]. Avtorji zbirke poročajo, da so se modeli, naučeni z zbirko RDD2019, na cestah zunaj Japonske bistveno poslabšali [3, str. 1], in sklepajo, da dodajanje podatkov iz drugih držav izboljša posplošljivost modelov [3, str. 3]. Izziv GRDDC 2020 s 121 sodelujočimi ekipami je dobil najboljši rezultat F1 0,67 z ansamblom na osnovi YOLOv5 [3, str. 2].

3.4 Izziv CRDDC 2022 in njegove rešitve

Zbirka RDD2022 poleg držav iz prejšnjih različic vsebuje še Norveško, ZDA in Kitajsko in je bila objavljena za izziv CRDDC 2022 [3, str. 1–2]. Izziv je rešitve vrednotil z merilom F1 pri IoU nad 0,5 na skritih testnih slikah, ločeno za pet lestvic: vse države skupaj, Indijo, Japonsko, Norveško in ZDA [1, str. 4]. Prvih pet ekip je večinoma z ansambli modelov in številnimi postopki razširjanja podatkov doseglo povprečen F1 nad 65 % [1, str. 4]; podatke o njihovih rešitvah in rezultatih na skupni lestvici povzema tabela 2; ekipa MDPT je skupina avtorjev dela [4].

Mesto	Ekipa	Rešitev	F1
1	ShiYu SeaView	ansambel modelov YOLO in dvostopenjskih detektorjev (Cascade R-CNN z hrbtenico Swin)	0,770
2	DongjunJeong	ansambel YOLOv5x (P5 in P6) z rezanjem slik na izreze	0,743
3	MDPT	YOLOv7 s koordinatno pozornostjo in glajenjem oznak	0,741
4	SGG-RS-Group	ansambel YOLOv5 z moduli pozornosti	0,727
5	IRCV-URV	ansambel na osnovi YOLOv7	0,726

Tabela 2: Prvih pet ekip izziva CRDDC 2022 na skupni lestvici (F1 na skritih testnih slikah vseh šestih držav)

Vir: povzeto po [1, str. 5–6]

Zmagovalna ekipa je združila enostopenjske detektorje YOLOv5 in YOLOv7 z dvostopenjskimi Cascade R-CNN; po analizi najuspešnejših rešitev izziva GRDDC 2020 je sklepala, da modeli YOLO dajejo višji priklic, modeli Cascade R-CNN pa višjo natančnost in

so boljši pri majhnih poškodbah [1, str. 6]. To je sklep ekipe, ki ga naša primerjava ne more potrditi ali ovreči, ker je ne izvajamo z ansamblom, a nakazuje, da je smiselno primerjati obe družini. Rezultati zmagovalne ekipe so po državah zelo različni: na lestvici za ZDA 0,844 in za Japonsko 0,789, za Indijo 0,583 in za Norveško 0,595 [1, str. 5]. Razlike med državami so torej večje kot razlike med ekipami, kar potrjuje pomen razčlenitve po podmnožicah.

Pham idr. [4] so z modeli YOLOv7, dopolnjenimi s koordinatno pozornostjo (angl. *coordinate attention*), glajenjem oznak, prilagojenimi parametri razširjanja in ansamblom, dosegli F1 74,1 % na vseh testnih slikah in 81,7 % na slikah iz ZDA. Za slike z Norveške, ki so velike (do 4040 × 2044 točk) in imajo majhne poškodbe, so ugotovili težavo pri zmanjšanju na 640 × 640 točk, zato so jih obrezali na kvadrat 1824 × 1824 točk in odstranili označbe zunaj njega [4, str. 3–4]. Zbirko so za izbor modelov delili po državah v razmerju 90 % proti 10 % in se pri izbiri najboljšega modela opirali na mAP@0,5 na validacijski množici [4, str. 4–5]; podoben postopek uporabimo v naslednjem poglavju.

Li idr. [35] so model YOLOv8 spremenili z mehanizmom pozornosti SimAM, zamenjavo konvolucij z GhostConv in boljšim povečevanjem ločljivosti v vratu ter po svojem poročilu dosegli mAP@0,5 62,5 % na validacijski množici zbirke RDD2022, 2,5 odstotne točke več kot izhodiščni model, in F1 69,6 % na testni množici. Povzemamo jih le po izvlečku članka, zato ne vemo, katero podmnožico zbirke in kakšno delitev so uporabili; njihovih števil zato ne moremo primerjati z našimi kot enakovrednih.

3.5 Umestitev dela

Pregled pokaže tri značilnosti sorodnih del. Prvič, rešitve izziva CRDDC so ansambli z več modeli, ovrednoteni z enim samim merilom F1 na skritih slikah, zato iz njih ne moremo razbrati, koliko k rezultatu prispeva posamezen sestavni del, niti kakšna je hitrost sklepanja. Drugič, rezultati so zelo odvisni od države, zato je razčlenitev po podmnožicah zbirke nujna. Tretjič, v pregledanih delih nismo našli primerjave dvostopenjskega in dveh enostopenjskih detektorjev ob istem protokolu skupaj z ablacijo začetnih uteži in razširjanja podatkov.

To vrzel zapolnjuje naše delo. Uporabimo en protokol za tri modele, merimo točnost po razredih, hitrost in napake po državah ter velikostih, ablacijo pa izvedemo samo za enega modela, da ostane poskus obvladljiv. Rezultatov ne primerjamo z lestvicami izziva, ker označb njegovih testnih slik ni; zato ne trdimo, da dosegamo ali presegamo katero koli objavljeno vrednost.

4 Zasnova sistema za zaznavanje poškodb

Poglavje predstavi podatke, modele in merila. Vse odločitve so zasnovane tako, da jih je mogoče ponoviti brez dostopa do skritih testnih označb zbirke, in so utemeljene z literaturo iz poglavij 2 in 3 ali s praktično omejitvijo, ki jo navedemo.

4.1 Zbirka RDD2022

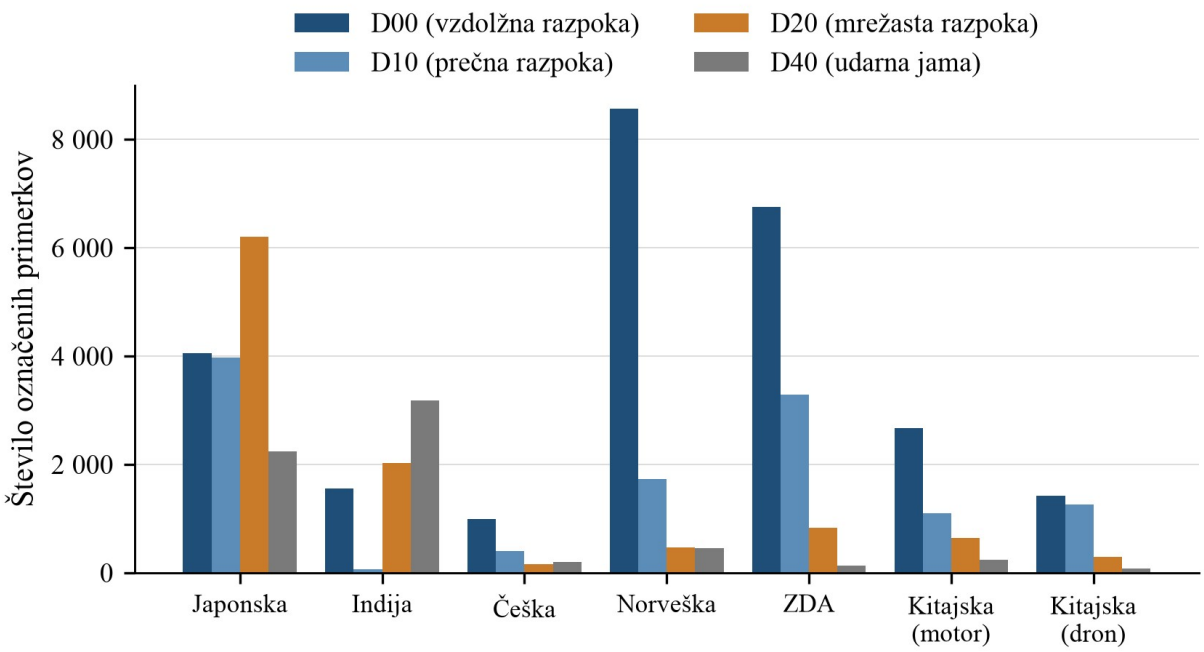
Zbirka RDD2022 vsebuje slike cest iz Japonske, Indije, Češke, Norveške, ZDA in Kitajske, označene s pravokotniki v štirih razredih: vzdolžna razpoka (D00), prečna razpoka (D10), mrežasta razpoka (D20) in udarna jama (D40) [3, str. 1–2]. Definicije razredov so v prilogi A (tabela 12). Razred D20 pri avtorjih zbirke imenujejo aligatorska razpoka (angl. *alligator crack*); v nalogi uporabljamo izraz mrežasta razpoka. Slike so posnete na različne načine: s pametnimi telefoni v avtomobilih na Japonskem, v Indiji in na Češkem, s kamerama v posebnem vozilu na Norveškem, z ulične panorame Google Street View v ZDA ter s kamero na motorju in z brezpilotnikom na Kitajskem [3, str. 8]. Označbe so shranjene v datotekah XML v obliki, podobni zbirki PASCAL VOC [3, str. 10], zbirka pa je objavljena z licenco CC BY-SA 4.0 [1, str. 3].

Tabela 3 povzema obseg podmnožic (stolpci D00 do D40 podajajo število označenih primerkov) in velikost delitve, slika 2 pa razporeditev označenih primerkov po razredih. Razredi so zelo neenakomerno zastopani: razred D00 obsega 47,3 % vseh označenih primerkov, udarne jame (D40) pa le 11,9 %. Posamezne države se razlikujejo tudi po sestavi razredov; v indijski podmnožici je na primer le 68 prečnih razpok proti 3187 udarnim jamam [3, str. 5].

Podmnožica	Slike	D00	D10	D20	D40	Ločljivost	Učne	Valid.	Testne
Japonska	10 506	4 049	3 979	6 199	2 243	600×600	8 404	1 051	1 051
Indija	7 706	1 555	68	2 021	3 187	720×720	6 164	771	771
Češka	2 829	988	399	161	197	600×600	2 263	283	283
Norveška	8 161	8 570	1 730	468	461	3650×2044	6 529	816	816
ZDA	4 805	6 750	3 295	834	135	640×640	3 845	480	480
Kitajska (motor)	1 977	2 678	1 096	641	235	512×512	1 581	198	198
Kitajska (dron)	2 401	1 426	1 263	293	86	512×512	2 401	0	0
Skupaj	38 385	26 016	11 830	10 617	6 544		31 187	3 599	3 599

Tabela 3: Javno dostopni učni del zbirke RDD2022 po podmnožicah in naša delitev na učno, validacijsko in testno množico

Vir: število slik in označenih primerkov po [3, str. 4–5], delitev lastna



Slika 2: Število označenih primerkov po razredih in podmnožicah zbirke RDD2022

Vir: izdelano po podatkih iz [3, str. 5]

4.2 Delitev in priprava podatkov

Označbe testnih slik zbirke RDD2022 niso objavljene [3, str. 12], zato smo javno dostopni učni del razdelili sami. Vsako od šestih podmnožic s testnim delom smo naključno razdelili po slikah v razmerju 80 : 10 : 10 na učno, validacijsko in testno množico; podmnožica Kitajska (dron) testnega dela nima in je v celoti v učni množici, kar je enako kot pri Pham idr. [4, str. 4]. Seznami slik posameznih množic so shranjeni skupaj s kodo. Učna množica ima 31 187 slik, validacijska in testna pa po 3 599 slik; testna množica vsebuje 5 255 označenih primerkov. Delitev po državah je v tabeli 3.

Delitev po slikah ima znano šibkost: zbirka vsebuje posnetke z vozil v gibanju, zato se lahko posnetki istega odseka ceste znajdejo hkrati v učni in testni množici, ocene pa so zaradi tega lahko optimistične. Oznak odseka v zbirki ni, zato tega ne moremo preveriti; težavo navajamo med omejitvami (razdelek 5.8). Pham idr. [4, str. 4] slike vsake države delijo v razmerju 90 % proti 10 %, vendar za izbiro modela in ne za poročanje.

Označbe smo iz formata PASCAL VOC pretvorili v format YOLO, ki podaja središče in velikost pravokotnika kot deleže slike, in preverili, da pravokotniki ležijo znotraj slike; vsebine označb nismo spreminjali. Slike smo pri učenju in sklepanju prilagodili na velikost 640 × 640 točk z ohranjanjem razmerja stranic. Slik z Norveške, ki imajo ločljivost okoli 3650 × 2044 točk, pri tem nismo obrezali, čeprav so Pham idr. [4, str. 4] to storili, ker bi moral biti izrez prilagojen vsaki državi, mi pa želimo enak postopek za vse podmnožice;

posledica je, da majhne poškodbe na teh slikah po zmanjšanju postanejo zelo majhne, kar preučimo v razdelku 5.5.

Testna množica ima 33,7 % slik brez označenih poškodb. Po velikosti razdelimo označene primerke na majhne (manj kot 1 % ploščine slike, 21,5 % primerkov), srednje (od 1 % do 5 %, 50,5 %) in velike (vsaj 5 %, 28,0 %). Po razredih je v testni množici 2 416 primerkov D00, 1 127 primerkov D10, 1 061 primerkov D20 in 651 primerkov D40.

4.3 Izbira modelov in začetnih uteži

Primerjamo tri detektorje, ki zastopajo različne rešitve iz poglavja 2. Faster R-CNN z ResNet-50 in FPN iz knjižnice torchvision je dvostopenjski detektor; njegova različica z utežmi na naboru COCO ima 41,8 milijona parametrov in 134,4 milijarde operacij na sliko ter na naboru COCO doseže 37,0 AP [9]. YOLOv7 (36,9 milijona parametrov, 104,7 milijarde operacij [23]) je enostopenjski detektor s sidrnimi pravokotniki, uporabljen iz uradnega repozitorija avtorjev [36]. YOLOv8m (25,9 milijona parametrov, 78,9 milijarde operacij [24]) je enostopenjski detektor brez sider, uporabljen prek knjižnice Ultralytics [25]. Vse tri smo implementirali v ogrodju PyTorch [37].

Vsi modeli začnejo z utežmi, naučenimi na naboru COCO [14], razen v ablaciji, kjer tudi YOLOv8m učimo iz naključnih uteži. Izbira začetnih uteži je pri vseh treh modelih enaka, zato odločilna razlika ni v predhodnem znanju, ampak v arhitekturi in postopku učenja.

4.4 Razširjanje podatkov

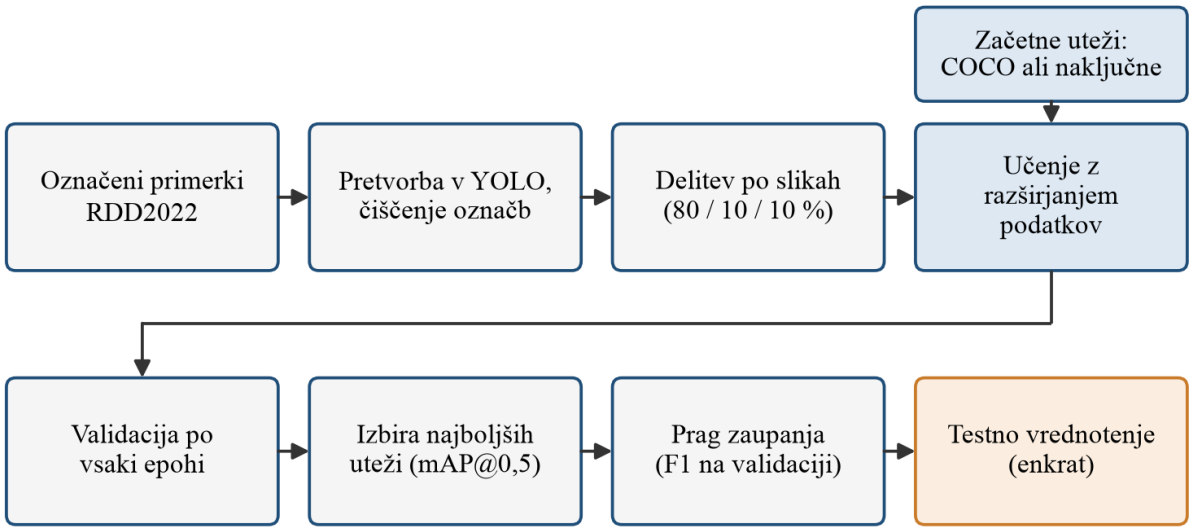
Razširjanje podatkov je pri vsakem modelu določeno s privzeto nastavitvijo njegove izvedbe: pri YOLOv8m so to mozaično združevanje štirih slik, spreminjanje odtenka, nasičenosti in svetlosti, premik, vodoravno zrcaljenje in izklop mozaika v zadnjih desetih epohah [26]; pri YOLOv7 so to privzete vrednosti iz konfiguracijske datoteke avtorjev, med njimi mozaik, mešanje slik (angl. *mixup*), premik, skaliranje in vodoravno zrcaljenje [36]; pri Faster R-CNN smo uporabili le vodoravno zrcaljenje. Pham idr. [4, str. 5] navajajo, da so privzete vrednosti za YOLOv7 pri poškodbah vozišča premočne, ker skaliranje, mozaik, mešanje slik in lepljenje delov slik ustvarijo nerealistične slike, in so jih zmanjšali. Mi privzetih vrednosti nismo ročno uglaševali, saj želimo primerjati modele pri enakem trudu; to je omejitev, ki jo omenjamo v razdelku 5.8.

Za ablacijo (RV3) pri YOLOv8m izklopimo vse spremembe slik, torej mozaik, barvne spremembe, premike, skaliranje in zrcaljenje. Ta izklop je skrajni primer; delne različice razširjanja v nalogi niso preizkušene.

4.5 Protokol učenja

Vse modele učimo na isti učni množici s hiperparametri, zbranimi v prilogi B (tabela 13). YOLOv8m učimo 100 epoh z optimizatorjem Adam [11] pri začetnem koraku 0,001, ki je privzeta vrednost za ta optimizator v knjižnici [26], z utežnim razpadom 0,0005, velikostjo paketa 16 in ogrevanjem prve tri epohe. YOLOv7 učimo 100 epoh z optimizatorjem SGD in privzetimi hiperparametri avtorjev (začetni korak 0,01, gibalna količina 0,937, utežni razpad 0,0005) [36], Faster R-CNN pa 24 epoh z optimizatorjem SGD. Po vsaki epohi izmerimo mAP@0,5 na validacijski množici in obdržimo uteži z najvišjo vrednostjo.

Vsako konfiguracijo učimo trikrat z različnim naključnim semenom (11, 22 in 33), da ocenimo, koliko rezultat niha med ponovitvami. Testno množico uporabimo šele po izbiri uteži in praga zaupanja, in sicer enkrat za vsako ponovitev. Postopek ponazarja slika 3.



Slika 3: Potek učenja in vrednotenja

Vir: lastna ponazoritev

4.6 Protokol vrednotenja

Zaznavo štejemo za pravilno, če ima isti razred kot označeni primerek in je prekrivanje IoU z njim vsaj 0,5, vsak označeni primerek pa lahko ujame največ ena zaznava. Zaznave obravnavamo po padajoči oceni zaupanja in vsako povežemo z še neujetim označenim primerkom istega razreda z največjim prekrivanjem, kar je običajen postopek po vzoru izziva PASCAL VOC [1, str. 2]. Za vsak razred izračunamo AP kot ploščino pod interpolirano krivuljo natančnosti glede na priklic, upoštevajoč vse točke krivulje, mAP@0,5 pa je povprečje AP za štiri razrede.

Delovno točko izberemo na validacijski množici: izmed pragov zaupanja od 0,05 do 0,95 s korakom 0,05 izberemo tistega, ki daje najvišji povprečni F1 po razredih, in ga nespremenjenega uporabimo na testni množici. Poročamo o natančnosti, priklicu in F1 po razredih ter o skupnem (mikro) F1, v katerem so zaznave in označeni primerki vseh razredov združeni; ta vrednost je podobna merilu izziva CRDDC [1, str. 4], vendar ni z njim neposredno primerljiva, ker testne množice niso iste. Rezultate podajamo kot povprečje treh ponovitev z oznako ± za vzorčni standardni odklon (SD).

Razlike med konfiguracijami ocenimo z vzorčenjem po testnih slikah: iz testne množice 1000-krat naključno izberemo slike z vračanjem, za vsako vzorčenje izračunamo mAP@0,5 vsake konfiguracije (povprečje treh ponovitev) in iz razlik določimo 95-odstotni interval zaupanja s percentili. Vzorčenje je parjeno, torej vse konfiguracije v posameznem koraku dobijo iste slike. Interval izraža le nihanje zaradi izbire testnih slik, ne pa nihanja zaradi učenja ali zaradi razdelitve podatkov.

Hitrost merimo na istem grafičnem procesorju (Nvidia A100, 40 gigabajtov) z velikostjo paketa 1 in vhodom 640 × 640 točk, in sicer kot povprečni čas za predobdelavo, sklepanje in izločanje podvojenih zaznav (NMS) pri 500 testnih slikah, po izločitvi 50 ogrevalnih slik.

Napake razvrstimo po tem, kaj je z zaznavo narobe. Pri lažnih pozitivih (zaznava, ki ni ujeta) ločimo zamenjavo razreda (IoU vsaj 0,5 z označenim primerkom drugega razreda), slabo lokalizacijo (IoU od 0,1 do 0,5 z označenim primerkom istega razreda), podvojeno zaznavo (IoU vsaj 0,5 z že ujetim primerkom istega razreda) in lažni alarm na ozadju (vse ostalo). Pri lažnih negativih (spregledan označeni primerek) ločimo zamenjan razred, slabo lokalizacijo in popolnoma spregledan primerek. Za razdelitev napak, matriko zmede, razčlenitev po podmnožicah in velikostih uporabimo ponovitev z najvišjim mAP@0,5 na validacijski množici, da testne množice ne uporabimo za izbiro.

5 Implementacija in vrednotenje

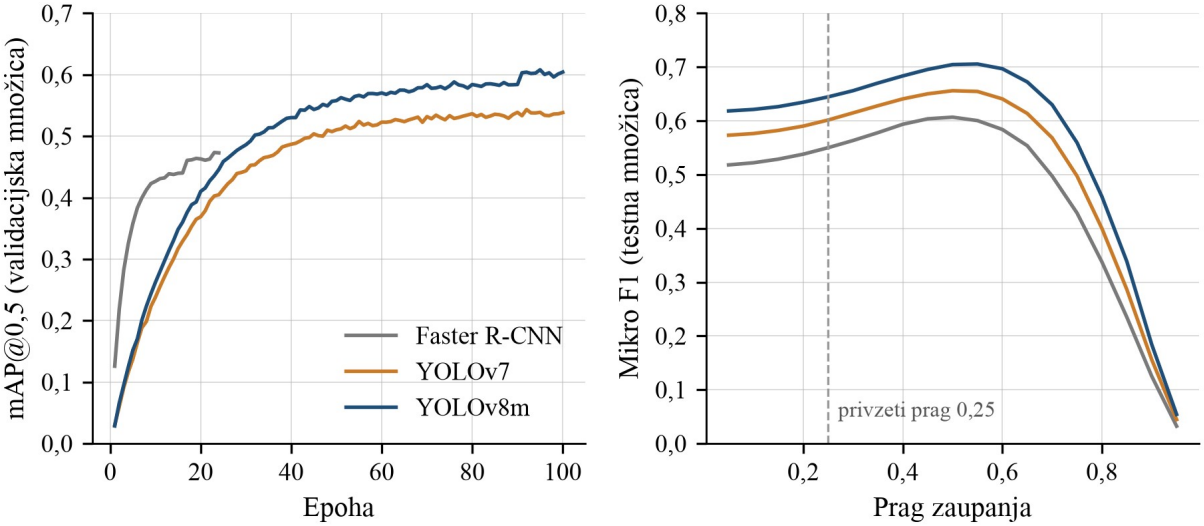
Poglavje opisuje izvedbo poskusov in njihove rezultate. Rezultate podajamo po raziskovalnih vprašanjih, od točnosti po razredih (RV1) prek hitrosti (RV2) in ablacije (RV3) do analize napak (RV4); sledi primerjava z literaturo in obravnava omejitev. Vse številke v poglavju izračuna ena skripta iz surovih zaznav in označb, zato se v besedilu, tabelah in slikah ujemajo.

5.1 Razvojno okolje in potek učenja

Modele smo učili in merili na enem grafičnem procesorju Nvidia A100 s 40 gigabajti pomnilnika. Programska podlaga so ogrodje PyTorch [37], knjižnica torchvision za Faster R-CNN [9], uradna izvedba YOLOv7 [36] in knjižnica Ultralytics za YOLOv8m [25]. Vrednotenje je izvedeno z lastno skripto, ki iz datotek z označbami in zaznavami izračuna ujemanja, AP, F1, intervale zaupanja in razčlenitev napak; izsek skripte je v prilogi C. Zaznave vseh konfiguracij in vseh ponovitev se vrednotijo z isto kodo, zato razlike med njimi ne morejo izvirati iz različnih izvedb merila.

Na validacijski množici je mAP@0,5, povprečen čez tri ponovitve, 0,47 za Faster R-CNN, 0,55 za YOLOv7 in 0,61 za YOLOv8m. Slika 4 (levo) kaže potek te vrednosti po epohah. Po desetih epohah je mAP@0,5 pri YOLOv7 0,24 in pri YOLOv8m 0,26, najvišjo vrednost pa oba dosežeta pozno, med 92. in 95. epoho; pri YOLOv8m skok proti koncu sovpada z izklopom mozaika v zadnjih desetih epohah, ki ga predvideva privzeta nastavitev knjižnice [26]. Faster R-CNN doseže najvišjo vrednost v 23. ali 24. epohi, po znižanjih koraka učenja.

Prag zaupanja izberemo na validacijski množici, povprečni izbrani pragovi so 0,52 (Faster R-CNN), 0,53 (YOLOv7) in 0,55 (YOLOv8m). Desni del slike 4 kaže, kako se F1 na testni množici spreminja s pragom: pri YOLOv8m ostane v razponu 0,02 od najvišje vrednosti (0,71) za pragove od 0,45 do 0,60, izbrani prag pa je v tem območju. Pri privzetem pragu zaupanja knjižnice Ultralytics 0,25 [26] bi bil F1 le 0,64, pri Faster R-CNN 0,55 namesto 0,61. Izbira praga na validacijski množici torej ni formalnost: privzeta vrednost bi F1 zmanjšala pri vseh treh modelih.



Slika 4: Potek validacijskega mAP@0,5 po epohah (levo) in mikro F1 na testni množici glede na prag zaupanja (desno); povprečje treh ponovitev

Vir: lastni rezultati

5.2 Točnost po razredih poškodb (RV1)

Tabela 4 prikazuje uspešnost modela YOLOv8m po razredih, tabela 5 pa AP in F1 vseh treh modelov s standardnim odklonom treh ponovitev; v njej so razredi označeni z oznakami D00 (vzdolžna razpoka), D10 (prečna razpoka), D20 (mrežasta razpoka) in D40 (udarna jama). YOLOv8m doseže mAP@0,5 0,62 in mikro F1 0,71, YOLOv7 0,56 in 0,66, Faster R-CNN pa 0,49 in 0,60.

Razred poškodbe	Natančnost	Priklic	F1	AP@0,5	Št. primerkov
Vzdolžna razpoka (D00)	0,84	0,62	0,72	0,63	2 416
Prečna razpoka (D10)	0,78	0,62	0,69	0,61	1 127
Mrežasta razpoka (D20)	0,79	0,64	0,71	0,64	1 061
Udarna jama (D40)	0,80	0,60	0,69	0,60	651
Skupaj (mAP / mikro F1)	0,81	0,62	0,71	0,62	5 255

Tabela 4: Uspešnost modela YOLOv8m po razredih poškodb na testni množici (povprečje treh ponovitev, pri pragu zaupanja iz validacije; v zadnji vrstici mikro vrednosti in mAP@0,5)

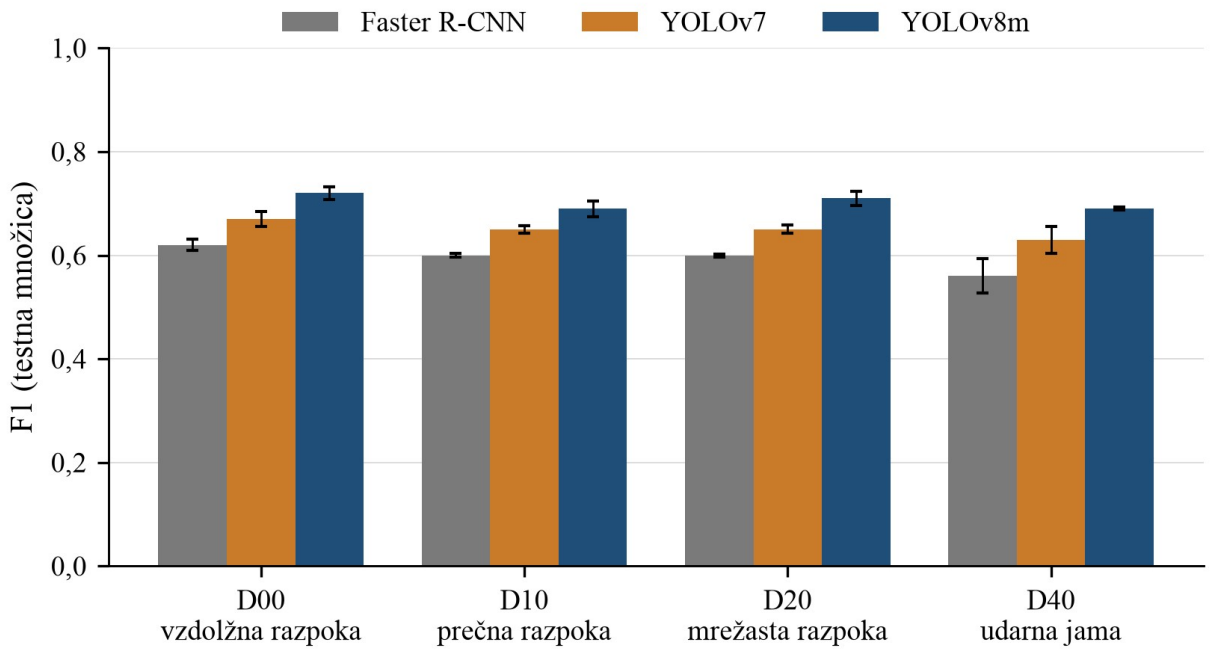
Vir: lastni rezultati

Razred	FRCNN AP	FRCNN F1	YOLOv7 AP	YOLOv7 F1	YOLOv8m AP	YOLOv8m F1
D00	0,52 ± 0,008	0,62 ± 0,011	0,57 ± 0,014	0,67 ± 0,015	0,63 ± 0,013	0,72 ± 0,012
D10	0,49 ± 0,007	0,60 ± 0,004	0,56 ± 0,006	0,65 ± 0,007	0,61 ± 0,018	0,69 ± 0,015
D20	0,50 ± 0,006	0,60 ± 0,003	0,57 ± 0,013	0,65 ± 0,008	0,64 ± 0,016	0,71 ± 0,014
D40	0,46 ± 0,042	0,56 ± 0,033	0,54 ± 0,033	0,63 ± 0,026	0,60 ± 0,010	0,69 ± 0,003
Skupaj	0,49 ± 0,013	0,60 ± 0,009	0,56 ± 0,011	0,66 ± 0,009	0,62 ± 0,013	0,71 ± 0,008

Tabela 5: AP@0,5 in F1 po razredih za tri modele na testni množici (povprečje ± standardni odklon treh ponovitev; v zadnji vrstici mAP@0,5 in mikro F1)

Vir: lastni rezultati

Razlike med modeli so večje od nihanja med ponovitvami. Pri mAP@0,5 je YOLOv8m boljši od YOLOv7 za 6,1 odstotne točke (95-odstotni interval od 5,0 do 7,3), YOLOv7 pa od Faster R-CNN za 6,5 točke (od 5,2 do 7,9); razlika med YOLOv8m in Faster R-CNN je 12,6 točke (od 11,3 do 13,8). Vsi trije intervali ležijo nad ničlo, a izražajo le nihanje zaradi izbire testnih slik. Vrstni red modelov je enak pri vseh štirih razredih, kot kaže tabela 5 in slika 5. Natančnost in priklic po razredih za vse tri modele sta v prilogi E (tabela 15), rezultati posameznih ponovitev pa v tabeli 16.



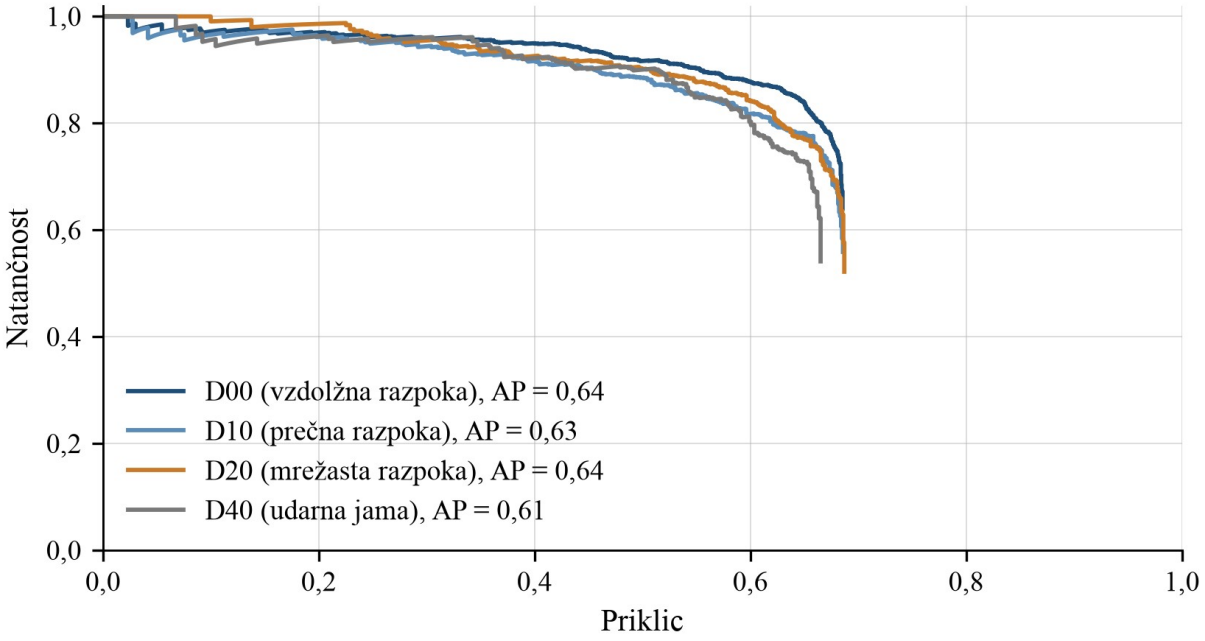
Slika 5: F1 po razredih poškodb za tri modele na testni množici (navpične črte so standardni odklon treh ponovitev)

Vir: lastni rezultati

Med razredi so razlike znatno manjše kot med modeli. Pri YOLOv8m je AP največji pri mrežastih razpokah (D20, 0,64), najmanjši pri udarnih jamah (D40, 0,60), vendar je razpon vrednosti, od 0,60 do 0,64, le nekajkrat večji od standardnega odklona ponovitev (0,018 pri D10). Razredov zato ne moremo zanesljivo razvrstiti. Natančnost je najvišja pri vzdolžnih razpokah (0,84), priklic pa se med razredi razlikuje za največ 0,04. Razred D40 ima s 651 primerki najmanj podatkov, pri modelu Faster R-CNN pa je pri njem nihanje med ponovitvami največje (standardni odklon AP 0,042).

Krivulje natančnosti glede na priklic na sliki 6 se pri vseh razredih držijo visoko do priklica okoli 0,5, kjer je natančnost pri vzdolžnih razpokah še 0,92, nato pa strmo padejo. Največji priklic, ki ga model sploh doseže, je 0,69 pri vzdolžnih razpokah in 0,67 pri udarnih jamah; preostali označeni primerki ostanejo spregledani tudi pri nizkem pragu zaupanja. To je glavni

razlog za mAP@0,5 okoli 0,62, in ne napačna razvrstitev zaznanih poškodb, kar podrobneje preučimo v razdelku 5.5.



Slika 6: Krivulje natančnosti glede na priklic za model YOLOv8m na testni množici (ponovitev z najvišjim mAP@0,5 na validacijski množici)

Vir: lastni rezultati

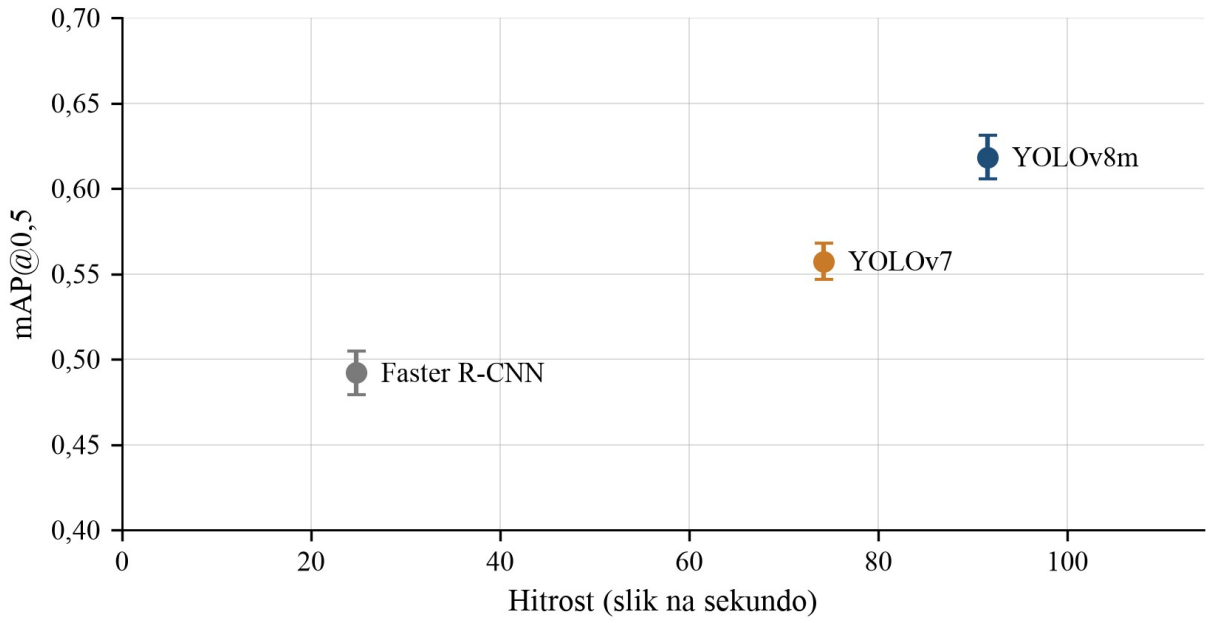
5.3 Točnost in hitrost (RV2)

Hitrost smo izmerili pri velikosti paketa 1 na 500 testnih slikah. Tabela 6 prikazuje, da Faster R-CNN potrebuje v povprečju 40,3 ms na sliko (25 slik na sekundo), YOLOv7 13,5 ms (74 slik na sekundo) in YOLOv8m 10,9 ms (92 slik na sekundo). Od tega je pri Faster R-CNN za sklepanje samo porabljenih 30,9 ms, pri YOLOv7 8,1 ms in pri YOLOv8m 6,2 ms; preostanek odpade na predobdelavo in izločanje podvojenih zaznav.

Model	Parametri (mio.)	Operacije (mrd.)	Čas na sliko (ms)	Slik na sekundo	mAP@0,5
Faster R-CNN	41,8	134,4	40,3 ± 2,1	25	0,49 ± 0,013
YOLOv7	36,9	104,7	13,5 ± 0,8	74	0,56 ± 0,011
YOLOv8m	25,9	78,9	10,9 ± 0,6	92	0,62 ± 0,013

Tabela 6: Velikost, hitrost in točnost treh detektorjev (čas je povprečje ± standardni odklon čez 500 slik, mAP@0,5 pa povprečje ± standardni odklon treh ponovitev)

Vir: lastne meritve; število parametrov in operacij po [9, 23, 24]



Slika 7: Točnost in hitrost treh detektorjev (navpične črte so standardni odklon treh ponovitev)

Vir: lastni rezultati

Razmerje hitrosti je 3,7-kratno med YOLOv8m in Faster R-CNN ter 1,23-kratno med YOLOv8m in YOLOv7. Slika 7 pokaže, da je YOLOv8m hkrati najhitrejši in najtočnejši, zato med obravnavanimi modeli ni kompromisa, ki bi ga bilo treba tehtati: YOLOv8m je boljši po obeh merilih, hkrati pa ima manj parametrov in operacij kot druga modela (tabela 6). Za obdelavo video toka s 25 slikami na sekundo je na sliko na voljo 40 ms; Faster R-CNN s povprečjem 40,3 ms te omejitve ne dosega z rezervo, 95. percentil njegovega časa pa je 43,7 ms. Oba modela YOLO imata za to dovolj rezerve. Merjenje velja za našo izvedbo, velikost paketa 1 in grafični procesor A100; na mobilni napravi, ki jo omenjajo Maeda idr. [2], bi bili časi bistveno daljši, česar nismo preizkušali.

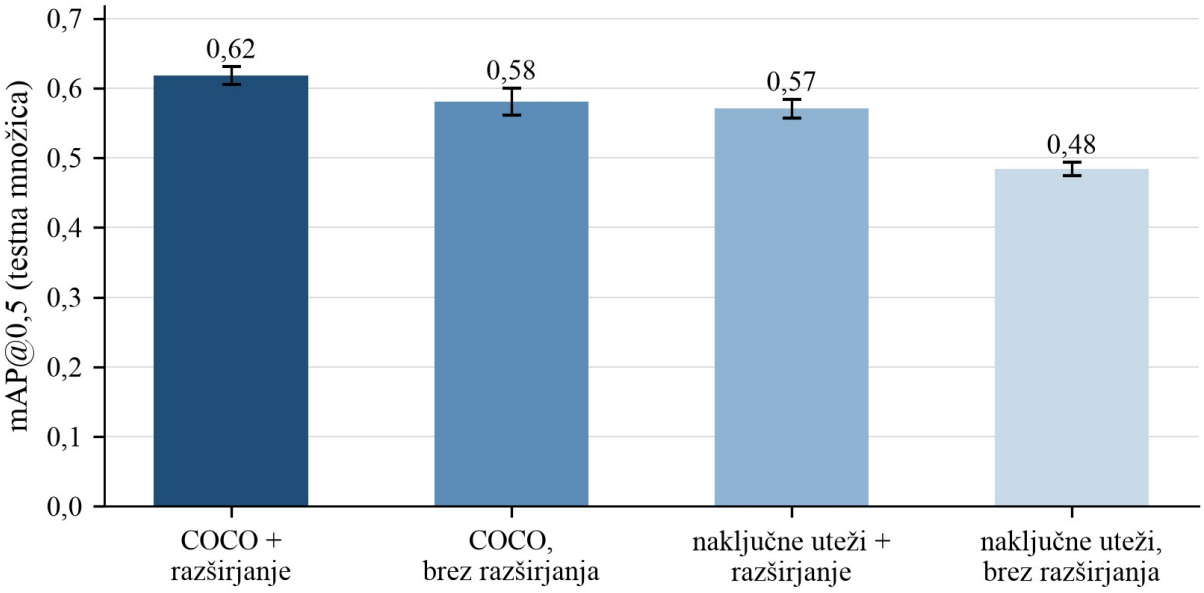
5.4 Prispevek začetnih uteži in razširjanja podatkov (RV3)

Tabela 7 in slika 8 prikazujeta štiri različice modela YOLOv8m. Izhodiščna različica (uteži COCO in polno razširjanje) doseže mAP@0,5 0,62. Brez razširjanja pade na 0,58, brez uteži COCO na 0,57, brez obojega pa na 0,48.

Začetne uteži	Razširjanje	mAP@0,5	Mikro F1	AP D00	AP D10	AP D20	AP D40
COCO	polno	0,62 ± 0,013	0,71	0,63	0,61	0,64	0,60
COCO	brez	0,58 ± 0,019	0,68	0,60	0,57	0,60	0,56
naključne	polno	0,57 ± 0,013	0,66	0,57	0,58	0,59	0,55
naključne	brez	0,48 ± 0,010	0,60	0,51	0,47	0,49	0,46

Tabela 7: Ablacija modela YOLOv8m: začetne uteži in razširjanje podatkov (mAP@0,5 kot povprečje ± standardni odklon treh ponovitev)

Vir: lastni rezultati



Slika 8: mAP@0,5 modela YOLOv8m pri štirih kombinacijah začetnih uteži in razširjanja podatkov

Vir: lastni rezultati

Razširjanje podatkov prispeva 3,7 odstotne točke (interval od 2,6 do 4,9), začetne uteži COCO pa 4,8 točke (od 3,7 do 5,8). Intervala se prekrivata, zato prispevkov ne moremo razvrstiti. Obe izboljšavi skupaj dasta 13,4 točke (od 12,2 do 14,5), kar je več od vsote posameznih prispevkov (8,5 točke); zanesljivosti tega presežka nismo ocenjevali. Prispevek posameznega postopka je večji, kadar drugega ni: prenos uteži COCO pomeni brez razširjanja 9,7 točke, z razširjanjem pa 4,8; razširjanje brez uteži COCO pomeni 8,6 točke, z utežmi pa 3,7. Podatki torej kažejo, da se postopka deloma nadomeščata. Pri vseh štirih različicah ima razred D40 najnižji AP, razlike med razredi pa so manjše od razlik med različicami.

Opomba k razlagi: ablacija je izvedena na enem modelu in na enem naboru podatkov, interval zaupanja pa zajema le izbiro testnih slik. Sklep, da sta oba postopka koristna, zato velja za YOLOv8m na zbirki RDD2022; za druge modele ali podatke ga ne moremo prenesti.

5.5 Analiza napak (RV4)

Najprej pogledamo, kakšne napake sploh nastajajo. Pri YOLOv8m je na testni množici 1 941 spregledanih označenih primerkov in 733 lažnih pozitivov, torej 2,6-krat več spregledanih kot napačnih zaznav. Od spregledanih primerkov jih 84,2 % model ni zaznal nikjer v bližini, pri 8,0 % je zaznal poškodbo istega razreda s premajhnim prekrivanjem, pri 7,8 % pa jo je zaznal z napačnim razredom. Med lažnimi pozitivi je 73,4 % alarmov na ozadju, 20,5 % zamenjav razreda in 6,1 % slabo lokaliziranih zaznav; podvojenih zaznav ni, ker jih izloči NMS. Tabela 8 podaja število napak za vse tri modele. Boljši model ima manj napak vseh vrst, razmerja

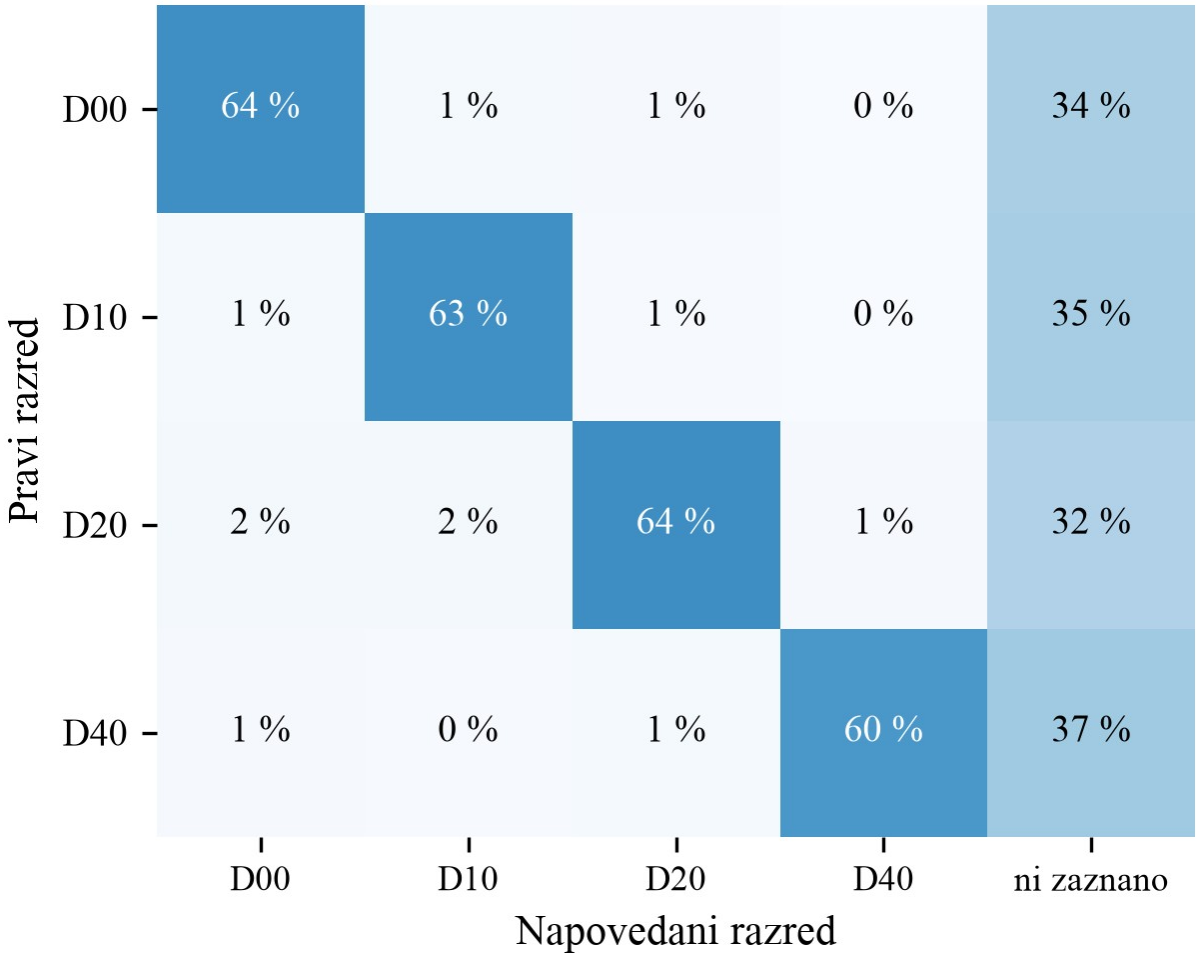
med vrstami pa so podobna: pri vseh treh prevladujejo spregledani primerki, med lažnimi pozitivi pa alarmi na ozadju.

Model	TP	FP skupaj	FP: ozadje	FP: lokalizacija	FP: razred	FN skupaj	FN: spregledan
Faster R-CNN	2 760	1 080	771	105	204	2 495	2 081
YOLOv7	2 973	803	555	83	164	2 282	1 924
YOLOv8m	3 314	733	538	45	150	1 941	1 634

Tabela 8: Število pravilnih zaznav (TP), lažnih pozitivov (FP) in spregledanih primerkov (FN) po vrstah pri treh modelih na testni množici

Vir: lastni rezultati

Zamenjave razredov so redke glede na spregledane primerke, vendar imajo značilen vzorec. Matrika zmede na sliki 9 kaže, da je razred pravilno zaznan v 63,6 % primerov D00, 63,0 % primerov D10, 63,8 % primerov D20 in 60,1 % primerov D40, v preostalih pa model poškodbe ne zazna, jo zazna s premajhnim prekrivanjem ali jo zamenja z drugim razredom. Zamenjave med pari razredov razpok (D00 in D10, D00 in D20, D10 in D20) tvorijo 114 od 151 zamenjav, zamenjave z udarno jamo pa le 37. Najpogostejši par je vzdolžna in prečna razpoka (49 zamenjav). Pham idr. [4, str. 6] navajajo, da so lahko prečne razpoke zaradi perspektive ali rahlega zasuka kamere napačno označene kot vzdolžne in obratno ter da se aligatorske razpoke pogosto zamenjujejo z vzdolžnimi in prečnimi; naša razporeditev zamenjav je s tem opisom skladna, a naši podatki ne omogočajo ločiti napak modela od napačnih označb. Števila, iz katerih so izračunani deleži na sliki, so v tabeli 14 v prilogi D.



Slika 9: Matrika zmede modela YOLOv8m (deleži označenih primerkov po pravih razredih)

Vir: lastni rezultati

Razlike po podmnožicah so bistveno večje od razlik po razredih. Tabela 9 kaže, da je F1 najvišji v ZDA (0,80), nizek pa na Norveškem (0,57); razpon med podmnožicami je 0,23. Priklic je na Norveškem 0,46, v drugih državah pa od 0,65 do 0,72. Za Češko je v testni množici le 182 primerkov, zato je ocena za to državo negotova.

Podmnožica	Primerki	Natančnost	Priklic	F1 (YOLOv8m)	F1 (YOLOv7)	F1 (Faster R-CNN)
Japonska	1 705	0,83	0,65	0,73	0,68	0,64
Indija	696	0,76	0,66	0,70	0,65	0,58
Češka	182	0,68	0,66	0,67	0,64	0,57
Norveška	1 088	0,75	0,46	0,57	0,49	0,43
ZDA	1 121	0,90	0,72	0,80	0,75	0,71
Kitajska (motor)	463	0,89	0,69	0,78	0,71	0,67

Tabela 9: Uspešnost po podmnožicah zbirke na testni množici (natančnost, priklic in F1 modela YOLOv8m ter F1 modelov YOLOv7 in Faster R-CNN)

Vir: lastni rezultati

Del razlike lahko pojasnimo z velikostjo poškodb. Na Norveškem je 70,4 % označenih primerkov majhnih (manj kot 1 % ploščine slike), v drugih državah pa največ 18,2 %; slike z Norveške so zelo velike in jih pri učenju zmanjšamo na 640 × 640 točk, pri čemer tudi poškodbe postanejo majhne, kot so ugotovili že Pham idr. [4, str. 4]. Priklic je pri majhnih primerkih 0,44, pri srednjih 0,66 in pri velikih 0,72 (tabela 10). Velikost sama pa Norveške ne pojasni v celoti: tudi pri srednje velikih primerkih je priklic na Norveškem 0,59, drugje pa 0,67. Preostanek razlike bi lahko pripisali drugim lastnostim teh slik, na primer drugačni kameri in perspektivi, česar naši podatki ne ločijo.

Razred	Majhna	Srednja	Velika
Vzdolžna razpoka (D00)	0,42 (631)	0,69 (1 182)	0,76 (603)
Prečna razpoka (D10)	0,47 (202)	0,64 (655)	0,72 (270)
Mrežasta razpoka (D20)	0,50 (36)	0,62 (450)	0,66 (575)
Udarna jama (D40)	0,48 (262)	0,67 (367)	0,86 (22)
Vsi razredi	0,44 (1 131)	0,66 (2 654)	0,72 (1 470)

Tabela 10: Priklic modela YOLOv8m po razredih in velikostih označenih primerkov (v oklepaju število primerkov; majhna: manj kot 1 % ploščine slike, srednja: od 1 % do 5 %, velika: vsaj 5 %)

Vir: lastni rezultati

Podrobnejši pogled na razrede po podmnožicah daje tabela 11. Na Norveškem je priklic nizek pri vseh razredih, pri vzdolžnih razpokah 0,46 proti 0,75 v ZDA in 0,70 na Japonskem. Pri manj kot 30 primerkih, na primer D10 v Indiji (3 primerki) ali D40 v ZDA (13), je primerjava preveč nezanesljiva.

Podmnožica	D00	D10	D20	D40
Japonska	0,70 (399)	0,66 (417)	0,63 (648)	0,61 (241)
Indija	0,74 (168)	0,67 (3)	0,63 (205)	0,62 (320)
Češka	0,66 (100)	0,69 (49)	0,63 (19)	0,57 (14)
Norveška	0,46 (836)	0,45 (181)	0,56 (34)	0,38 (37)
ZDA	0,75 (647)	0,67 (379)	0,70 (82)	0,54 (13)
Kitajska (motor)	0,71 (266)	0,64 (98)	0,71 (73)	0,58 (26)

Tabela 11: Priklic modela YOLOv8m po podmnožicah in razredih (v oklepaju število označenih primerkov v testni množici)

Vir: lastni rezultati

Skupaj gledano so najpomembnejši dejavniki napak velikost poškodbe (razpon priklica 0,28) in podmnožica zbirke (razpon priklica 0,26), razred poškodbe pa je z razponom priklica 0,04 precej manj pomemben. Pri razredu D20 je majhnih primerkov le 36, pri razredu D40 je velikih le 22, zato razlik med razredi znotraj velikostnih razredov ne interpretiramo.

5.6 Pomen rezultatov za uporabo

Pri pregledu cest sta pomembnejši številki na ravni slike kot na ravni označenih primerkov. Med 2 386 testnimi slikami, ki vsebujejo vsaj eno označeno poškodbo, model YOLOv8m

zazna vsaj eno pravilno na 78,1 % slik, YOLOv7 na 72,6 % in Faster R-CNN na 69,6 % slik. Med 1 213 slikami brez označenih poškodb jih 13,3 % dobi vsaj en lažni alarm pri YOLOv8m (19,9 % pri Faster R-CNN). Preračunano na 1 000 slik YOLOv8m spregleda 539 označenih primerkov in sproži 204 lažnih pozitivov, Faster R-CNN pa 693 in 300.

Te vrednosti veljajo za prag, ki na validacijski množici največ da F1, in za testno množico z 33,7 % slik brez označene poškodbe; pri drugačni sestavi cest bi bile drugačne. Upravljavec, ki bi želel manj spregledanih poškodb, lahko prag zniža na račun več lažnih alarmov, vendar se priklic modela YOLOv8m tudi pri nizkem pragu ustavi pri 0,69. Po naših rezultatih bi bil sistem zato primeren kot pomoč pri izboru slik za ročni pregled, ne pa kot popolna zamenjava pregleda, in še posebej mora biti previden na slikah z majhnimi poškodbami.

5.7 Primerjava z literaturo

Neposredna primerjava z objavljenimi številkami ni možna, ker označb testnih slik izziva CRDDC ni. Za orientacijo: zmagovalna ekipa je na skritem testu dosegla F1 0,770 [1, str. 5], Pham idr. [4, str. 7] pa 0,741, naš mikro F1 na lastni testni množici pa je 0,71. Vrednosti so primerljive po velikosti, a ne po vsebini, ker se testne množice razlikujejo.

Vzorec po državah v naših rezultatih je podoben tistemu v literaturi. Pham idr. [4, str. 7] poročajo, da so lestvice za Japonsko in ZDA dosegle najvišjo, za Indijo in Norveško pa nižjo vrednost F1, in jo povežejo s številom uporabnih učnih slik; pri zmagovalni ekipi izziva je F1 na lestvici za ZDA 0,844, za Norveško 0,595 [1, str. 5]. Pri nas so ZDA najboljše in Norveška najslabša, Indija pa je s F1 0,70 v sredini, kar se od literature razlikuje.

Zmagovalna ekipa je iz analize prejšnjega izziva sklepala, da modeli YOLO dajejo višji priklic, dvostopenjski detektorji pa višjo natančnost [1, str. 6]. V naši primerjavi tega ni mogoče potrditi: Faster R-CNN ima mikro natančnost 0,73 in priklic 0,52, YOLOv8m pa 0,81 in 0,62, torej je YOLOv8m boljši pri obeh. Razlika se lahko deloma pojasni s tem, da smo uporabili osnovni Faster R-CNN s krajšim učenjem in brez ansambla, ne pa različice Cascade R-CNN.

Li idr. [35] poročajo za svojo izboljšano različico YOLOv8 mAP@0,5 62,5 % na validacijski množici zbirke RDD2022; naš YOLOv8m ima na validacijski množici 0,61. Ker ne vemo, katere podmnožice in kakšno delitev so uporabili, to ni utemeljena primerjava, kaže pa na enak velikostni razred.

5.8 Omejitve veljavnosti

Rezultati imajo več omejitev. Delitev po slikah je naključna, zato lahko zaporedni posnetki istega odseka ceste pristanejo v različnih množicah in so ocene optimistične. Zbirka zajema le

eno domeno, ceste in kamere iz šestih držav, in rezultati za druge ceste niso preizkušeni. Tri ponovitve dajo le grobo oceno nihanja zaradi učenja, intervali zaupanja pa upoštevajo samo vzorec testnih slik. Hiperparametrov nismo uglaševali, privzete nastavitve razširjanja za YOLOv7 so po Pham idr. [4, str. 5] pri poškodbah vozišča lahko premočne; boljši rezultati so zato mogoči pri vseh modelih. Faster R-CNN je imel krajše učenje in le zrcaljenje, kar primerjavo naredi v njegovo škodo. Hitrost smo izmerili na enem grafičnem procesorju. Izbrana različica YOLOv8m ni najnovejša različica družine, saj je izdelovalec medtem objavil novejše [24]. Označbe so pravokotniki, zato naše meritve ne zajamejo oblike ali dolžine razpoke.

6 Sklep

V nalogi smo na javno dostopnem delu zbirke RDD2022 primerjali tri detektorje, Faster R-CNN, YOLOv7 in YOLOv8m, pri zaznavanju štirih razredov poškodb vozišča in izmerili, koliko k točnosti prispevata začetne uteži, naučene na naboru COCO in razširjanje podatkov. Ker označb testnih slik zbirke ni, smo učni del razdelili sami in vse rezultate podali na lastni testni množici, kar zahteva previdnost pri primerjavi z objavljenimi vrednostmi. Cilje dela štejemo za dosežene: protokol je opisan in ponovljiv, trije detektorji so primerjani pri enakih pogojih, ablacija je izvedena in napake so razčlenjene.

6.1 Odgovori na raziskovalna vprašanja

RV1. Najboljši model YOLOv8m je dosegel mAP@0,5 0,62 in mikro F1 0,71, YOLOv7 0,56 in 0,66, Faster R-CNN pa 0,49 in 0,60. Vrstni red modelov je enak pri vseh štirih razredih, razlike med razredi pa so zaradi nihanja med ponovitvami za razvrstitev premajhne.

RV2. YOLOv8m je s 10,9 ms na sliko (92 slik na sekundo) najhitrejši in hkrati najtočnejši, YOLOv7 potrebuje 13,5 ms, Faster R-CNN pa 40,3 ms. Za obdelavo video toka v vozilu je na podlagi naših meritev primeren YOLOv8m ali YOLOv7, medtem ko Faster R-CNN doseže le okoli 25 slik na sekundo.

RV3. Če razširjanja podatkov ne uporabimo, je mAP@0,5 modela YOLOv8m za 3,7 odstotne točke nižji, brez začetnih uteži COCO za 4,8 točke, brez obojega pa za 13,4 točke. Intervala posameznih prispevkov se prekrivata, zato njunega vrstnega reda ne zagovarjamo; prispevek vsakega postopka je večji, če drugega ni.

RV4. Največ napak povzročijo spregledani primerki: pri YOLOv8m jih je 1 941, lažnih pozitivov pa 733. Napake so mnogo bolj odvisne od velikosti poškodbe in podmnožice zbirke kot od razreda: priklic je za majhne poškodbe 0,44, za velike 0,72, na Norveški 0,46, v ZDA pa 0,72. Zamenjave razredov so redke in nastajajo večinoma med razredi razpok.

6.2 Prispevki in omejitve

Prispevek dela je primerjava, ki loči vpliv arhitekture od vpliva začetnih uteži in razširjanja podatkov in ki rezultate podaja po razredih, državah in velikostih, ne le kot eno število. Glavne omejitve so naključna delitev po slikah, ki lahko napihne ocene, ena zbirka, tri ponovitve ter neuglaševani hiperparametri.

6.3 Smernice za nadaljnje delo

Naravno nadaljevanje je preveriti, ali bi razrez velikih slik na manjše izreze povečal priklic pri majhnih poškodbah, zlasti na slikah z Norveške. Delitev po odsekih cest, če bi bili podatki o odsekih na voljo, bi odpravila možnost uhajanja podatkov med množicami. Ker označbe zbirke niso podane po točkah, bi bilo za oceno oblike in dolžine razpok treba nabor dopolniti z maskami in preizkusiti segmentacijske modele. Primerjavo je smiselno ponoviti z novejšimi različicami družine YOLO.

Literatura

[1] D. Arya idr., „Crowdsensing-based road damage detection challenge (CRDDC'2022),“ *arXiv:2211.11362*, 2022. doi: 10.48550/arXiv.2211.11362

[2] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, H. Omata, „Road damage detection and classification using deep neural networks with smartphone images,“ *Computer-Aided Civil and Infrastructure Engineering*, let. 33, št. 12, str. 1127–1141, 2018. doi: 10.1111/mice.12387

[3] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, Y. Sekimoto, „RDD2022: a multi-national image dataset for automatic road damage detection,“ *arXiv:2209.08538*, 2022. doi: 10.48550/arXiv.2209.08538

[4] V. Pham, D. Nguyen, C. Donan, „Road damages detection and classification with YOLOv7,“ *arXiv:2211.00091*, 2022. doi: 10.48550/arXiv.2211.00091

[5] Y. LeCun, Y. Bengio, G. Hinton, „Deep learning,“ *Nature*, let. 521, št. 7553, str. 436–444, 2015. doi: 10.1038/nature14539

[6] A. Krizhevsky, I. Sutskever, G. E. Hinton, „ImageNet classification with deep convolutional neural networks,“ *Communications of the ACM*, let. 60, št. 6, str. 84–90, 2017. doi: 10.1145/3065386

[7] K. Simonyan, A. Zisserman, „Very deep convolutional networks for large-scale image recognition,“ v *Proc. 3rd Int. Conf. Learning Representations (ICLR)*, 2015. https://arxiv.org/abs/1409.1556

[8] K. He, X. Zhang, S. Ren, J. Sun, „Deep residual learning for image recognition,“ v *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, str. 770–778. doi: 10.1109/CVPR.2016.90

[9] PyTorch, „Models and pre-trained weights.“ 2026. [Splet; dostop 23. 10. 2026]. https://docs.pytorch.org/vision/stable/models.html

[10] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, D. Ramanan, „Object detection with discriminatively trained part-based models,“ *IEEE Transactions on Pattern Analysis and Machine Intelligence*, let. 32, št. 9, str. 1627–1645, 2010. doi: 10.1109/TPAMI.2009.167

[11] D. P. Kingma, J. Ba, „Adam: a method for stochastic optimization,“ v *Proc. 3rd Int. Conf. Learning Representations (ICLR)*, 2015. https://arxiv.org/abs/1412.6980

[12] S. Ioffe, C. Szegedy, „Batch normalization: accelerating deep network training by reducing internal covariate shift,“ v *Proc. 32nd Int. Conf. Machine Learning (ICML)*, 2015, str. 448–456. https://proceedings.mlr.press/v37/ioffe15.html

[13] S. J. Pan, Q. Yang, „A survey on transfer learning,“ *IEEE Transactions on Knowledge and Data Engineering*, let. 22, št. 10, str. 1345–1359, 2010. doi: 10.1109/TKDE.2009.191

[14] T.-Y. Lin idr., „Microsoft COCO: common objects in context,“ v *Computer Vision – ECCV 2014, Lecture Notes in Computer Science, let. 8693*, 2014, str. 740–755. doi: 10.1007/978-3-319-10602-1_48

[15] C. Shorten, T. M. Khoshgoftaar, „A survey on image data augmentation for deep learning,“ *Journal of Big Data*, let. 6, članek 60, 2019. doi: 10.1186/s40537-019-0197-0

[16] R. Girshick, „Fast R-CNN,“ v *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2015, str. 1440–1448. doi: 10.1109/ICCV.2015.169

[17] S. Ren, K. He, R. Girshick, J. Sun, „Faster R-CNN: towards real-time object detection with region proposal networks,“ *IEEE Transactions on Pattern Analysis and Machine Intelligence*, let. 39, št. 6, str. 1137–1149, 2017. doi: 10.1109/TPAMI.2016.2577031

[18] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, „Feature pyramid networks for object detection,“ v *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, str. 936–944. doi: 10.1109/CVPR.2017.106

[19] M. Tan, R. Pang, Q. V. Le, „EfficientDet: scalable and efficient object detection,“ v *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2020, str. 10778–10787. doi: 10.1109/CVPR42600.2020.01079

[20] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, „You only look once: unified, real-time object detection,“ v *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, str. 779–788. doi: 10.1109/CVPR.2016.91

[21] W. Liu idr., „SSD: single shot multibox detector,“ v *Computer Vision – ECCV 2016, Lecture Notes in Computer Science, let. 9905*, 2016, str. 21–37. doi: 10.1007/978-3-319-46448-0_2

[22] J. Redmon, A. Farhadi, „YOLO9000: better, faster, stronger,“ v *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, str. 6517–6525. doi: 10.1109/CVPR.2017.690

[23] C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, „YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,“ v *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2023, str. 7464–7475. doi: 10.1109/CVPR52729.2023.00721

[24] Ultralytics, „YOLOv8.“ 2026. [Splet; dostop 23. 10. 2026]. https://docs.ultralytics.com/models/yolov8/

[25] G. Jocher, J. Qiu, A. Chaurasia, „Ultralytics YOLO, programska oprema, različica 8.0.0.“ Ultralytics, 2023. https://github.com/ultralytics/ultralytics

[26] Ultralytics, „Configuration.“ 2026. [Splet; dostop 23. 10. 2026]. https://docs.ultralytics.com/usage/cfg/

[27] Y. Shi, L. Cui, Z. Qi, F. Meng, Z. Chen, „Automatic road crack detection using random structured forests,“ *IEEE Transactions on Intelligent Transportation Systems*, let. 17, št. 12, str. 3434–3445, 2016. doi: 10.1109/TITS.2016.2552248

[28] Y.-J. Cha, W. Choi, O. Büyüköztürk, „Deep learning-based crack damage detection using convolutional neural networks,“ *Computer-Aided Civil and Infrastructure Engineering*, let. 32, št. 5, str. 361–378, 2017. doi: 10.1111/mice.12263

[29] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, S. Wang, „DeepCrack: learning hierarchical convolutional features for crack detection,“ *IEEE Transactions on Image Processing*, let. 28, št. 3, str. 1498–1512, 2019. doi: 10.1109/TIP.2018.2878966

[30] Y. Liu, J. Yao, X. Lu, R. Xie, L. Li, „DeepCrack: a deep hierarchical feature learning architecture for crack segmentation,“ *Neurocomputing*, let. 338, str. 139–153, 2019. doi: 10.1016/j.neucom.2019.01.036

[31] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei, H. Ling, „Feature pyramid and hierarchical boosting network for pavement crack detection,“ *IEEE Transactions on Intelligent Transportation Systems*, let. 21, št. 4, str. 1525–1535, 2020. doi: 10.1109/TITS.2019.2910595

[32] D. Dais, İ. E. Bal, E. Smyrou, V. Sarhosis, „Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning,“ *Automation in Construction*, let. 125, članek 103606, 2021. doi: 10.1016/j.autcon.2021.103606

[33] D. Tabernik, S. Šela, J. Skvarč, D. Skočaj, „Segmentation-based deep-learning approach for surface-defect detection,“ *Journal of Intelligent Manufacturing*, let. 31, št. 3, str. 759–776, 2020. doi: 10.1007/s10845-019-01476-x

[34] J. Božič, D. Tabernik, D. Skočaj, „Mixed supervision for surface-defect detection: from weakly to fully supervised learning,“ *Computers in Industry*, let. 129, članek 103459, 2021. doi: 10.1016/j.compind.2021.103459

[35] Y. Li, C. Yin, Y. Lei, J. Zhang, Y. Yan, „RDD-YOLO: road damage detection algorithm based on improved You Only Look Once version 8,“ *Applied Sciences*, let. 14, št. 8, članek 3360, 2024. doi: 10.3390/app14083360

[36] C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, „YOLOv7: uradna izvedba (programska oprema).“ GitHub, 2022. https://github.com/WongKinYiu/yolov7

[37] A. Paszke idr., „PyTorch: an imperative style, high-performance deep learning library,“ v *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, 2019. https://arxiv.org/abs/1912.01703

Priloge

Priloga A: Razredi poškodb zbirke RDD2022

Zbirka RDD2022 razlikuje štiri razrede poškodb [3, str. 2]. Tabela 12 povzema njihove oznake in imena; izvorne definicije iz zbirke RDD2018, iz katere so razredi izpeljani, so navedene v članku [3, str. 3]: vzdolžne razpoke (D00) so razpoke vzdolž smeri vožnje (v zbirki RDD2018 ločene na del v kolesnici in del ob delovnem stiku), prečne razpoke (D10) potekajo prečno (enakomerno razporejene ali ob delovnem stiku), aligatorske razpoke (D20) so mrežasto pokanje dela ali celotne površine, razred D40 pa je v zbirki RDD2018 širši razred drugih poškodb (kolesnice, grbine, udarne jame, ločitve), v zbirki RDD2022 pa vsebuje udarne jame.

Oznaka	Razred	Angleško ime v zbirki
D00	vzdolžna razpoka	longitudinal crack
D10	prečna razpoka	transverse crack
D20	mrežasta (aligatorska) razpoka	alligator crack
D40	udarna jama	pothole

Tabela 12: Razredi poškodb v zbirki RDD2022

Vir: povzeto po [3, str. 2–3]

Priloga B: Hiperparametri učenja

Tabela 13 navaja hiperparametre, ki smo jih izbrali ali privzeli. Za YOLOv8m so vse nastavitve razen navedenih privzete v knjižnici Ultralytics [26], za YOLOv7 so privzete vrednosti iz datoteke hyp.scratch.p5 uradnega repozitorija [36].

Nastavitev	Faster R-CNN	YOLOv7	YOLOv8m
Optimizator	SGD	SGD	Adam
Začetni korak učenja	0,01	0,01	0,001
Gibalna količina / β1	0,9	0,937	privzeta
Utežni razpad	0,0001	0,0005	0,0005
Število epoh	24	100	100
Velikost paketa	8	16	16
Ogrevanje	500 iteracij	3 epohe	3 epohe
Vhodna slika	640 × 640	640 × 640	640 × 640
Razširjanje	zrcaljenje	privzeto (mozaik, mešanje slik, premik, skaliranje, zrcaljenje)	privzeto (mozaik, barvne spremembe, premik, zrcaljenje)
Začetne uteži	COCO	COCO	COCO

Tabela 13: Hiperparametri učenja treh modelov

Vir: lastna izbira; privzete vrednosti po [26] in [36]

Priloga C: Postopek vrednotenja v obliki psevdokode

Postopek, po katerem skripta za vrednotenje iz zaznav in označb izračuna AP, je naslednji (vsi koraki veljajo za en razred poškodbe in eno ponovitev):

1. Zaznave razreda uredi po padajoči oceni zaupanja.
2. Za vsako zaznavo najdi označene primerke istega razreda na isti sliki, ki še niso ujeti.
3. Izračunaj IoU zaznave z vsakim od njih; če je največji IoU vsaj 0,5, označi zaznavo kot pravilno (TP) in označeni primerek kot ujet, sicer kot lažno pozitivno (FP).
4. Po vrsti izračunaj kumulativno število TP in FP; priklic je kumulativni TP, deljen s številom označenih primerkov razreda, natančnost pa kumulativni TP, deljen z vsoto TP in FP.
5. Natančnost popravi tako, da je vsaka vrednost enaka največji natančnosti pri enakem ali večjem priklicu.
6. AP je vsota ploščin pravokotnikov pod popravljeno krivuljo natančnosti glede na priklic.
7. mAP@0,5 je povprečje AP po štirih razredih.

Pri vzorčenju po slikah se v koraku 4 vsaka zaznava in vsak označeni primerek šteje tolikokrat, kolikokrat je bila slika izbrana v vzorec. Celotna skripta je del priloženih datotek naloge (skripta 02_ovrednoti).

Priloga D: Matrika zmede modela YOLOv8m v številkah

Tabela 14 podaja število označenih primerkov po pravih razredih (vrstice) in napovedanih razredih (stolpci), iz katerih so izračunani deleži na sliki 9. Stolpec »ni zaznano« vsebuje primerke, ki jih model ni zaznal ali jih je zaznal s premajhnim prekrivanjem, vrstica »ozadje« pa lažne alarme na ozadju po napovedanem razredu.

Pravi razred	D00	D10	D20	D40	ni zaznano	Skupaj
D00	1 536	35	19	7	819	2 416
D10	14	710	8	1	394	1 127
D20	21	17	677	10	336	1 061
D40	7	3	9	391	241	651
ozadje	193	125	145	75	–	–

Tabela 14: Matrika zmede modela YOLOv8m na testni množici (število primerkov)

Vir: lastni rezultati

Priloga E: Natančnost, priklic in posamezne ponovitve

Tabela 15 podaja natančnost in priklic po razredih za vse tri modele (povprečje treh ponovitev, pri pragu iz validacije), tabela 16 pa rezultate vsake ponovitve posebej.

Razred	FRCNN natančnost	FRCNN priklic	YOLOv7 natančnost	YOLOv7 priklic	YOLOv8m natančnost	YOLOv8m priklic
D00	0,77	0,52	0,81	0,57	0,84	0,62
D10	0,71	0,52	0,75	0,57	0,78	0,62
D20	0,69	0,53	0,74	0,58	0,79	0,64
D40	0,67	0,48	0,73	0,55	0,80	0,60
Mikro	0,73	0,52	0,77	0,57	0,81	0,62

Tabela 15: Natančnost in priklic po razredih za tri modele na testni množici

Vir: lastni rezultati

Konfiguracija	Seme	Prag	mAP validacija	mAP test	Mikro F1 test
Faster R-CNN	11	0,50	0,48	0,50	0,61
Faster R-CNN	22	0,50	0,48	0,50	0,61
Faster R-CNN	33	0,55	0,47	0,48	0,59
YOLOv7	11	0,50	0,55	0,57	0,66
YOLOv7	22	0,55	0,55	0,56	0,66
YOLOv7	33	0,55	0,54	0,55	0,64
YOLOv8m	11	0,55	0,63	0,63	0,71
YOLOv8m	22	0,55	0,60	0,62	0,71
YOLOv8m	33	0,55	0,59	0,60	0,70
YOLOv8m brez razširjanja	11	0,50	0,57	0,60	0,69
YOLOv8m brez razširjanja	22	0,55	0,55	0,58	0,67
YOLOv8m brez razširjanja	33	0,55	0,55	0,56	0,67
YOLOv8m brez prenosa	11	0,50	0,56	0,58	0,68
YOLOv8m brez prenosa	22	0,50	0,55	0,57	0,66
YOLOv8m brez prenosa	33	0,50	0,54	0,56	0,65
YOLOv8m brez obojega	11	0,50	0,45	0,48	0,59
YOLOv8m brez obojega	22	0,55	0,47	0,48	0,60
YOLOv8m brez obojega	33	0,55	0,48	0,49	0,60

Tabela 16: Rezultati posameznih ponovitev (prag zaupanja izbran na validacijski množici)

Vir: lastni rezultati

Priloga F: Uporaba orodij umetne inteligence

Navodila fakultete zahtevajo, da navedemo načine uporabe orodij umetne inteligence. Pri pripravi naloge so bila orodja umetne inteligence uporabljena za jezikovne popravke in preoblikovanje lastnih osnutkov besedila, za iskanje virov ter za pripravo osnutkov programske kode za obdelavo podatkov in izris grafov. Vsebino besedila, rezultate, kodo in vse navedene vire je avtor pregledal in preveril; vsi viri so bili odprti in primerjani z navedbami v besedilu. Orodja umetne inteligence niso navedena kot avtor ali soavtor, odgovornost za vsebino dela nosi avtor.