

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
ELEKTROTEHNIČKI FAKULTET

Raspoznavanje uzoraka i strojno učenje

Projektni zadatak – Detektor objekata u prometu

Antonio Antunović

Osijek, 2023.

SADRŽAJ

1. UVOD	1
1.1. Pregled postojećih rješenja	1
1.2. Opis rješenja	2
2. YOLOv5 DETEKTOR	3
2.1. Evaluacija detektora na Kitti skupu podataka	4
2.2. Evaluacija detektora na BDD10k skupu podataka	7
3. DETR DETEKTOR	9
3.1. Evaluacija detektora na BDD10k skupu podataka	9
3.2. Evaluacija detektora na Kitti skupu podataka	10
4. ZAKLJUČAK	12

1. UVOD

Tehnološki napredak u području računalnog vida i strojnog učenja otvorio je nove mogućnosti razvoja tehnologije koja bi učinila svakodnevni život jednostavnijim i sigurnijim. Autonomna vožnja predstavlja jedno od najznačajnijih područja istraživanja i razvoja u posljednjim godinama. Autonomna vožnja, obuhvaća širi spektar tehnologija i postignuća. Kombinacija računalnog vida, senzora i algoritama strojnog učenja omogućuje vozilu da autonomno percipira okolinu, donosi odluke i upravlja vozilom. Važan napredak ostvaren je u području percepcije, gdje su razvijene tehnike za prepoznavanje prometnih znakova, semantičku segmentaciju cesta i praćenje vozila. Stoga razvoj detektora objekata za prometne situacije je od iznimne važnosti za uspješnu implementaciju autonomne vožnje.

1.1. Pregled postojećih rješenja

Trenutno postoje brojna istraživanja koja se bave različitim aspektima autonomne vožnje i tehnologija koje se koriste. Neki od trenutno najnaprednijih istraživačkih radova u području autonomne vožnje uključuje sljedeće:

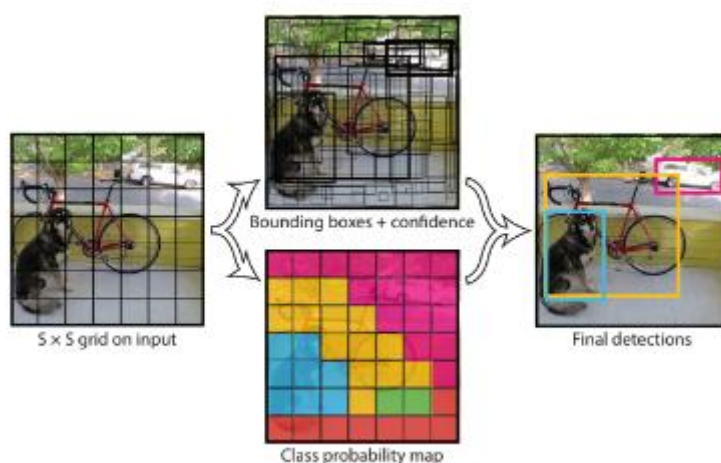
- 1) Percepcija: Unapređenje sposobnosti autonomnih vozila za precizno razumijevanje okoline. To uključuje napredak u detekciji objekata, semantičkoj segmentaciji i 3D percepciji pomoću tehnologija poput LiDAR-a, radara i kamera[1].
- 2) Donošenje odluka i upravljanje: Uključuje planiranje sigurnih i učinkovitih trajektorija za autonomna vozila, razvoj algoritama koji mogu donositi sigurne odluke u složenim prometnim situacijama u stvarnom vremenu[1].
- 3) Mapiranje i lokalizacija: Odnosi se na razvoj kvalitetnih karata, algoritama mapiranja i preciznih tehnika lokalizacije koje se mogu nositi sa stvarnim scenarijima[2].
- 4) Simulacija i virtualno testiranje: Simulacije pružaju siguran i ekonomičan način testiranja i validacije algoritama i sustava za autonomnu vožnju. Istraživanja se fokusiraju na razvoj realističnih virtualnih okruženja koja mogu simulirati različite prometne scenarije, vremenske uvjete i izazovne situacije kako bi se osigurala pouzdanost i sigurnost[3].

1.2. Opis rješenja

Zadatak ovog rada jest razviti dva detektora objekata u prometu, trenirati ih na različitim bazama podataka, evaluirati i testirati razvijene modele. Prvi model treniran na kitti 2D object detection bazi podataka temelji se na YOLOv5 arhitekturi konvolucijske neuronske mreže. Drugi model je treniran na BDD10k bazi podataka, a temelji na DETR (*engl. DEtection TRansformer*) arhitekturi neuronske mreže.

2. YOLOv5 DETEKTOR

YOLOv5 predstavlja 5. verziju YOLO(*engl. You Only Look Once*) detekcijskog modela. Neuronske mreže za detekciju koje se temelje na *region proposal* algoritmu prvo generiraju potencijalne granične okvire (*engl. Bounding box*) na slici, nakon toga klasificiraju dobivene okvire i postprocesiraju rezultate. Ovakve metode su često spore i teške za optimizaciju, jer se svaka komponenta mora tretirati zasebno. YOLO algoritam s druge strane umjesto složenih postupaka izdvajanja značajki i klasifikacije koristi samo jedan prolaz kroz mrežu kako bi izvršio detekciju objekata. Glavna značajka YOLO algoritma jest da se ulazna slika dijeli na $S \times S$ ćelija, gdje se za svaku ćeliju predviđa N bounding box-ova, postotak sigurnosti da se u ćeliji nalazi objekt i vjerojatnost pripadnosti određenoj klasi. Na slici 2.1. shematski je prikazan rad YOLO algoritma.



SI.2.1. Yolo algoritam [6]

Svaki bounding box je definiran s pet brojeva: x i y predstavljaju koordinate centra okvira, w i h predstavljaju širinu i visinu okvira i C vjerojatnost pripadnosti određenoj klasi.

Prije početka treniranja potrebno je preprocesirati podatke. Ovaj postupak uključuje agumentaciju slika, podjelu slika u skupove za treniranje, validaciju i testiranje i prilagoditi oznake slika formatu koji model očekuje. Kitti baza podataka se sastoji od otprilike 7500 slika. Međutim, da bi se postigli zadovoljavajući rezultati rada detektora, preporučuje se imati više slika na raspolaganju. Iz tog razloga i da bi se uvela raznolikost u podatke, a samim time i povećala robusnost modela slike su

agumentirane. Na originalne slike dodan je šum, zamućenje i pretvorene su u slike sivih tonova, što je prikazano na slici 2.2.

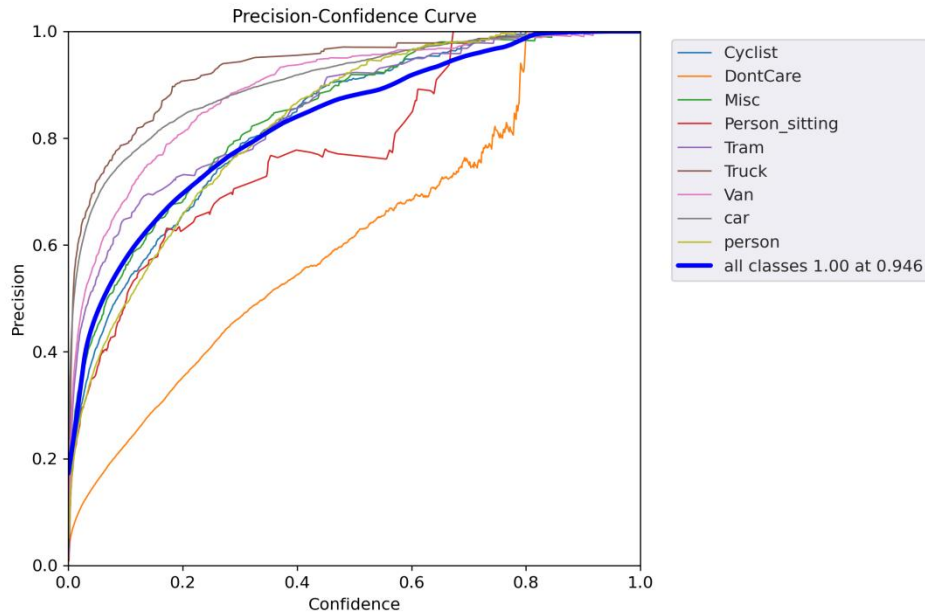


Sl.2.2. Agumentacija slika

Nakon agumentacije slika i prilagodbe njihovih oznaka pokrenuto je učenje modela.

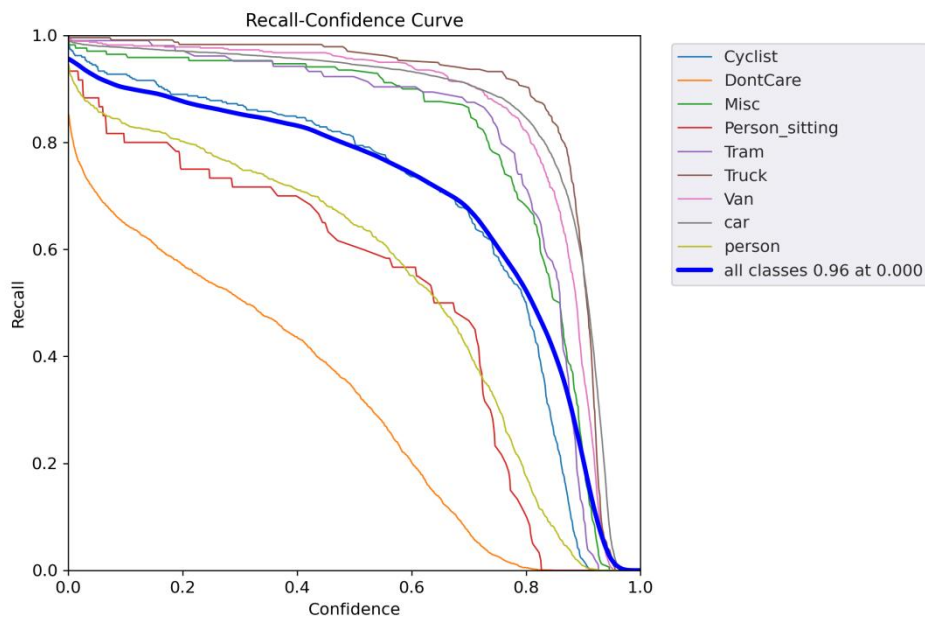
2.1. Evaluacija detektora na Kitti skupu podataka

Evaluacija modela ključni je korak u procjeni performansi modela, njegove sposobnosti za detekciju objekata u stvarnim situacijama i identifikaciji mogućnosti poboljšanja. Evaluacija je napravljena na kitti i BDD10k skupu podataka i zabilježena je preciznost, odziv i srednja prosječna preciznost. Na sljedećim slikama prikazane su metrike evaluacije za kitti skup podataka.



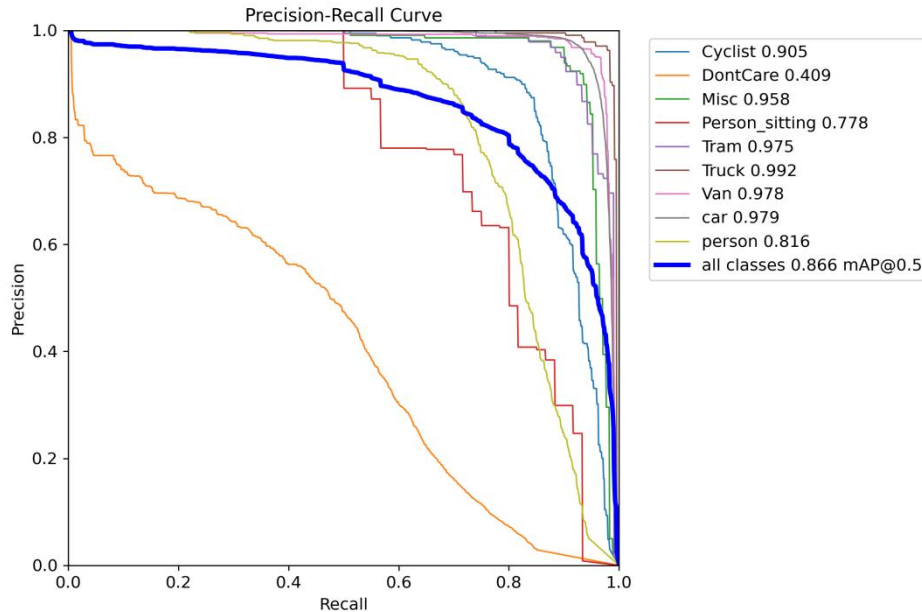
SI.2.3. Precision-Confidence krivulja

Preciznost se definira formulom: $Precision = \frac{True\ positive}{True\ positive + False\ positive}$, tj. govori koliki postotak detektiranih objekata zaista pripada određenoj klasi. Iz gornjeg grafa uočljivo je da model imao najbolji rezultat preciznosti za detekciju kamiona, kombi vozila i automobila, zato što su upravo te tri klase bile najzastupljenije u podatkovnom skupu. S druge strane, manje zastupljene klase su imale lošiju preciznost i negativno su utjecale na prosječnu preciznost modela (plava krivulja).



SI.2.4. Recall-Confidence krivulja

Odziv je definiran formulom: $Recall = \frac{True\ positive}{True\ positive + False\ negative}$, a predstavlja mjeru koja govori koliko je instanci određene klase detektirano od svih prisutnih instance iste klase na slici. Za vrijednosti sigurnosti između 0.8 i 0.9 model pokazuje vrlo dobre rezultate odziva kod klasa koje su zastupljenije u podatkovnom skupu, dok slabije zastupljene klase imaju lošiji odziv, kao i kod grafa preciznosti.



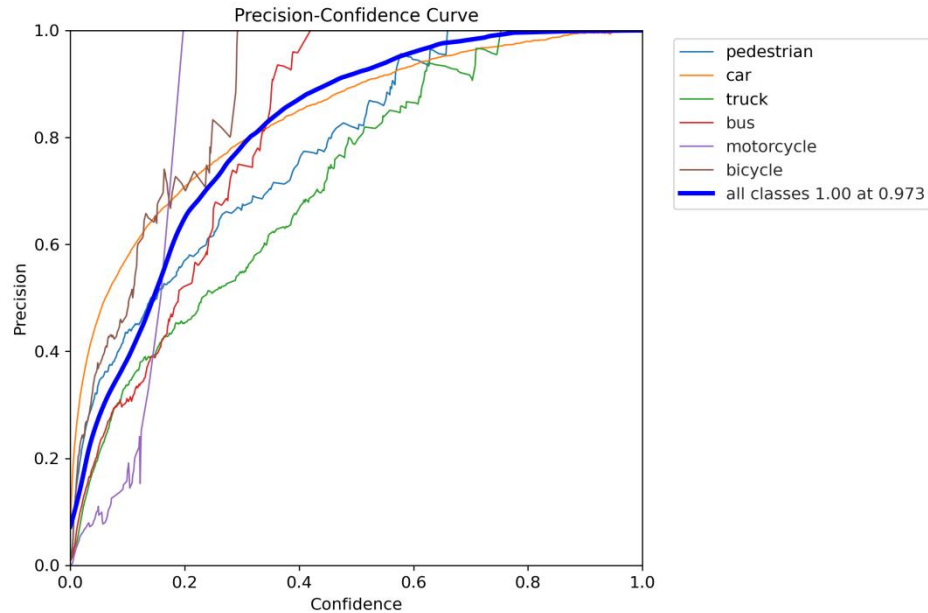
Sl.2.5. Precision-Recall krivulja

Precision-Recall krivulja predstavlja odnos preciznosti i odziva za postavljenu granicu IoU (*Intersection over Union*), u ovom slučaju 0.5. mAP@50 jest srednja prosječna preciznost i računa se kao površina ispod precision-recall krivulje. Ova metrika je vrlo korisna i često se upotrebljava jer može vrlo dobro opisati rad modela jer u sebi sadrži informaciju o preciznosti i o odzivu.

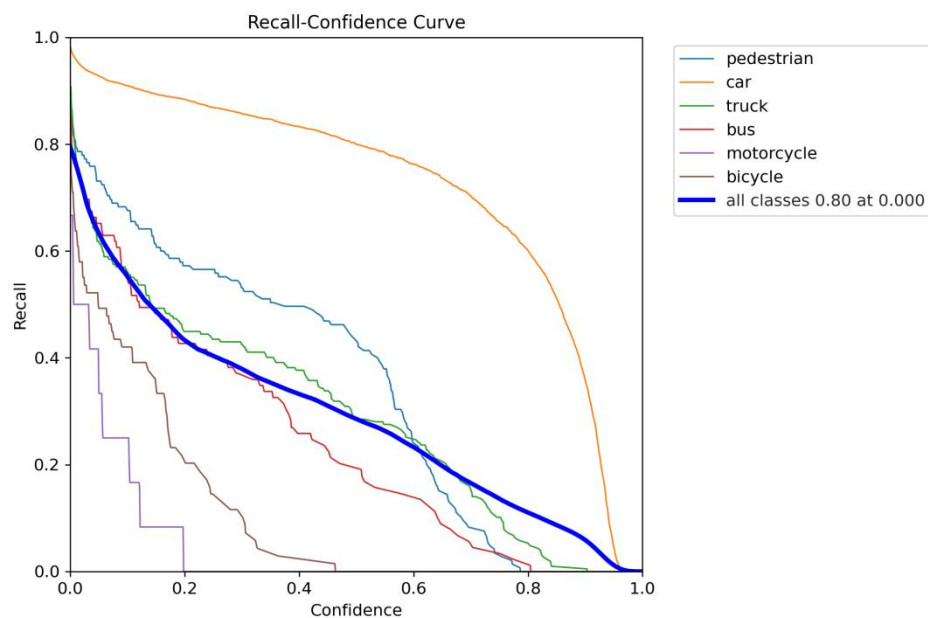
2.2. Evaluacija detektora na BDD10k skupu podataka

Isti postupak evaluacije YOLOv5 modela proveden je i na BDD10k skupu podataka.

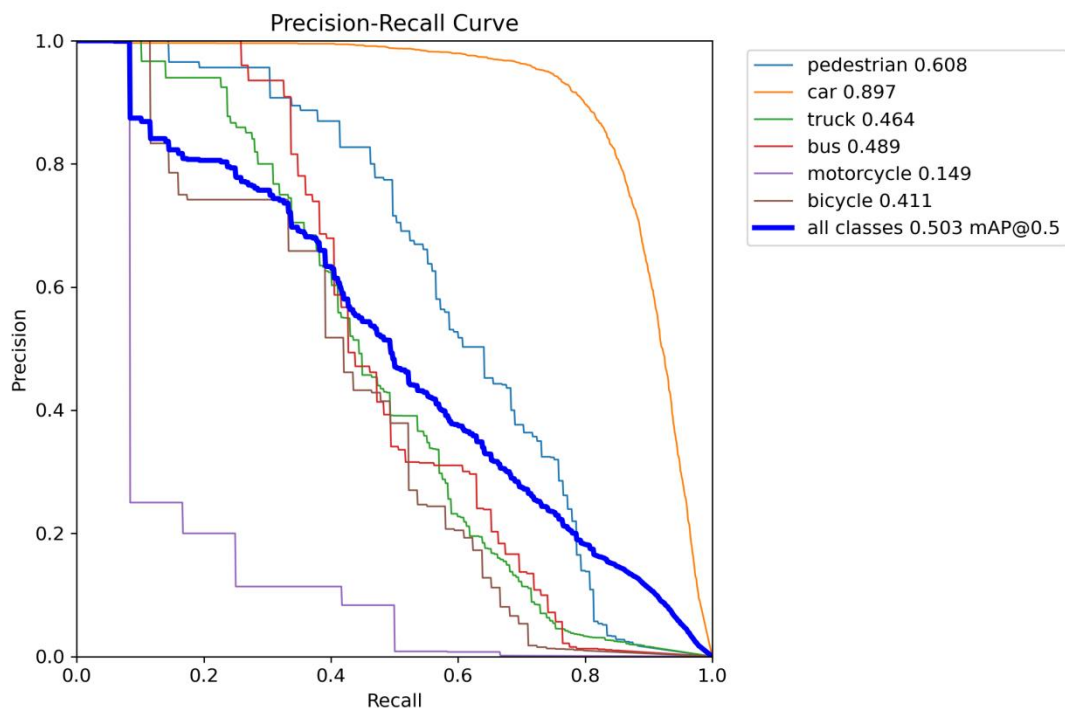
Na sljedećim slikama prikazani su grafovi evaluacije za navedeni skup.



Sl.2.6. Precision-Confidence krivulja

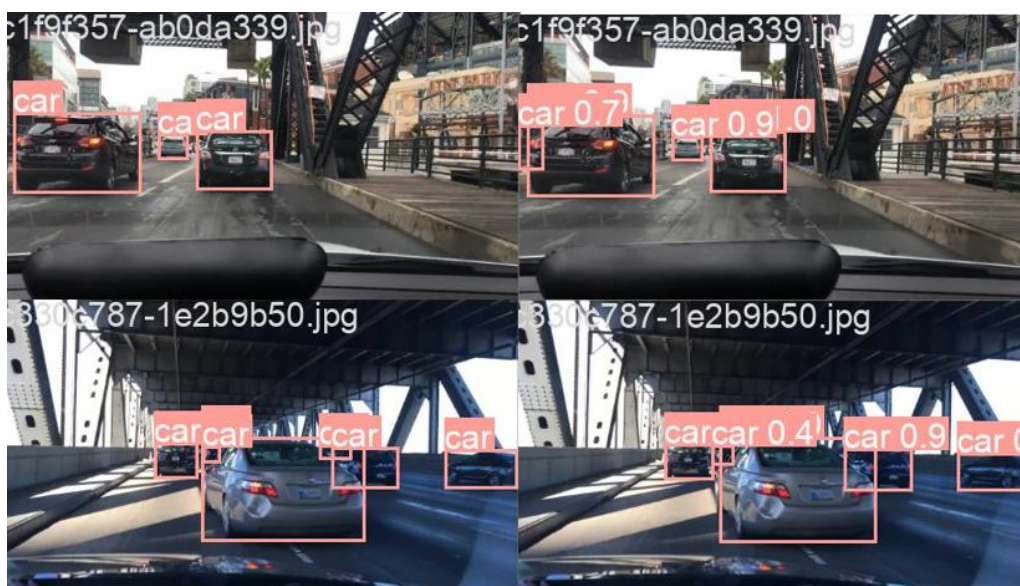


Sl.2.7. Recall-Confidence krivulja



SI.2.8. Precision-Recall krivulja

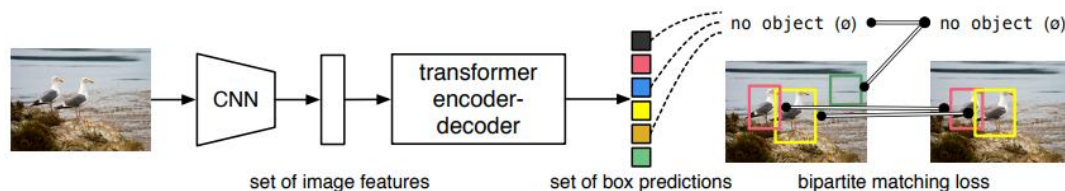
Klase prisutne u ovom skupu podataka razlikuju se od klasa prisutnih u kitti skupu, također zbog različitosti skupova podataka model je pokazao lošije performanse na BDD10k skupu. Problem kod BDD10k skupa jest što klase nisu ravnomjerno raspoređene, tj. klasa automobil je prezastupljena u skupu, zato je model na toj klasi pokazao najbolje rezultate. Kada se promatraju rezultati za klasu automobil zasebno, mogu se usporediti s evaluacijskim rezultatima dobivenim na kitti skupu. Na slici 2.9. prikazane su slike s ground truth oznakama i predviđanje modela za te slike.



SI.2.9. Ground truth slike i predviđanje modela

3. DETR DETEKTOR

DETR (*DEtection TRansformer*) detektor, za razliku od klasičnih detektora objekata, detekciji objekata pristupa tako da direktno predviđa skup detektiranih objekata koristeći konvolucijske slojeve i transformer arhitekturu. Arhitektura DETR mreže prikazana je na slici 3.1.



SI.3.1. DETR arhitektura [7]

Transformer se sastoji od enkodera i dekodera, glavni mehanizam ove arhitekture neuronske mreže jest da kodira međusobni odnos ulaznih vrijednosti, u ovom slučaju značajki dobivenih konvolucijom i ulazne vrijednosti promatra kao cjelinu. Ovakav mehanizam omogućuje neuronskoj mreži da napravi predviđanje na temelju odnosa između objekata na slici[7]. Zbog toga se ovakva arhitektura neuronskih mreža često koristi za NLP (*Natural Language Processing*).

DETR detektor je treniran na BDD10k bazi podataka. Postupak pretprocesiranja slika je jednak kao i u slučaju YOLO detektora. Nakon učenja modela provedena je evaluacija na BDD10k i kitti bazi podataka. Za evaluaciju rada DETR koristi COCO-Evaluator koji za evaluaciju pruža malo drugačije sučelje od YOLO api-a.

3.1. Evaluacija detektora na BDD10k skupu podataka

Na slici 3.2. prikazane su metrike evaluacije DETR modela na BDD10k validacijskom skupu podataka.

```

IoU metric: bbox
Average Precision (AP) @[ IoU=0.50:0.95 | area= all | maxDets=100 ] = 0.659
Average Precision (AP) @[ IoU=0.50 | area= all | maxDets=100 ] = 0.740
Average Precision (AP) @[ IoU=0.75 | area= all | maxDets=100 ] = 0.585
Average Precision (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.353
Average Precision (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.632
Average Precision (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.839

Average Recall (AR) @[ IoU=0.50:0.95 | area= all | maxDets= 1 ] = 0.359
Average Recall (AR) @[ IoU=0.50:0.95 | area= all | maxDets= 10 ] = 0.502
Average Recall (AR) @[ IoU=0.50:0.95 | area= all | maxDets=100 ] = 0.685
Average Recall (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.425
Average Recall (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.718
Average Recall (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.885

```

Sl.3.2. DETR evaluacijske metrike

Za evaluaciju ovog modela korištene su metrike prosječne preciznosti i prosječnog odziva za različite vrijednosti IoU parametra. Za usporedbu s YOLO modelom može se promatrati prosječna preciznost za IoU=0.5, koja iznosi 0.74, što je malo lošije nego ista vrijednost YOLO modela. Uz to DETR pruža metrike prosječne preciznosti i odziva za različite vrijednosti IoU i za različite veličine detektiranih graničnih okvira objekta. Iz toga se može zaključiti koliko dobro model detektira male objekte na slikama. Mali predmeti na slikama bi za ovaj problem najvjerojatnije predstavljali vozila, ljudi i druge objekte koji se nalaze daleko od kamere. Za poboljšanje ovog parametra bilo bi potrebno u bazu podataka dodati više slika s udaljenijim objektima. Manjkavost COCO-Evaluatora je što ne pruža metrike evaluacije po određenim klasama, pa se ne može odrediti koliko dobro model detektira objekte koji su manje zastupljeni u bazi podataka.

3.2. Evaluacija detektora na Kitti skupu podataka

Na slici 3.3. prikazane su evaluacijske metrike DETR modela na kitti skupu podataka.

IoU metric: bbox

Average Precision	(AP) @[IoU=0.50:0.95 area= all maxDets=100]	= 0.532
Average Precision	(AP) @[IoU=0.50 area= all maxDets=100]	= 0.613
Average Precision	(AP) @[IoU=0.75 area= all maxDets=100]	= 0.399
Average Precision	(AP) @[IoU=0.50:0.95 area= small maxDets=100]	= 0.302
Average Precision	(AP) @[IoU=0.50:0.95 area=medium maxDets=100]	= 0.600
Average Precision	(AP) @[IoU=0.50:0.95 area= large maxDets=100]	= 0.712

Average Recall	(AR) @[IoU=0.50:0.95 area= all maxDets= 1]	= 0.412
Average Recall	(AR) @[IoU=0.50:0.95 area= all maxDets= 10]	= 0.425
Average Recall	(AR) @[IoU=0.50:0.95 area= all maxDets=100]	= 0.556
Average Recall	(AR) @[IoU=0.50:0.95 area= small maxDets=100]	= 0.401
Average Recall	(AR) @[IoU=0.50:0.95 area=medium maxDets=100]	= 0.618
Average Recall	(AR) @[IoU=0.50:0.95 area= large maxDets=100]	= 0.722

SI.3.3. DETR evaluacijske metrike

Evaluacije modela na Kitti skupu je pokazala lošije rezultate nego na skupu na kojem je model naučen, zato što se skupovi podataka razlikuju. Razlike u podatkovnim skupovima mogu biti specifičnost prometnih okruženja, osvijetljenost, različite zastupljenosti instanci klasa itd. Na slici 3.4. prikazana su slike s ground truth oznakama i predviđanje modela za te slike.



SI.3.4. Ground truth slike i predviđanje modela

4. ZAKLJUČAK

Razvojem tehnologija u području računalnog vida i stajnog učenja otvorile su se mogućnosti njihove primjene u raznim područjima znanosti, pa tako i za potrebe autonomne vožnje. Detekcija objekata u prometnim situacijama izazovan je zadatak koji bi uvelike doprinio razvoju autonomne vožnje. U sklopu ovog projekta razvijena su dva detektora objekata za prometne situacije temeljena na različitim arhitekturama neuronskih mreža. Svaki detektor je treniran na zasebnoj bazi podataka kako bi se osigurala raznolikost i povećala sposobnost modela da se nosi s različitim situacijama. U tu svrhu, primijenjena je augmentacija podataka na slikama, uključujući dodavanje šuma, zamagljivanje i pretvaranje u slike sivih tonova. Ovaj pristup je bio koristan za povećanje robusnosti modela i dostupnosti većeg broja slika za učenje. Prvi detektor se temelji na YOLO arhitekturi, koja je poznata po svojoj brzini i sposobnosti detektiranja objekata u stvarnom vremenu. Ovaj model je optimiziran za brzu obradu podataka i omogućava detekciju objekata u stvarnom vremenu bez gubitka preciznosti. Drugi detektor se temelji na neuronskoj mreži koja sadrži transformer arhitekturu. Ova arhitektura omogućuje modelu da nauči složene odnose između objekata. Korištenje transformer arhitekture omogućuje modelu da obuhvati šire kontekstualne informacije i bolje razumije odnose između objekata na slici. Oba detektora su pokazala zadovoljavajuće rezultate detekcije na testnim skupovima baza podataka na kojima su trenirani. Međutim, kada su primijenjeni na slikama iz drugih baza podataka, rezultati detekcije nisu bili jednako dobri. Ovo može biti posljedica različitosti baza podataka, različitih prometnih okruženja i varijacija u oznakama objekata. Važno je razumjeti da je generalizacija detektora na različite skupove podataka i scenarije izazovan problem koji zahtijeva daljnje istraživanje i optimizaciju modela.

LITERATURA

- [1] You, Dingyi & Haiyan, Wang & Kaiming, Yang. (2018). State-of-the-art and trends of autonomous driving technology. 1-8. 10.1109/TEMS-ISIE.2018.8478449.
- [2] T. Qin, Y. Zheng, T. Chen, Y. Chen and Q. Su, "A Light-Weight Semantic Map for Visual Localization towards Autonomous Driving," 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 2021
- [3] T. Qin, Y. Zheng, T. Chen, Y. Chen and Q. Su, "A Light-Weight Semantic Map for Visual Localization towards Autonomous Driving," 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 2021
- [4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
- [5] Carion, Nicolas, et al. "End-to-end object detection with transformers." Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16. Springer International Publishing, 2020.