Evaluacija ograničene perspektive za segmentaciju prometnog okruženja

Babić, Lucija

Sveučilište Jurja Dobrile u Puli

Fakultet informatike u Puli

LUCIJA BABIĆ

EVALUACIJA OGRANIČENE PERSPEKTIVE ZA SEGMENTACIJU PROMETNOG OKRUŽENJA

Diplomski rad

Pula, rujan 2023. godine

Sveučilište Jurja Dobrile u Puli

Fakultet informatike u Puli

LUCIJA BABIĆ

EVALUACIJA OGRANIČENE PERSPEKTIVE ZA SEGMENTACIJU PROMETNOG OKRUŽENJA

Diplomski rad

JMBAG: 0054045963, redoviti student

Studijski smjer: Informatika

Kolegij: Neuronske mreže i duboko učenje

Znanstveno područje: Društvene znanosti

Znanstveno polje: Informacijske i komunikacijske znanosti

Znanstvena grana: Informacijski sustavi i informatologija

Mentor: doc. dr. sc. Goran Oreški

Pula, rujan 2023. godine

IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani Lucija Babić, kandidat za magistra informatike ovime izjavljujem da je ovaj Diplomski rad rezultat isključivo mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio Diplomskog rada nije napisan na nedozvoljeni način, odnosno da je prepisan iz kojega necitiranog rada, te da ikoji dio rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Student _____

U Puli, 1. rujna 2023.

IZJAVA O KORIŠTENJU AUTORSKOG DJELA

Ja, Lucija Babić dajem odobrenje Sveučilištu Jurja Dobrile u Puli, kao nositelju prava iskorištavanja, da moj diplomski rad pod nazivom Evaluacija ograničene perspektive za segmentaciju prometnog okruženja

koristi na način da gore navedeno autorsko djelo, kao cjeloviti tekst trajno objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te kopira u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu s Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnoga pristupa znanstvenim informacijama.

Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

U Puli, 1. rujna 2023.

Potpis

Evaluacija ograničene perspektive za segmentaciju prometnog okruženja

Lucija Babić

Sažetak: Kako autonomna vožnja brzo postaje stvarnost, najveći izazov s autonomnim vozilima ostaje sigurnost, budući da automobil mora obraditi okolinu kako bi donio ispravne predikcije. Ovaj rad bavi se detekcijom i klasifikacijom dinamičkih objekata u prometu, poput vozila i pješaka, na razini piksela. Predstavljena su dva eksperimenta koja se razlikuju samo po podacima korištenim tijekom procesa treninga: jedan koristi prednju kameru, dok drugi kombinira prednju i lijevo-usmjerenu kameru kako bi uhvatio dijagonalni pogled. Prikupljanje podataka oslanja se na CARLA simulator, namjerno isključujući standardne tehnologije poput lidara ili radara, fokusirajući se isključivo na monokularne podatke. Cilj ovog istraživanja je ispitati sposobnosti generalizacije modela obučenih na ograničenim perspektivama kada su izloženi podacima iz kamera koje hvataju potpun 360° pogled na njihovu okolinu. Navedeni problem adresira se zadatkom segmentacije instanci koristeći najpoznatije algoritme za segmentaciju instanci: Mask Region-based Convolutional Neural Network i YOLOv7. Rezultati pokazuju do koje mjere ograničeni monokularni sustav može biti korišten za segmentaciju instanci perspektiva koje se razlikuju od perspektiva uključenih u fazu treninga.

Ključne riječi: dinamički objekti u prometu; vozila; prolaznici; segmentacija instanci; autonomna vožnja; YOLO; Mask R-CNN

Evaluating the Limited Perspective Training for Full-View Traffic Instance Segmentation

Lucija Babić

Abstract: As autonomous driving is quickly becoming a reality, the most considerable challenge with self-driving remains safety, as the car must process surroundings to make the correct predictions. This thesis delves into the detection and classification of dynamic objects in traffic, such as vehicles and pedestrians, on a pixel level. Two experiments are presented, distinguished only by the data used during the training process: one utilizing a front-facing camera and the other a combination of front and left-facing cameras to capture a diagonal view. Data collection relied on the CARLA simulator, purposely omitting standard technologies like lidar or radar, and focusing exclusively on monocular data. The objective of this research is to examine the generalization capabilities of models trained on restricted perspectives when exposed to inputs from cameras capturing a full 360° view of their surroundings. The described challenge is tackled with the task of instance segmentation using the most famous instance segmentation algorithms: Mask Region-based Convolutional Neural Network and YOLOv7. The results show to what extent a restricted monocular setup can be used for the instance segmentation of the perspectives that differ from the perspectives included in the training phase.

Keywords: dynamic traffic objects; vehicles; pedestrians; instance segmentation; autonomous driving; YOLO; Mask R-CNN

Sadržaj

1 Uvod

Danas je uobičajena pojava da su vozila opremljena ugrađenim kamerama. Kamere služe različitim svrhama, od pomoći vozačima tijekom parkiranja i vožnje unatrag do pružanja sigurnosnih snimaka. Integracija sustava kamera potaknula je razvoj naprednih sustava pomoći vozačima, pružajući vozačima upozorenja i podrška za sigurnije iskustvo vožnje [CHL08]. Primjena takvih sustava značajno poboljšava sigurnost prometa, s obzirom na to da ljudska pogreška i dalje predstavlja glavni uzrok prometnih nesreća, kako izvješćuje Europska komisija (2019, 2020).

Sve te tehnološke napretke otvorile su put razvoju najkompleksnijih sustava koji imaju ogroman potencijal za transformaciju prometa u budućnosti - sustavima autonomne vožnje. Stoga se u skoroj budućnosti očekuje široka primjena autonomnih vozila (AV) na gradskim cestama [Che+23]. Jedan od najvećih izazova u razvoju bilo kojeg produkcijski spremnog modela dubokog učenja je kvaliteta i količina podataka. To posebno vrijedi za AV, gdje se ogromna sredstva dodjeljuju za prikupljanje bogatih stvarnih skupova podataka. Tvrtke snimaju promet koristeći različite senzore koji prikupljaju petabajte podataka. Prikupljanje i pohrana te količine, kao i senzori koji se koriste za snimanje, su skupi.

Osnovni cilj ovog diplomskog rada je istražiti učinkovitost treninga modela na ograničenim perspektivama kamera umjesto na potpunom 360-stupanjskom pogledu kako bi se točno segmentirali pokretni objekti u prometu poput vozila i prolaznika. Motivacija za korištenje ovih ograničenih perspektiva je potencijalno smanjenje resursa potrebnih za prikupljanje podataka i trening modela. To bi mogao učiniti cijeli proces ekonomičnijim, bržim i jednostavnijim za izvođenje. Dakle, fokus je na dinamičnom prometnom okruženju, točnije vozilima i prolaznicima, te se uspoređuju performanse algoritama Mask R-CNN [He+17] i YOLO [Red+16].

U tijelu rada prvo je pružen duboki kontekst problema s kojim se suočavamo i specifičnog područja istraživanja na koji se fokusiramo, praćen detaljnim opisima arhitektura modela koji se koriste u ovom istraživanju. Zatim se bavimo metodologijom istraživanja, što je posebno važno s obzirom da oba eksperimenta koriste slične pristupe. Nakon toga

provodimo dva odvojena eksperimenta koristeći različite skupove podataka, ali primjenjujući iste modele. Naposljetku pažljivo procjenjujemo razlike u rezultatima izvedbe modela i zaključujemo rad sveobuhvatnim zaključkom i sažetkom.

2 Kontekstualni pregled

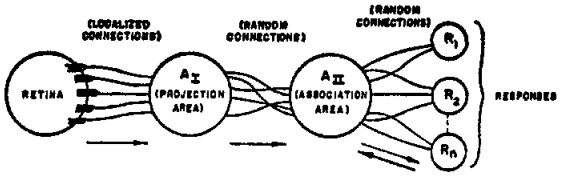
Ovo poglavlje pruža kratki povijesni pregled razvoja umjetne inteligencije, s posebnim naglaskom na evoluciju neuronskih mreža i dubokog učenja. Pružen je dublji pogled u osnovne teorije i ključne prekretnice koje su oblikovale polje, od Alan Turingovog seminalnog rada do Dartmouth radionice itd. Kako se tematika razvija, sve više se usmjerava prema dubokom učenju kao specifičnoj podgrani strojnog učenja, posebno ističući njegovu ključnu ulogu u napretku računalnog vida. Poglavlje završava predstavljanjem primarnog predmeta istraživanja ovog rada: segmentacijom instanci unutar domene računalnog vida. Osim toga, pregled dotiče povezana istraživanja, konkretno raspravljajući o tome kako su tehnike segmentacije instanci primijenjene u polju autonomne vožnje.

2.1 Porijeklo umjetne inteligencije i neuronskih mreža

Iako je teško odrediti početak povijesti umjetne inteligencije *(eng. AI)*, značajan početak mogla bi biti prva pomalo fizička forma AI-a koja je potaknula šire rasprave o inteligenciji strojeva. Godine 1950., engleski matematičar Alan Turing objavio je seminalni članak pod naslovom "Computing Machinery and Intelligence", koji je postavio temelje za procjenjivanje inteligencije strojeva. Ovaj okvir postao je poznat kao Turingov test. Turingov test postao je prekretnica u polju AI-a, služeći kao referenca za procjenu inteligencije strojeva. Prema ovom testu, ako bi stroj mogao komunicirati s čovjekom na takav način da čovjek ne može razlikovati stroj od drugog čovjeka, stroj bi se smatrao inteligentnim. Turingov test pružio je početni standard za procjenu potencijala i ograničenja umjetne inteligencije, potičući rasprave koje su i danas relevantne o tome što čini inteligenciju i kako bi se ona mogla umjetno replicirati. [HK19]

Godine 1956., na Dartmouth Collegeu u New Hampshireu, održana je radionica koja je nazvana Dartmouth radionicom *(eng. Dartmouth workshop)*. Cilj ovog događaja bio je istražiti potencijal strojeva da oponašaju ljudsku inteligenciju i bila je ključna u stvaranju izraza "Umjetna Inteligencija". Radionica je okupila istraživače iz različitih područja s ciljem stvaranja novog interdisciplinarnog područja istraživanja usmjerenog na razvoj strojeva sposobnih za simuliranje ljudske inteligencije. Iako sama radionica nije

odmah proizvela tehnološke proboje, njezin značaj ležao je u uspostavljanju zajednice i postavljanju istraživačke agende za budućnost. Povezala je vodeće figure u polju, koji su zajedno s njihovim studentima i kolegama na istaknutim institucijama poput MIT-a, CMU-a, Stanforda i IBM-a, oblikovali disciplinu AI-a u sljedeća dva desetljeća i dalje. U ranim godinama istraživanja AI-a, čak su i skromna postignuća smatrana revolucionarnima zbog računalnih ograničenja računala tog doba, koja su bila primarno dizajnirana za aritmetičke zadatke. Tijekom tog razdoblja, istraživači su se uglavnom fokusirali na simbolički AI, koristeći sustave zasnovane na pravilima kako bi oponašali ljudske procese razmišljanja. Međutim, ovaj pristup imao je svoje nedostatke - nije bio sposoban *učiti*[1] iz podataka i borio se s kompleksnim problemima koji se nisu mogli pojednostavniti u skup pravila. [Rus10].



Slika 1: Organizacija Perceptrona [Ros58].

Paralelno s tim, znanstvenici su istraživali algoritme koji su crpili inspiraciju iz bioloških neuralnih procesa. Osnovna ideja o umjetnom neuronu ima svoje korijene u 1940-ima, posebno u modelu neurona McCulloch-Pitts. Ovaj model je rezultirao jednostavnim logičkim funkcijama koje su prikazivale "sve-ili-ništa" prirodu neuralne aktivnosti i postavljale temelje za razvoj umjetnih neuronskih mreža. Godine 1957., psiholog Frank Rosenblatt dao je značajan doprinos polju razvijajući Perceptron, elektronički uređaj konstruiran na temelju bioloških principa koji je pokazao sposobnost učenja (vidi Sliku 1). Ova inovacija označila je prvi val neuronskih mreža, koje su privukle značajnu pažnju. Unatoč početnom entuzijazmu, ove rane neuronske mreže bile su ograničenog opsega, obično sastavljene od samo jednog ili dva sloja neurona. Njihov razvoj bio je dodatno ograničen dostupnom računalnom snagom tog doba, što je dovelo do smanjenog interesa do kraja 1960-ih [Mac16].

Interes za neuronske mreže obnovljen je 1980-ih s izumom algoritma propagacije

[1]Učenje je sposobnost mijenjanja prema vanjskim podražajima i pamćenja većine prethodnih iskustava [Bon17]

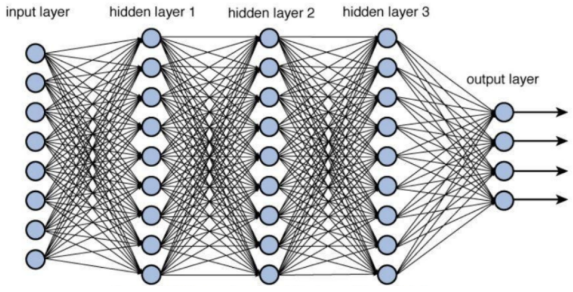
greške unazad *(eng. backpropagation)*. Ova nova metoda učenja omogućila je učinkovit trening višeslojnih neuronskih mreža, uključujući prilagodbu *skrivenih* slojeva koji su ključni za hvatanje složenosti različitih domenskih zadataka. Algoritam propagacije greške unazad bio velik korak naprijed jer je omogućio neuronskim mrežama da se bolje prilagode specifičnim zadacima razvijajući prikladnije unutarnje strukture. Unatoč ovim napretcima, ove rane neuronske mreže još uvijek su smatrane plitkima u usporedbi s današnjim dubokim mrežama [RHW86].

Duboko učenje doživjelo je procvat krajem 2000-ih, posebno s Geoffreyem Hintonovim probojem iz 2006. u treniranju dubokih mreža vjerovanja koristeći metodu poznatu kao slojevito-pohlepno učenje *(eng. layer-wise-greedy-learning)*. Ova tehnika uključivala je nekontrolirano prethodno treniranje kako bi mreža naučila značajke iz podataka prije prilagodbe *(eng. fine-tuning)* s označenim podacima. Hintonov rad je revolucionirao polje smanjujući problem preprilagodbe i omogućavajući bržu konvergenciju treniranja. Ovi napreci, kombinirani s rastom analize velike količine podataka *(eng. Big data)* i sve većom dostupnošću jeftinih računala s većom računalnom snagom, ključni su za široku prihvaćenost i uspjeh dubokog učenja [Liu+17].

Izazovi i ograničenja ranih godina poslužili su kao osnovica za današnje napretke i znanje, oblikujući disciplinu koja je temeljito promijenila naš pristup rješavanju složenih problema. Kroz desetljeća istraživanja, neuspjeha i uspjeha, AI zajednica pružila je snažan intelektualni okvir za intenzivno proučavanje ključne tematike ovog diplomskog rada.

2.2 Duboko učenje

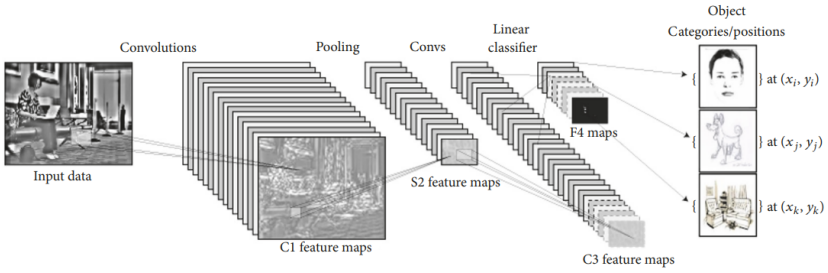
Duboko učenje je podskup strojnog učenja koji omogućuje računalnim modelima da uče iz podataka kroz višestruke slojeve obrade, što dopušta složenije reprezentacije i donošenje odluka. Specijalizirano je za prepoznavanje složenih struktura u velikim skupovima podataka oponašajući kako ljud-



Slika 2: Duboka arhitektura mreže s više slojeva [Par18].

ski mozak obrađuje multi-modalne informa-
cije. U srži dubokog učenja su neuronske mreže, posebno duboke neuronske mreže, koje su dizajnirane s tri glavne vrste slojeva: ulazni, skriveni i izlazni slojevi (Slika 2). Podaci se kroz ove slojeve prenose hijerarhijski, omogućujući modelu da izvodi i jednostavne i složene zadatke prepoznavajući osnovne obrasce ili značajke u podacima.

Među najznačajnijim arhitekturama dubokog učenja su konvolucijske neuronske mreže *(eng. CNNs)*. CNNs (Slika 3) su posebno sposobne za obradu prostornih hijerarhija vizualnih podataka, čineći ih idealnima za zadatke raspoznavanja slika *(eng. image recognition)*. Uključuju specijalizirane konvolucijske slojeve koji filtriraju ulazne podatke kako bi izvukli korisne značajke poput rubova, kutova ili tekstura. Proces učenja znatno je ubrzan paralelnim računarstvom, posebno kroz upotrebu grafičkih procesorskih jedinica *(eng. GPUs)*, što omogućuje ovim dubokim modelima da uče iz velikih skupova podataka učinkovitije. Napredak u tehnikama regularizacije kao što su *dropout*, *batch normalization* i augmentacija podataka također je pridonio sposobnosti generalizacije modela dubokog učenja [Vou+18].



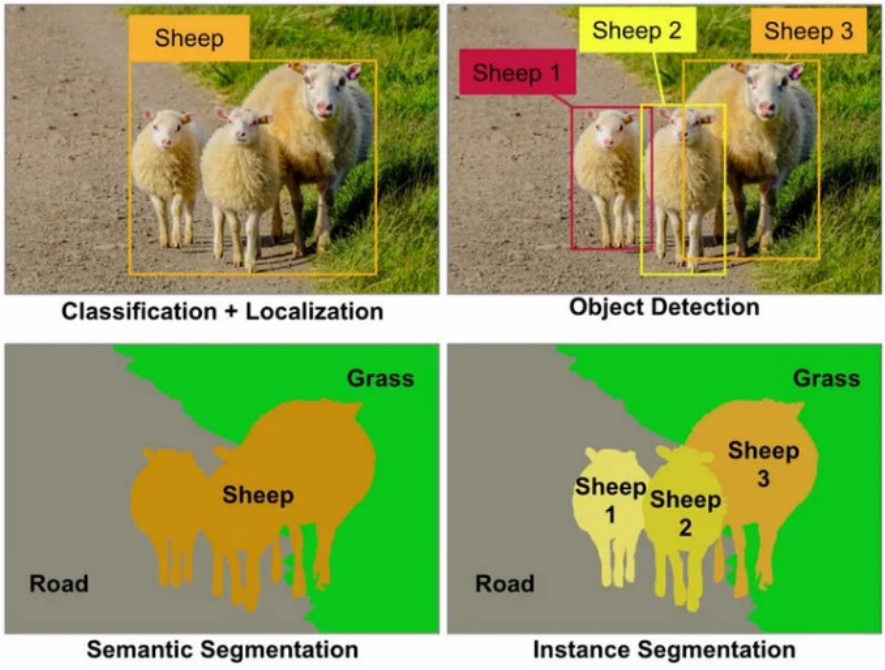
Slika 3: Primjer arhitekture CNN-a za zadatak računalnog vida (detekcija objekata). [Vou+18]

2.2.1 Računalni vid

U kontinuirano mijenjajućem okruženju umjetne inteligencije, računalni vid se ističe kao posebno utjecajna i sveprisutna tehnologija, suptilno utječući na brojne aspekte našeg svakodnevnog života. Ova interdisciplinarna domena teži oponašanju elemenata ljudskog vizualnog sustava, osposobljavajući računalne strojeve da analiziraju, tumače i donose odluke temeljene na vizualnim podacima. Dok su rane implementacije raču-

nalnog vida bile ograničene tehnološkim preprekama, polje je doživjelo transformacijske skokove u posljednjim godinama, zahvaljujući probojima u dubokom učenju i arhitekturama neuronskih mreža. Ovi pristupi čak su nadmašili ljudske performanse u određenim zadacima detekcije i segmentacije objekata. Značajan poticaj ovog ubrzanog razvoja bio je nagli rast dostupnih vizualnih podataka; više od 3 milijarde slika se dijeli online svakodnevno. U kombinaciji s povećanom pristupačnošću visokoj računalnoj snazi, ova ogromna količina podataka služi kao i obrazovna i validacijska baza za evoluirajuće algoritme. Posljedično, stope točnosti za identifikaciju objekata u sustavima računalnog vida su eksponencijalno rasle, povećavajući se s 50% na 99% u periodu manje od desetljeća, nadmašujući tako ljudske sposobnosti u brzoj interpretaciji vizualnih podataka. [Mih19]

U brzorastućem polju računalnog vida, razni su zadaci dizajnirani kako bi pomogli računalima razumjeti vizualne podatke (vidi Sliku 4). Klasifikacija slika je osnovni zadatak u računalnom vidu koji uključuje kategorizaciju cijele slike u jednu od nekoliko unaprijed definiranih klasa. U suštini, cilj je dodijeliti oznaku slici temeljenu na njenom sadržaju. Zadatak obično koristi tehnike poput konvolucijskih neuronskih mreža i može se izvoditi kroz nadzirane ili nenadzirane pristupe učenju.



Slika 4: Zadaci računalnog vida [Mur21]

Prelazeći s klasifikacije cijelih slika na analizu specifičnih objekata unutar njih, susrećemo zadatke poput lokalizacije i detekcije objekata. Lokalizacija ima za cilj pronaći položaj jednog objekta na slici stvarajući okvir oko njega. Nasuprot tome, detekcija objekata ne samo da identificira više objekata na slici već također pruža okvire za svakoga od njih, označavajući ih sukladno. Zatim postoji semantička segmentacija koja detekciju objekata dovodi korak dalje. Označava svaki piksel na slici s klasom objekta kojoj pripada, pružajući klasifikaciju na razini piksela.

U ovoj diplomskom radu, primarni fokus bit će na segmentaciji instanci *(eng. Instance Segmentation)*, zadatku koji predstavlja visoku razinu kompleksnosti u području računalnog vida. Zadatak zahtijeva identifikaciju svakog pojedinačnog objekta unutar tih kategorija, sve do razine piksela. U suštini, segmentacija instanci kombinira sposobnosti detekcije objekata i semantičke segmentacije. Ne samo da locira pojedinačne objekte unutar slike, već svakom pikselu koji pripada tom objektu dodjeljuje jedinstveni ID objekta. Tijekom godina, s napretkom tehnologije računalnog vida, performanse segmentacije instanci značajno su poboljšane. Modeli za segmentaciju instanci već se koriste za identifikaciju i segmentaciju različitih objekata [JVO21; Per+22; Pol+21], što rezultira automatizacijom procesa u različitim područjima.

2.3 Povezana istraživanja

U kontekstu autonomne vožnje, segmentacija instanci predstavlja aktivno područje istraživanja. Jedno od glavnih područja primjene je detekcija traka i oznaka na cesti, gdje se segmentacija instanci koristi za poboljšanje preciznosti detekcije [Ko+21; Zha+20; Cha+19]. Slično ovom istraživanju, ostala istraživanja primijenila su segmentaciju instanci na dinamičke prometne elemente poput vozila [ZZ20; Car+22] i pješaka [Mal19; Lys+21]. Na primjer, autori u [OSD21] predlažu implementaciju najsuvremenije metode Mask R-CNN koristeći tehniku transfernog učenja za detekciju vozila putem segmentacije instanci, koja istodobno proizvodi ograničavajući okvir i masku objekta. Autori [Mal19] također su primijenili okvir temeljen na Mask R-CNN-u, za segmentaciju instanci primijenjenu na pješačke prijelaze, i postigli impresivnu točnost. U [Tse+21] predložena je brza

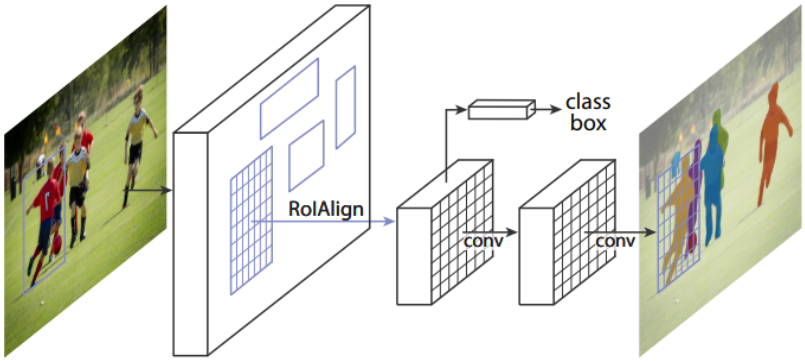
neuronska mreža s višestrukim zadacima u jednoj fazi, za segmentaciju instanci, koja može zadovoljiti zahtjeve za obradom u stvarnom vremenu s dovoljnom točnošću, što je važna značajka za aplikacije autonomnih vozila. Treći smjer istraživanja odnosi se na senzore korištene na vozilu; neki radovi koriste samo kamere [Den+22; Zha+22], dok drugi uključuju lidar [RKS22; Li+22]. Značajan napor usmjeren je prema spajanju podataka prikupljenih različitim tipovima senzora [JSZ22; WZY23].

3 Athitekture modela

Mask R-CNN i YOLO u suštini se razlikuju jer predstavljaju dvije vodeće klase detektora objekata u području računalnog vida: prvi je dvostupanjski detektor koji generira regije za prijedloge i klasificira objekte unutar njih, dok drugi primjenjuje jednostavan jednostupanjski pristup tretirajući problem kao zadatak regresije za predviđanje koordinata okvira i vjerojatnosti klasa izravno sa slike [Car+21]. U ovom poglavlju detaljno ćemo istražiti oba modela, objašnjavajući kako svaki model funkcionira unutar svojeg okvira i njihovu učinkovitost u zadatku segmentacije instanci.

3.1 Mask R-CNN

Mask R-CNN je najsuvremeniji *(eng. state-of-the-art)* algoritam dubokog učenja za detekciju objekata i segmentaciju instanci. Proširuje popularni okvir Faster R-CNN dodavanjem dodatnog sloja mreži koji vraća binarnu masku za svaki detektirani objekt, označavajući koji pikseli pripadaju objektu, a koji ne.



Slika 5: Mask R-CNN okvir za segmentaciju instanci.

Kao što je prikazano na Slici 5, Mask R-CNN koristi dvostupanjsku arhitekturu, slično kao Faster R-CNN. U prvom stupnju, mreža generira skup prijedloga regija koristeći mrežu za prijedlog regija *(eng. RPN - Region Proposal Network)*. Ovi prijedlozi se zatim dovode u drugi stupanj, gdje mreža obavlja klasifikaciju, regresiju okvira i predviđanje maske. Mask grana Mask R-CNN-a je potpuna konvolucijska mreža koja uzima regiju

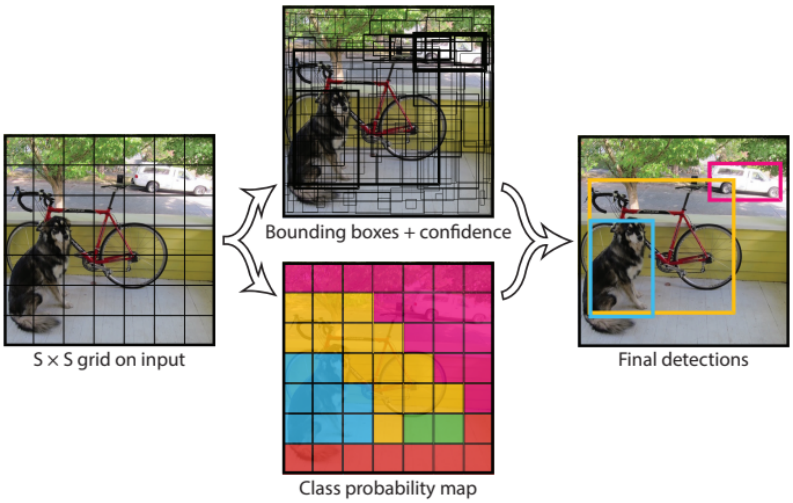
interesa *(eng. ROI - Region of Interest)* kao ulaz i generira binarnu masku za svaki objekt u ROI-u. Ova maska se koristi za segmentaciju objekta od pozadine i dobivanje preciznijih granica oko objekta.

Mask R-CNN je dizajniran da bude fleksibilan i može se instancirati s različitim arhitekturama za svoju osnovu *(eng. Backbone)* i glavu mreže *(eng. Network head)*. Navedeno omogućava optimizaciju za brzinu i preciznost, ovisno o zahtjevima aplikacije. Na primjer, može koristiti naprednu mrežu značajki *(eng. FPN - Feature Pyramid Network)* kao osnovu za bolje performanse. Integriranjem naprednih osnova poput FPN-a i efikasnih glava, Mask R-CNN ne samo da poboljšava točnost, već može pružiti i računalnu učinkovitost, posebno u usporedbi s arhitekturama koje nisu toliko optimizirane.

Jedna od ključnih prednosti Mask R-CNN-a je sposobnost istovremene detekcije objekata i segmentacije instanci, što je korisno u aplikacijama poput autonomne vožnje, medicinske slikovne dijagnostike i robotike. Algoritam je postigao najbolje performanse na nekoliko referentnih skupova podataka, uključujući COCO i Cityscapes.

3.2 YOLO

You Only Look Once (YOLO) je jednostupanjski algoritam dubokog učenja za detekciju objekata u slikama i videima. Temelji se na konvolucijskim neuronskim mrežama (CNN) i koristi jednu neuronsku mrežu za predviđanje kako lokacija tako i klase objekata na slici. Za razliku od tradicionalnih algoritama za detekciju objekata koji koriste pristup kliznog prozora *(eng. sliding-window)* za klasifikaciju regija slike, YOLO dijeli sliku na rešetku ćelija i donosi predviđanja na temelju sadržaja svake ćelije. Za svaku ćeliju u rešetki, YOLO predviđa vjerojatnosti klasa za svaki objekt koji se nalazi u toj ćeliji, kao i ogradne okvire koji okružuju objekte (vidi Sliku 6).



Slika 6: YOLO model za detekciju objekata.

Model za detekciju sastoji se od niza od 24 sloja konvolucije, koje slijede dva potpuno povezana sloja. Isprepleteni konvolucijski slojevi veličine 1x1 koriste se za smanjenje prostora značajki generiranih od prethodnih slojeva. Svako predviđanje okvira sastoji se od pet elemenata: x i y koordinate centra okvira, širine i visine okvira i ocjene pouzdanosti. Osim toga, svaka ćelija rešetke predviđa uvjetne vjerojatnosti klasa koje su uvjetovane prisutnošću objekta u ćeliji. Tijekom testiranja, ove vjerojatnosti klasa kombiniraju se s pojedinačnim predviđanjima pouzdanosti okvira kako bi se dobile pouzdanosti specifične za klasu.

Za treniranje YOLO algoritma koristi se veliki skup označenih slika. Tijekom treniranja, parametri mreže prilagođavaju se korištenjem algoritma propagacije greške unazad kako bi se minimizirala pogreška predviđanja.

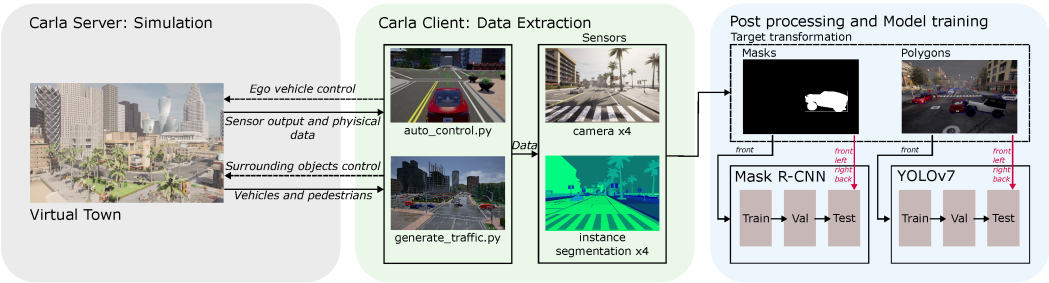
Jedna od prednosti YOLO-a je njegova izvedba u stvarnom vremenu. Budući da obrađuje cijelu sliku u jednom prolazu, može se koristiti za primjene poput autonomnih vozila, dronova i video nadzora u stvarnom vremenu. Osim toga, YOLO može detektirati objekte različitih veličina i oblika, što ga čini prikladnim za širok spektar primjena.

Algoritam YOLO prošao je brz razvoj, pri čemu svaka nova verzija nadograđuje svoje prethodnike putem kontinuiranih poboljšanja [Jia+22]. Najnovija verzija YOLO-a, YOLO verzija 7 [WBL22], nadmašuje sve postojeće modele za detekciju objekata u smislu

brzine i preciznosti, koristeći manje parametara i manje računanja. YOLO ima bržu i snažniju arhitekturu mreže koja učinkovito integrira značajke, pruža precizniju detekciju objekata i bolje performanse segmentacije instanci, te ima robusniju funkciju gubitka i poboljšanu učinkovitost treniranja. To YOLO čini ekonomičnijim i omogućava brže treniranje na manjim skupovima podataka bez prethodno istreniranih težina.

4 Metodologija istraživanja

Ovo poglavlje nudi temeljnu analizu eksperimentalnog okvira i metodologije, što je ilustrirano na Slici 7. Prvi dio raspravlja o prikupljanju podataka korištenjem simulatora CARLA, opisujući virtualno okruženje i konfiguracije senzora. Sljedeći odjeljak detaljno objašnjava korake predobrade podataka potrebne za dva različita algoritma: Mask R-CNN i YOLO. Konkretno, objašnjene su tehnike za generiranje binarnih maski i poligona za instance objekata. Na kraju, opisuju se metrike evaluacije koje se koriste za procjenu učinkovitosti modela računalnog vida. Cijeli postupak ima za cilj pružiti sveobuhvatno razumijevanje kako su provedeni eksperimenti, kako su podaci obrađeni i kako su rezultati procijenjeni.



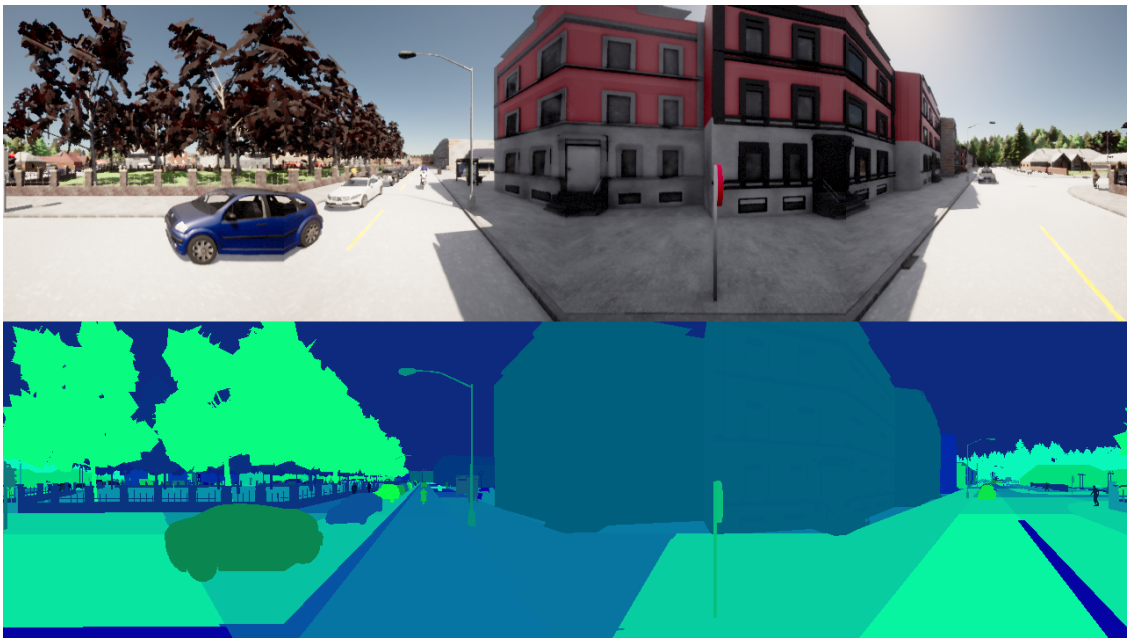
Slika 7: Grafički prikaz eksperimenta.

4.1 Simulator CARLA i prikupljanje podataka

Za razvoj modela prvo nam je potrebno dovoljno podataka za trening, validaciju i testiranje. Podaci se dobivaju korištenjem simulatora CARLA [Dos+17], simulatora otvorenog koda *(eng. open source)* dizajniranog za istraživanje autonomne vožnje, posebno kako bi olakšao razvoj, trening i validaciju autonomnih sustava za vožnju. Platforma nudi otvorene digitalne resurse poput urbanog okruženja, zgrada i vozila, te značajke kao što su specificiranje senzorskih sustava i uvjeta okoline, potpunu kontrolu nad statičkim i dinamičkim akterima, te mogućnost generiranja karata. U ovom radu, simulator se koristi za snimanje visokokvalitetnih RGB slika pomoću prednjih, lijevih, desnih i stražnjih kamera, pri čemu svaka kamera pruža vidno polje od 90 stupnjeva. Za svaku sliku generira se slika segmentacije instanci koja se koristi za izdvajanje ciljanih podataka o vozilima i

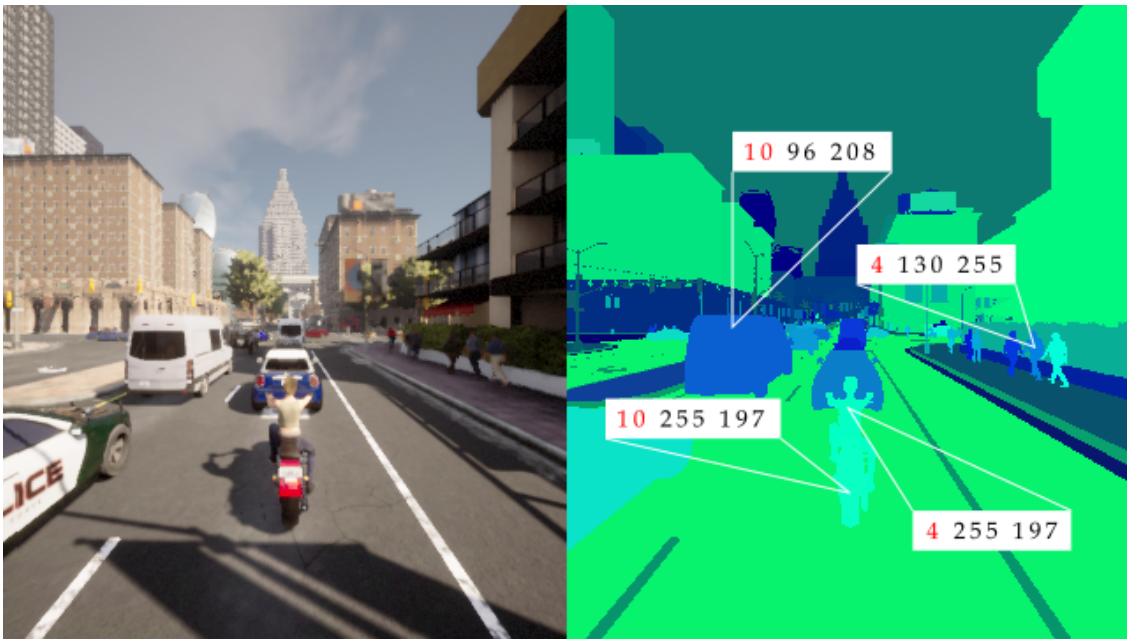
prolaznicima na slici, posebno za svaki model.

Što se tiče virtualnog okruženja, generirano je oko stotinu vozila i pješaka koji slobodno šeću ulicama zadane virtualne lokacije. Zatim su senzori inicijalizirani na pokretnom automobilu kako bi prikupljali podatke. Iako CARLA knjižnica nudi širok izbor senzora, bili su potrebni samo RGB i kamere za segmentaciju instanci. Za postavljanje kamera stvoreno je četiri različite kamere s pogledom iz različitih kutova i pričvršćene za dva tipa senzora na tzv. *Ego* vozilu, koje je bilo postavljeno da vozi po gradu u autopilot načinu (vidi sliku 8). Nakon što je sve bilo postavljeno, svi senzori čekali su signal za promjenu okvira, s razmakom od 10 okvira, kako bi snimili slike i spremili ih na disk, čime je stvoren anotirani skup podataka.



Slika 8: RGB & prikazi segmentacije instanci: Lijeva, Prednja, Desna, Stražnja Kamera.

Razlikovanje različitih instanci prolaznika i vozila u konačnom skupu podataka bilo je jednostavno. Slika segmentacije instanci, spremljena na disk, sadržavala je identifikacijske brojeve instanci kodirane u G i B kanalima RGB slikovne datoteke, dok je R kanal sadržavao standardni ID klase. prolaznicii vozila imali su ID klase 4 odnosno 10, što je olakšalo pripremu ciljanih podataka za modele (vidi sliku 9).



Slika 9: Primjer RGB kodiranja za instance vozila i prolaznika.

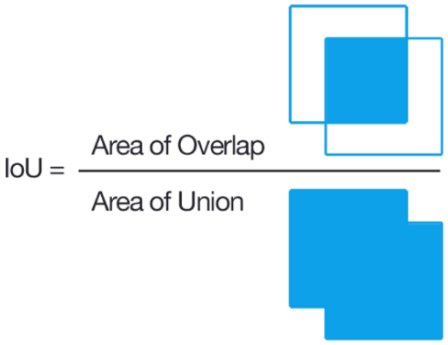
4.2 Predobrada podataka

Za stvaranje ciljnih labela *(eng. target)* za trening Mask R-CNN-a, bilo je potrebno generirati binarne maske za svaki objekt na slici. Srećom, identifikacijski brojevi u G i B kanalima slike segmentacije instanci mogli su se iskoristiti za razdvajanje objekata. Svaka maska predstavlja pojedinačni objekt u slici kanala s istom razlučivošću kao i originalni ulaz. Binarna maska sadrži samo jedinice i nule, označavajući piksele koji pripadaju određenom objektu. Istodobno je stvorena lista ID-ova klasa odnosno kategorije, gdje broj 1 odgovara prolaznicima, a broj 2 vozilima; razred 0 se koristio za pozadinu. Dodatno, generirane su koordinate okvira oko objekata, a izračunana je i odgovarajuća površina za svaku instancu. Ova površina služila je za filtriranje šuma, posebno objekata s niskim brojem piksela. Nakon opisanog procesa, dataset za Mask R-CNN bio je spreman.

Ciljne labele za algoritam YOLO značajno se razlikuju. Umjesto da model zahtijeva masku za svaku instancu, YOLO koristi poligon oko objekta. Poligon je skup proizvoljnog broja točaka koje stvaraju kompleksan oblik oko objekta. Primjeri maski i poligona prikazani su na slici 7. Trening varira za svaki model, a posebne postavke i prilagodbe korištene u eksperimentima bit će raspravljene zasebno.

4.3 Metrike evaluacije

Za evaluaciju rezultata modela za računalni vid koristit će se standardna metrika, srednja prosječna preciznost *(eng. mAP - Mean Average Precision)*, kako bi se ocijenila učinkovitost sustava. Navedena metrika mjeri koliko dobro sustav podudara predviđene okvire objekata s istinskim okvirima objekata za svaku klasu objekta, koristeći metriku presjeka nad unijom *(IoU - Intersection over Union)* kako bi se izračunalo preklapanje okvira (vidi sliku 10). mAP se izračunava računanjem prosječne površine ispod krivulje *precision-recall* za svaku klasu [Sha22]. Viši rezultat mAP ukazuje na bolju točnost sustava za detekciju objekata. Za evaluaciju mAP za segmentaciju, umjesto okvira objekta, prati se podudaranje pojedinačnih piksela.

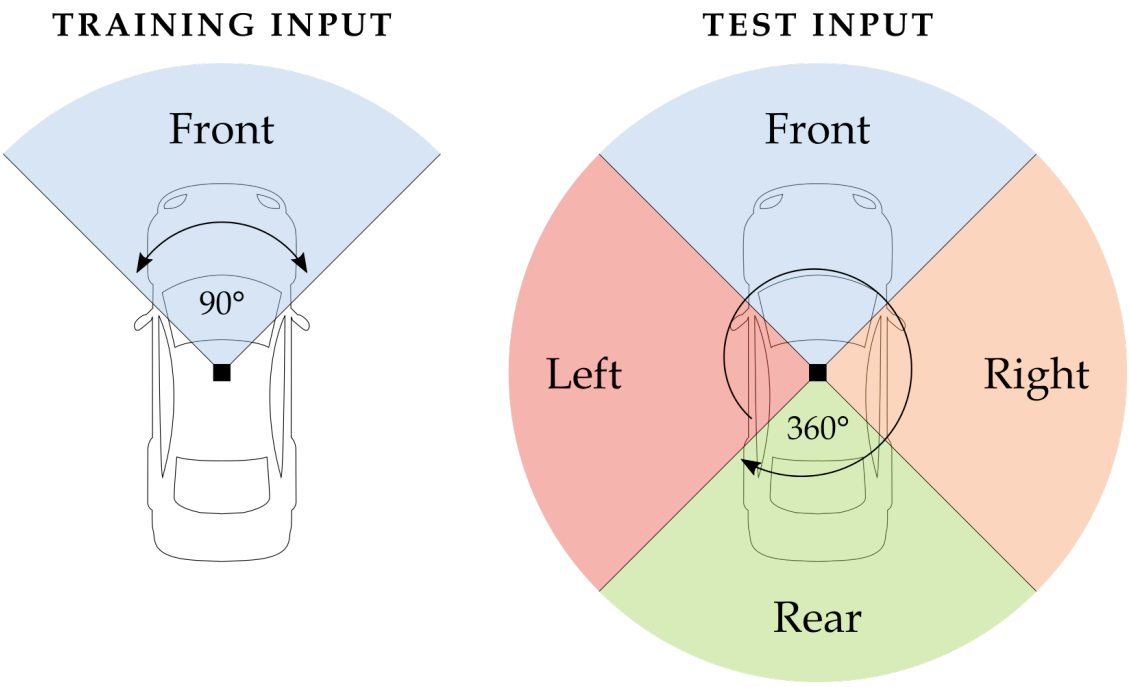


Slika 10: IoU metrika. [**map**].

U oba eksperimenta rezultati mAP-a analizirat će se koristeći točan prag od 0,5 i promjenjiv prag IoU-a, uz prosječno uzimanje više pragova u rasponu od 0,5 do 0,95. Obje mjere primijenit će se na rezultate okvira označene kao *Boxset* u rezultatima, što predstavlja performanse detekcije objekata, i na predikcije maski označene kao *Maskset*, koje se odnose na točnost segmentacije instanci. Budući da je cilj ocijeniti segmentaciju, fokus će uglavnom biti na analizi rezultata *maskset-a*.

5 Procjena treninga prednje kamere za segmentaciju instanci prometnog okruženja [2]

Ovo poglavlje istražuje u kojoj mjeri se monokularna prednja kamera može koristiti za treniranje modela koji, tijekom inferencije, uzimaju ulaz cijelog prikaza okoline vozila, kako je prikazano na slici 11. Testirat će se sposobnosti generalizacije modela treniranih isključivo na slikama prednje kamere na dodatnim kamerama koje gledaju u različitim smjerovima i pokrivaju prikaz od 360 stupnjeva. Ovo istraživanje analizira kako promjena perspektive iz različitih perspektiva kamera utječe na performanse. U nastavku, rezultati će odgovoriti na pitanje kako se snimke s prednjih kamera, široko dostupne na internetu, mogu ponovno upotrijebiti u treniranju AV modela koji moraju zabilježiti cijelo okruženje automobila, čime se smanjuju troškovi prikupljanja podataka.



Slika 11: Grafički prikaz pozicija kamera u prvom eksperimentu.

[2]Ovo poglavlje je objavljeno u časopisu *Engineering Proceedings* izdavača *MDPI* pod naslovom "Using a Monocular Camera for 360° Dynamic Object Instance Segmentation in Traffic" [OB23].

5.1 Postavke treninga

Tablica 1 navodi osnovnu konfiguraciju i hiperparametre modela korištenih u eksperimentu. Parametri su zadržani na zadanim vrijednostima definiranim u izvornim repozitorijima. Cilj eksperimenta nije bio postizanje *state-of-the-art* rezultata, već usporedba sposobnosti generalizacije dvaju arhitektura na neviđenim perspektivama. Generalizaciju u ovom istraživanju definiramo kao razliku u performansama segmentacije instanci na prednjoj kameri (referentna vrijednost) i segmentacije instanci na kamerama s različitim perspektivama.

Tablica 1: Vrijednosti ključnih hiperparametara korištenih u treningu u Eksperimentu 1.

Name	Mask R-CNN [30]	YOLOv7 [28]
Optimizer	Adam	SGD+M
Learning rate	1×10^{-4}	1×10^{-3}
Number of epochs	32	30
Pretrained weights	None	COCO
Backbone	Resnet50	E-ELAN
Batch size	2	4

Generirani set podataka od 10,000 slika korišten je za treniranje oba modela, koji je kasnije podijeljen na trening i validacijske skupove prema omjeru 80:20. Vrijedi napomenuti da su modeli trenirani koristeći slike snimljene isključivo prednjom kamerom, dok je testni proces iskoristio potpuni 360-stupanjski pogled na okolinu vozila. Nakon treninga svakog modela, skup od 1000 neviđenih testnih slika korišten je za evaluaciju performansi. Pri početnom pregledu, kako je prikazano na Slici 12, oba modela su proizvela usporedive i precizne rezultate, ispravno segmentirajući objekte snimljene bočnim i stražnjim kamerama, uključujući vozila i prolaznike. Modeli su koristili izlazne maske segmentacije i omeđujuće okvire za svaku ulaznu sliku.

Slika 12: Prikaz rezultata za prvi eksperiment na istoj 360° sceni.

5.2 Rezultati i rasprava

Ovo poglavlje podijeljeno je na tri glavna dijela, prvo se fokusira na ukupne rezultate, a zatim ih dodatno razlaže kako bi se ispitale performanse specifično za vozila i prolaznike. Uz tekstualnu analizu, tablice i grafikoni pružaju konkretniji pogled na metrike performansi, dajući potpuniju sliku o performansama modela. Cilj ovog odjeljka je pružiti sveobuhvatan pregled prednosti i slabosti eksperimenta.

5.2.1 Ukupni rezultati

Tablica 2 prikazuje rezultate modela na različitim ulaznim perspektivama. Prvo, analizirajmo kako modeli djeluju na slikama s prednje kamere, osnovne perspektive na kojoj su trenirani. YOLO je precizniji od Mask R-CNN-a jer ima viši ukupni mAP rezultat. Međutim, postoji primjetna razlika u dosljednosti između okvira i maske predikcija na mAP .5 između arhitektura. Dok Mask R-CNN detektira i segmentira objekte na približno 50% mAP, YOLO bolje detektira objekte nego što izvodi segmentaciju instanci. Takvi

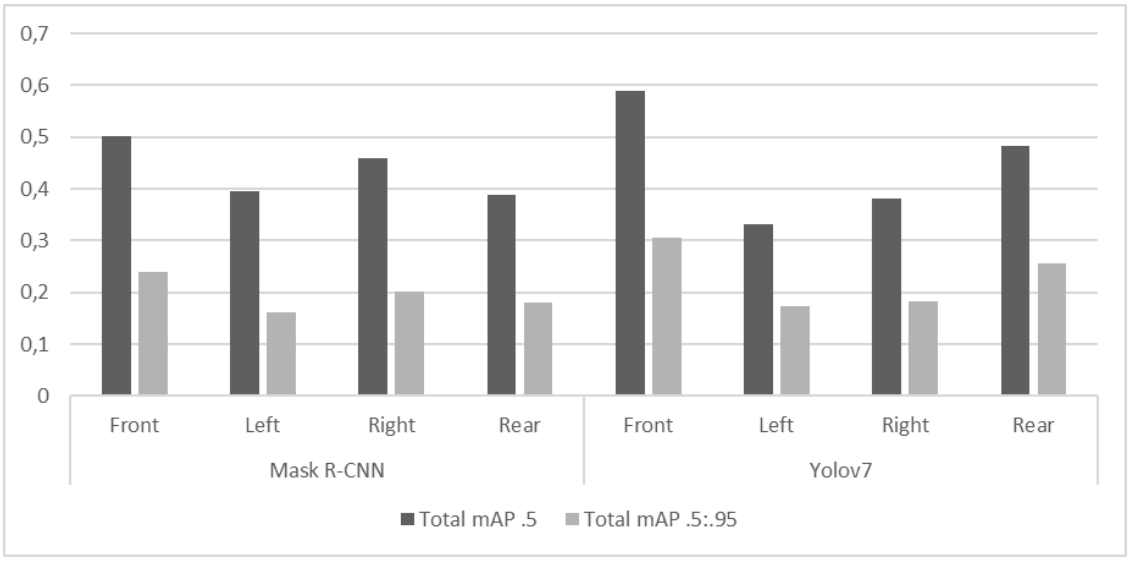
rezultati potvrđuju da je YOLO prvenstveno algoritam za detekciju. Iako YOLO pokazuje superiornu izvedbu, pada u segmentaciji instanci, dok Mask R-CNN ostaje dosljedan u detekciji i segmentaciji.

Zatim se fokusirajmo na izvedbu modela na kamerama s perspektivama koje se razlikuju od osnovne, posebno predikciju maske (*Maskset*), kako bismo procijenili koliko dobro modeli izvode segmentaciju instanci. Primjećujemo zanimljive rezultate. Mask R-CNN ima bolji mAP .5 na lijevoj i desnoj kameri, unatoč boljim osnovnim rezultatima YOLO-a, što znači da Mask R-CNN ima znatno bolju generalizaciju na neviđenim bočnim perspektivama od YOLO-a. Ako analiziramo mAP .5:.95 na istim bočnim kamerama, iako su apsolutne vrijednosti slične, YOLO znatno pada u izvedbi od osnovne perspektive. Rezultati za stražnju kameru, koja ima drugačiju perspektivu od prednje kamere, ali se može smatrati njenim odrazom, idu u korist YOLO-a. S obzirom na ovu karakteristiku, oba modela dobro generaliziraju na slikama stražnje kamere.

Model	Camera	Boxset		Maskset	
		mAP .5	mAP .5:.95	mAP .5	mAP .5:.95
Mask R-CNN	Front (baseline)	0.526	0.309	0.502	0.239
	Left	0.439	0.213	0.396	0.162
	difference	*-0.087*	*-0.096*	*-0.106*	*-0.077*
	Right	0.462	0.247	0.459	0.202
	difference	*-0.064*	*-0.062*	*-0.043*	*-0.037*
	Rear	0.406	0.209	0.389	0.18
	difference	*-0.12*	*-0.1*	*-0.113*	*-0.059*
Yolov7	Front (baseline)	0.687	0.494	0.59	0.305
	Left	0.402	0.251	0.331	0.172
	difference	*-0.285*	*-0.243*	*-0.259*	*-0.133*
	Right	0.45	0.289	0.382	0.183
	difference	*-0.237*	*-0.205*	*-0.208*	*-0.122*
	Rear	0.597	0.371	0.483	0.257
	difference	*-0.09*	*-0.123*	*-0.107*	*-0.048*

Tablica 2: Ukupni rezultati prvog eksperimenta.

Slika 13 služi kao vizualna potvrda gore navedenih iskaza, sažimajući ukupnu izvedbu preko različitih mAP vrijednosti za oba modela.



Slika 13: Grafički prikaz ukupnih rezultata preko mAP vrijednosti prvog eksperimenta.

5.2.2 Rezultati segmentacije vozila

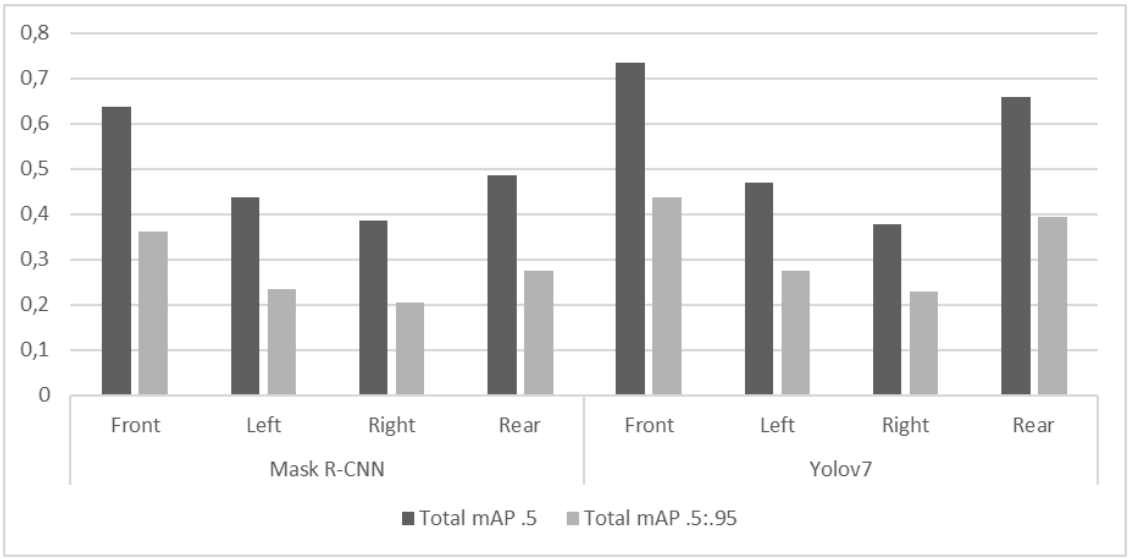
Kako bismo bolje razumjeli rezultate, razmotrimo izvedbu za svaku klasu zasebno. Prvo je fokus na vozilima; rezultati su navedeni u tablici 3. Vozila generirana od strane CARLA simulatora variraju od različitih vrsta automobila i kamiona do manjih vozila poput motocikala i bicikala. Vozila mogu biti pozicionirana bliže i dalje na horizontu. Ako je vozilo blizu, obično to znači da *Ego* automobil vozi iza njega u redu i prikuplja slike stražnjeg dijela vozila. U većini slučajeva, prednja kamera može snimati prednju i lijevu stranu vozila kada je dalje, posebno ako automobil vozi u suprotnom smjeru, ali u rijetkim slučajevima vidimo desnu stranu vozila. Ove točke su bitne kako bismo razumjeli podatke na kojima su modeli trenirani i kako se mogu razlikovati od drugih perspektiva.

YOLO bolje izvodi oba zadatka i iznimno je uspješan u detekciji vozila. Oba modela pokazuju pad u izvedbi kada je riječ o sposobnostima generalizacije *maskset-a* na lijevim i desnim kamerama. Takvo ponašanje može se pripisati nedostatku izbliza snimljenog pogleda na strane vozila, budući da ih prednja kamera obično ne snima. Razmatranja skupa podataka, o kojima se raspravljalo ranije, objašnjavaju zašto desna perspektiva daje najslabije rezultate. Iako Mask R-CNN ima izvedbu bližu osnovnoj perspektivi nego YOLO, manje je precizan. Rezultati za stražnju kameru su očekivani jer stražnja kamera

Model	Camera	Boxset		Maskset	
		mAP .5	mAP .5:.95	mAP .5	mAP .5:.95
Mask R-CNN	Front (baseline)	0.67	0.448	0.636	0.361
	Left	0.452	0.263	0.437	0.236
	difference	*-0.218*	*-0.185*	*-0.199*	*-0.125*
	Right	0.381	0.215	0.386	0.204
	difference	*-0.289*	*-0.233*	*-0.25*	*-0.157*
	Rear	0.48	0.284	0.485	0.275
	difference	*-0.19*	*-0.164*	*-0.151*	*-0.086*
Yolov7	Front (baseline)	0.83	0.641	0.735	0.438
	Left	0.52	0.346	0.469	0.275
	difference	*-0.31*	*-0.295*	*-0.266*	*-0.163*
	Right	0.411	0.28	0.377	0.23
	difference	*-0.419*	*-0.361*	*-0.358*	*-0.208*
	Rear	0.733	0.503	0.659	0.395
	difference	*-0.097*	*-0.138*	*-0.076*	*-0.043*

Tablica 3: Rezultati segmentacije vozila prvog eksperimenta.

uglavnom snima prednji dio vozila izbliza, dok je udaljeni pogled zrcalna slika prednje kamere. YOLO je precizniji u segmentaciji vozila sa stražnje perspektive nego Mask R-CNN i bliže je osnovnoj perspektivi, kako je vidljivo u obje mAP metrike (pogledajte Sliku 14).



Slika 14: Grafički prikaz rezultata vozila preko mAP vrijednosti prvog eksperimenta.

5.2.3 Rezultati segmentacije prolaznika

Tablica 4 sadrži rezultate za prolaznike. Prolaznici su uglavnom vidljivi na nogostupu na slikama snimljenim prednjom kamerom, ali mogu biti prisutni i na pješačkom prijelazu. Važno je napomenuti da se prolaznici obično mogu uočiti na znatnoj udaljenosti od naše točke gledišta, što implicira da su modeli trenirani da prepoznaju relativno male značajke prolaznika. Što se tiče ukupne izvedbe osnovne perspektive, YOLO ponovno nadmašuje Mask R-CNN.

Model	Camera	Boxset		Maskset	
		mAP .5	mAP .5:.95	mAP .5	mAP .5:.95
Mask R-CNN	Front (baseline)	0.383	0.17	0.368	0.117
	Left	0.426	0.163	0.355	0.089
	difference	*0.043*	*-0.007*	*-0.013*	*-0.028*
	Right	0.544	0.278	0.532	0.2
	difference	*0.161*	*0.108*	*0.164*	*0.083*
	Rear	0.332	0.135	0.293	0.084
	difference	*-0.051*	*-0.035*	*-0.075*	*-0.033*
Yolov7	Front (baseline)	0.543	0.348	0.444	0.172
	Left	0.283	0.156	0.194	0.0697
	difference	*-0.26*	*-0.192*	*-0.25*	*-0.1023*
	Right	0.488	0.298	0.387	0.136
	difference	*-0.055*	*-0.05*	*-0.057*	*-0.036*
	Rear	0.46	0.239	0.308	0.119
	difference	*-0.083*	*-0.109*	*-0.136*	*-0.053*

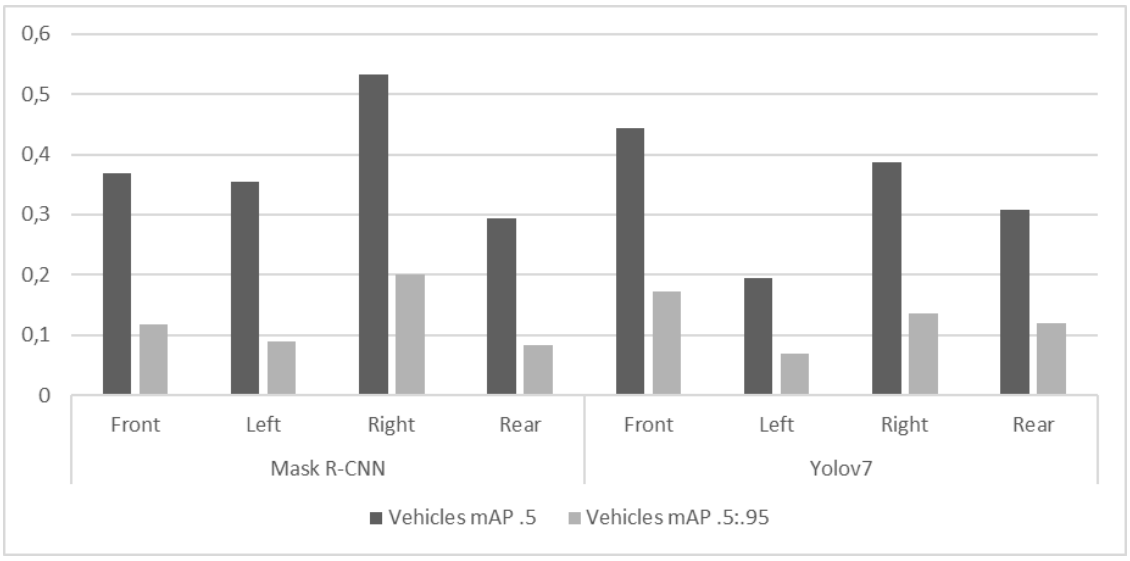
Tablica 4: Rezultati segmentacije prolaznika prvog eksperimenta.

Lijeva kamera obično snima ljude na lijevom nogostupu. Na temelju te opservacije, lijeva kamera tipično bilježi slike prolaznika u većoj rezoluciji. Mask R-CNN ne gubi mAP izvedbu kao što je to bio slučaj pri segmentaciji vozila. Nadmašuje YOLO u detekciji i segmentaciji. Razlika u izvedbi predviđanja Mask R-CNN-a za prolaznike u odnosu na osnovnu perspektivu u pogledu detekcije je značajno pozitivna i gotovo ista u pogledu segmentacije instanci. Stoga rezultati ukazuju na to da je Mask R-CNN iznimno dobar u učenju značajki prolaznika s te perspektive. Nasuprot tome, YOLO-ova izvedba značajno je smanjena u usporedbi s osnovnom perspektivom.

Desna kamera promatra prolaznike koji hodaju desnim nogostupom. To implicira

da je desna kamera ta koja snima prolaznike u znatno većoj veličini, omogućavajući bolje prepoznavanje značajki. Može se primijetiti da Mask R-CNN poboljšava rezultate osnovne perspektive mAP detekcije i segmentacije značajnom marginom. Sve razlike u odnosu na osnovnu perspektivu su pozitivne, a ako se usredotočimo na segmentaciju, možemo primijetiti porast od 0,16 postotnih bodova na mAP .5. Izvedba YOLO algoritma ponovno opada prema svim mjerilima evaluacije.

Što se tiče stražnje kamere, primijećeni su sljedeći rezultati: Budući da je točka gledišta ista kao kod prednje kamere, ali zrcaljena, očekuje se da bi rezultati segmentacije prolaznika trebali biti slični. YOLO je opet precizniji od Mask R-CNN-a, ali razlika je manje značajna nego kod segmentacije vozila. Uspoređujući razliku u segmentaciji prolaznika u odnosu na osnovnu perspektivu, Mask R-CNN ima bolju generalizaciju (manja razlika) od YOLO-a po prvi puta (pogledajte Sliku 15).



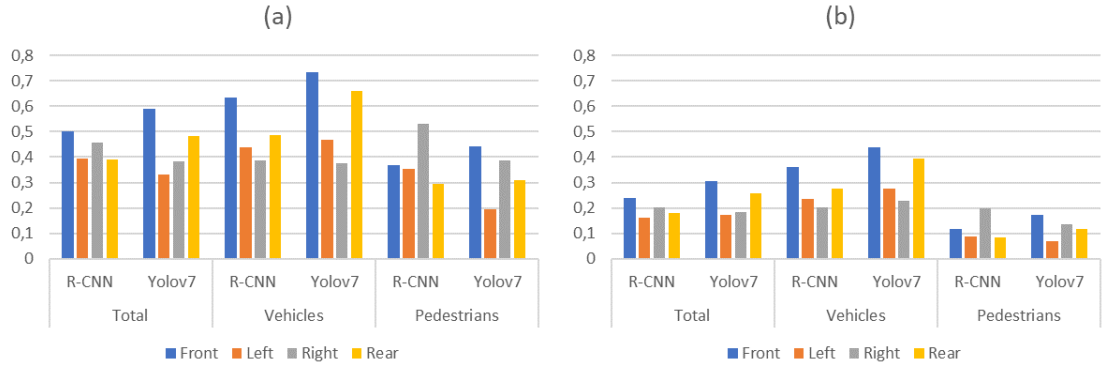
Slika 15: Grafički prikaz rezultata prolaznika preko mAP vrijednosti prvog eksperimenta.

5.3 Zaključci

Ovo poglavlje istražuje svojstva generalizacije dvaju popularnih algoritama segmentacije instanci na nepromatrane perspektive pokretnih objekata u prometu. Cilj je bio utvrditi mogu li se uobičajene snimke prednjom kamerom koristiti kao podaci za treniranje kako bi se dobila potpuna slika okoline vozila. Konkretno, uspoređivane su

sposobnosti generalizacije, definirane kao razlika u mAP mjerenju između osnovne perspektive (prednja kamera) i drugih perspektiva (lijeva, desna i stražnja kamera).

Na Slici 16 prikazana je grafička reprezentacija rezultata segmentacije instanci u pogledu metrika mAP .5 i mAP .5:.95. Iz Slike 16 jasno je da YOLO ima bolju mAP izvedbu u većini slučajeva. Kada je riječ o segmentaciji vozila, oba modela slično se ponašaju za lijevu i desnu perspektivu, iako su YOLO-ovi rezultati osnovne perspektive bolji. Mask R-CNN je jedina arhitektura s pozitivnim razlikama u usporedbi s osnovnom perspektivom, kako je prikazano u rezultatima za prolaznike. Ako posebno pogledamo desnu kameru, Mask R-CNN je bio u stanju izuzetno dobro prepoznati prolaznike koji su vrlo blizu točki gledišta. Nasuprot tome, YOLO ne uspijeva bolje izvesti od osnovne perspektive kada se suoči s prolaznicima iz neposredne blizine. S druge strane, YOLO ima bolju generalizaciju na stražnjoj kameri.



Slika 16: Rezultati segmentacije instanci (*Maskset*) mjereni metrikom mAP koristeći dvije granice: (a) mAP .5 i (b) mAP .5:.95 prvog eksperimenta.

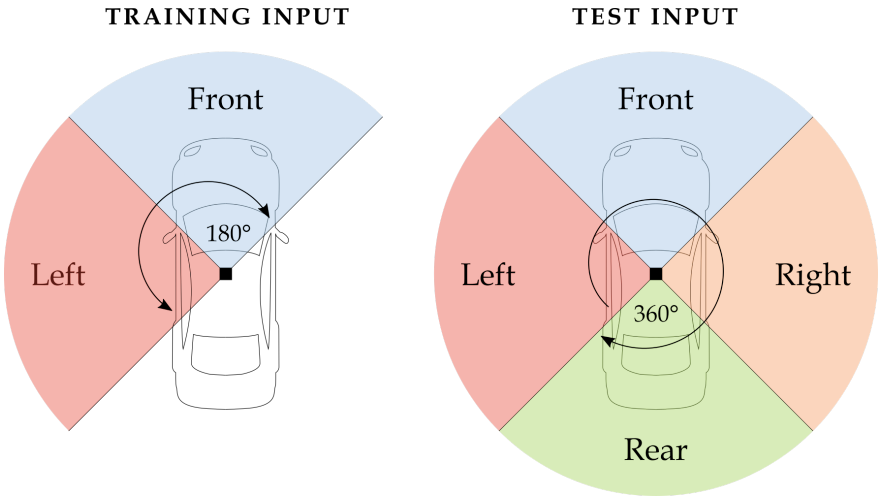
Iz rezultata je vidljivo da YOLO bolje izvodi na prednjoj kameri, tj. daje točnije rezultate na osnovnoj perspektivi. Vjerojatno je to zato što YOLO intenzivno koristi tehnike augmentacije, dok Mask R-CNN korišten u ovom eksperimentu ne koristi. Glavni cilj eksperimenta nije bio uspoređivati apsolutne mAP vrijednosti već sposobnost generalizacije, kako je ranije opisano. YOLO također obično ima bolje rezultate generalizacije na testovima sa stražnjom kamerom. Dakle, ako podaci iz perspektive ne odstupaju puno od podataka korištenih prilikom treniranja, arhitektura YOLO je bolji izbor. No, u isto vrijeme, ne uspijeva izvući smislene značajke iz većih, drugačije orijentiranih objekata koji su

se u treniranju pojavljivali manji. S druge strane, Mask R-CNN ima bolju generalizaciju za strane perspektive, što znači da mu performanse manje padaju na slikama koje više odstupaju od osnovne. U slučaju segmentacije prolaznika, čak i poboljšava performanse osnovne perspektive.

6 Procjena treninga prednje i lijeve kamere za poboljšanu segmentaciju instanci prometnog okruženja

U prethodnom poglavlju fokus je bio na procjeni izvedbe modela treniranih isključivo na podacima kamere usmjerene prema naprijed kako bi se prepoznalo potpuno 360-stupanjsko okruženje oko vozila. Rezultati su otkrili određena ograničenja, posebno prilikom interpretiranja bočnih pogleda za vozila. Nadograđujući se na te rezultate, ovaj eksperiment u trening nadodajte podatke s kamere usmjerene prema lijevo kako bi se istražilo može li ovaj dodatak poboljšati sposobnost modela da točno segmentira instance vozila i prolaznika kroz potpuni 360-stupanjski pogled.

Kamera usmjerena prema naprijed uglavnom snima objekte ispred, poput nadolazećih vozila. Nasuprot tome, lijeva kamera pruža drugačiju perspektivu i nove informacije o bočnim stranama vozila i prolaznicima u blizini. Desna kamera također snima prolaznike, ali pruža manje instanci vozila, budući da se vožnja odvija s desne strane ceste. Kombinirajući podatke iz prednje i lijeve kamere, cilj je proizvesti neku vrstu "dijagonalnog" pogleda, s ciljem poboljšavanja rezultata modela. Ova organizacija perspektiva kamera za drugi eksperiment vizualno je prikazana na Slici 17.



Slika 17: Grafički prikaz pozicija kamera u drugom eksperimentu.

Slično prethodnom istraživanju, ovaj će eksperiment koristiti algoritme Mask R-CNN i YOLO kako bi se procijenila učinkovitost dodatnih podataka s lijeve kamere. Ako

bude uspješna, ova strategija dijagonalnog pogleda ne samo da bi ponudila bolje razumijevanje okoliša i veću sigurnost, već bi također mogla značajno optimizirati resurse potrebne za prikupljanje i obradu podataka, čineći proces efikasnijim i ekonomičnijim.

6.1 Postavke treninga

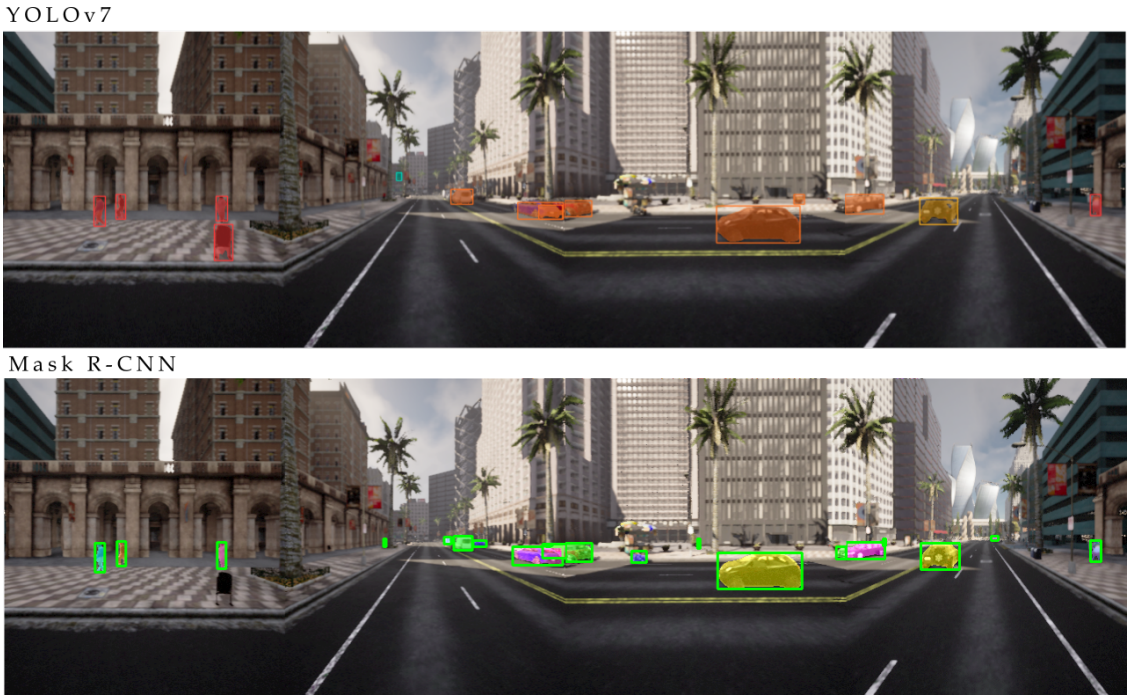
Tablica 5 prikazuje osnovne postavke i hiperparametre modela korištenih u ovom eksperimentu. Ovi parametri ostali su dosljedni njihovim zadanima vrijednostima, kako su navedene u originalnim izvornim repozitorijima, te se nisu mijenjali od onih korištenih u našem prethodnom eksperimentu, osim promjene u veličini *batch-a* kako bi se prilagodili većem skupu podataka. Primarni cilj ovog eksperimenta nije bio pokazati optimalne performanse, već procijeniti koliko dobro dvije različite arhitekture generaliziraju kada se susreću s novim scenarijima. Ovdje se "generalizacija" odnosi na razliku u performansama prilikom izvođenja segmentacije instanci pomoću prednje i lijeve kamere (koje služe kao naše nove osnovne perspektive) u usporedbi s performansama s desne i stražnje kamere. Osim toga, proučavamo varijacije u performansama ovih novih osnovnih perspektiva.

Tablica 5: Vrijednosti ključnih hiperparametara korištenih u treningu u Eksperimentu 2

Name	Mask R-CNN [30]	YOLOv7 [28]
Optimizer	Adam	SGD+M
Learning rate	1×10^{-4}	1×10^{-3}
Number of epochs	32	30
Pretrained weights	None	COCO
Backbone	Resnet50	E-ELAN
Batch size	6	8

Stvoren je skup podataka koji je dvostruko veći u odnosu na prethodni eksperiment, sastoji se od 20,000 slika s jednakim količinama podataka prednje i lijeve kamere. Ovaj skup podataka korišten je za trening oba modela te je ponovo podijeljen na podskupove za trening i validaciju prema omjeru 80:20, zadržavajući omjer 50:50 slika prednje i lijeve kamere. Važno je napomenuti da je faza treninga koristila slike snimljene isključivo s prednjih i lijevo usmjerenih kamera, dok je testni proces koristio cjeloviti pogled od

360 stupnjeva oko vozila. Nakon što su modeli istrenirani, za procjenu performansi korištena je zasebna serija od 2,000 ranije neviđenih slika za svaku kameru. Preliminarna promatranja, kako je prikazano na slici 18, ukazuju da su oba modela dala slične i točne rezultate, uglavnom uspješno segmentirajući objekte snimljene bočnim i stražnjim kamerama, uključujući i vozila i prolaznike. Zanimljiv izuzetak na slici je kanta za smeće koja je pogrešno segmentirana kao prolaznik kod YOLO modela. Modeli su generirali maske segmentacije i granične okvire za svaku evaluiranu sliku.



Slika 18: Prikaz rezultata za drugi eksperiment na istoj 360° sceni.

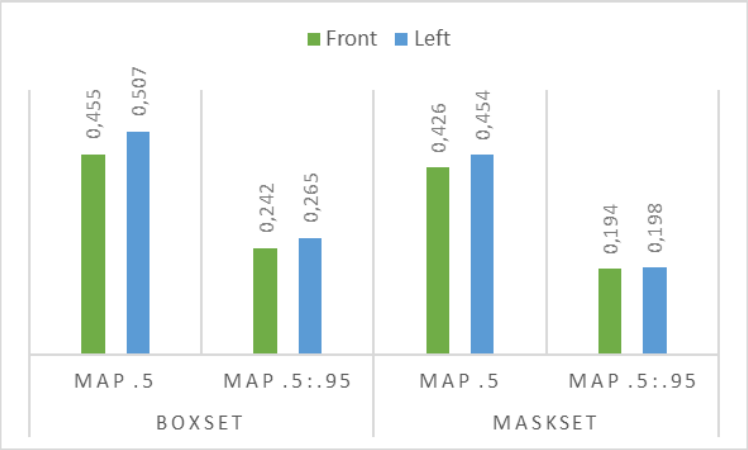
6.2 Rezultati i rasprava

Ovo poglavlje razvija se u sličnoj trodijelnoj strukturi, prvo predstavljajući ukupne rezultate prije ulaska u specijalizirane evaluacije za vozila i prolaznike. Uz tekst i tablice, grafički prikazi ponovno se koriste kako bi potvrdili metrike izvedbe modela. Ključna novost ovog poglavlja je ispitivanje vodi li uključivanje dodatne perspektive kamere uniformnijoj i sličnijoj izvedbi segmentacije instanci iz svih perspektiva. Stoga će se pažljivo analizirati razlika u izvedbi za obje perspektive kamere. Glavni cilj ostaje pružiti sveobuhvatnu analizu učinkovitosti i ograničenja testiranih pristupa.

6.2.1 Ukupni rezultati

U ovom odjeljku koji se fokusira na ukupne rezultate, neophodno je prvo procijeniti metrike izvedbe novih osnovnih perspektiva - prednje i lijeve kamere - za svaki model. Počevši s Mask R-CNN-om, model pokazuje dosljednu izvedbu u oba zadatka: detekciji objekata (*boxset*) i segmentaciji instanci (*maskset*) (vidi Sliku 19). Nema značajne razlike između ove dvije metrike, potvrđujući uravnotežene sposobnosti modela u oba zadatka.

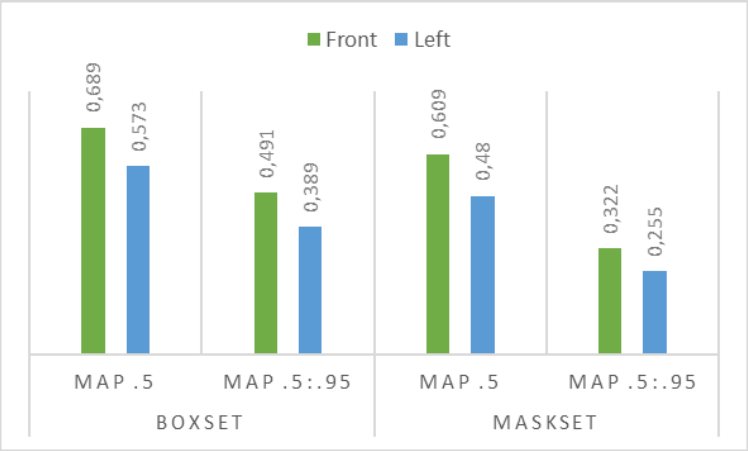
Zanimljivo je da se primjećuje suptilan, ali primjetan trend: izvedba lijeve kamere je marginalno superiorna u odnosu na prednju kameru. Konkretno, lijeva kamera pokazuje viši mAP za 5 postotna boda na .5 *boxset-u* i za 3 postotna boda viši mAP na .5 *maskset-u* u usporedbi s prednjom kamerom. Iako ove razlike nisu velike, postoje i ne smiju se zanemariti. Proširujući prag na mAP .5:.95, obje perspektive pokazuju pad u izvedbi, međutim, lijeva kamera i dalje marginalno nadmašuje prednju. S obzirom na ova promatranja, postaje ključno uzeti u obzir ove trendove prilikom evaluacije metrika izvedbe za druge perspektive poput desne i stražnje kamere.



Slika 19: Usporedba performansi osnovnih perspektiva na ukupnim rezultatima modela Mask R-CNN.

Prelazeći na YOLO, komparativna analiza između perspektiva prednje i lijeve kamere ilustrirana je na Slici 20. Vrijedi istaknuti da YOLO nadmašuje Mask R-CNN u svim mAP metrikama, kao što je to bio slučaj i u prethodnom eksperimentu. Detaljnijom analizom performansi YOLO-a, model ponovno potvrđuje svoju primarnu snagu u detekciji objekata umjesto segmentaciji instanci, s vidljivom razlikom od 10 postotna boda

između ta dva zadatka. Zanimljivo je da se jasan trend pojavljuje pri usporedbi podataka prednje i lijeve kamere - prednja kamera daje preciznije rezultate u ovom slučaju. Ova opservacija potkrepljena je povećanjem od 11,5 postotnih bodova u mAP na .5 u *boxset-u* te povećanjem od 7 postotnih bodova u mAP na .5 u *maskset-u*, čineći razliku značajnom. Proširujući prag na mAP .5:.95, ovaj trend se nastavlja, a također je vrijedno napomenuti da je pad performansi YOLO-a nešto manje izražen u usporedbi s Mask R-CNN-om.



Slika 20: Usporedba performansi osnovnih perspektiva na ukupnim rezultatima YOLOv7 modela.

Sljedeće, usmjeravamo pažnju prema Tablici 6 kako bi vidjeli kako se nove kamere, koje nisu bile dio naših izvornih osnovnih vrijednosti, nose s rezultatima u odnosu na osnovne perspektive. Budući da je naša glavna evaluacija na segmentaciji instanci, fokusirat ćemo se na te metrike. S Mask R-CNN-om, primjećujemo da su mAP rezultati za desnu i stražnju kameru gotovo identični osnovnim perspektivama, razlikujući se za samo oko 1 postotni bod. Unatoč tome, zanimljivo je vidjeti da je performansa desne kamere više u skladu s prednjom kamerom nego s lijevom, posebno uzimajući u obzir da lijeva kamera ima bolje rezultate od desne. Slično tome, stražnja kamera također se više naginje prema osnovnoj vrijednosti prednje kamere nego prema lijevoj. Iako govorimo o malim varijacijama ovdje, ukupna performansa kroz različite poglede s Mask R-CNN-om čini se prilično konzistentnom, čak i kada povećamo prag preciznosti.

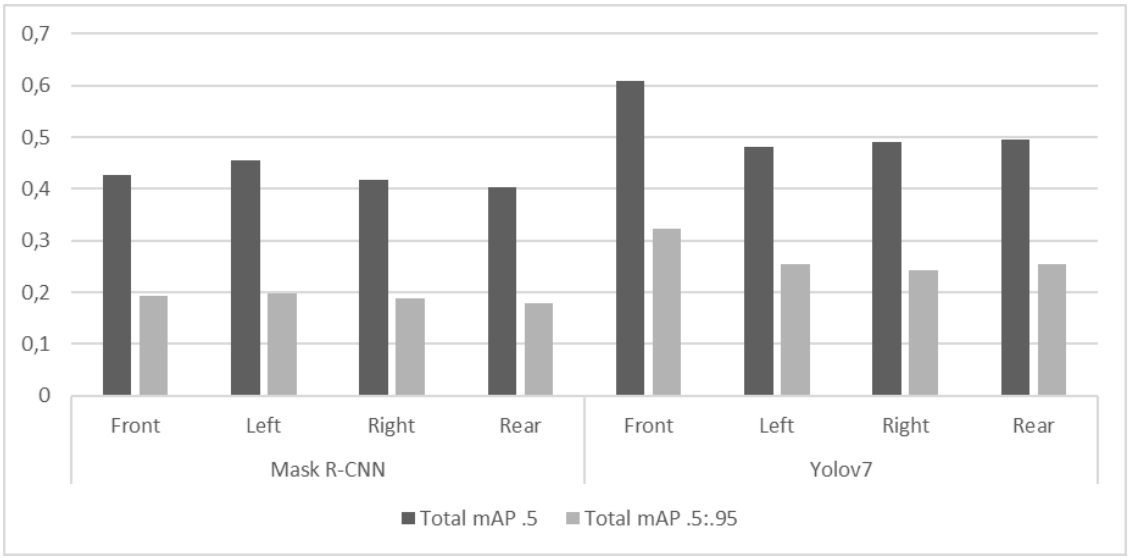
Prelazeći na YOLO, nalazimo još čvršće usklađivanje između desne i stražnje kamere. Razlika nije veća od 0.4 postotnih bodova na mAP .5 razini i 1.1 postotnih

bodova na višim pragovima. Vrijedi napomenuti da je performansa desne kamere više usklađena s osnovnom vrijednosti lijeve kamere, čak je i blago nadmašuje kada se prag povećava. Slično tome, stražnja kamera također se više naginje prema osnovnoj vrijednosti lijeve kamere, nadmašujući je za 1.5 postotnih bodova na mAP .5. Male razlike u performansama samo naglašavaju koliko su slični rezultati kroz perspektive. Međutim, ova uniformnost je donekle zasjenjena izvanrednom izvedbom prednje kamere, koja je bolja za otprilike 11 postotnih bodova od ostalih pogleda. Iako je cilj vidjeti može li dijagonalna perspektiva biti jednako učinkovita, znatno bolja izvedba prednje kamere narušava ovu uniformnost.

Model	Camera	Boxset		Maskset	
		mAP .5	mAP .5:.95	mAP .5	mAP .5:.95
Mask R-CNN	Front	0.455	0.242	0.426	0.194
	Left	0.507	0.265	0.454	0.198
	Right	0.435	0.231	0.417	0.187
	difference Front	*-0.02*	*-0.011*	*-0.009*	*-0.007*
	difference Left	*-0.072*	*-0.034*	*-0.037*	*-0.011*
	Rear	0.435	0.22	0.404	0.178
	difference Front	*-0.02*	*-0.022*	*-0.022*	*-0.016*
	difference Left	*-0.072*	*-0.045*	*-0.05*	*-0.02*
Yolov7	Front	0.689	0.491	0.609	0.322
	Left	0.573	0.389	0.48	0.255
	Right	0.564	0.375	0.491	0.243
	difference Front	*-0.125*	*-0.116*	*-0.118*	*-0.079*
	difference Left	*0.009*	*0.014*	*-0.011*	*0.012*
	Rear	0.619	0.386	0.495	0.254
	difference Front	*-0.07*	*-0.105*	*-0.114*	*-0.068*
	difference Left	*0.046*	*-0.003*	*0.015*	*-0.001*

Tablica 6: Ukupni rezultati drugog eksperimenta.

Vizualni pregled performansi svakog modela kroz različite perspektive može se pronaći na slici 21. Ovdje je vidljivo da Mask R-CNN pokazuje konzistentne performanse kroz sve kutove kamere, dok se YOLO posebno ističe s prednjom kamerom. Ova konzistentnost u performansama sugerira da su oba modela sada razmjerno sposobna za generalizaciju kroz cijelu 360-stupanjsku perspektivu.



Slika 21: Grafički prikaz ukupnih rezultata preko mAP vrijednosti drugog eksperimenta.

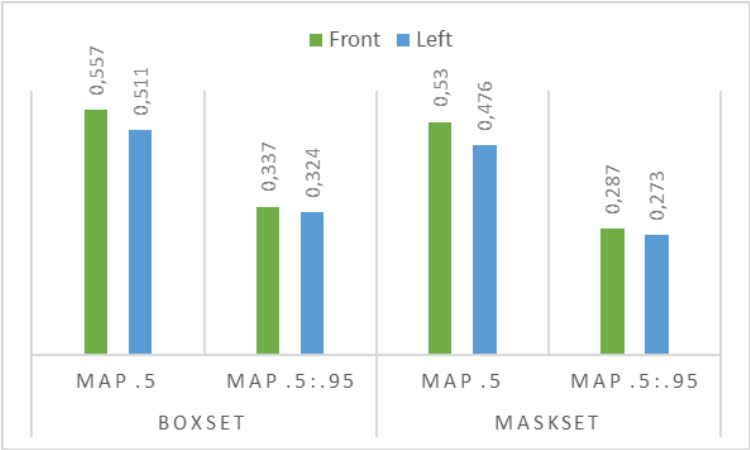
6.2.2 Rezultati segmentacije vozila

Dok ukupni rezultati sugeriraju da su oba modela prilično sposobna generalizirati preko različitih pogleda (osim YOLO-ve prednje kamere koja značajno bolje segmentira objekte), to nije konačan dokaz. Naša evaluacija obuhvaća segmentaciju i vozila i prolaznika, te je bitno ispitati svaku kategoriju zasebno.

Što se tiče dodatne lijeve kamere, ona nam obično pruža izbliza, sa strane, ili dijagonalne poglede na vozila koja voze pored nas. U rijetkim prilikama, kao što je prelazak ceste, također možemo vidjeti prednji ili stražnji dio vozila iz daljine, pogled koji češće hvata prednja kamera. Ovaj pristup dijagonalnog pogleda, što znači uključivanje prednje i lijeve kamere za trening, općenito pruža sveobuhvatan pregled dimenzija i neposredne okoline vozila.

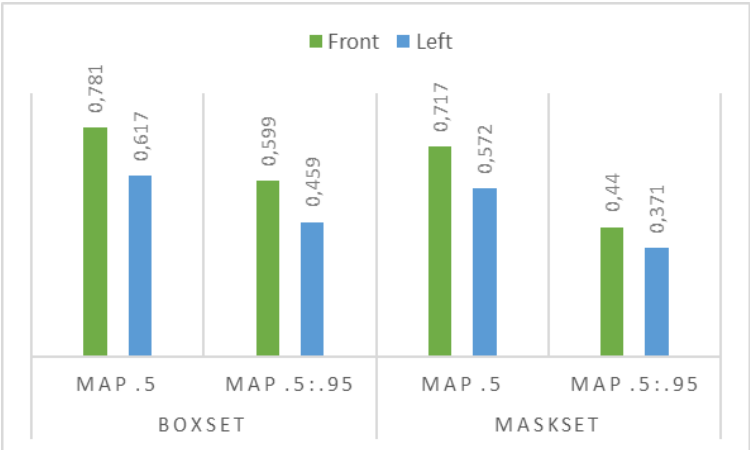
Da bismo ispitivali osnovne performanse, počnimo s Mask R-CNN-om, koji već pokazuje neke značajne prednosti. Za detekciju objekata, metrika mAP .5 prednje kamere je za otprilike 4.5 postotnih bodova viša, a za segmentaciju instanci je za oko 5.5 postotnih bodova bolja. Ovo su značajne razlike na ovoj razini praga. Međutim, kada povećavamo prag, prednost se smanjuje na zanemarivu razinu, ne prelazeći 1.5 postotnih bodova. Uključivanje lijeve kamere za segmentaciju vozila čini se da može imati značajan utjecaj, potencijalno poboljšavajući sposobnost modela da generalizira preko različitih perspek-

tiva, posebno bočnih perspektiva koje su bile problematične u prošlom eksperimentu. Ovi nalazi su grafički predstavljeni na Slici 22.



Slika 22: Usporedba performansi osnovnih perspektiva na rezultatima vozila modela Mask R-CNN.

Promatrajući YOLO, prednja kamera se primjetno ističe u usporedbi. Za detekciju objekata, prednost je značajnih 16,5 postotnih bodova, a za segmentaciju instanci 14,5 postotnih bodova. Štoviše, ova značajna prednost održava se čak i kada se prag podigne, ostajući na 14 postotnih bodova za detekciju i 7 postotnih bodova za segmentaciju. To bi nas moglo navesti na zaključak da čak i uz uključivanje lijeve perspektive, sposobnosti generalizacije bočnih kutova možda i dalje zaostaju u usporedbi s prednjom kamerom. Uvidi su vidljivi na Slici 23.



Slika 23: Usporedba performansi osnovnih perspektiva na rezultatima vozila Yolov7 modela.

Sada ćemo se osvrnuti na Tablicu 7 za detaljan pregled učinkovitosti segmentacije vozila. Počevši s analizom za Mask R-CNN, naša početna hipoteza sugerirala je da bi desna kamera mogla pokazati poboljšane performanse, s obzirom na potencijalne sličnosti s lijevom kamerom, posebno u snimanju vozila izbliza. Međutim, rezultati ne potvrđuju tu pretpostavku. Iako je performansa donekle bliža onoj lijeve kamere, postoji značajan pad od približno 12 postotnih bodova. U usporedbi s prednjom kamerom, ova razlika postaje još izraženija, dosežući značajan pad od 17,5 postotnih bodova. Povećanje praga za mAP evaluacijski kriterij ne mijenja ovaj trend. Međutim, performansa stražnje kamere, koju bi se moglo očekivati da će paralelno pratiti prednju kameru, pokazuje samo skroman pad od približno 5,5 postotnih bodova. Zanimljivo je da je njena performansa rame uz rame s lijevom kamerom. Ovaj obrazac sugerira da uključivanje lijeve kamere u proces treninga nije dovelo do uravnoteženog poboljšanja sposobnosti segmentacije bočnih perspektiva.

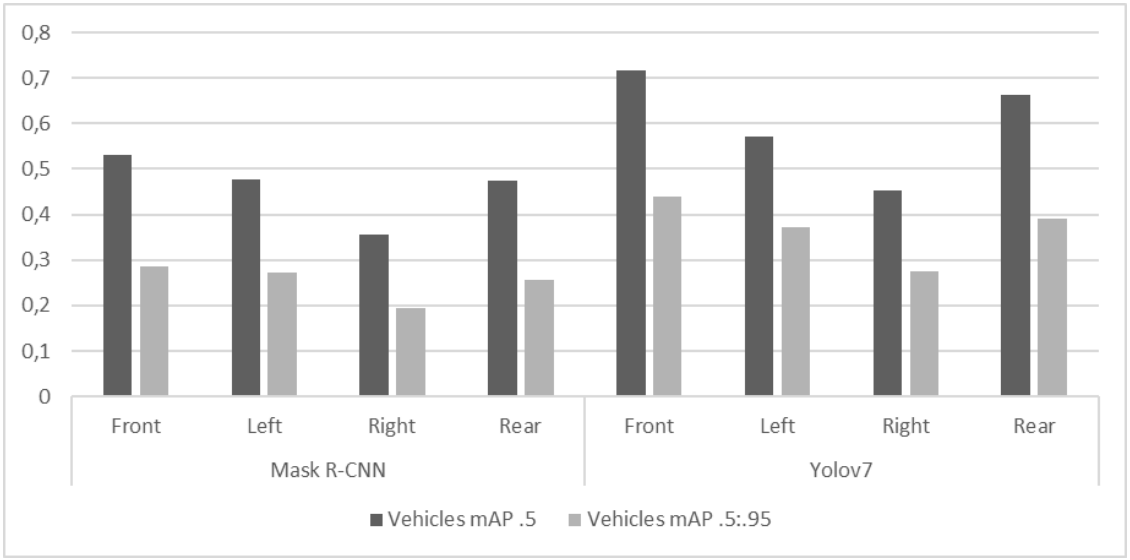
Model	Camera	Boxset		Maskset	
		mAP .5	mAP .5:.95	mAP .5	mAP .5:.95
Mask R-CNN	Front	0.557	0.337	0.53	0.287
	Left	0.511	0.324	0.476	0.273
	Right	0.375	0.22	0.355	0.195
	difference Front	*-0.182*	*-0.117*	*-0.175*	*-0.092*
	difference Left	*-0.136*	*-0.104*	*-0.121*	*-0.078*
	Rear	0.494	0.283	0.474	0.256
	difference Front	*-0.063*	*-0.054*	*-0.056*	*-0.031*
	difference Left	*-0.017*	*-0.041*	*-0.002*	*-0.017*
Yolov7	Front	0.781	0.599	0.717	0.44
	Left	0.617	0.459	0.572	0.371
	Right	0.512	0.347	0.454	0.276
	difference Front	*-0.269*	*-0.252*	*-0.263*	*-0.164*
	difference Left	*-0.105*	*-0.112*	*-0.118*	*-0.095*
	Rear	0.731	0.507	0.663	0.391
	difference Front	*-0.05*	*-0.092*	*-0.054*	*-0.049*
	difference Left	*0.114*	*0.048*	*0.091*	*0.02*

Tablica 7: Ukupni rezultati segmentacije vozila drugog eksperimenta.

U slučaju YOLO-a, zapaženi obrazac donekle je sličan onome kod Mask R-CNN-a.

Desna kamera pokazuje značajan zaostatak u performansama, prikazujući jaz od 26 postotnih bodova u usporedbi s prednjom kamerom i jaz od 11,8 postotnih bodova s lijevom kamerom. Ova suboptimalna performansa se održava, čak i povećavanjem praga. Stoga je očito da ni YOLO nije optimizirao svoje sposobnosti segmentacije za neviđenu bočnu perspektivu vozila, čak ni kada je treniran s lijevom stranom. Što se tiče stražnje kamere, ona pokazuje razinu performansi sličnu prednjoj kameri, zaostajući za približno 5,5 postotnih bodova. Međutim, nadmašuje performansu osnovne lijeve kamere za značajnih 9 postotnih bodova. Ovaj obrazac se održava čak i kada se prag poveća.

Iako su ukupni rezultati iz prethodnog odjeljka na prvi pogled sugerirali prilično ravnomjernu distribuciju performansi preko različitih pogleda za segmentaciju instanci, rezultati za segmentaciju vozila (vidi Sliku 24) sugeriraju suprotno. Oba modela podbacuju kada je riječ o bočnim perspektivama. Međutim, Mask R-CNN pokazuje bolje sposobnosti generalizacije - performanse lijeve i stražnje perspektive usko su usklađene, a prednja perspektiva nije daleko ispred. Desna perspektiva, s druge strane, značajno zaostaje. Uz navedeno, YOLO doživljava značajan pad u performansama preko bočnih perspektiva, pri čemu desni pogled pokazuje najslabije rezultate.



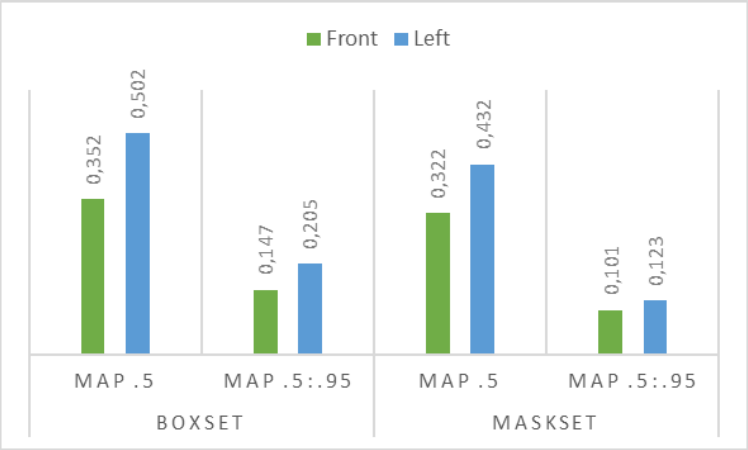
Slika 24: Grafički prikaz rezultata vozila preko mAP vrijednosti drugog eksperimenta.

6.2.3 Rezultati segmentacije prolaznika

S obzirom na uočene nedostatke u segmentaciji vozila i očitu nesposobnost modela da se učinkovito generaliziraju preko različitih perspektiva za detekciju vozila, očekujemo poboljšane rezultate u detekciji prolaznika, s obzirom da ukupni rezultati pokazuju prilično konzistentnu distribuciju.

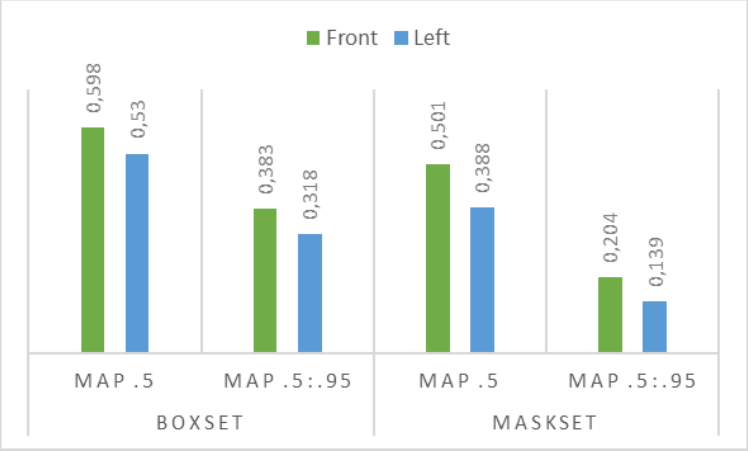
Uključivanje lijeve kamere jednostavno pruža detaljnije poglede na prolaznike, koji se često nalaze kako hodaju po nogostupima ili voze vozila poput motocikla i bicikle. Pretpostavljamo da ova dodatna perspektiva značajno poboljšava skup podataka za prolaznike, pogotovo s obzirom na to da prednja kamera obično snima prolaznike ili na daljinu ili rijetko blizu na pješačkim prijelazima.

Uspoređujući osnovne rezultate, ustanovili smo da Mask R-CNN daje bolje rezultate s lijevom kamerom nego s prednjom kamerom. Ovi zaključci mogu se pripisati činjenici da lijeva kamera vidi prolaznike na znatno većoj i detaljnijoj skali, kako je već spomenuto. U detaljnijoj analizi, lijeva kamera pokazuje poboljšanje od 15 postotnih bodova u detekciji objekata i poboljšanje od 11 postotnih bodova u segmentaciji instanci, mjerenom mAP od 0.5. Kako povećavamo prag, trendovi performansi ostaju konzistentni. Međutim, važno je naglasiti da su vrijednosti mAP-a pri višim pragovima znatno niže, s segmentacijom instanci koja čak pada na samo 10% za prednju kameru. To može biti zato što prolaznici predstavljaju veću složenost od vozila, s obzirom na varijabilnost pozicija udova i kretanja kroz različite scene. Slika 25 vizualno predstavlja ove nalaze.



Slika 25: Usporedba performansi osnovnih perspektiva na rezultatima prolaznika modela Mask R-CNN.

Prebacujući fokus na YOLO, primjećujemo da prednja kamera još uvijek nadmašuje lijevu kameru, iako je razlika u performansama uža u usporedbi s segmentacijom vozila. Za detekciju objekata, razlika je 7 postotnih bodova, a za segmentaciju instanci iznosi 11 postotnih bodova. Ove razlike ostaju čak i kada se pragovi povećavaju. Međutim, vrijedno je napomenuti da se jaz u performansama znatno proširuje za segmentaciju instanci na višim pragovima. To dokazuje da segmentacija instanci postaje sve izazovnija kako zadatak zahtijeva preciznije definiranje pojedinačnih značajki, pogotovo za algoritam poput YOLO-a koji je primarno detektor objekata. Ovi uvidi prenose se na slici 26.



Slika 26: Usporedba performansi osnovnih perspektiva na rezultatima prolaznika YOLOv7 modela.

Preusmjeravajući pažnju na performanse iz ne treniranih perspektiva, pogledajmo

Tablicu 8. U početku, s Mask R-CNN-om, očito je da desna i stražnja kamera općenito nadmašuju osnovne perspektive. Superiorna izvedba desne kamere posebno je impresivna - nadmašuje prednju za gotovo 16 postotnih bodova i lijevu za 5 postotnih bodova u segmentaciji instanci procijenjenoj na mAP od 0,5. Ova tendencija se održava čak i kada se prag povećava, iako se vrijednosti smanjuju, vjerojatno zbog inherentnih složenosti ljudskih oblika. Logično je da prolaznici koji se pojavljuju u najvećim veličinama kada se gledaju s desne strane - budući da hodaju pored nas po pločniku - doprinose ovom ishodu. Kao rezultat, desna kamera postiže najviši mAP za segmentaciju prolaznika s Mask R-CNN-om, unatoč tome što model nije bio obučen za ovu točku gledišta. Stražnja kamera je usporediva s prednjom, ali zaostaje za lijevom kamerom za oko 10 postotnih bodova.

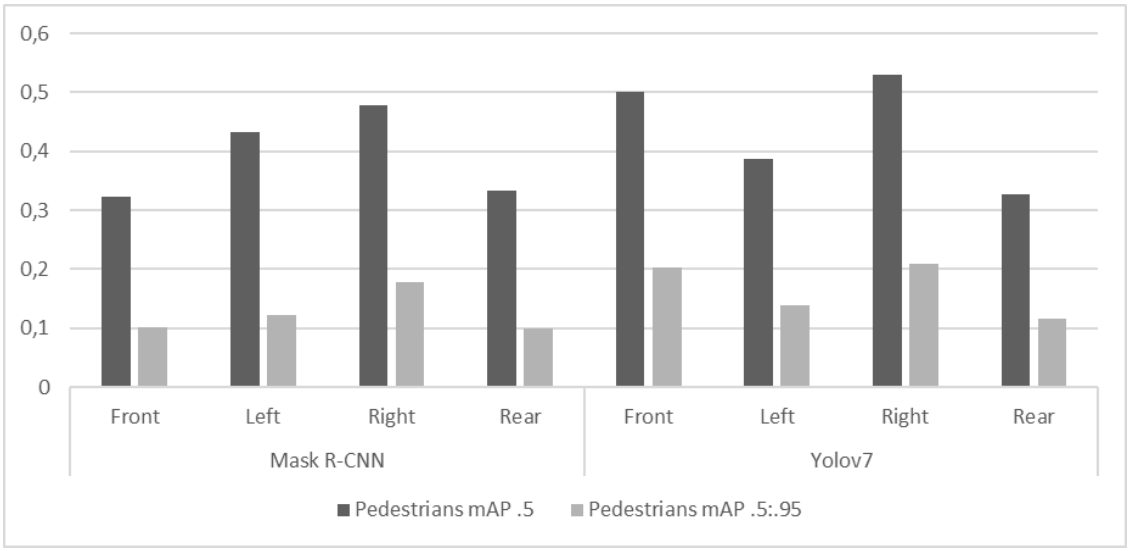
Što se tiče YOLO modela, nailazimo na neke zanimljive rezultate. Zapaženo, desna kamera, neobučena točka gledišta, nadmašuje obje osnovne perspektive. Iako nadmašuje prednju kameru za samo 3 postotnih bodova, to je značajno s obzirom na to da nijedna druga perspektiva ne postiže to, uključujući lijevu perspektivu, koja je bila dio trening podataka. Desna kamera nadmašuje osnovnu lijevu kameru impresivnih 14 postotnih bodova. Ovaj obrazac je istinit pri višim pragovima, iako su mAP rezultati skromni. Među četiri kamere, stražnja kamera je najslabija karika, zaostajući za prednjom za gotovo 17,5 postotnih bodova i lijevom za 6 postotnih bodova. To bi moglo biti zato što stražnja kamera obično snima prolaznike s veće udaljenosti ili hvata samo djelomičan pogled na njih na pločniku, s obzirom na to da automobili obično ne staju izravno ispred pješačkih prijelaza. Ukupno gledano, možemo zaključiti da je dodavanje krupnih snimaka prolaznika s lijeve kamere značajno poboljšalo YOLO-ovu izvedbu.

mAP rezultati segmentacije prolaznika otkrivaju nekonzistentne performanse preko različitih kamerinih perspektiva, kako je ilustrirano na Slici 27. U slučaju Mask R-CNN-a, čini se da prednja i stražnja kamera nisu tako efikasne, vjerojatno zato što bočne perspektive izvrsno rade s većim, bližim instancama prolaznika. Desna kamera posebno pokazuje snažne rezultate u rezultatima oba modela. Za YOLO, prednja i desna kamera postižu bolje rezultate, dok lijeva i stražnja kamera zaostaju, unatoč uključivanju podataka lijeve kamere tijekom treninga. Sveukupno, ovi podaci impliciraju da dok je generalizacija

Model	Camera	Boxset		Maskset	
		mAP .5	mAP .5:.95	mAP .5	mAP .5:.95
Mask R-CNN	Front	0.352	0.147	0.322	0.101
	Left	0.502	0.205	0.432	0.123
	Right	0.495	0.242	0.479	0.178
	difference Front	*0.143*	*0.095*	*0.157*	*0.077*
	difference Left	*-0.007*	*0.037*	*0.047*	*0.055*
	Rear	0.377	0.158	0.334	0.1
	difference Front	*0.025*	*0.011*	*0.012*	*-0.001*
	difference Left	*-0.125*	*-0.047*	*-0.098*	*-0.023*
Yolov7	Front	0.598	0.383	0.501	0.204
	Left	0.53	0.318	0.388	0.139
	Right	0.616	0.403	0.529	0.21
	difference Front	*0.018*	*0.02*	*0.028*	*0.006*
	difference Left	*0.086*	*0.085*	*0.141*	*0.071*
	Rear	0.508	0.265	0.327	0.117
	difference Front	*-0.09*	*-0.118*	*-0.174*	*-0.087*
	difference Left	*-0.022*	*-0.053*	*-0.061*	*-0.022*

Tablica 8: Ukupni rezultati segmentacije prolaznika drugog eksperimenta.

Mask R-CNN-a poboljšana, YOLO-ova je samo za izrazito velike prikaze prolaznika na desnoj kameri.



Slika 27: Grafički prikaz rezultata prolaznika preko mAP vrijednosti drugog eksperimenta.

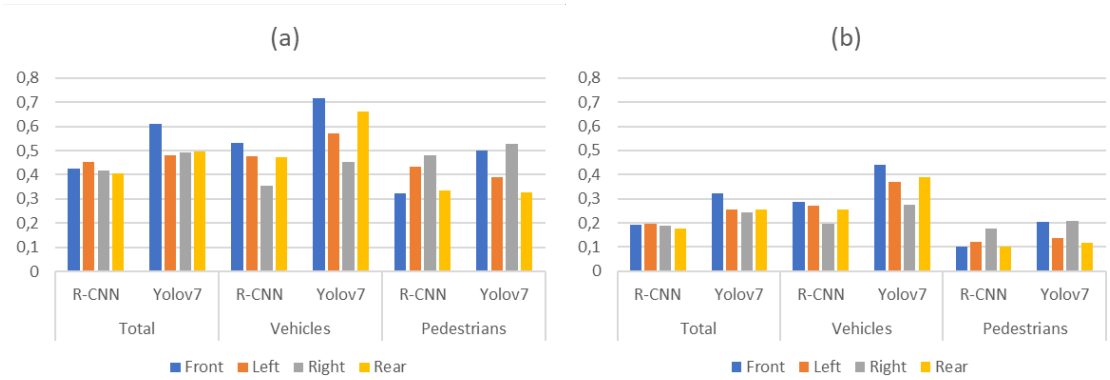
6.3 Zaključci

Primarni cilj ovog poglavlja bio je procijeniti hoće li uključivanje dodatne lijeve kamere uz prednju kameru tijekom faze treninga rezultirati uravnoteženim performansama preko različitih perspektiva—prednje, lijeve, desne i stražnje kamere. Na prvi pogled, ukupni rezultati mogu ostaviti dojam da je ovaj cilj postignut. Međutim, dublja analiza segmentacije vozila i prolaznika otkriva razlike koje ukazuju na to da ova ravnoteža nije u potpunosti ostvarena.

Kao što je prikazano na Slici 28, rezultati segmentacije instanci prikazani su kroz metrike mAP .5 i mAP .5:.95. Slika naglašava da YOLO i dalje postiže bolje rezultate u smislu viših mAP rezultata u većini scenarija. Kada je riječ o generalizaciji, posebno sada kada su modeli trenirani koristeći perspektive obje prednje i lijeve kamere, zaključci su manje jasni. Mask R-CNN se muči sa segmentacijom vozila kada se promatra s desne kamere, iako je relativno bolji s lijevim i stražnjim pogledima, približavajući se rezultatima prednje kamere. Ovo je superiorno u odnosu na YOLO, gdje performanse značajno opadaju za bočne perspektive, iako su rezultati stražnjeg pogleda gotovo jednako dobri kao i prednji. Govoreći o segmentaciji prolaznika, bočne perspektive izrazito dobro segmentiraju. Kod YOLO-a situacija gubi smisao, jer desna kamera segmentira izuzetno dobro kao i prednja, dok lijeva i desna značajno padaju. Vrijedi napomenuti da, kada se prag mAP povećava kako bi se procijenila preciznost modela u finijim detaljima, i YOLO i Mask R-CNN pokazuju izrazit pad u svojoj sposobnosti točne segmentacije prolaznika. Nasuprot tome, pogoršanje performansi segmentacije vozila pod ovim strožim mAP kriterijima je usporedno skromno za oba modela.

Nekonzistentnosti preko različitih perspektiva možda su uvjetovane različitom prirodom kretanja kamera, što značajno utječe na skup podataka snimljen s 360 stupnjeva. Bočne strane vozila izgledaju izrazito različito zbog raznolikosti vozila i stalnog kretanja. Kao rezultat toga, lijevi pogled je promjenjiviji od statičnijih prednjih i stražnjih perspektiva. Možemo zaključiti da je uključivanje lijeve kamere u trening dovelo do nekih poboljšanja u generalizaciji, poput YOLO-ove pojačane sposobnosti prepoznavanja prolaznika iz bliza. Međutim, performanse ostaju neujednačene, ostavljajući prostor za

daljnje istraživanje i optimizaciju.



Slika 28: Rezultati segmentacije instanci (*Maskset*) mjereni metrikom mAP koristeći dvije granice: (a) mAP .5 i (b) mAP .5:.95 drugog eksperimenta.

7 Proučavanje varijabilnosti klasa između dva eksperimentalna uvjeta

Nakon što smo zaključili rezultate, naša preporuka je da uključivanje dodatne lijeve kamere tijekom faze treninga jest poboljšalo sposobnost generalizacije modela u nekim aspektima, iako ne u svima. Jasno je da zadaci segmentacije vozila i prolaznika zahtijevaju individualne procjene. Stoga, cilj ovog odjeljka je analizirati razlike u segmentaciji vozila i prolaznika te ispitati kako dodavanje lijeve kamere u drugom eksperimentu utječe na rezultate u odnosu na prvi eksperiment. Za potrebe ovog brzog sažetka, ograničit ćemo našu raspravu na mAP .5 metrike dobivene iz skupa maski, koncentrirajući se isključivo na osnovnu mjeru segmentacije instanci.

7.1 Detaljniji pogled na zadatak segmentacije vozila

Prvo, važno je napomenuti da CARLA generira širok spektar vozila, uključujući različite vrste automobila, kamiona i manjih oblika poput motocikala i bicikala. Ova raznolikost je ključna jer ovi tipovi vozila pokazuju značajne razlike. Na primjer, kamion i bicikl dijele malo zajedničkog, možda samo zaobljeni oblik njihovih kotača.

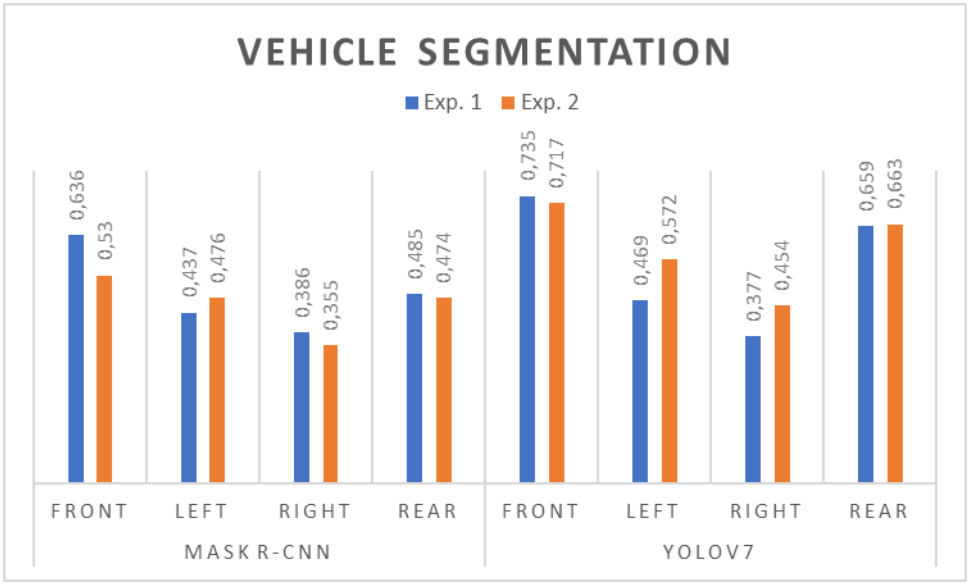
Nadalje, vrijedi istaknuti kontrast između bočnih gledišta i prednjih/zadnjih gledišta prilikom rasprave o dinamici vozila. Prednji i zadnji pogledi često pokazuju sličnosti. Obično ili prikazuju automobil u prometu među ostalim vozilima ili dok je kretanje zaustavljeno na raskrižjima, stop znakovima ili crvenim svjetlima. Rijetko vidimo automobil koji skreće u lijevu ili desnu traku, što je jedini trenutak kada možemo vidjeti nešto drugačiju perspektivu. U većini slučajeva promatramo prednji ili stražnji dio obližnjih vozila, kao i dijagonalni pogled na susjedna vozila koja nisu u našoj traci. Primjere ovoga možete vidjeti na Slici 29.

Slika 29: Primjeri slika vozila snimljenih prednjom i stražnjom kamerom.

Što se tiče bočnih perspektiva, obično pružaju raznolike poglede na vozila dok automobil ide naprijed i scenografija se neprekidno mijenja. Lijeve i desne kamere snimaju krupne planove strana vozila, pri čemu lijeva kamera pokriva više traka kada se vozi s desne strane, a desna snima parkirana ili susjedna vozila ovisno o položaju na cesti. Vrijedi napomenuti da kamere često snimaju nepotpuna vozila i scene bez bilo kakvih vozila, posebno desna kamera. Uzimajući u obzir raznovrsnost vrsta vozila prisutnih u ovoj simulaciji, kao i u stvarnom svijetu, zaključujemo da je ova perspektiva izrazito dinamična i nekonzistentna. Neki ilustrativni primjeri za svaku bočnu perspektivu mogu se naći na Slici 30.

Slika 30: Primjeri slika vozila snimljenih lijevom i desnom kamerom.

Kako bismo usporedili dva eksperimenta, pogledamo graf na Slici 31 koji prikazuje rezultate mAP .5 *maskset-a* za svaku kameru. U slučaju Mask R-CNN-a, performanse prednje kamere smanjile su se za 10 postotnih bodova u drugom eksperimentu, vjerojatno zato što su oba eksperimenta imala isti broj epoha a značajno različitu veličinu skupa podataka. Međutim, i dalje možemo procijeniti generalizaciju, i zaključujemo da su rezultati kroz sve perspektive bili sličniji u drugom eksperimentu. Konkretno, možemo vidjeti da desna kamera najlošije segmentira, s razlikom od 25 postotnih bodova od prednje u prvom eksperimentu i 17,5 postotnih bodova u drugom eksperimentu što je već bolje. S YOLO modelom, rezultati za prednju i zadnju kameru ostali su blizu u mAP-u u oba eksperimenta. Dodavanje lijeve kamere napravilo je jasnu razliku, povećavajući rezultate za bočne poglede čak za 12 postotnih bodova za lijevi i 7 postotnih bodova za desni pogled.



Slika 31: Usporedba rezultata segmentacije vozila u oba eksperimenta.

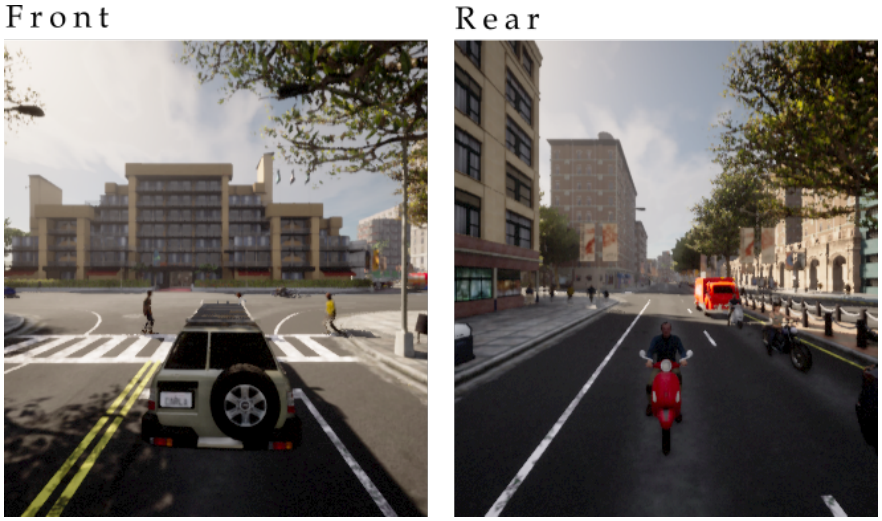
Iako je dobro vidjeti da oba modela napreduju s dodatnim pogledom iz lijeve kamere, razočaravajuće je što bočni pogledi nisu nadmašili prednje i zadnje poglede. To je iznenađujuće jer se očekuje da bi bočni pogledi trebali bolje segmentirati s obzirom na to da se obližnja vozila mogu približiti prilično blizu. Ova nekonzistentnost može biti zbog široke raznolikosti vozila i znatno manjeg broja slika u setu podataka na kojima se oni nalaze, što otežava modelima da uče jednako učinkovito kao što su to činili s prednjom

kamerom.

7.2 Dublji pogled na zadatak segmentacije prolaznika

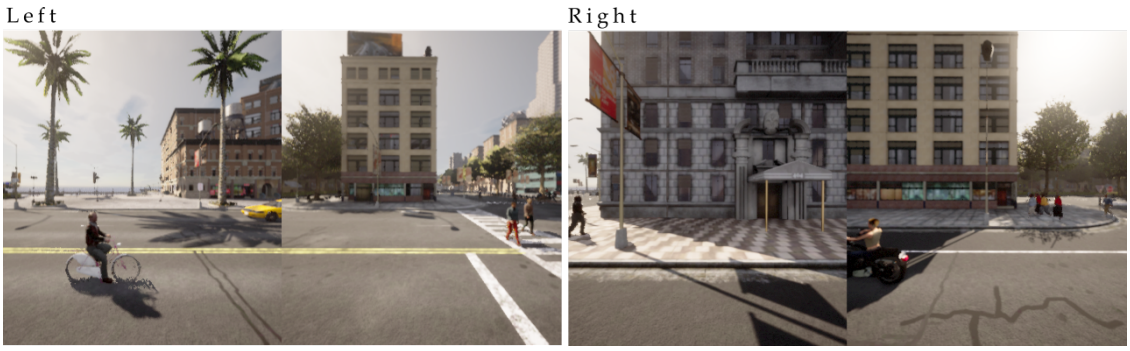
Da bismo bolje razumjeli koliko je svaki eksperiment prepoznavao prolaznike, prvo ćemo pojasniti što mislimo pod "prolaznicima" u ovom kontekstu. Simulator CARLA omogućuje nam da razlikujemo ljude koji voze motocikle i bicikle. Dakle, za naše potrebe, kada govorimo o prolaznicima, mislimo na osobe koje hodaju po nogostupima, prelaze ulice, ali i one koje voze vozila.

Promatrajući perspektive kamere, primjećujemo značajne razlike u tome kako prolaznici izgledaju ovisno o kutu. S prednjih i zadnjih kamera, prolaznici obično kratko ulaze u vidno polje dok hodaju po nogostupima sve dok ne pređu u bočni okvir. Ako su na nogostupu, obično se pojavljuju vrlo malih dimenzija. Prolaznici se pojavljuju s bliže udaljenosti u pogledu prednje kamere kada prelaze ulice, što se ne događa često za zadnju kameru jer obično nismo zaustavljeni ispred pješačkih prijelaza za taj pogled. Stoga je rijetko vidjeti detaljni pogled na prolaznika koji hoda po nogostupu, ali često možemo bolje vidjeti nekoga tko vozi vozilo ako je odmah do nas na cesti. U tim slučajevima obično vidimo dijelove njihove glave i ruku ako su izravno ispred ili iza, te cijelo tijelo kada ih možemo vidjeti dijagonalno. Ove karakteristike mogu se vidjeti na Slici 32.



Slika 32: Primjeri slika prolaznika snimljenih prednjom i stražnjom kamerom.

Kada govorimo o bočnim pogledima, situacija je drugačija. Bilo da su prolaznici na nogostupu, usred pješačkog prijelaza ili na vozilu, vjerojatnije je da ćemo vidjeti cijela njihova tijela — i često s bliže udaljenosti. Desna kamera, posebno, obično prikazuje prolaznike mnogo veće nego što to čini lijeva kamera. Snimamo raznovrsniji i veći set slika prolaznika s bočnih perspektiva u usporedbi s onim što dobivamo s prednjih i zadnjih kamera. Za vizualnije razumijevanje, pogledajte Sliku 33.

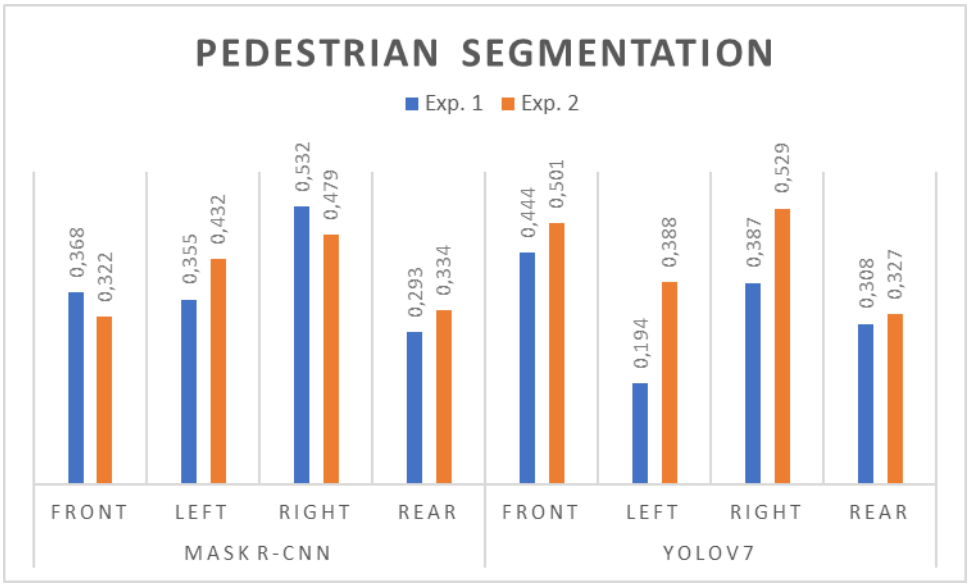


Slika 33: Primjeri slika prolaznika snimljenih lijevom i desnom kamerom.

Preusmjeravajući pažnju na performanse svakog modela u eksperimentima, nailazimo na neka zanimljiva saznanja (vidi Sliku 34). Za Mask R-CNN, drugi eksperiment pokazao je pad performansi od 4,5 postotnih bodova s prednjom kamerom, ali gotovo 8 postotnih bodova poboljšanja s lijevom kamerom. Zanimljivo je da desna kamera bolje izvodi od osnovnih perspektiva u oba eksperimenta, ali je zapravo bila neznatno lošija u drugom u usporedbi s prvim. Zadnja kamera konzistentno zaostaje, najvjerojatnije zato što rijetko snima velike, detaljne slike prolaznika. Iako je drugi eksperiment dao konzistentnije rezultate, ukazujući na bolje sposobnosti generalizacije, Mask R-CNN je već dokazao svoju sposobnost učinkovite generalizacije koristeći samo prednju kameru, posebno za poglede s desne strane, i pokazao je samo marginalna poboljšanja za lijevu i zadnju kameru.

Što se tiče YOLO-a, jasno je da je uključivanje dodatne kamere tijekom treninga dovelo do znatno boljih rezultata, posebno za bočne poglede. Performanse lijeve kamere dramatično su se poboljšale, povećavajući mAP za impresivnih 19,5 postotnih bodova. U međuvremenu, desna kamera premašila je sve prethodne rezultate u oba eksperimenta i čak se približila razini performansi Mask R-CNN-a u prvom eksperimentu. To ukazuje

na to da YOLO sada ima poboljšanu sposobnost generalizacije, posebno kada ima pristup bližim slikama prolaznika. Sve u svemu, i performanse i pouzdanost YOLO-a doživjele su značajna poboljšanja.



Slika 34: Usporedba rezultata segmentacije prolaznika u oba eksperimenta.

8 Sažetak i zaključak

Kroz ovaj rad, istražili smo područje tehnologije temeljene na kamerama u vozilima, rastuće područje koje ima implikacije za sve, od pomoći vozačima do potpuno autonomnih vozila. Naša istraga usredotočena je na učinkovitost korištenja ograničenih perspektiva kamere umjesto sveobuhvatnog pogleda od 360 stupnjeva za zadatak točne segmentacije dinamičkih prometnih elemenata, poput vozila i prolaznika. Temeljna logika ovog suženog fokusa bila je procijeniti je li moguće minimizirati opsežne resurse, financijske i računalne, koji su trenutno potrebni za prikupljanje podataka i trening modela u području autonomnih vozila. Kroz detaljnu kontekstualnu pozadinu, objašnjenja arhitekture modela i metodologije istraživanja, proveli smo dva eksperimenta koristeći različite skupove podataka, ali iste modele kako bismo postigli naše ciljeve.

Rezultati su otkrili ohrabrujuće i neočekivane uvide. Modeli trenirani isključivo na prednjoj kameri postižu najbolje rezultate kada se testiraju na toj istoj kameri, osim u jednom slučaju. Desna kamera modela Mask R-CNN jedina je koja premašuje rezultate osnovne kamere pri segmentaciji prolaznika, što potvrđuje njenu sposobnost generalizacije. Uključivanje dodatne lijeve kamere tijekom treninga donijelo je određena poboljšanja u sposobnostima generalizacije modela, naročito kod YOLO modela. To potvrđuje hipotezu da modeli, koristeći raznovrsniji skup podataka za trening, mogu postići uravnoteženije performanse iz različitih kuteva gledišta. No, jasno je da postizanje optimalnih performansi iz određenih perspektiva još uvijek predstavlja izazov, budući da rezultati nisu bili konzistentno bolji iz svih kutova kamere.

Nakon usporedbe eksperimenta možemo zaključiti da je segmentacija prolaznika bila uspješna, dok segmentacija vozila nije. U prometu težimo boljoj segmentaciji objekata koji su nam bliži, zbog čega bočne kamere oba modela nisu postigle značajnije rezultate. Jedan od mogućih načina za poboljšanje mogao bi dotreniravanje modela s fotografijama vozila snimljenih bočno iz blizine. Sve u svemu, potvrđujemo da trening modela s ograničenim perspektivama donosi značajne rezultate u segmentaciji cjelokupnog prometnog okruženja, i ovakav pristup s određenim prilagodbama, može se primjenjivati u istraživanju i implementaciji modela za segmentaciju dinamičkog okruženja u prometu.

Literatura

[Ros58] Frank Rosenblatt. "The perceptron: A probabilistic model for information storage and organization in the brain". In: *Psychological Review* 65.6 (1958), pp. 386–408. DOI: 10.1037/h0042519.

[RHW86] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323 (1986), pp. 533–536. URL: https://api.semanticscholar.org/CorpusID:205001834.

[CHL08] Y.M. Chiang, N.Z. Hsu, and K.L. Lin. "Driver assistance system based on monocular vision". In: *Proceedings of the New Frontiers in Applied Artificial Intelligence: 21st International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2008*. Wrocław, Poland: Springer: Berlin/Heidelberg, Germany, 2008, pp. 1–10.

[Rus10] Stuart J Russell. *Artificial intelligence a modern approach*. 3rd ed. Pearson Education, Inc., 2010, pp. 17–18.

[Mac16] Bohdan Macukow. "Neural Networks – State of Art, Brief History, Basic Models and Architecture". In: *Computer Information Systems and Industrial Management*. Ed. by Khalid Saeed and Władysław Homenda. Cham: Springer International Publishing, 2016, pp. 4–5. ISBN: 978-3-319-45378-1.

[Red+16] J. Redmon et al. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA, 2016, pp. 779–788.

[Bon17] Giuseppe Bonaccorso. *Machine Learning Algorithms: A Reference Guide to Popular Algorithms for Data Science and Machine Learning*. Packt Publishing, 2017. ISBN: 1785889621.

[Dos+17] A. Dosovitskiy et al. "CARLA: An open urban driving simulator". In: *Proceedings of the Conference on Robot Learning (PMLR)*. Mountain View, CA, USA, 2017, pp. 1–16.

[He+17] K. He et al. “Mask r-cnn”. In: *Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy, 2017, pp. 2961–2969.

[Liu+17] W. Liu et al. “A survey of deep neural network architectures and their applications”. In: *Neurocomputing* 234 (2017), pp. 11–26. URL: http://dx.doi.org/10.1016/j.neucom.2016.12.038.

[Par18] Ravindra Parmar. *Training Deep Neural Networks*. Accessed: 2023-08-15. 2018. URL: https://towardsdatascience.com/training-deep-neural-networks-9fdb1964b964.

[Vou+18] A. Voulodimos et al. “Deep learning for computer vision: A brief review”. In: *Comput. Intell. Neurosci.* 2018 (2018), p. 7068349. URL: http://dx.doi.org/10.1155/2018/7068349.

[Cha+19] D. Chang et al. “Multi-lane detection using instance segmentation and attentive voting”. In: *Proceedings of the 2019 19th International Conference on Control, Automation and Systems (ICCAS)*. 2019, pp. 1538–1542.

[HK19] Michael Haenlein and Andreas Kaplan. “A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence”. In: *California Management Review* 61.4 (2019), pp. 5–14. DOI: 10.1177/0008125619864925. eprint: https://doi.org/10.1177/0008125619864925. URL: https://doi.org/10.1177/0008125619864925.

[Mal19] M.A. Malbog. “MASK R-CNN for pedestrian crosswalk detection and instance segmentation”. In: *Proceedings of the 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*. Kuala Lumpur, Malaysia, 2019, pp. 1–5.

[Mih19] Ilija Mihajlovic. *Everything You Ever Wanted To Know About Computer Vision*. Accessed: 20-08-2023. Apr. 2019. URL: https://towardsdatascience.com/everything-you-ever-wanted-to-know-about-computer-vision-heres-a-look-why-it-s-so-awesome-e8a58dfb641e.

[ZZ20] B. Zhang and J. Zhang. "A traffic surveillance system for obtaining comprehensive information of the passing vehicles based on instance segmentation". In: *IEEE Trans. Intell. Transp. Syst.* 22 (2020), pp. 7040–7055. URL: http://dx.doi.org/10.1109/TITS.2020.3001154.

[Zha+20] H. Zhang et al. "A virtual-real interaction approach to object instance segmentation in traffic scenes". In: *IEEE Trans. Intell. Transp. Syst.* 22 (2020). [CrossRef], pp. 863–875. URL: http://dx.doi.org/10.1109/TITS.2019.2961145.

[Car+21] M. Carranza-García et al. "On the Performance of One-Stage and Two-Stage Object Detectors in Autonomous Vehicles Using Camera Data". In: *Remote Sensing* 13 (2021), p. 89. URL: http://dx.doi.org/10.3390/rs13010089.

[JVO21] P. Jaikumar, R. Vandaele, and V. Ojha. "Transfer learning for instance segmentation of waste bottles using Mask R-CNN algorithm". In: *Proceedings of the Intelligent Systems Design and Applications: 20th International Conference on Intelligent Systems Design and Applications (ISDA 2020)*. Springer: Berlin/Heidelberg, Germany, 2021, pp. 140–149.

[Ko+21] Y. Ko et al. "Key points estimation and point instance segmentation approach for lane detection". In: *IEEE Trans. Intell. Transp. Syst.* 23 (2021). [CrossRef], pp. 8949–8958. URL: http://dx.doi.org/10.1109/TITS.2021.3088488.

[Lys+21] M. Lyssenko et al. "Instance Segmentation in CARLA: Methodology and Analysis for Pedestrian-oriented Synthetic Data Generation in Crowded Scenes". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal, QC, Canada, 2021, pp. 988–996.

[Mur21] Nirmala Murali. *Image Classification vs Semantic Segmentation vs Instance Segmentation*. Accessed: 20-08-2023. Apr. 2021. URL: https://nirmalamurali.medium.com/image-classification-vs-semantic-segmentation-vs-instance-segmentation-625c33a08d50.

[OSD21] A. Ojha, S.P. Sahu, and D.K. Dewangan. "Vehicle detection through instance segmentation using mask R-CNN for intelligent vehicle system". In: *Proceed-*

ings of the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). Madurai, India, 2021, pp. 954–959.

[Pol+21] P. Polewski et al. "Instance segmentation of fallen trees in aerial color infrared imagery using active multi-contour evolution with fully convolutional network-based intensity priors". In: *ISPRS J. Photogramm. Remote Sens.* 178 (2021), pp. 297–313. URL: http://dx.doi.org/10.1016/j.isprsjprs.2021.06.016.

[Tse+21] K.K. Tseng et al. "A fast instance segmentation with one-stage multi-task deep neural network for autonomous driving". In: *Computers and Electrical Engineering* 93 (2021), p. 107194. URL: http://dx.doi.org/10.1016/j.compeleceng.2021.107194.

[Car+22] O.L.F. de Carvalho et al. "Bounding box-free instance segmentation using semi-supervised iterative learning for vehicle detection". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022), pp. 3403–3420. URL: http://dx.doi.org/10.1109/JSTARS.2022.3169128.

[Den+22] Z. Deng et al. "TrafficCAM: A Versatile Dataset for Traffic Flow Segmentation". In: *arXiv* arXiv:2211.09620 (2022).

[Jia+22] P. Jiang et al. "A Review of Yolo Algorithm Developments". In: *Procedia Computer Science* 199 (2022), pp. 1066–1073.

[JSZ22] Q. Jiang, H. Sun, and X. Zhang. "SemanticBEVFusion: Rethink LiDAR-Camera Fusion in Unified Bird's-Eye View Representation for 3D Object Detection". In: *arXiv* arXiv:2212.04675 (2022).

[Li+22] X. Li et al. "LWSIS: LiDAR-guided Weakly Supervised Instance Segmentation for Autonomous Driving". In: *arXiv* arXiv:2212.03504 (2022).

[Per+22] M.I. Perez et al. "Precision silviculture: Use of UAVs and comparison of deep learning models for the identification and segmentation of tree crowns in pine crops". In: *Int. J. Digit. Earth* 15 (2022), pp. 2223–2238. URL: http://dx.doi.org/10.1080/17538947.2022.2152882.

[RKS22] P. Rotter, M. Klemiato, and P. Skruch. "Automatic Calibration of a LiDAR–Camera System Based on Instance Segmentation". In: *Remote Sensing* 14 (2022), p. 2531. URL: http://dx.doi.org/10.3390/rs14112531.

[Sha22] Deval Shah. *Mean Average Precision (mAP) Explained: Everything You Need to Know*. Accessed: 2023-02-10. 2022. URL: https://www.v7labs.com/blog/mean-average-precision.

[WBL22] C.Y. Wang, A. Bochkovskiy, and H.Y.M. Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors". In: *arXiv* arXiv:2207.02696 (2022).

[Zha+22] T. Zhang et al. "Vehicle detection and tracking for 511 traffic cameras with U-shaped dual attention inception neural networks and spatial-temporal map". In: *Transportation Research Record* 2676 (2022), pp. 613–629.

[Che+23] J. Chen et al. "Safe, Efficient, and Comfortable Autonomous Driving Based on Cooperative Vehicle Infrastructure System". In: *Int. J. Environ. Res. Public Health* 20 (2023), p. 893. URL: http://doi.org/10.3390/ijerph20010893.

[OB23] Goran Oreski and Lucija Babic. "Using a Monocular Camera for 3608728; Dynamic Object Instance Segmentation in Traffic". In: *Engineering Proceedings* 41.1 (2023). ISSN: 2673-4591. DOI: 10.3390/engproc2023041006. URL: https://www.mdpi.com/2673-4591/41/1/6.

[WZY23] M. Wang, L. Zhao, and Y. Yue. "PA3DNet: 3-D Vehicle Detection with Pseudo Shape Segmentation and Adaptive Camera-LiDAR Fusion". In: *IEEE Transactions on Industrial Informatics* (2023), pp. 1–11. URL: http://dx.doi.org/10.1109/TII.2023.3241585.

Popis slika

Popis tablica