SVEUČILIŠTE U SPLITU
FAKULTET ELEKTROTEHNIKE, STROJARSTVA I BRODOGRADNJE

Ivana Marin

MODEL DUBOKOG UČENJA ZA PRAĆENJE PLOVILA S FOKUSOM NA RJEŠAVANJE PROBLEMA DUGOTRAJNIH OKLUZIJA

DOKTORSKA DISERTACIJA

Split, 2025.

S V E U Č I L I Š T E U S P L I T U
FAKULTET ELEKTROTEHNIKE, STROJARSTVA I BRODOGRADNJE

Ivana Marin

Model dubokog učenja za praćenje plovila s fokusom na rješavanje problema dugotrajnih okluzija

DOKTORSKA DISERTACIJA

Split, 2025.

Doktorska disertacija je izrađena na Zavodu za elektroniku i računarstvo, Fakulteta elektrotehnike, strojarstva i brodogradnje

Mentor: prof. dr. sc. Sven Gotovac
Rad br. yyy

PODACI ZA BIBLIOGRAFSKU KARTICU

Ključne riječi: praćenje više objekata, pomorski nadzor, okluzija objekata, detekcija plovila, reidentifikacija, skup podataka za praćenje brodova, asocijacija podataka, Kalmanov filter, ekstrakcija značajki
Znanstveno područje: Tehničke znanosti
Znanstveno polje: Računarstvo
Znanstvena grana: Umjetna inteligencija
Institucija na kojoj je rad izrađen: Sveučilište u Splitu, Fakultet elektrotehnike, strojarstva i brodogradnje
Mentor rada: prof. dr. sc. Sven Gotovac
Broj stranica: 206
Broj slika: 68
Broj tablica: 27
Broj korištenih bibliografskih jedinica: 323

Povjerenstvo za ocjenu doktorske disertacije:

1. Xxxx. dr. sc. Xxxx Yyyyy, Naziv ustanove i grad
2. Xxxx. dr. sc. Xxxx Yyyyy, Naziv ustanove i grad
3. Xxxx. dr. sc. Xxxx Yyyyy, Naziv ustanove i grad

Povjerenstvo za obranu doktorske disertacije:

1. Xxxx. dr. sc. Xxxx Yyyyy, Naziv ustanove i grad
2. Xxxx. dr. sc. Xxxx Yyyyy, Naziv ustanove i grad
3. Xxxx. dr. sc. Xxxx Yyyyy, Naziv ustanove i grad
4. Xxxx. dr. sc. Xxxx Yyyyy, Naziv ustanove i grad
5. Xxxx. dr. sc. Xxxx Yyyyy, Naziv ustanove i grad

Disertacija obranjena dana: xx. mjesec 20xx.

MODEL DUBOKOG UČENJA ZA PRAĆENJE PLOVILA S FOKUSOM NA RJEŠAVANJE PROBLEMA DUGOTRAJNIH OKLUZIJA

Sažetak:

Automatska detekcija i praćenje plovila predstavljaju jedno od važnih istraživačkih područja u razvoju suvremenih sustava za sigurnost i upravljanje pomorskim prometom. Ograničenja postojećih rješenja, zajedno s izraženim utjecajem ljudskog faktora, ukazuju na potrebu za razvojem pouzdanijih i visoko autonomnih pristupa. U tom se kontekstu metode računalnog vida i dubokog učenja, osobito konvolucijske neuronske mreže, nameću kao obećavajuće rješenje za analizu RGB videozapisa. Ipak, praćenje plovila ostaje složen zadatak zbog dinamičnih uvjeta morskog okruženja, velikih varijacija u izgledu plovila te čestih i dugotrajnih okluzija, tijekom kojih detekcije izostaju, a održavanje kontinuiteta putanja i identiteta plovila postaje znatno otežano. Dodatni izazov predstavlja nedostatak javno dostupnih i standardiziranih skupova podataka za praćenje plovila. Stoga je u sklopu disertacije kreiran *Split Ship MOT* (*SSMOT*) skup podataka, koji obuhvaća komplementarne podskupove za detekciju, reidentifikaciju i praćenje plovila, uključujući realistične scenarije mimoilaženja i dugotrajnih okluzija. Na temelju SSMOT-a trenirani su i evaluirani modeli detekcije i reidentifikacije, te je evaluiran predloženi VOc-SORT algoritam praćenja s dvostupanjskom asocijacijom koja integrira geometrijske, vizualne i dinamičke informacije radi smanjenja zamjena identiteta i fragmentacije putanja. Dobiveni eksperimentalni rezultati ukazuju na poboljšane performanse praćenja u pogledu stabilnosti praćenja i očuvanja identiteta u odnosu na postojeće pristupe, osobito u scenarijima dugotrajnih okluzija, uz zadržavanje mogućnosti rada u stvarnom vremenu. Time se ukazuje na potencijal predloženog pristupa za daljnji razvoj autonomnih sustava pomorskog nadzora, kao i za moguća proširenja prema višesenzorskim i robusnijim rješenjima u otežanim uvjetima vidljivosti.

Ključne riječi:
praćenje više objekata, pomorski nadzor, okluzija objekata, detekcija plovila, reidentifikacija, skup podataka za praćenje brodova, asocijacija podataka, Kalmanov filter, ekstrakcija značajki

DEEP LEARNING MODEL FOR VESSEL TRACKING ADDRESSING LONG-TERM OCCLUSION CHALLENGES

Abstract:

Automatic vessel detection and tracking represent one of the important research areas in the development of modern systems for maritime safety and traffic management. The limitations of existing solutions, together with the substantial role of the human factor, indicate the need for the development of more reliable and highly autonomous approaches. In this context, computer vision and deep learning methods, particularly convolutional neural networks, have emerged as promising solutions for the analysis of RGB video data. Nevertheless, vessel tracking remains a complex task due to the dynamic nature of the maritime environment, large variations in vessel appearance, and frequent and long-term occlusions, during which detections are missing and maintaining trajectory continuity and vessel identity becomes significantly more difficult. An additional challenge is the lack of publicly available and standardized datasets for vessel tracking. To address this issue, the *Split Ship MOT* (*SSMOT*) dataset was created as part of this dissertation, comprising complementary subsets for vessel detection, re-identification, and tracking, including realistic scenarios of vessel encounters and long-term occlusions. Based on the SSMOT dataset, detection and re-identification models were trained and evaluated, and the proposed VOc-SORT tracking algorithm with a two-stage association strategy was evaluated, integrating geometric, visual, and dynamic information to reduce identity switches and trajectory fragmentation. The experimental results indicate improved tracking performance in terms of stability and identity continuity compared to existing approaches, particularly in scenarios with long-term occlusions, while maintaining real-time capability. These findings suggest the potential of the proposed approach for further development of autonomous maritime surveillance systems, as well as for extensions toward multi-sensor and more robust solutions under challenging visibility conditions.

Keywords:
multiple object tracking, maritime surveillance, object occlusion, vessel detection, reidentification, ship tracking dataset, data association, Kalman filter, feature extraction

Zahvala

Prvenstveno se od srca želim zahvaliti svom mentoru, prof. dr. sc. Svenu Gotovcu, na povjerenju i podršci tijekom cijelog doktorskog studija. Neizmjerno sam zahvalna na strpljenju i svoj pruženoj pomoći, jasnom usmjeravanju i poticajima u ključnim trenutcima kada mi je to bilo najpotrebnije.

Zahvaljujem se kolegicama i kolegama s Odjela za informatiku na prijateljstvu, podršci i ugodnom druženju, zbog čega je ovih šest godina proteklo u iznimno lijepom i pozitivnom radnom ozračju. Posebno hvala prof. dr. sc. Saši Mladenoviću na ohrabrenjima, kontinuiranoj podršci i savjetima tijekom doktorskog studija, ali i ranije.

Veliko hvala mojoj obitelji, Kisićima i Marinima, koji su bili neizmjerna potpora čitavo ovo vrijeme. Bez vas bi sve ovo bilo znatno teže.

Za kraj, najveće hvala mom suprugu Anti na bezuvjetnoj podršci, razumijevanju i strpljenju tijekom cijelog doktorskog studija, ali i van toga. Petre Ilija i Antea, iako ste još mali i toga niste svjesni, velika zasluga za ovaj rad ide vama. Vi ste bili moja najveća motivacija i snaga.

Zahvalna sam dragom Bogu što je, po zagovoru časnog sluge Božjega oca Ante Antića, ova disertacija privedena kraju.

Sadržaj

Popis tablica

Popis slika

Popis oznaka

AIS – *Automatic Identification System*
AP – *Average Precision*
CLI – *Command-Line Interface*
CNN – *Convolutional Neural Networks*
CPU – *Central Processing Unit*
CVAT – *Computer Vision Annotation Tool*
EKF – *Extended Kalman Filter*
FN – *False Negative*
FP – *False Positive*
FPS – *Frames Per Second*
Frag – *Fragment*
GFLOPS – *Giga FLoating point OPeration*
GPU – *Graphics Processing Unit*
HOG - *Histogram of Oriented Gradients*
HOTA – *Higher Order Tracking Accuracy*
IDFN – *IDentity False Negative*
IDFP – *IDentity False Positive*
IDSW – *IDentity SWitch*
IDTN – *IDentity True Negative*
IDTP – *IDentity True Positive*
IMO – *International Maritime Organisation*
IoU – *Intersection over Union*
JDT – *Joint-Detection and Tracking*
KF – *Kalman Filter*
LSTM – *Long Short-Term Memory*
MARVEL – *MARitime VEsseLs*
mAP – *mean Average Precision*
ML – *Mostly Lost*
MOT – *Multiple Object Tracking*
MOTA – *Multi-Object Tracking Accuracy*
MOTP – *Multi-Object Tracking Precision*
MT – *Mostly Tracked*
NSA – *Noise Scale Adaptive*
OCM – *Observation-Centric Momentum*
OCR – *Observation-Centric Recovery*

ORB – *Sea Situational Awareness*

ORU – *Observation-Centric Re-Update*

OSNet – *Omni-Scale Network*

P – *Precision*

PCA – *Principal Component Analysis*

R – *Recall*

R-CNN – *Region-based Convolutional Neural Network*

RAM – *Random Access Memory*

ReID – *You Only Look Once*

ReID – *Re-identification*

RNN – *ecurrent Neural Network*

SAR – *Synthetic Aperture Radar*

SeaSAw – *Sea Situational Awareness*

SIFT - *Scale-Invariant Feature Transform*

SMD – *Singapore Maritime Dataset*

SORT – *Simple Online and Realtime Tracking*

SOT – *Single Object Tracking*

SPSCD – *The Split Port Ship Classification Dataset*

SSD – *Single Shot Detector*

SSMOT – *Split Ship MOT* skup podataka

SSMOT – *Split Ship Multiple Object Tracking*

SURF - *Speeded-Up Robust Features*

TBD – *Tracking-By-Detection*

TN – *True Negative*

TP – *True Positive*

UKF – *Unscented Kalman Filter*

USV – *Unmanned Surface Vehicle*

VD – *Vizualno Dinamička*

VG – *Vizualno Geometrijska*

VHF – *Very High Frequency*

VRAM – *Video Random Access Memory*

YOLO – *You Only Look Once*

1. UVOD

Automatska detekcija i praćenje plovila važni su elementi sigurnosti i učinkovitosti globalnog pomorskog prometa. S obzirom na kontinuiran rast pomorskih aktivnosti, raste i potreba za naprednim tehnologijama koje omogućuju pravovremeno prepoznavanje i praćenje plovila radi smanjenja rizika od nesreća, učinkovitijeg upravljanja pomorskim prometom, suzbijanja ilegalnih aktivnosti te zaštite teritorijalnih voda i morskog okoliša [1, 2].

Sustavi za upravljanje pomorskim prometom uvelike se oslanjaju na tehnologije poput radara i automatskog identifikacijskog sustava za detekciju, identifikaciju i praćenje plovila [3, 4]. *Radar* (*RAdio Detecting And Ranging*) radi na principu odašiljanja radio valova i analize njihovih refleksija, što omogućuje precizno određivanje udaljenosti, smjera i brzine objekata unutar njegovog dometa [5, 6]. *Automatski identifikacijski sustav* (*AIS*) je autonomni brodski primopredajnik koji koristi VHF (engl. *Very High Frequency*) radio kanale za kontinuiranu razmjenu podataka između plovila i lučkih vlasti [7]. AIS se povezuje sa senzorima za navigaciju te prenosi: 1) dinamičke podatke poput pozicije broda, kursa, brzine i navigacijskog statusa broda, 2) statičke podatke koji su uneseni prilikom ugradnje uređaja poput imena broda, pozivnog znaka, dužine preko svega, širine i IMO (engl. *International Maritime Organisation*) broja, 3) podatke o plovidbi kao što su gaz broda, vrsta tereta, luka odredišta i vrijeme dolaska, 4) poruke vezane za sigurnost [8, 7].

Unatoč širokoj primjeni, sustavi temeljeni na radarima i AIS-u imaju određena ograničenja. Pouzdanost radara može biti smanjena zbog vremenskih uvjeta poput magle, kiše i visokih valova, kao i zbog interferencija uzrokovanih prisutnošću drugih radarskih sustava [9, 10]. Njegov domet osobito je ograničen u priobalnim područjima i lukama s brojnim preprekama te u prometnim pomorskim zonama [3]. Nadalje, radari nisu u mogućnosti detektirati manje objekte s lošim reflektirajućim svojstvima [11]. S druge strane, sva plovila ne moraju nužno imati ugrađen AIS, primjerice brodovi za sport i razonodu te ribarski i ratni brodovi. Također, AIS pojedinih plovila može biti neispravan ili isključen [7]. Pored toga, AIS podaci ranjivi su na manipulacije i namjerno lažiranje [12].

Ljudska pogreška odgovorna je za otprilike 70% pomorskih nesreća [8]. Trenutna rješenja koja se oslanjaju na AIS i radarske sustave prilagođena su poluautomatskoj navigaciji plovila te i dalje zahtijevaju značajnu ljudsku intervenciju [13]. Velika potreba za ručnim nadzorom i analizom povećava rizik od ljudske pogreške, kašnjenja u reakciji i nedosljednosti u interpretaciji podataka. Kako bi se postigla veća razina autonomije i smanjila ovisnost o ljudskom faktoru, sve se više istražuju napredne metode percepcije temeljene na računalnom vidu. U tom kontekstu, duboko učenje igra ključnu ulogu. Za razliku od tradicionalnih metoda strojnog učenja koje zahtijevaju ručno dizajnirane značajke, tehnike dubokog učenja automatski izdvajaju značajke iz sirovih podataka [14]. Konvolucijske neuronske mreže (engl. *Convolutional Neural Networks*, CNNs) [15] jedan su od ključnih pokretača uspjeha metoda dubokog učenja u obradi vizualnih podataka poput slika i videozapisa. Arhitektura konvolucijskih neuronskih mreža inspirirana je receptivnim poljima neurona u vizualnom korteksu životinja [16, 17], a dizajnirana je tako da se značajke uče hijerarhijski, komponirajući jednostavnije značajke u one složenije.

Prema vrsti podataka koji se koriste, algoritmi za automatsku detekciju i praćenje plovila se svrstavaju u tri kategorije: algoritme koji koriste slike (videozapise) snimljene standardnim kamerama, one koji se oslanjaju na podatke radara sa sintetičkom aperturom (engl. *Synthetic Aperture Radar*, SAR), te algoritme koji koriste satelitske snimke daljinskih istraživanja [18, 19]. Zbog velike veličine i duljeg vremena prikupljanja, što rezultira relativno dugim vremenom obrade, SAR i satelitski snimci nisu optimalan izbor za primjene koje zahtijevaju obradu u stvarnom vremenu. Nasuprot tome, široka dostupnost, niski troškovi i dovoljna razina vizualne informativnosti čine standardne RGB kamere obećavajućim rješenjem za ove vrste aplikacija [19, 4, 13]. Integracijom RGB kamera s algoritmima dubokog učenja otvara se put ka razvoju potpuno autonomnih sustava za praćenje u pomorstvu. Ova doktorska disertacija fokusirana je na implementaciju algoritama za praćenje plovila u RGB videozapisima, pri čemu se koriste metode dubokog učenja za pouzdanu detekciju plovila te ekstrakciju relevantnih vizualnih značajki koje doprinose učinkovitijem praćenju u dinamičnim i promjenjivim pomorskim uvjetima.

Iako automatska detekcija i praćenje plovila imaju brojne praktične primjene, ovo je područje istraživanja znatno manje zastupljeno u literaturi u usporedbi s istraživanjima usmjerenima na praćenje pješaka, automobila i sličnih objekata [4, 20]. Praćenje plovila na videozapisima složen je zadatak obilježen brojnim izazovima koji proizlaze iz složenih uvjeta morskog okruženja, tehničkih ograničenja i velike varijabilnosti izgleda plovila. Ono zahtijeva robustan sustav koji je otporan na promjenjive vremenske uvjete i uvjete na moru. Loša vidljivost uzrokovana maglom, kišom, sumrakom ili noćnim uvjetima može značajno smanjiti kvalitetu videozapisa. Odsjaji i refleksije sunčeve svjetlosti na vodi, kao i veliki valovi te drugi plutajući objekti, dodatno otežavaju detekciju plovila. Nadalje, u zadacima praćenja koji uključuju objekte poput pješaka ili automobila, varijacije u veličini, izgledu i kategorijama objekata znatno su manje u usporedbi s varijacijama prisutnima kod plovila [4]. Dodatan izazov predstavlja potreba za izvršavanjem složenih algoritama praćenja u stvarnom vremenu uz ograničene hardverske resurse [21]. Uz navedene specifične izazove, algoritmi praćenja plovila suočavaju se i s općim izazovima detekcije malih objekata poput plovila na velikim udaljenostima od kamere, okluzije i reidentifikacije izgubljenih objekata.

Jedan od najvećih izazova u praćenju objekata je problem ***okluzije***, odnosno djelomične ili potpune zaklonjenosti objekta nekim drugim objektom ili pozadinom. Već djelomična zaklonjenost objekta otežava ekstrakciju kvalitetnih vizualnih značajki za njegovu reidentifikaciju [22], dok za vrijeme potpune zaklonjenosti objekta vizualna informacija u potpunosti nedostaje. S obzirom na duljinu razdoblja zaklonjenosti, okluzije se obično dijele na kratkotrajne i dugotrajne, pri čemu ne postoji jasno i jednoznačno definirana granica između tih dviju kategorija [23, 24, 25]. U slučaju praćenja plovila, dugotrajne okluzije se češće javljaju nego što je to slučaj kod uobičajenog praćenja pješaka, primjerice u situacijama zaklonjenosti manjeg plovila nekim većim plovilom koje se sporije kreće. Kada detekcija u trenutnom okviru videozapisa nedostaje zbog zaklonjenosti, putanju objekta moguće je nadopuniti graničnim okvirom predviđenim algoritmom za predviđanje sljedeće pozicije objekta. U mnogim algoritmima praćenja, za predviđanje sljedeće pozicije koristi se Kalmanov filter [26] koji pretpostavlja linearno kretanje objekata. Navedena pretpostavka je prihvatljiva za vrijeme kratkotrajnih okluzija. Međutim, u slučaju dulje zaklonjenosti objekta, linearna procjena može postati izrazito nepouzdana zbog akumulacije pogreške uzrokovane izostankom novih mjerenja (detekcija) za ažuriranje parametara Kalmanovog filtera [27]. Iako su performanse algoritama u situacijama kratkih zaklonjenosti znatno poboljšane, i dalje imaju ograničenu učinkovitost u slučaju dugotrajnih okluzija, pri čemu se uspješno rješava manje od 10 % okluzija duljih od tri sekunde [23]. To ukazuje na potrebu za daljnjim istraživanjima usmjerenima na poboljšanje performansi algoritama praćenja u takvim scenarijima.

Javno dostupni referentni skupovi podataka (engl. *benchmark datasets*) pružaju standardiziranu osnovu za usporedbu različitih modela i tehnika, potičući transparentnost, reproduktibilnost i napredak u istraživanju [28]. Fokus velikog broja referentnih skupova podataka za praćenje više objekata je na detekciji i praćenju pješaka [29, 30, 31, 32, 33]. Uvriježila se evaluacija novih metoda praćenja upravo na takvim skupovima podataka, pri čemu se ekstenzivno koriste skupovi iz MOT izazova [30, 31, 32, 33][1]. Zbog aktualne popularnosti autonomne vožnje, kreirani su i referentni skupovi podataka koji omogućuju razvoj i evaluaciju algoritama za praćenje pješaka i vozila [34, 35, 36, 37, 38, 39, 40]. Evaluacija na navedenim skupovima podataka često ne uspijeva adekvatno odraziti ograničenja algoritma za praćenje objekata u slučajevima dugotrajnih okluzija, budući da su takvi scenariji statistički rijetki [23]. S druge strane, nedostatak javno dostupnog i opće prihvaćenog skupa podataka za praćenje plovila predstavlja prepreku daljnjem napretku u ovom području istraživanja, ograničavajući mogućnost objektivne usporedbe postojećih metoda i razvoja novih.

Prethodna razmatranja ukazuju na ograničenja u području praćenja plovila, posebice u kontekstu dugotrajnih okluzija i nedostatka odgovarajućih referentnih skupova podataka. Upravo ta ograničenja određuju istraživački okvir ove disertacije, čiji je cilj adresirati navedene izazove i ponuditi doprinos u njihovom rješavanju.

[1] https://motchallenge.net/

1.1. Hipoteza i znanstveni doprinos

Polazeći od identificiranih izazova u području praćenja plovila u ovoj disertaciji formulirana je sljedeća, glavna, istraživačka hipoteza:

H: ***Moguće je razviti i implementirati novi algoritam za praćenje plovila koji će imati poboljšane performanse u odnosu na dosadašnja istraživanja te biti otporniji na dugotrajne okluzije.***

Kako bi se glavna istraživačka hipoteza mogla sustavno i empirijski provjeriti, ona je razložena na pet pomoćnih hipoteza koje se odnose na identificirane izazove u području praćenja plovila te na očekivane učinke mogućih poboljšanja.

H_1: *Postojeće metode praćenja plovila pokazuju smanjenu sposobnost očuvanja identiteta u uvjetima dugotrajnih okluzija.*

H_2: *Unaprjeđenje kvalitete detekcija poboljšava opće performanse praćenja, ali ne rješava u potpunosti problem reidentifikacije plovila nakon okluzija.*

H_3: *Domenska prilagodba modela za reidentifikaciju plovila, provedena na podacima iz pomorskog okruženja, doprinosi pouzdanijem očuvanju identiteta plovila u algoritmima praćenja u odnosu na korištenje modela treniranih isključivo na općim skupovima podataka, osobito u uvjetima dugotrajnih okluzija.*

H_4: *Integracija vizualnih ReID značajki s informacijama o kretanju unutar algoritma praćenja omogućuje pouzdaniju asocijaciju detekcija i putanja nakon dugotrajnih okluzija nego korištenje isključivo prostornih i kinematičkih informacija.*

H_5: *Unaprjeđenje postupka asocijacije detekcija i putanja može dovesti do poboljšanog očuvanja identiteta plovila tijekom dugotrajnih okluzija.*

U kontekstu postavljenih hipoteza i definiranih istraživačkih ciljeva, provedeno istraživanje obuhvatilo je analizu postojećih pristupa, razvoj odgovarajuće metodologije te implementaciju i eksperimentalnu evaluaciju novog algoritamskog rješenja. Kao rezultat tog procesa, ostvarena su sljedeća dva glavna znanstvena doprinosa ove disertacije:

1. ***Izrada referentne baze označenih slika i videozapisa za detekciju, reidentifikaciju i praćenje plovila.***
2. ***Razvoj algoritma za praćenje plovila na videozapisima s poboljšanim performansama u slučaju dugotrajnih okluzija.***

1.2. Organizacija rada

U uvodnom poglavlju istaknuti su motivacija i kontekst istraživanja, razmatraju se ograničenja postojećih sustava za nadzor pomorskog prometa te se obrazlaže potreba za primjenom metoda računalnog vida i dubokog učenja u zadacima automatske detekcije i praćenja plovila. Također, formulirane su istraživačke hipoteze i definiran je znanstveni doprinos disertacije. Zajedno s uvodnim poglavljem, doktorska disertacija je podijeljena na ukupno osam poglavlja.

Drugo poglavlje daje teorijsku osnovu problema praćenja više objekata. U poglavlju su opisani osnovni pojmovi te je predstavljena kategorizacija algoritama za praćenje više objekata. Detaljno su razrađeni ključni koraci MOT algoritama, te su opisani neki od popularnih algoritama praćenja temeljenih na detekciji. Poseban dio poglavlja posvećen je standardnim evaluacijskim metrikama za praćenje više objekata koje se koriste za objektivnu procjenu performansi algoritama. Na kraju poglavlja razmotreni su ključni izazovi praćenja, uključujući problem okluzije, zamjene identiteta i praćenje malih objekata.

Sustavan pregled relevantne znanstvene literature prikazan je u trećem poglavlju. U tom poglavlju razmatraju se dosadašnja istraživanja iz područja automatske detekcije i praćenja plovila, s posebnim naglaskom na problem okluzije praćenih objekata. Najprije se analiziraju postojeći skupovi podataka za detekciju i praćenje plovila, nakon čega slijedi pregled radova usmjerenih na razvoj i primjenu algoritama za detekciju i praćenje plovila. Poglavlje završava kritičkom analizom postojećih pristupa i identifikacijom otvorenih izazova koji motiviraju daljnje istraživanje.

U četvrtom poglavlju opisan je dizajn i implementacija SSMOT (*Split Ship MOT*) skupa podataka. Poglavlje detaljno prikazuje strukturu skupa podataka, koji se sastoji od detekcijskog podskupa, podskupa za reidentifikaciju te skupa videozapisa za praćenje plovila. Analizirane su statističke karakteristike pojedinih podskupova, pri čemu je posebna pažnja posvećena analizi okluzija u podskupu za praćenje.

Peto poglavlje opisuje odabrani YOLO11 model za detekciju plovila. Poglavlje obuhvaća detaljan opis procesa treniranja detektora na SSMOT detekcijskom podskupu, uključujući korištene hiperparametre te tehničke karakteristike računalnog okruženja. Nadalje, provedena je evaluacija dobivenog modela primjenom standardnih metrika, te su analizirani dobiveni rezultati.

U šestom poglavlju opisani su reidentifikacijski modeli korišteni za ekstrakciju vizualnih značajki (ResNet50 i OSNet), te postupak njihove implementacije i treniranja. Poglavlje završava evaluacijom dobivenih ReID modela i analizom postignutih rezultata.

Sedmo poglavlje predstavlja središnji doprinos ove disertacije. U njemu se uvodi predloženi VOc-SORT algoritam praćenja s poboljšanim performansama u slučaju dugotrajnih okluzija. Detaljno je opisan dizajn algoritma, ključne prilagodbe u odnosu na postojeće metode te pojedini koraci njegova rada. Nadalje, provedena je usporedba performansi s postojećim metodama, prikazani su rezultati evaluacije i ablacijske analize, te je dana opsežna rasprava o utjecaju pojedinih komponenti algoritma, ulozi ReID modela i ograničenjima predloženog pristupa.

Osmo, završno poglavlje donosi zaključak rada, u kojem se sažimaju glavni rezultati i ostvareni znanstveni doprinosi disertacije. Također se razmatraju i ograničenja predloženog pristupa te se predlažu smjernice za buduća istraživanja.

2. PROBLEM PRAĆENJA VIŠE OBJEKATA

Zahvaljujući izraženom komercijalnom i akademskom potencijalu, praćenje objekata postalo je jednim od važnijih istraživačkih područja u domeni računalnog vida [41, 42]. Proces praćenja uključuje detekciju i identifikaciju objekata u svakom okviru (engl. *frame*) videozapisa, uz očuvanje konzistentnog identifikatora objekta tijekom cijelog snimka. Iako detekcija objekta iz okvira u okvir videozapisa pruža osnovne informacije o lokaciji i klasi objekta, sama po sebi ne daje kontinuiranu informaciju o kretanju objekta kroz vrijeme koja je nužna za mnoge praktične primjene. Kako bi se steklo dublje razumijevanje dinamike kretanja objekata, nužno je primijeniti algoritme praćenja.

2.1. Kategorizacija algoritama praćenja

S obzirom na broj objekata koji se prati, razlikujemo metode praćenja jednog objekta (engl. *Single Object Tracking, SOT*) i **metode praćenja više objekata** (engl. *Multiple Object Tracking, **MOT***). Metode praćenja jednog objekta, usmjerene su na praćenje jednog konkretnog objekta tijekom cijelog videozapisa. Objekt, definiran u prvom okviru, detektira se i prati kroz sve naredne okvire [41]. S druge strane, metodama za praćenje više objekata cilj je locirati ***sve*** objekte od interesa te pratiti njihov identitet i putanje kroz dani videozapis. U odnosu na praćenje jednog objekta, praćenje više objekata uključuje dva dodatna zadatka: 1) *određivanje broja objekata koji se prate*, a koji se mijenja tijekom vremena, 2) *održavanje konzistentnog identiteta objekata* tokom cijelog videozapisa [43].

Algoritmi praćenja više objekata mogu se podijeliti prema različitim kriterijima. S obzirom na način obrade videozapisa, MOT algoritmi dalje se dijele na [43]:

- ***online*** algoritme, koji prilikom obrade trenutnog okvira koriste isključivo informacije iz prethodnih okvira,
- ***offline*** (*batch*) algoritme, koji uzimaju u obzir informacije iz prošlih, ali i iz budućih okvira videozapisa.

Offline metode generalno daju bolje rezultate zbog dostupnosti globalne informacije iz cijelog videozapisa. Ipak, njihova primjena nije moguća u aplikacijama koje zahtijevaju izvršavanje u stvarnom vremenu jer tada budući okviri nisu dostupni [44, 45]. Stoga se one uglavnom primjenjuju za analizu snimljenih videozapisa [46]. Za aplikacije koje zahtijevaju obradu u stvarnom vremenu, koriste se isključivo online algoritmi praćenja.

MOT algoritmi mogu se kategorizirati i prema načinu inicijalizacije objekata koji se prate. Prema tom kriteriju, razlikuju se algoritmi koji **koriste detekcije za inicijalizaciju** i oni koji ne koriste detekcije, već **zahtijevaju ručnu inicijalizaciju** fiksnog broja objekata u prvom okviru videozapisa [47, 43]. Algoritmi koji koriste detekcije mogu automatski otkriti nove objekte koji ulaze u scenu te prestati pratiti one koji iz nje izlaze. Međutim, njihova učinkovitost uvelike ovisi o kvaliteti detekcija koje se koriste. S druge strane, algoritmi koji zahtijevaju ručnu inicijalizaciju ne ovise o detekciji, ali su manje fleksibilni jer zahtijevaju unaprijed definirani broj objekata i ne mogu se automatski prilagoditi novim objektima koji se u sceni pojave naknadno. Zbog ovih ograničenja nisu prikladni za složenije scenarije s promjenjivim brojem objekata.

Prema načinu na koji pristupaju zadatku praćenja objekata, MOT algoritmi najčešće se dijele u tri osnovne skupine [48]:

- algoritme **temeljene na detekciji** (engl. *Tracking-By-Detection*, *TBD*),
- algoritme **zajedničke detekcije i praćenja** (engl. *Joint-Detection and Tracking*, *JDT*),
- algoritme **temeljene na transformerima**.

Algoritmi temeljeni na detekciji odvajaju detekciju i praćenje u dvije nezavisne komponente, što ih čini intuitivnijima i jednostavnijima za implementaciju [49]. Zahvaljujući naprednim modelima za detekciju, koji se mogu samostalno optimizirati, algoritmi temeljeni na detekciji često postižu visoku razinu točnosti [50] zadržavajući računalnu učinkovitost nužnu za primjene u stvarnom vremenu [51]. Međutim, njihova osjetljivost na pogrešne i nedostajuće detekcije može dovesti do značajne degradacije u performansama algoritma praćenja. S druge strane, algoritmi zajedničke detekcije i praćenja integriraju detekciju i praćenje u jedinstveni model koji istovremeno obavlja obje zadaće, čime se postiže veća koherentnost i poboljšane performanse u složenijim scenarijima, poput onih s nedostajućim detekcijama objekata [52]. Ipak, ovakav pristup ima manju fleksibilnost te zahtijeva složenije modele, koji su računalno zahtjevniji i sporije konvergiraju tijekom treniranja [53].

U posljednjem desetljeću, transformerske neuronske mreže [54], poznate kao transformeri, privukle su značajnu pažnju i unutar područja računalnog vida, te su pokazale potencijal za primjenu u zadacima praćenja više objekata [49]. Izvorno razvijeni za zadatke obrade prirodnog jezika, transformeri su pokazali izvanredne rezultate u prepoznavanju dalekih ovisnosti i složenih odnosa u sekvencijalnim podacima. Takve karakteristike čine ih prikladnima i za MOT zadatke, gdje je nužno razumijevanje prostornog i vremenskog konteksta za precizno praćenje više objekata. Primjenom mehanizama pozornosti (engl. *attention*), transformeri mogu učinkovito procesirati i povezati informacije iz različitih okvira videozapisa, čime se adresiraju izazovi poput okluzija i interakcija među objektima u složenim scenama [55]. Međutim, glavni nedostaci transformera su njihova visoka računalna složenost i nedovoljna prilagođenost specifičnostima računalnog vida [56, 49, 57]. Performanse sustava koji koriste transformere još uvijek zaostaju za najsuvremenijim metodama temeljenim na detekciji, kako u aspektu točnosti tako i u pogledu vremenske učinkovitosti [27]. Usporedba karakteristika MOT algoritama temeljenih na transformerima, u odnosu na TBD i JDT pristupe, prikazana je u Tablici 2.1.

Tablica 2.1: Usporedba karakteristika tri različite kategorije MOT algoritama.

Vrsta	Predstavnici	Prednosti	Nedostaci
TBD	SORT [58] DeepSORT [59] ByteTrack [60] BoT-SORT [61] StrongSORT [62] OC-SORT [27] BoostTrack [63]	modularnost, fleksibilnost, jednostavnost, efikasniji za online praćenje	značajna ovisnost o kvaliteti detekcija
JDT	Tracktor++ [64] CenterTrack [65] FairMOT [66] JDE [67]	bolja integracija informacija, robusniji na gubitak detekcija, višemodularno zajedničko učenje	manja fleksibilnost, složenija implementacija, veći računalni zahtjevi
Transformeri	TransTrack [68] TrackFormer [69] TransCenter [57] TransMOT [70]	značajna paralelizacija, bolje performanse u složenim scenama, bogata globalna i kontekstualna informacija	visoka računska složenost, potreban velik broj podataka i resursa za treniranje, složena implementacija i optimizacija, nisu potpuno prilagođeni području računalnog vida

2.2. Osnovni koraci MOT algoritma

Zahvaljujući značajnom napretku i iznimnim postignućima u području detekcije objekata, metode koje koriste detekcije postale su standardnim pristupom u praćenju više objekata [48]. Stoga je u nastavku fokus isključivo na njima. Unatoč velikoj raznolikosti, većina algoritama za praćenje više objekata na neki način kombinira sljedeće korake, bilo djelomično ili u cijelosti [44, 47]:

(a) **Detekcija objekata**: U danom ulaznom okviru, detektor identificira i lokalizira objekte od interesa koristeći pravokutne granične okvire.

(b) **Predviđanje sljedeće pozicije**: Za svaku aktivnu putanju, buduća pozicija objekta procjenjuje se temeljem analize njegovog trenutnog stanja i prethodnih kretanja.

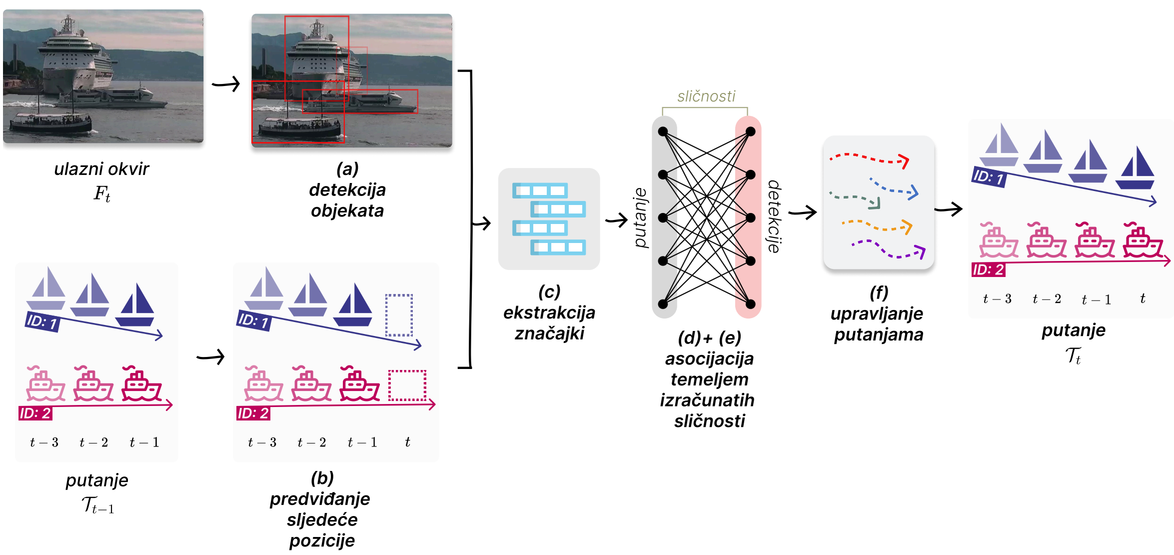
(c) **Ekstrakcija značajki**: Vizualne karakteristike i značajke kretanja izdvajaju se iz detekcija i putanja objekata primjenom jednog ili više različitih ekstraktora.

(d) **Izračun sličnosti**: Dobivene značajke i predviđene pozicije objekta koriste se za izračun sličnosti (udaljenosti) između novih detekcija i postojećih putanja.

(e) **Asocijacija**: Temeljem dobivenih vrijednosti sličnosti (udaljenosti) detekcije se povezuju s putanjama, čime se detekcijama dodjeljuje identifikator odgovarajućeg objekta.

(f) **Upravljanje putanjama**: Ažuriraju se stanja postojećih putanja, incijaliziraju se nove putanje i završavaju neaktivne putanje.

Korak (a) i koraci (b) - (f) mogu se promatrati kao dvije nezavisne komponente algoritama temeljenih na detekciji: komponenta za *detekciju* i komponenta za *praćenje*. Kvaliteta detekcija koje se koriste za praćenje ima značajan utjecaj na performanse algoritma temeljenog na detekciji. Kako bi se omogućila transparentna usporedba različitih komponenti za praćenje, neki MOT izazovi pružaju pristup javnim detekcijama [31, 33]. Time se stavlja fokus na razvoj inovativnih komponenti za praćenje, umjesto na implementaciju moćnih detektora. Slika 2.1 ilustrira uobičajen redoslijed navedenih koraka u algoritmu za praćenje. S druge strane, u algoritmima zajedničke detekcije i praćenja, određeni koraci (b) - (f) se integriraju s detekcijom iz koraka (a). Najčešće je to integracija detekcije i ekstrakcije značajki [71, 67, 66] ili detekcije i predviđanja sljedeće pozicije objekta [64, 72, 65].



Slika 2.1: *Uobičajeni koraci MOT algoritma temeljenog na detekciji.*

2.2.1. Detekcija objekata

Detekcija objekata važan je zadatak računalnog vida koji obuhvaća prepoznavanje, precizno lociranje i klasifikaciju relevantnih objekata unutar digitalne slike ili videozapisa. Svaka detekcija opisana je **trima atributima**: ***klasom*** kojoj pripada prepoznati objekt, ***pravokutnim graničnim okvirom*** koji određuje njegov položaj unutar slike i ***razinom pouzdanosti*** (engl. *confidence score*) detektora u dano predviđanje.

Suvremeni algoritmi za praćenje u velikoj mjeri se oslanjaju na metode dubokog učenja u fazi detekcije, napuštajući tako tradicionalne pristupe koji se temelje na ručno definiranim značajkama. Nasuprot tome, koriste duboke konvolucijske neuronske mreže [15], koje automatski uče složene semantičke reprezentacije iz podataka, čime se postiže veća preciznost i robusnost detekcija [73].

Detektori zasnovani na dubokom učenju dijele se na **dvostupanjske** (engl. *two-stage*) i **jednostupanjske** (engl. *one-stage*) detektore. Dvostupanjski detektori prvo generiraju prijedloge regija unutar slike koje bi mogle sadržavati objekte, a zatim te regije klasificiraju i regresijom prilagođavaju granične okvire objekata. Nasuprot tome, jednostupanjski detektori direktno iz ulaza predviđaju klase i granične okvire, izostavljajući fazu generiranja prijedloga regija, što rezultira bržim izvođenjem [74]. Pregled najpoznatijih predstavnika ovih dviju kategorija detektora dan je u Tablici 2.2.

Izbor odgovarajućeg detektora ključan je za učinkovit rad sustava za praćenje objekata. Različiti detektori primjenjuju se ovisno o zahtjevima točnosti i vremena izvođenja. Nekoliko metoda praćenja temelji se na dvostupanjskom Faster R-CNN detektoru [75, 58, 59, 76], koji obično postiže veću točnost detekcije u usporedbi s jednostupanjskim pristupima. S druge strane, neki radovi [77, 78, 79] koriste jednostupanjski SSD detektor, koji omogućava bržu detekciju uz nešto manju preciznost. Također, CenterNet detektor sve se češće primjenjuje [80, 66, 65] zahvaljujući svojoj efikasnosti i jednostavnosti, što ga čini pogodnim za različite scenarije praćenja. Zahvaljujući optimalnom balansu između točnosti i brzine izvršavanja, YOLO detektor se afirmirao kao najpopularniji detektor za praćenje objekata, što pokazuju brojne recentne studije [81, 62, 60, 61, 82, 83].

2.2.2. Predviđanje sljedeće pozicije

Algoritmi za detekciju objekata koji se koriste u okviru praćenja više objekata nisu bez nedostataka. Često se suočavaju s izazovima poput lažno pozitivnih i nepreciznih detekcija, koje su obično posljedica loših uvjeta okoline poput lošeg osvjetljenja, prisutnosti sjena, djelomične zaklonjenosti objekta ili drugih faktora. Osim toga, moguće je da detektor uopće ne registrira objekt, osobito u situacijama potpune zaklonjenosti, što može rezultirati prekidom putanje tog objekta. Kako bi se prevladali navedeni problemi i poboljšali rezultati praćenja, primjenjuju se različite metode za predviđanje budućeg stanja objekta. Te metode omogućuju nadopunu putanje objekta u slučajevima kada detekcija nedostaje i korekciju putanje u

Tablica 2.2: Pregled popularnih dvostupanjskih i jednostupanjskih detektora.

Vrsta	Predstavnici	Godina	Kratki opis
Two-stage	R-CNN [84]	2014.	Selektivna pretraga generira prijedloge regija iz ulazne slike, iz kojih CNN ekstrahira značajke za klasifikaciju objekata i korekciju predloženih graničnih okvira.
	Fast R-CNN [85]	2015.	CNN prvo izdvaja značajke iz cijele slike, a zatim selektivna pretraga generira prijedloge regija, koje *RoI pooling* svodi na fiksne dimenzije.
	Faster R-CNN [86]	2015.	Selektivnu pretragu zamjenjuje neuronskom mrežom za generiranje prijedloga regija (engl. *Region Proposal Network, RPN*), čime se znatno ubrzava proces detekcije.
	Mask R-CNN [87]	2017.	Proširuje Faster R-CNN dodavanjem segmentacije objekata uz detekciju, koristeći dodatni izlaz za semantičku masku objekta.
One-stage	YOLO [88]	2016.	Dijeli ulaznu sliku na mrežu ćelija i za svaku ćeliju predviđa granične okvire i vjerojatnosti klasa, koristeći samo jedan prolazak kroz neuronsku mrežu.
	SSD [89]	2016.	Također koristi jedan prolazak kroz mrežu, ali koristi više slojeva različitih rezolucija za predviđanje graničnih okvira i klasa, što omogućuje učinkovitije prepoznavanje objekata različitih veličina.
	RetinaNet [90]	2017.	Koristi piramidalnu mrežu značajki (engl. *Feature Pyramid Network*, FPN) za izdvajanje značajki na više razina i primjenjuje fokusni gubitak (engl. *focal loss*) kojim se rješava problem neravnoteže klasa.
	CenterNet [91]	2019.	Pristup detekciji u kojem je svaki objekt predstavljen samo jednom središnjom točkom. Model generira mapu ključnih točaka koja sadrži informacije o lokaciji središta objekata, njihovom pomaku te visini i širini graničnog okvira.
	EfficientDet [92]	2020.	Skalabilni model za detekciju koji kombinira dvosmjernu piramidalnu mrežu značajki (engl. *Bidirectional Feature Pyramid Network*, BiFPN) za učinkovitu fuziju značajki na različitim razinama i skalirajuću arhitekturu temeljenu na EfficentNetu.

slučajevima kada je lokalizacija objekta neprecizna [93].

Osnovna svrha ovog segmenta algoritma za praćenje jest procjena vjerojatne pozicije objekta u narednom okviru videozapisa, oslanjajući se pri tom na njegovo trenutno stanje i prethodne obrasce kretanja. U ovom kontekstu, Kalmanov filter i njegove modifikacije najčešće se koriste za predviđanje buduće pozicije objekta [81, 58, 59, 62, 60, 61, 94, 75, 66, 67]. Alternativno, umjesto Kalmanovog filtera, moguće je koristiti čestični filter (engl. *particle filter)* [95, 96, 97] ili modele temeljene na dubokom učenju, kao što su RNN i LSTM neuronske mreže [98, 99, 100, 93], kao i arhitekture temeljene na transformerima [68, 101]. Međutim, ovi pristupi uvode dodatnu složenost u algoritam praćenja u vidu potrebe za dodatnim treniranjem kompleksnih modela i povećane računske složenosti, u usporedbi s Kalmanovim filterom koji je poznat po svojoj jednostavnosti, računalnoj efikasnosti i skromnim zahtjevima za resursima.

Kalmanov filter

Kalmanov filter [26] rekurzivni je algoritam za procjenu stanja diskretnog dinamičkog sustava temeljem matematičkog modela sustava i niza mjerenja koja u sebi sadrže šum. Termin *"filter"* proizlazi iz činjenice da ovaj algoritam prilikom procjene stanja filtrira šum iz podataka [102]. Tijekom vremena, Kalmanov filter kontinuirano poboljšava procjene stanja, smanjujući utjecaj šuma i nesigurnosti.

Matematički model formalno opisuje kako se stanje sustava mijenja tijekom vremena te je obično definiran s:

$$\mathbf{x}_t = F_t\mathbf{x}_{t-1} + B_t\mathbf{u}_t + \mathbf{w}_t, \tag{2.1}$$

gdje je $F_t \in \mathbb{R}^{n_x \times n_x}$ tranzicijska matrica iz stanja u trenutku $t-1$ do stanja u trenutku t, $\mathbf{u}_t \in \mathbb{R}^{n_u}$ kontrolni, vanjski ulaz koji može neposredno utjecati na stanje sustava, $B_t \in \mathbb{R}^{n_x \times n_u}$ matrica koja opisuje kako kontrolni ulaz utječe na stanje sustava, $\mathbf{w}_t \in \mathbb{R}^{n_x}$ procesni šum koji u obzir uzima nesavršenosti modela sustava. Mjerenje $\mathbf{z}_t \in \mathbb{R}^{n_z}$ u trenutku t opisano je s:

$$\mathbf{z}_t = H_t\mathbf{x}_t + \mathbf{v}_t, \tag{2.2}$$

pri čemu $H_t \in \mathbb{R}_{n_z \times n_x}$ označava matricu koja preslikava stanje sustava u prostor mjerenja, a $\mathbf{v}_t \in \mathbb{R}^{n_z}$ šum mjerenja. Pretpostavka je da su procesni šum i šum mjerenja nezavisne, normalno distribuirane slučajne varijable za koje vrijedi [102, 103]:

$$\mathbf{v}_t \sim N(\mathbf{0}, Q_t), \quad \mathbf{w}_t \sim N(\mathbf{0}, R_t), \tag{2.3}$$

gdje su $Q_t \in \mathbb{R}^{n_x \times n_x}$ i $R_t \in \mathbb{R}^{n_z \times n_z}$ pozitivno definitne matrice kovarijance procesnog šuma i šuma mjerenja, redom.

Napomena. *$n_x, n_z, n_u \in \mathbb{N}$, gdje n_x označava broj stanja u vektoru stanja, n_z broj različitih mjerenja koja pristižu u svakom koraku, a n_u broj elemenata kontrolnog, vanjskog ulaza.*

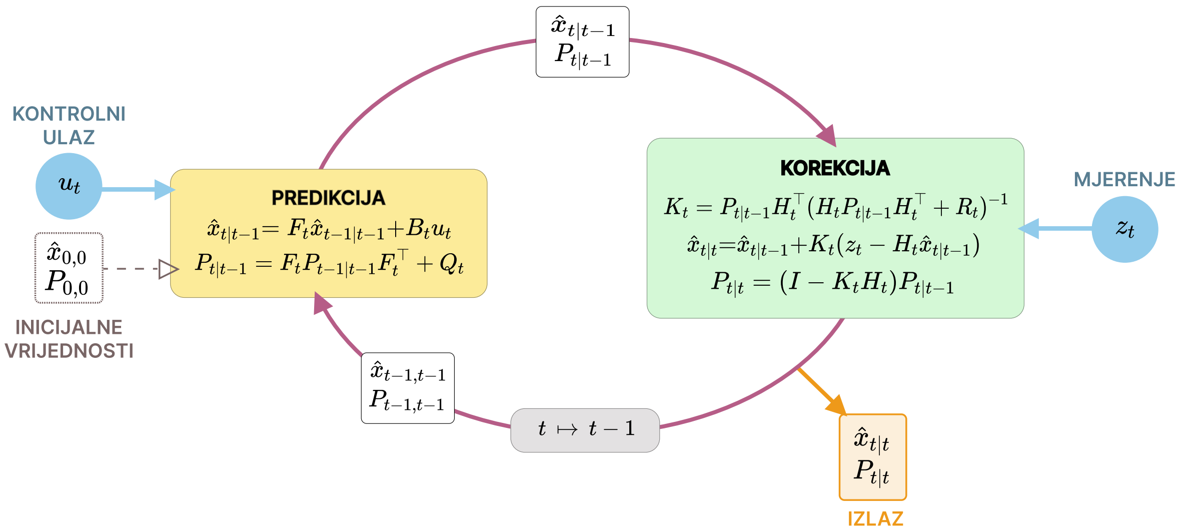
Kalmanov filter u svakoj iteraciji stanje dinamičkog sustava procjenjuje u dva koraka. Prvo se, temeljem prethodnog stanja iz koraka $t-1$ i matematičkog modela sustava, računa ***a priori procjena*** $\hat{\mathbf{x}}_{t|t-1} \in \mathbb{R}^{n_x}$ ***sljedećeg stanja sustava*** i odgovarajuća ***matrica kovarijance a priori pogreške procjene*** $P_{t|t-1} \in \mathbb{R}^{n_x \times n_x}$. Zatim se, u drugom koraku, apriori pretpostavke korigiraju temeljem novih, pristiglih mjerenja što rezultira poboljšanom ***a posteriori procjenom stanja*** $\hat{\mathbf{x}}_{t|t} \in \mathbb{R}^{n_x}$ i ***matricom kovarijance a posteriori pogreške procjene*** $P_{t|t} \in \mathbb{R}^{n_x \times n_x}$.

Jednadžbe Kalmanovog filtera stoga se mogu podijeliti na dvije grupe:

$$\textbf{\textit{predikcija}} \quad \begin{cases} \hat{\mathbf{x}}_{t|t-1} = F_t \hat{\mathbf{x}}_{t-1|t-1} + B_t \mathbf{u}_t \\ P_{t|t-1} = F_t P_{t-1|t-1} F_t^\top + Q_t \end{cases} \tag{2.4}$$

$$\textbf{\textit{korekcija}} \quad \begin{cases} K_t = P_{t|t-1} H_t^\top (H_t P_{t|t-1} H_t^\top + R_t)^{-1} \\ \hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + K_t(\mathbf{z}_t - H_t \hat{\mathbf{x}}_{t|t-1}) \\ P_{t|t} = (I - K_t H_t) P_{t|t-1} \end{cases} \tag{2.5}$$

Kalmanovo pojačanje $K_t \in \mathbb{R}^{n_x \times n_z}$ iz (2.5) je odabrano tako da minimizira kovarijancu a posteriori pogreške, osiguravajući tako optimalnost Kalmanovog filtera uz pretpostavke linearnosti i Gaussove distribucije šuma. Ono određuje utjecaj mjerenja $\mathbf{z}_t$ u odnosu na predviđanje $\hat{\mathbf{x}}_{t|t-1}$ pri izračunu ažuriranog stanja sustava $\hat{\mathbf{x}}_{t|t}$ u korektivnom koraku. Nakon svakog para koraka predikcije i korekcije, postupak se ponavlja uz korištenje dobivenih a posteriori procjena prilikom predviđanja novih a priori procjena. Cijeli proces rada Kalmanovog filtera shematski je prikazan na Slici 2.2.



Slika 2.2: Dijagram Kalmanovog algoritma.

U kontekstu praćenja više objekata, mjerenja u trenutku t odgovaraju detekcijama objekata u okviru t danog videozapisa. Detekcije se obično opisuju vektorom $\mathbf{z} = (x_c, y_c, w, h, c)$, gdje (x_c, y_c) predstavljaju koordinate središta, w i h širinu i visinu graničnog okvira, a c pouzdanost u danu detekciju [58, 27]. Vektor stanja sustava Kalmanovog filtera, koji se koristi u algoritmima praćenja više objekata, obično sadrži informacije o položaju i dimenzijama objekta, kao i o brzini promjene tih vrijednosti. Na primjer, SORT [58] koristi vektor stanja $\mathbf{x} = (x_c, y_c, a, r, \dot{x}_c, \dot{y}_c, \dot{a})$, dok algoritmi poput DeepSORT-a [59], ByteTrack-a [60], StrongSORT [62] i FairMOT-a [66] koriste $\mathbf{x} = (x_c,\ y_c,\ r,\ h,\ \dot{x}_c,\ \dot{y}_c,\ \dot{r},\ \dot{h})$. S druge strane, BoT-SORT [61] vektor stanja definira kao $\mathbf{x} = (x_c,\ y_c,\ h,\ w,\ \dot{x}_c,\ \dot{y}_c,\ \dot{h},\ \dot{w})$. Ovdje $a = w \cdot h$ predstavlja površinu graničnog okvira, a $r = \frac{w}{h}$ omjer širine i visine okvira.

Generalizacije Kalmanovog filtera

U sklopu GIAOTracker [104] algoritma praćenja predstavljena je modifikacija Kalmanovog filtera, **NSA** (*Noise Scale Adaptive*) **Kalman**, koji prilagođava kovarijancu šuma mjerenja u skladu s pouzdanošću detekcija:

$$\widetilde{R}_t = (1 - c_t)R_t, \tag{2.6}$$

gdje je R_t unaprijed postavljena konstantna kovarijanca šuma, a c_t pouzdanost detekcije. Što je pouzdanost c_t veća, to $\widetilde{R}_t$ poprima manje vrijednosti, što implicira da će u korektivnom koraku, tijekom ažuriranja stanja sustava, veća težina biti stavljena na mjerenje, odnosno detekciju [62].

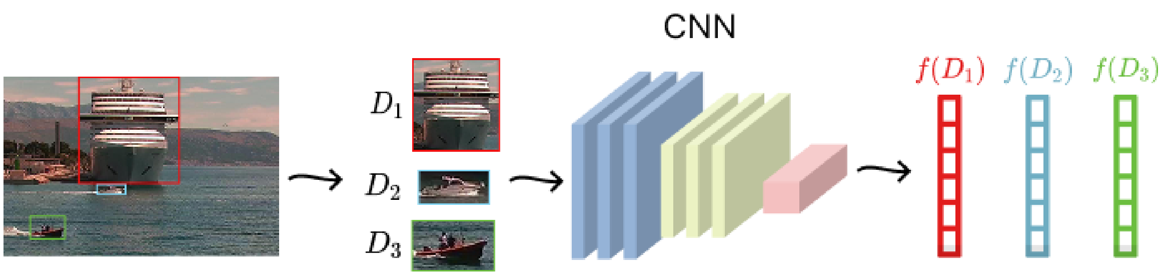
S obzirom na to da Kalmanov filter podrazumijeva linearno kretanje objekta, što ne mora uvijek biti slučaj, razvijene su inačice originalnog algoritma koja su prikladne za procjenu stanja nelinearnih sustava: **prošireni Kalmanov filter** (engl. *Extended Kalman Filter*, EKF) [105] i ***Unscented* varijanta Kalmanovog filtera** (UKF) [106]. EKF linearizira nelinearne modele sustava i mjerenja koristeći Taylorov razvoj prvog reda. Međutim, ovaj pristup može dovesti do loše aproksimacije u sustavima s jako izraženom nelinearnosti budući da se članovi višeg reda zanemaruju [107, 108]. S druge strane, UKF zaobilazi derivacije i koristi transformaciju koja se temelji na preciznom odabiru sigma točaka iz distribucije stanja i njihovoj propagaciji kroz modele sustava i mjerenja, što je čini robusnijom u slučajevima jakih nelinearnosti [109], ali uz puno veće računalne zahtjeve. No, navedeni pristupi pretpostavljaju Gaussovu distribuciju i zahtijevaju unaprijed definirane obrasce kretanja [27]. S druge strane, **čestični filter** (engl. *particle filters*) generalizira Kalmanov filter na probleme procjene stanja nelinearnih sustava čija distribucija ne mora nužno biti Gaussova [110] korištenjem skupa čestica i pridruženih normaliziranih težina za procjenu funkcije gustoće vjerojatnosti stanja sustava, ali su računski vrlo zahtjevni [111]. Zbog toga se navedene metode rijetko koriste u zadacima vizualnog praćenja više objekata, u kojima i dalje prevladava standardni Kalmanov filter [27].

2.2.3. Ekstrakcija značajki

Ekstrakcija značajki ključan je korak MOT algoritma koji omogućuje identifikaciju i razlikovanje objekata, što je neophodno za njihovo praćenje kroz vrijeme. Ovaj korak uključuje izdvajanje relevantnih informacija koje opisuju objekt, a koje se zatim koriste za asocijaciju detekcija iz različitih okvira videozapisa s istim objektom. Stoga je od presudne važnosti kreirati značajke koje su robusne na promjene izgleda istog objekta tijekom vremena, a istovremeno dovoljno diskriminativne da jasno razlikuju različite objekte. Značajke koje se najčešće koriste uključuju one vezane uz *vizualne karakteristike* objekta poput boje, oblika ili teksture, te značajke koje kodiraju *svojstva kretanja* objekta kao što su brzina, pozicija i akceleracija, zajedno s *prostornim značajkama* poput orijentacije, udaljenosti i dimenzije objekta. Iako različite vrste značajki pružaju komplementarne informacije, one se obično promatraju zasebno te se naknadno kombiniraju prilikom izračuna sličnosti [112, 75, 59].

Vizualne značajke

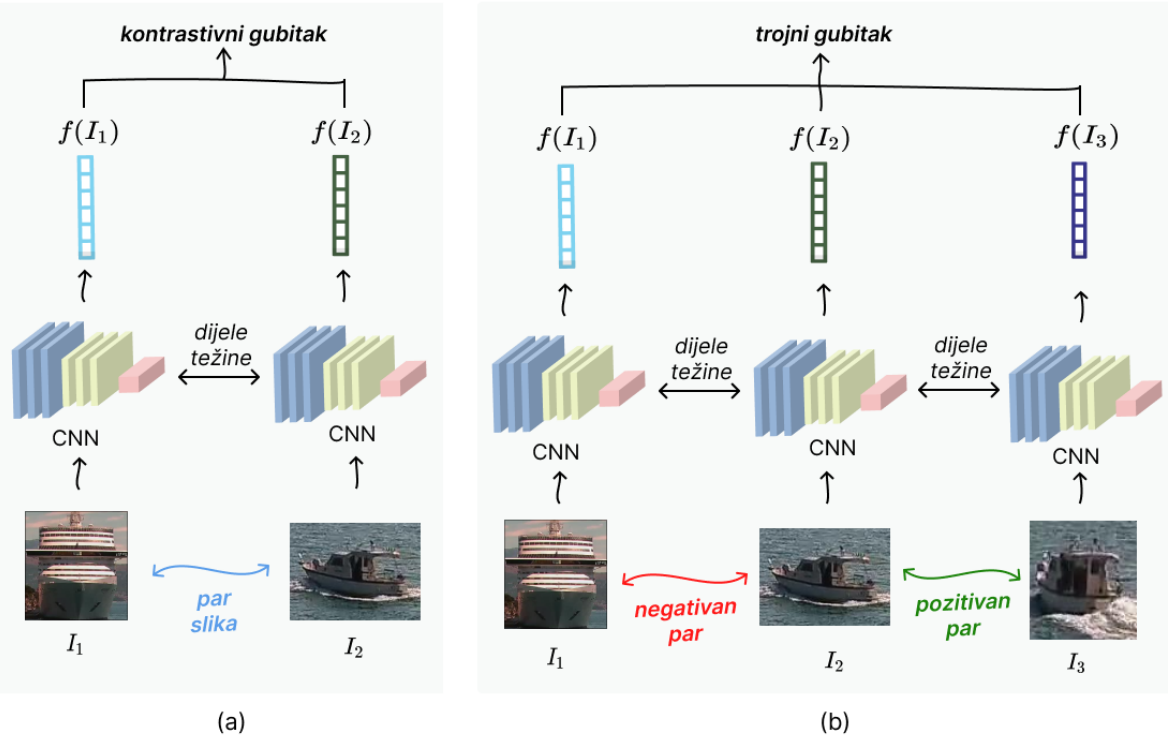
Zbog svoje sposobnosti da automatski izvlače složene značajke iz ulaznih podataka, **konvolucijske neuronske mreže** postale su vodeća metoda za ekstrakciju vizualnih značajki u MOT algoritmima [44]. Slika 2.3 prikazuje proces ekstrakcije vizualnih značajki detekcija jednog okvira videozapisa pomoću konvolucijske neuronske mreže. Detekcija D_i šalje se na ulaz konvolucijske mreže, koja koristi niz konvolucijskih slojeva i slojeva sažimanja za izdvajanje značajki iz slike, postepeno napredujući od jednostavnijih ka složenijima. Na kraju mreže, potpuno povezani sloj [112, 113, 114] ili, alternativno, sloj globalnog prosječnog sažimanja (engl. *Global Average Pooling*) [115, 116] pretvara konvolucijske značajke u realni vektor $f(D_i) \in \mathbb{R}^n$ fiksne duljine, koji numerički reprezentira ključne karakteristike ulazne slike.



Slika 2.3: Ekstrakcija vizualnih značajki detekcija pomoću konvolucijske neuronske mreže.

Jedan od prvih primjera primjene konvolucijskih značajki za praćenje objekata opisan je u [112], gdje autori koriste predtreniranu konvolucijsku neuronsku mrežu kako bi izvukli 4096-dimenzionalni vektor značajki iz svakog graničnog okvira. Dobiveni vektori se zatim reduciraju na 256 dimenzija korištenjem PCA algoritma [117]. Pojedini istraživači za ekstrakciju vizualnih značajki u MOT algoritmima koriste standardne arhitekture konvolucijskih neuronskih mreža, poput GoogLeNet [75, 118, 119, 120], ResNet [100, 121, 122] i VGG [123, 124, 93] arhitektura. Većinom se radi o neuronskim mrežama predtreniranim na velikim skupovima podataka, poput ImageNet [125] skupa podataka za klasifikaciju [121, 123, 124], koje se naknadno prilagođavaju specifičnostima zadatka od interesa, odnosno primjenjuju metodu učenja prijenosom znanja (engl. *transfer learning*).

Drugi pristup, koji se također zasniva na konvolucijskim neuronskim mrežama, je ekstrakcija značajki pomoću sijamskih neuronskih mreža koje se koriste za učenje sličnosti. **Sijamske neuronske mreže** [126] najčešće se sastoje od dvije ili tri identične podmreže koje se zajedno treniraju i kojima se težine za vrijeme treniranja zrcalno ažuriraju. Ilustracija sijamskih neuronskih mreža prikazana je na Slici 2.4.



Slika 2.4: Sijamska neuronska mreža s dvije podmreže (a) i s tri podmreže (b).

Kada se koriste dvije podmreže, sijamska neuronska mreža na ulaz prima par slika i izračunava kontrastivni gubitak (engl. *contrastive loss*) [127]:

$$\mathcal{L}(I_1, I_2) = y \cdot \frac{1}{2} \|f(I_1) - f(I_2)\|_2^2 + (1-y) \cdot \frac{1}{2} \max\{0,\ m - \|f(I_1) - f(I_2)\|_2\}^2, \tag{2.7}$$

gdje je $m > 0$ unaprijed zadana margina koja određuje radijus sličnosti, $y = 1$ za slike istog objekta (slične slike), dok je $y = 0$ za različite. U slučaju korištenja tri podmreže, na ulaz sijamske mreže šalju se tri slike: temeljna slika (engl. *anchor*) (I_2), pozitivni primjer (I_3) sličan temeljnoj slici i negativan primjer (I_1) koji se od nje razlikuje. Tijekom treniranja minimizira se trojni gubitak (engl. *triplet loss*) [128]:

$$\mathcal{L}(I_1, I_2, I_3) = \max\{0,\ \|f(I_2) - f(I_3)\|_2^2 - \|f(I_2) - f(I_1)\|_2^2 + m\} \tag{2.8}$$

U oba slučaja, cilj je isti: smanjiti udaljenost vektora značajki sličnih (pozitivnih) primjera i istovremeno povećati udaljenost vektora značajki različitih (negativnih) primjera.

U kontekstu algoritama za praćenje više objekata, Kim i suradnici [129] koriste konvolucijsku podmrežu sijamske neuronske mreže, treniranu uz kontrastivni gubitak, kako bi iz ulazne slike izdvojili vektor značajki. Varijante sijamskih neuronskih mreža koje također primaju dva ulaza, ali ne koriste klasični kontrastivni gubitak, primijenjene su u radovima [130, 131]. S druge strane, Zhou i suradnici [132] koriste podmrežu sijamske neuronske mreže trenirane s trojnim gubitkom kako bi iz ulazne slike izdvojili 128-dimenzionalni vektor značajki. U radu [133], za ekstrakciju vektora značajki korištena je GoogLeNet podmreža sijamske neuronske mreže uz trojni gubitak. Drugi pristup predlaže korištenje podmreže sijamske mreže s poboljšanom inačicom trojnog gubitka (*SymTriplet*) [134], koja uzima u obzir i udaljenost između vektora značajki negativnog i pozitivnog primjera, odnosno $\|f(I_1) - f(I_3)\|_2$.

Kombinacija različitih vrsta značajki

U cilju poboljšanja robusnosti i preciznosti algoritama za praćenje objekata, istraživači u radovima [75, 94, 135] kombiniraju vizualne značajke dobivene konvolucijskim neuronskim mrežama s informacijama o kretanju i obliku objekta. Yu i suradnici [75] integriraju značajke dobivene GoogLeNet konvolucijskom mrežom s prostornim značajkama koje opisuju kretanje i oblik objekta dobivenih pomoću atributa graničnih okvira predviđenih Kalmanovim filterom i novih detekcija. U [94], vizualne značajke dobivene konvolucijskom neuronskom mrežom koriste se zajedno sa značajkama koje opisuju veličinu, poziciju i dinamiku kretanja objekta. Bae i Yoon [135] također predlažu kombinaciju konvolucijskih značajki sa značajkama modela kretanja i oblika objekta koji se prati. Vizualne značajke dobivene rezidualnom konvolucijskom mrežom se u [59] koriste zajedno s informacijama o kretanju. Nešto drugačiji pristup predstavljen je u [93], gdje se vizualne značajke i značajke kretanja izdvajaju zasebno korištenjem konvolucijske VGG16 mreže i LSTM mreže, a zatim se integriraju u jedinstveni vektor značajki pomoću Metric-Net mreže trenirane s trojnim gubitkom.

2.2.4. Mjere sličnosti

Kako bi se detekcije iz novog okvira videozapisa mogle uspješno povezati s odgovarajućim prethodno praćenim objektima u koraku asocijacije, potrebno je procijeniti njihovu međusobnu sličnost ili, ekvivalentno, udaljenost. Ta se sličnost može temeljiti na jednoj relevantnoj komponenti, poput kretanja [58, 136], ili pak kombinirati više različitih faktora, uključujući vizualni izgled, dinamiku kretanja i oblik objekta [135, 75, 94, 59, 81]. Općenito, mjere sličnosti mogu se podijeliti u tri glavne kategorije: prostorne, vizualne i kombinirane mjere sličnosti [137].

Prostorne i dinamičke mjere sličnosti

Prostorne, odnosno geometrijske, mjere sličnosti kvantificiraju sličnost između dva objekta, koji su najčešće reprezentirani pravokutnim graničnim okvirima, uspoređujući njihov položaj, oblik ili veličinu. One najčešće procjenjuju koliko se blizu jedan drugome nalaze dva objekta u prostoru te u kojoj mjeri im se oblici podudaraju.

Jedna od najčešće korištenih mjera sličnosti je **IoU** (engl. *Intersection over Union*) [58, 68, 60, 136, 129, 59, 61, 81], koja računa omjer površine presjeka dvaju graničnih okvira A i B u odnosu na površinu njihove unije. Matematički, IoU se definira kao:

$$IoU(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}, \tag{2.9}$$

gdje $|\cdot|$ označava površinu. U algoritmima za praćenje više objekata, IoU se obično računa između graničnih okvira detekcija iz novog okvira videozapisa i graničnih okvira predviđenih stanja postojećih putanja [58, 68, 60, 59, 61, 81] ili između graničnih okvira detekcija iz susjednih okvira koje se povezuju u putanje [136, 129]. U radovima [58, 136, 68, 60], IoU se koristi kao jedina mjera sličnosti prilikom povezivanja putanja i detekcija.

Napomena. *Umjesto IoU vrijednosti direktno, u algoritmima praćenja koristi se i IoU udaljenost zadana s:* $d^{IoU}(A,B) = 1 - IoU(A,B)$.

U radu [65], kao mjera sličnosti između objekta A i objekta B koristi se **euklidska udaljenost** njihovih **centara** $C_A = (x_1, y_1)$ i $C_B = (x_2, y_2)$:

$$d_2(C_A, C_B) = \sqrt{(x_1 - x_2)^2 - (y_1 - y_2)^2}. \tag{2.10}$$

Kako bi integrirali informaciju o dinamici kretanju u izračun sličnosti, u radovima [67, 59, 114, 66, 62] koriste **Mahalanobisovu udaljenost** [138] koja uzima u obzir korelaciju promatranih varijabli i njihove varijance. Mahalanobisova udaljenost definirana je s

$$d^{MhD}(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^\top S^{-1} (\mathbf{x} - \mathbf{y})}, \tag{2.11}$$

gdje su $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, a S matrica kovarijance. Na primjer, u [59], računa se kvadrat Mahalanobisove udaljenosti gdje $\mathbf{x}$ predstavlja detekciju iz trenutnog okvira videozapisa, dok su $\mathbf{y}$ i S projekcije stanja i kovarijance putanje, predviđene Kalmanovim filterom, u prostor mjerenja.

Vizualne mjere sličnosti

Vizualne mjere sličnosti koriste se za prepoznavanje istog objekta kroz različite vremenske okvire na temelju njegovih vizualnih karakteristika. Ove mjere se obično oslanjaju na visokodimenzionalne vektore značajki generirane dubokim modelima poput konvolucijskih neuronskih mreža [139].

Za kvantificiranje vizualne sličnosti najčešće se koristi **kosinusna sličnost** vektora značajki [75, 132, 59, 119, 67, 66], definirana s:

$$\cos(\mathbf{f}_D, \mathbf{f}_T) = \frac{\langle \mathbf{f}_D,\, \mathbf{f}_T \rangle}{\|\mathbf{f}_D\|_2\,\|\mathbf{f}_T\|_2}, \tag{2.12}$$

gdje su $\mathbf{f}_D$ i $\mathbf{f}_T$ n-dimenzionalni vektori značajki detekcije D i putanje T.

Napomena. *Umjesto kosinusne sličnosti, obično se računa kosinusna udaljenost* $d^{\cos}(\mathbf{f}_D, \mathbf{f}_T)$ *definirana s* $d^{\cos}(\mathbf{f}_D, \mathbf{f}_T) = 1 - \cos(\mathbf{f}_A, \mathbf{f}_B)$.

Pored kosinusne sličnosti, još se često koristi i **euklidska udaljenost** vektora značajki [129, 134, 133]:

$$d_2(\mathbf{f}_D, \mathbf{f}_T) = \|\mathbf{f}_D - \mathbf{f}_T\|_2 = \sqrt{\sum_{i=1}^{n} (f_{D,i} - f_{T,i})^2}. \tag{2.13}$$

Prednost kosinusne sličnosti u odnosu na euklidsku udaljenost je u njenoj neosjetljivosti na skaliranje podataka i učinkovitosti u visokodimenzionalnim prostorima. Vektori se smatraju sličnima ako imaju istu orijentaciju u prostoru, bez obzira na njihovu veličinu. S druge strane, euklidska udaljenost mjeri stvarnu geometrijsku udaljenost između vektora, uzimajući u obzir razlike kako u njihovoj veličini, tako i u orijentaciji.

Kombinirane mjere sličnosti

Prostorne i vizualne mjere sličnost pružaju različite, ali međusobno komplementarne informacije u procesu praćenja objekata, a njihova kombinacija omogućuje značajno poboljšanje točnosti asocijacije i povećanje robusnosti sustava za praćenje [137, 139].

Najjednostavniji način kreiranja kombinirane mjere sličnosti temelji se na težinskoj sumi. Neka su dane dvije mjere sličnosti s_1 i s_2 zajedno s odgovarajućim težinama $w_1, w_2 \in \mathbb{R}$. Tada se kombinirana mjera sličnosti s jednostavno može definirati kao [137]:

$$s = w_1 s_1 + w_2 s_2. \tag{2.14}$$

Na ovaj način omogućuje se prilagodba doprinosa prostornih i vizualnih mjera ovisno o specifičnostima zadanog sustava.

2.2.5. Asocijacija

U koraku asocijacije cilj je odrediti kojoj od postojećih praćenih putanja pripada detekcija iz trenutnog okvira ili, alternativno, predstavlja li ta detekcija novi objekt koji tek treba početi pratiti. Problem optimalnog pridruživanja detekcija postojećim putanjama može se formulirati kao problem pridruživanja maksimalne težine u potpunom težinskom bipartitnom grafu. Ta se formulacija jednostavno može prilagoditi i problemu pridruživanja minimalne težine, kada se umjesto sličnosti koriste različite mjere udaljenosti kao cijene pridruživanja detekcija postojećim putanjama objekata.

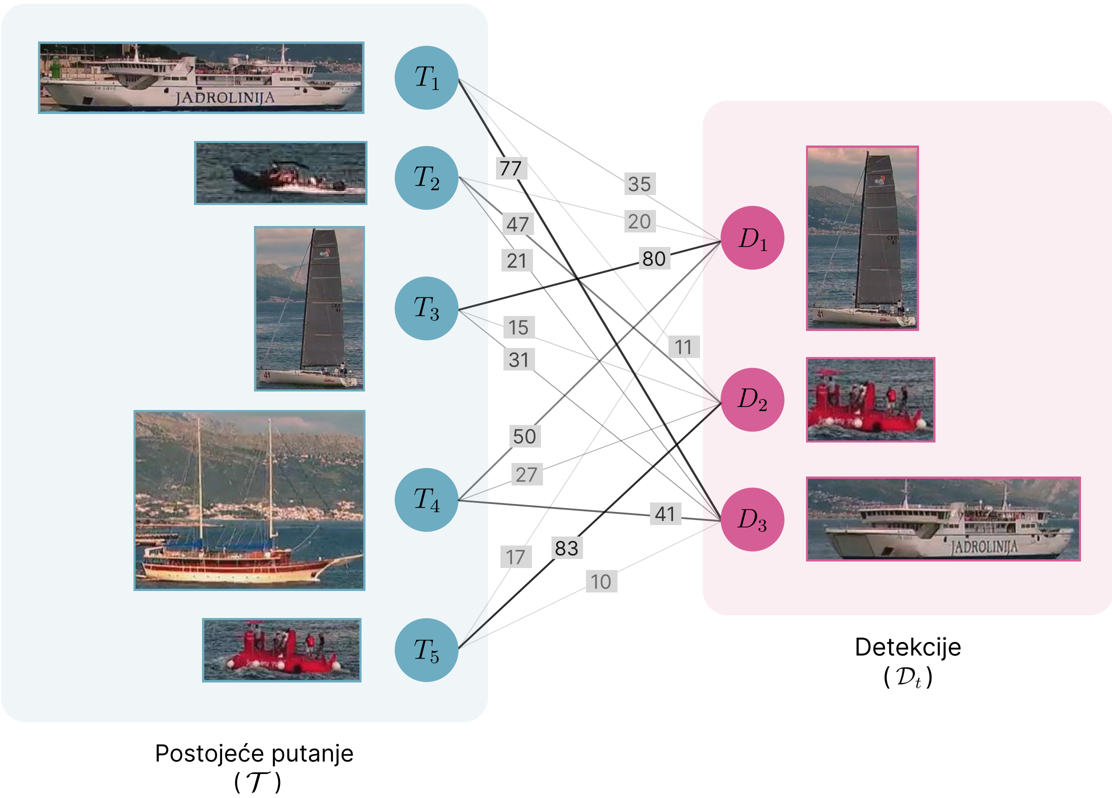
Mađarski algoritam

Neka je $\mathcal{T} = \{T_1, \ldots, T_n\}$ skup postojećih putanja objekata, a $\mathcal{D}_t = \{D_1, \ldots, D_m\}$ skup detekcija iz trenutnog okvira t videozapisa. Nadalje, neka je $G = (V, E)$, gdje je $V = \mathcal{T} \cup \mathcal{D}_t$ skup vrhova, a $E = \mathcal{T} \times \mathcal{D}_t$ skup bridova, potpuni težinski bipartitni graf s funkcijom težine $w : E \to \mathbb{R}_0^+$ koja svakom bridu (T_i, D_j) pridruži cijenu pridruživanja detekcije D_j putanji T_i, odnosno $w(T_i, D_j) = s(T_i, D_j)$, gdje je s odabrana mjera sličnosti. Jedan primjer takvog grafa prikazan je na Slici 2.5.

Pridruživanje M u grafu G je podskup bridova $M \subseteq E$ takav da za svaki vrh $v \in V$ vrijedi da je incidentan najviše jednom bridu iz M. Težina pridruživanja M jednaka je sumi težina svih bridova $e \in M$:

$$w(M) = \sum_{e \in M} w(e). \tag{2.15}$$

Cilj je za dani bipartitni graf G odrediti pridruživanje M maksimalne težine. Navedeno, optimalno pridruživanje može se pronaći pomoću mađarskog algoritma.



Slika 2.5: Ilustracija bipartitnog grafa kakav se koristi prilikom pridruživanja putanja i detekcija primjenom mađarskog algoritma.

Mađarski algoritam [140], također poznat i pod nazivom *Kuhn-Munkersov* algoritam, problem pridruživanja rješava u polinomijalnom vremenu $O(n^3)$ gdje je $n = |\mathcal{T}| = |\mathcal{D}|$ [141]. Iako datira još iz 1955. godine, mađarski algoritam je najčešće korišteni algoritam za asocijaciju detekcija i putanja u algoritmima praćenja [75, 134, 94, 135, 58, 61, 60, 77, 67, 93, 66, 68, 129]. U slučaju da pretpostavka $|\mathcal{T}| = |\mathcal{D}|$ ne vrijedi, u particiju vrhova manje kardinalnosti može se dodati odgovarajući broj fiktivnih čvorova koji se onda povezuju sa svim vrhovima iz druge particije bridovima težine 0, odnosno minimalne vrijednosti sličnosti. Dani bipartitni graf može se reprezentirati $n \times n$ matricom susjedstva $C = [c_{i,j}]$ kojoj retci odgovaraju putanjama, a stupci detekcijama pri čemu je $c_{i,j} = w(T_i, D_j)$. Tada je mađarska metoda dana Algoritmom 1. Ako matrica susjedstva C sadrži udaljenosti, a ne sličnosti, onda se traži pridruživanje minimalne težine i preskače se prvi korak algoritma.

Algoritam 1 Mađarski algoritam

Ulaz: $n \times n$ matrica susjedstva $C = [c_{i,j}]$
Izlaz: optimalno pridruživanje redaka i stupaca matrice

1. *Svođenje problema maksimizacije na problem minimizacije:*

   Vrijednost $c_{i,j}$ svake ćelije zamijeniti razlikom $C_{max} - c_{i,j}$ gdje je C_{max} maksimalna vrijednost dane matrice susjedstva.

2. *Redukcija redaka*:

   Od svake vrijednosti u retku oduzeti minimalnu vrijednost tog retka.

3. *Redukcija stupaca*:

   Od svake vrijednosti u stupcu oduzeti minimalnu vrijednost tog stupca.

4. *Minimalna pokrivenost*:

   Minimalnim brojem vertikalnih i horizontalnih linija precrtati retke i stupce matrice tako da sve nule budu precrtane.

5. **Ako** je broj nacrtanih linija jednak n:
6.   **vrati** $\{(i, j) : c_{i,j} = 0\}$
7. **Inače**:
8.   Pronađi najmanji element matrice koji nije precrtan linijama.
9.   Taj element **oduzmi** od svih elemenata redaka koji **nisu** precrtani.
10.   Taj element **dodaj** svim elementima stupaca koji **su** precrtani.
11.   Vrati se na liniju 4.

Pridruživanje detekcija postojećim putanjama može se provoditi u jednoj fazi [58, 62, 136] ili u više njih [59, 60, 61, 27]. DeepSORT [59] koristi **kaskadno pridruživanje** kod kojeg se detekcije ne povezuju sa svim postojećim putanjama odjednom, već se asocijacija provodi u više ciklusa, uzimajući u obzir starost putanja. U prvom ciklusu, detekcije se pokušavaju pridružiti najmlađim putanjama koje su u prethodnom okviru uspješno ažurirane ($dob = 1$), a zatim se u narednim ciklusima asocijacija proširuje na starije putanje ($dob = 2, \ldots, dob_{max}$) i preostale neuparene detekcije. U svakom ciklusu pridruživanja koristi se ista mjera sličnosti. Druga popularna varijanta asocijacije je ona u dvije faze [60, 61]. ByteTrack [60] i BoT-SORT [61] upotrebljavaju varijantu pridruživanja u dvije faze u kojoj se u prvoj fazi putanjama pridružuju detekcije *visoke pouzdanosti* temeljem vizualne sličnosti i/ili IoU mjere preklapanja graničnih okvira, dok se u drugoj fazi neuparene putanje pokušavaju povezati s detekcijama *niže pouzdanosti* isključivo temeljem IoU sličnosti.

Alternativne metode asocijacije

Umjesto mađarskog algoritma, u [129, 142, 65] koristi se jednostavni pohlepni algoritam, koji u svakom koraku pridružuje parove detekcija i putanja s najvećom vrijednošću izračunate sličnosti. Zbog svoje efikasnosti, pohlepna metoda asocijacije koristi se i u [143] za postizanje online praćenja u stvarnom vremenu.

Azizpour i suradnici [144] predstavljaju algoritam asocijacije koji se temelji na problemu pridruživanja grafa putanja i grafa detekcija, koristeći pritom kvadratno programiranje i graf neuronske mreže (engl. *graph neural networks*). S druge strane, u [145, 98] primjenjuje se LSTM mreža za asocijaciju, dok Yoon i suradnici [146] predlažu neuronsku mrežu koja se sastoji od enkodera s potpuno povezanim slojevima i dvosmjerne LSTM mreže u dekoderu. Navedene metode asocijacije unose znatno računsko opterećenje, što onemogućava izvođenje algoritama u stvarnom vremenu.

2.2.6. Upravljanje putanjama

U koraku upravljanja putanjama provode se sljedeće operacije: 1) ažuriranje stanja postojećih putanja kojima je uspješno pridružena detekcija u koraku asocijacije, 2) inicijalizacija novih putanja za detekcije koje nisu uspješno pridružene postojećim putanjama, 3) završavanje putanja objekata koji su napustili scenu.

1) **Ažuriranje stanja putanja**
   Ovaj proces obuhvaća ažuriranje vrijednosti promatranih varijabli putanja na temelju pridruženih detekcija, odnosno novih mjerenja. U praksi, to može uključivati korektivni korak Kalmanovog filtera, gdje se stanje putanje ažurira kombinacijom predviđenog stanja i inovacije, koja predstavlja razliku između predviđenih i stvarnih mjerenja. Također, može se koristiti prilagodba vizualnih značajki putanja temeljem vizualnih značajki novih, pridruženih detekcija [67].

2) **Inicijalizacija novih putanja**
   Nepridružena detekcija ne rezultira uvijek automatskim stvaranjem nove putanje. Nova putanja obično se ne dodaje odmah u skup postojećih i aktivnih putanja, budući da postoji mogućnost da se radi o lažno pozitivnoj detekciji. Uobičajena praksa je stvaranje

probnih putanja za nepridružene detekcije, koje se inicijaliziraju tek nakon što opstanu određeno testno razdoblje [58, 67]. Na primjer, u [67], nova putanja se stvara samo ako odgovarajuća detekcija preživi dva uzastopna okvira videozapisa. S druge strane, Zhang i suradnici [60] odmah inicijaliziraju putanje, ali isključivo za nepridružene detekcije visoke pouzdanosti.

3) ***Završavanje putanja***
Postojeće putanje obično se ne prekidaju odmah ako im se u jednom okviru nije uspjelo pridružiti detekciju, jer to može biti posljedica privremene zaklonjenosti praćenog objekta ili neuspješne detekcije. Umjesto toga, putanje se najčešće završavaju nakon određenog broja uzastopnih neuspješnih pokušaja pridruživanja detekcija [58, 142, 67, 61, 66, 68]. Na primjer, u [67, 61, 66], putanje se završavaju nakon 30 uzastopnih neuspješnih pridruživanja. Nasuprot tome, u radu [143], putanja se završava ako joj u nizu od N_{miss} uzastopnih okvira nije uspješno pridružena niti jedna detekcija ili ako je broj pridruženih detekcija manji od broja nedostajućih detekcija za tu putanju.

Mahmoudi i suradnici [94] kao izlaz prikazuju samo stabilne putanje koje ispunjavaju sljedeće kriterije: 1) najviše τ_{inv} okvira putanji nije uspješno pridružena detekcija, 2) omjer broja okvira tijekom postojanja putanje u kojima je putanji pridružena detekcija i broja okvira u kojima to nije bio slučaj veći je od τ_{vis}, 3) prosječna cijena pridruživanja detekcije putanji manja je od τ_{cost}. Rad [98] koristi rekurentnu neuronsku mrežu za predviđanje vjerojatnosti ε postojanja putanje u sljedećem okviru na temelju prethodno prikupljenih informacija. Vrijednost ε koristi se za odluku o inicijalizaciji ili završavanju putanje objekta.

2.3. Popularni algoritmi temeljeni na detekciji

Algoritmi praćenja temeljeni na detekciji predstavljaju vodeći pristup zadatku praćenja više objekata [147, 148]. Najpoznatiji predstavnik ove kategorije algoritama je SORT [58], brz i efikasan algoritam praćenja koji je pogodan za izvršavanje aplikacija u stvarnom vremenu. SORT je postavio temelje za razvoj mnogih popularnih algoritama koji dodatno unapređuju osnovni koncept SORT-a prilagođavajući ga različitim izazovima praćenja više objekata.

SORT (*Simple Online and Realtime Tracking*) [58] koristi Kalmanov filter za predviđanje sljedećeg stanja praćenih objekata, nakon čega se nove detekcije pokušavaju pridružiti odgovarajućim putanjama koristeći mađarski algoritam i IoU predviđenih i detektiranih graničnih okvira kao mjeru sličnosti. Kako bi se očuvala efikasnost algoritma, putanje se završavaju čim im u jednom koraku nije pridružena detekcija. Ako se objekt, čija je putanja završena, ponovo pojavi, praćenje se nastavlja s novim identitetom, što rezultira fragmentacijom putanja i gubitkom kontinuiteta praćenja.

Wojke i suradnici [59] predlažu **DeepSORT** algoritam, koji u SORT integrira vizualnu informaciju kodiranu 128-dimenzionalnim vektorom značajki dobivenih dubokom neuronskom mrežom. DeepSORT koristi kaskadno pridruživanje u kojem se detekcije putanjama pridružuju po starosti. Prilikom pridruživanja koristi se metrika koja kombinira informaciju o dinamici kretanju objekta i vizualnu informaciju: $d = \lambda d^{cos} + (1-\lambda)d^{MhD}$, gdje d^{MhD} predstavlja kvadrat Mahalanobisove udaljenosti predviđenog Kalmanovog stanja putanje i detekcije, a d^{cos} minimalnu kosinusnu udaljenost vizualnih značajki nove detekcije i značajki zadnjih 100 detekcija koje su pridružene danoj putanji. Koristeći unaprijed definirane granične vrijednosti θ_{MhD} i θ_{vis} za d^{MhD} i d^{cos}, filtriraju se neprihvatljiva pridruživanja. U eksperimentima iz [59], korištenje $\lambda = 1$ implicira da se informacija o kretanju d^{MhD} koristi ***isključivo za filtriranje neprihvatljivih pridruživanja***. Nakon kaskadnog pridruživanja, nepridružene detekcije i putanje se pokušavaju povezati koristeći IoU udaljenost. Za razliku od SORT-a, putanje se ne završavaju odmah, već nakon što im tijekom 30 uzastopnih okvira nije pridružena detekcija.

U [60], autori predlažu novu metodu asocijacije koja u obzir uzima gotovo sve detektirane granične okvire, čak i one s malom pouzdanošću, i implementiraju je u **ByteTrack** algoritam. Granični okviri s malom pouzdanošću mogu indicirati postojanje objekata koji su djelomično zaklonjeni, pa njihovo filtriranje može dovesti do fragmentacije putanja praćenih objekata. Predložena metoda asocijacije, koja se temelji na mađarskom algoritmu, odvija se u dvije faze. U prvoj fazi putanjama se pokušavaju pridružiti detekcije visoke pouzdanosti koristeći ili IoU ili udaljenost vektora vizualnih značajki detekcija i predviđanja Kalmanovog filtera. U drugoj fazi se neuparenim putanjama pridružuju detekcije koje imaju nižu pouzdanost koristeći isključivo IoU kao mjeru sličnosti, budući da takvi granični okviri obično sadrže zaklonjene objekte.

BoT-SORT algoritam [61] integrira sljedeće modifikacije u ByteTrack: (1) dodatno se upotrebljava kompenzacija pokreta kamere, (2) koristi se poboljšana verzija Kalmanovog filtera koja u vektoru stanja koristi direktno visinu i širinu graničnog okvira, umjesto visine i omjera širine i visine graničnog okvira, (3) u prvoj fazi asocijacije, u kojoj se putanjama pridružuju detekcije visoke pouzdanosti, koristi se mjera udaljenosti koja povezuje vizualne značajke i značajke kretanja. Cijena $C(T_i, D_j)$ pridruživanja detekcije D_j putanji T_i u prvoj fazi asocijacije dana je s $C(T_i, D_j) = \min\{d_{i,j}^{IoU}, \widehat{d}_{i,j}^{cos}\}$, gdje je $d_{i,j}^{IoU}$ IoU udaljenost predviđenog graničnog okvira za putanju T_i i detektiranog graničnog okvira D_j, a $\widehat{d}_{i,j}^{cos}$ predložena mjera vizualne udaljenosti definirana s:

$$\widehat{d}_{i,j}^{cos} = \begin{cases} 0.5 \cdot d_{i,j}^{cos}, & (d_{i,j}^{cos} < \theta_{vis}) \land (d_{i,j}^{IoU} < \theta_{IoU}) \\ 1, & \text{inače} \end{cases}, \tag{2.16}$$

pri čemu je $d_{i,j}^{cos}$ kosinusna udaljenost vektora značajki detekcije D_j i eksponencijalnog pomičnog prosjeka vektora značajki detekcija pridruženih putanji T_i, a $\theta_{IoU} = 0.5$ i $\theta_{vis} = 0.25$ granične vrijednosti kojima se odbacuju slabo vjerojatna pridruživanja.

StrongSORT [62] unaprjeđuje DeepSORT na različite načine: (1) koristi ECC [149] model za kompenzaciju pokreta kamere, (2) implementira NSA Kalmanov filter [104] umjesto običnog Kalmana, (3) za ekstrakciju vizualnih značajki koristi BoT ekstraktor [150] s ResNeSt50 [151] okosnicom, (4) za opis vizualnog izgleda putanje koristi eksponencijalni pomični prosjek vizualnih značajki pridruženih detekcija, (5) kao mjeru udaljenosti koristi linearnu kombinaciju $d = \lambda d^{cos} + (1-\lambda)d^{MhD}$, pri čemu je $\lambda = 0.98$, te se vrijednost d^{MhD} ne koristi samo za filtriranje neprihvatljivih pridruživanja već se i ***direktno integrira u konačnu udaljenost***, (6) za asocijaciju koristi jednostavno pridruživanje, umjesto kaskadnog pridruživanja. Dodatno, predložena su dva jednostavna i učinkovita algoritma za post-procesiranje: *AFLink* metoda za globalnu asocijaciju putanja koja koristi isključivo prostorno-vremenske informacije i *GSI* algoritam za interpolaciju putanja baziran na Gaussovoj regresiji procesa, koji se koristi za ublažavanje nepravilnosti nastalih zbog nedostajućih detekcija. *StrongSORT++* predstavlja nadogradnju StrongSORT algoritma u kojoj su implementirani navedeni post-procesni koraci.

OC-SORT (*Observation-Centric SORT*) [27] generalizirana je verzija klasičnog SORT algoritma koja povećava robusnost praćenja u situacijama okluzije objekata i njihovog nelinearnog kretanja. Ova metoda uvodi tri ključna modula koja nadograđuju standardni Kalmanov filter i metodu asocijacije putanja i detekcija. Prvi od njih je ***modul dodatne korekcije temeljene na opažanjima*** (engl. *Observation-Centric Re-Update,* ***ORU***) koji koristi virtualnu putanju $\tilde{\mathbf{z}}_t = \mathbf{z}_{t_1} + \frac{t-t_1}{t_2-t_1}(\mathbf{z}_{t_2} - \mathbf{z}_{t_1})$, $t_1 < t < t_2$, dobivenu linearnom interpolacijom posljednje poznate detekcije $\mathbf{z}_{t_1}$ izgubljenog objekta i nove detekcije $\mathbf{z}_{t_2}$ s kojom je objekt povezan nakon okluzije, kako bi se retroaktivno izvršili korektivni koraci Kalmanovog filtera u vremenskim trenutcima u kojima su stvarna mjerenja izostala, čime se smanjuje akumulirana pogreška predviđanja i poboljšava preciznost praćenja nakon završetka okluzije. Druga nadogradnja odnosi se na ***moment temeljen na opažanjima*** (engl. *Observation-Centric Momentum,* ***OCM***): dodatni kriterij zasnovan na konzistentnosti smjera kretanja $\Delta\theta$ koji se uvodi u cijenu pridruživanja, $d = -IoU + \Delta\theta$, što je naročito korisno kod nelinearnog kretanja objekata. Posljednje poboljšanje obuhvaća ***pokušaj ponovne asocijacije neuparenih putanja temeljem njihovih posljednjih opažanja*** (engl. *Observation-Centric Recovery,* ***OCR***), što pomaže u oporavku putanja objekata koji su kratkotrajno zaklonjeni ili izgubljeni, pri tom se kao cijena pridruživanja koristi samo $-IoU$.

Maggiolino i sur. [152] uvode adaptivnu integraciju vizualnih značajki u OC-SORT algoritam, koji se temelji isključivo na informacijama o kretanju objekata. Predloženi **Deep OC-SORT** algoritam za ažuriranje vizualnih značajki putanje koristi eksponencijalni pomični prosjek s dinamičkim faktorom težine α_t koji ovisi o pouzdanosti detekcija. Nadalje, kao mjera sličnosti u fazi asocijacije koristi se kombinacija IoU i kosinusne udaljenosti vizualnih vektora značajki, pri čemu se težina vizualnog dijela prilagođava ovisno o diskriminativnosti izgleda objekta. Naime, ako je neka putanja ili detekcija jasno povezana samo s jednim kandidatom, tj. velika je razlika između najbolje i druge najbolje vrijednosti sličnosti, onda se težina vizualnog dijela za taj par povećava.

BoostTrack [63] algoritam koristi metodu asocijacije u jednoj fazi u kojoj se sve detekcije (uključujući i one s niskom pouzdanošću) koriste zajedno u jednom koraku. Mjera sličnosti koja se pri tom koristi dobije se zbrajanjem osnovne IoU sličnosti s_{base} i ponderiranih dodataka koji se koriste za *"boostanje"* sličnosti, a oni uključuju: 1) IoU udaljenost skaliranu umnoškom pouzdanosti detekcije i pouzdanosti putanje, pri čemu pouzdanost putanje ovisi o njenoj starosti i vremenu od posljednjeg ažuriranja, kojom se implicitno favoriziraju parovi visoke pouzdanosti bez uvođenja dodatne faze asocijacije, 2) softmax-normalizirane vrijednosti Mahalanobisove udaljenosti s^{Mhd}, 3) sličnosti oblika, visine i širine, graničnih okvira s^{shape}. Dodatkom vizualnih zančajki u *boostanu* sličnost osnovne verziju BoostTrack algoritma performanse algoritma dodatno se povećaju performanse i stabilnost praćenja. Konačna mjera sličnosti koja se koristi u asocijaciji tada je dana s $s_{boost} = s_{base} + \lambda_{IoU} \cdot c_D \cdot c_T \cdot IoU + \lambda_{MhD} \cdot s^{MhD} + \lambda_{shape} \cdot s^{shape} + \lambda_{app} \cdot s^{app}$, gdje su c_D i c_T pouzdanosti detekcije i putanje redom, a $\lambda_{IoU}, \lambda_{MhD}, \lambda_{shape}, \lambda_{app}$ hiperparametri koji se koriste za ponderiranje pojedinih komponenti.

2.4. Evaluacija MOT algoritama

Za efikasnu i objektivnu evaluaciju MOT algoritama potrebne su nam kvantitativne metrike koje mjere sposobnost algoritama da u svakom okviru videozapisa pronađu i precizno lokaliziraju sve objekte od interesa te konzistentno prate njihove putanje i jedinstvene identifikatore kroz vrijeme [153].

2.4.1. Metrike

Neka je $O = \{o_1, \ldots, o_n\}$ skup stvarnih putanja objekata, a $H = \{h_1, \ldots, h_m\}$ skup hipoteza, odnosno predviđenih putanja MOT algoritma. Stvarne putanje objekata i hipoteze reprezentirane su skupom detekcija O_{det} i H_{det} u svakom okviru videozapisa. Svakoj detekciji pridružen je jedinstveni identifikator za pojedini okvir, koji je konzistentan kroz vrijeme za sve detekcije iz iste putanje. Detekciju objekta o_i ili hipoteze h_j u okviru t označavamo s $o_i^{(t)}$ odnosno $h_j^{(t)}$, pri čemu indeksi i i j predstavljaju odgovarajuće identifikatore detekcije.

Napomena. *Pri praćenju objekata različitih klasa, svakoj putanji se dodatno pridjeljuje i oznaka klase. Kako se u višeklasnom scenariju kod većine evaluacijskih metrika konačna vrijednost dobivena izračunom srednje vrijednosti po pojedinim klasama [154], u nastavku se zbog jednostavnosti pretpostavlja da se prate objekti samo jedne klase.*

Pogreške algoritama za praćenje mogu se klasificirati u tri kategorije: 1) ***pogreške lokalizacije*** koje nastaju kada predviđene detekcije nedovoljno precizno određuju položaje stvarnih objekata, 2) ***pogreške detekcije*** koje se javljaju kada algoritam predviđa detekcije koje u stvarnosti ne postoje ili propušta detektirati stvarne objekte, 3) ***pogreške asocijacije*** koje su rezultat pogrešnog povezivanja detekcija između okvira videozapisa, bilo da algoritam isti identifikator dodjeljuje različitim stvarnim putanjama ili više različitih identifikatora jednoj stvarnoj putanji.

MOTA, MOTP

Bernardin i suradnici u [153] predstavljaju dvije **CLEAR MOT** metrike: **MOTA** (engl. *Multi-Object Tracking Accuracy*) metriku koja mjeri koliko dobro algoritam detektira objekte i predviđa putanje i **MOTP** (engl. *Multi-Object Tracking Precision*) metriku koja mjeri preciznost lokalizacije praćenih objekata. Unatoč svojim nedostacima, MOTA se ubrzo afirmirala kao primarna metrika za evaluaciju MOT algoritama [154, 43]. Kako bi se mogle izračunati vrijednosti MOTA i MOTP metrika, u svakom okviru videozapisa potrebno je pridružiti stvarne detekcije objekata detekcijama hipoteza. Navedeno se radi na sljedeći način:

(1) Ako je sličnost s između detekcija $o_i^{(t)}$ i $h_j^{(t)}$ manja od unaprijed definirane granične vrijednosti α, onda pridruživanje detekcija $(o_i^{(t)}, h_j^{(t)})$ ***nije valjano***.[1]

(2) Sva pridruživanja detekcija $(o_i^{(t)}, h_j^{(t)})$ iz okvira t koja su valjana u okviru $t+1$, odnosno za koja vrijedi $s\left(o_i^{(t+1)}, h_j^{(t+1)}\right) \geq \alpha$, ostaju očuvana i u okviru $t+1$.

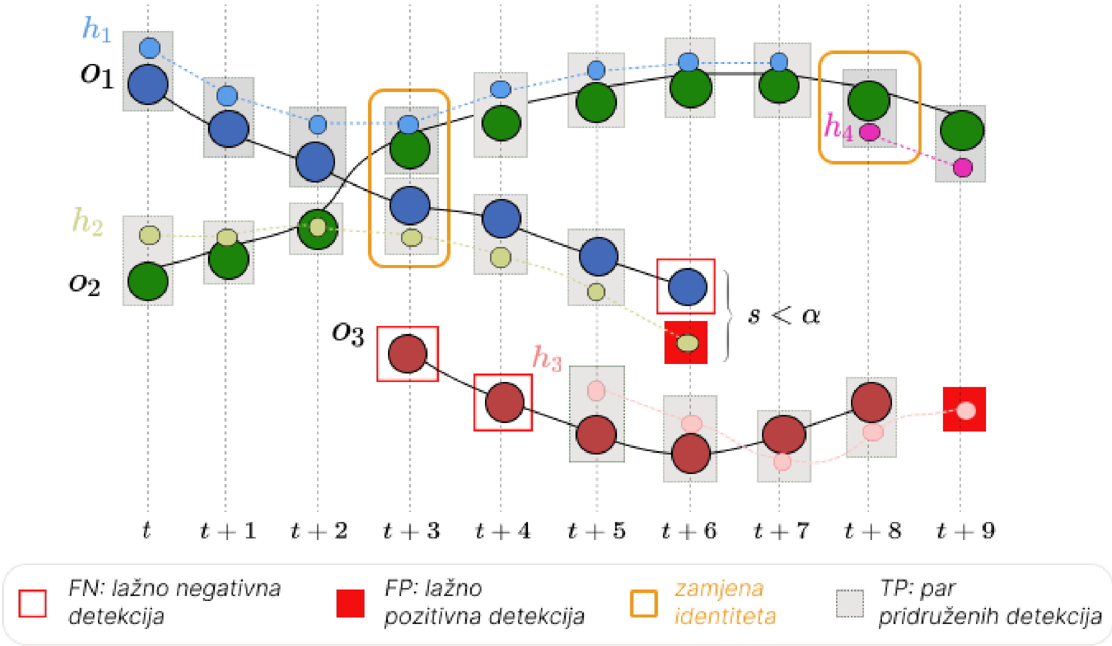
(3) Detekcije koje su ostale neuparene nakon (2), pokušavaju se pridružiti jedne drugima na način da se maksimizira njihova ukupna sličnost.[2] Navedeno se može napraviti pomoću *mađarskog algoritma*.

Neka je $TP \subseteq O_{det} \times H_{det}$ skup svih pridruženih parova detekcija stvarnih objekata i hipoteza, $FP \subseteq H_{det}$ skup **lažno pozitivnih** (preostalih, neuparenih) detekcija hipoteza te $FN \subseteq O_{det}$ skup **lažno negativnih** (preostalih, neuparenih) detekcija stvarnih objekata. Nadalje, neka $IDSW$ označava broj **zamjena identiteta** (engl. *IDentity SWitch*) tj. koliko je puta detekciji stvarnog objekta pridružena detekcija hipoteze čiji identifikator nije konzistentan identifikatoru hipoteze koja je tom istom objektu pridružena u prethodnim okvirima. Tada je

$$MOTA = 1 - \frac{|FN| + |FP| + IDSW}{|O_{det}|}, \quad MOTP = \frac{1}{|TP|} \sum_{(o,h) \in TP} s(o,h). \tag{2.17}$$

Prethodno definirani skupovi korišteni u izračunu CLEAR MOT metrika vizualizirani su na Slici 2.6.

[1] U slučaju 2D praćenja kao mjera sličnosti s najčešće se koristi *IoU* odgovarajućih graničnih okvira [44].
[2] Sva pridruživanja moraju biti valjana.



Slika 2.6: Vizualizacija CLEAR MOT koncepata. Na slici je prikazano deset uzastopnih okvira praćenja. o_1, o_2 i o_3 su stvarne putanje objekata, a $h_1,\dots,h_4$ hipoteze algoritma.

HOTA

Budući da MOTA ne mjeri pogrešku lokalizacije te prenaglašava važnost detekcije nauštrb asocijacije, u [154] je predložena nova **HOTA** (engl. *Higher Order Tracking Accuracy*) metrika koja na uravnotežen način kombinira sve aspekte evaluacije algoritama za praćenje.

U definicijama skupova (2.18), (2.19) i (2.20) koje slijede, zbog preglednosti se izostavlja $t' \in \{1,\dots,N\}$ gdje je N broj okvira danog videozapisa. Nadalje, ako je z skup, $\{x \in z \mid \psi\}$ označava skup $\{x \mid x \in z \wedge \psi\}$.

Za dani par pridruženih detekcija $(o_i^{(t)}, h_j^{(t)}) \in TP$ skupovi točnih (TPA), lažno negativnih (FNA) i lažno pozitivnih (FPA) asocijacija[3] definiraju se na sljedeći način:

$$TPA\left((o_i^{(t)}, h_j^{(t)})\right) = \left\{\left(o_i^{(t')}, h_j^{(t')}\right) \in TP\right\}, \tag{2.18}$$

$$FNA\left((o_i^{(t)}, h_j^{(t)})\right) = \left\{\left(o_i^{(t')}, h_k^{(t')}\right) \in TP \mid k \neq j\right\} \cup \left\{o_i^{(t')} \in FN\right\}, \tag{2.19}$$

$$FPA\left((o_i^{(t)}, h_j^{(t)})\right) = \left\{\left(o_k^{(t')}, h_j^{(t')}\right) \in TP \mid k \neq i\right\} \cup \left\{h_j^{(t')} \in FP\right\}. \tag{2.20}$$

Definirani skupovi vizualno su pojašnjeni na Slici 2.7. Tada je $HOTA_\alpha$ metrika za zadanu

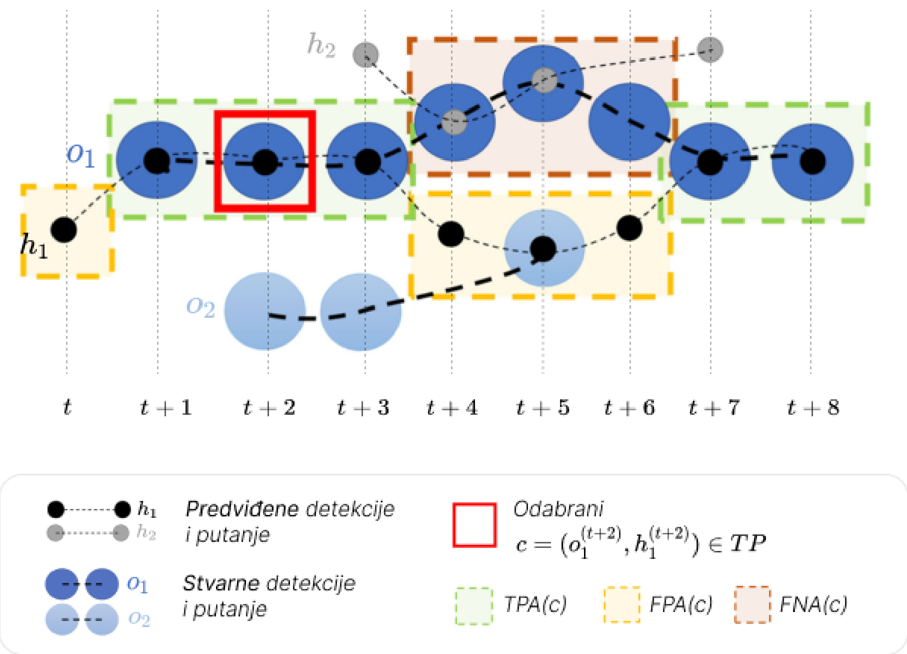
[3] TPA (engl. *True Positive Associations*), FNA (engl. *False Negative Associations*), FPA (engl. *False Positive Associations*).

graničnu (lokalizacijsku) vrijednost α dana s

$$HOTA_\alpha = \sqrt{\frac{\sum_{c \in TP} \mathcal{A}(c)}{|TP|+|FN|+|FP|}}, \quad \mathcal{A}(c) = \frac{|TPA(c)|}{|TPA(c)|+|FNA(c)|+|FPA(c)|}, \tag{2.21}$$

pri čemu TP, FN i FP mjere uspjeh, odnosno pogrešku detekcije, a TPA, FNA i FPA asocijacije. Kako bi se dodatno uzeo i aspekt lokalizacije, HOTA se definira kao vrijednost integrala po valjanim graničnim vrijednostima α, a u praksi se aproksimira aritmetičkom sredinom vrijednosti $HOTA_\alpha$ metrike za $\alpha \in \{0.05, 0.1, \ldots, 0.95\}$:

$$HOTA = \int_0^1 HOTA_\alpha \, d\alpha \approx \frac{1}{19} \sum_{\substack{\alpha=0.05 \\ \alpha+=0.05}}^{0.95} HOTA_\alpha. \tag{2.22}$$



Slika 2.7: HOTA: Ilustracija TPA, FPA i FNA skupova za odabrano ispravno pridruživanje $c \in TP$ predviđene detekcije hipoteze h_1 detekciji stvarne putanje o_1 iz trenutka $t+2$. Skup ispravnih pridruživanja $TPA(c)$ predviđenih detekcija hipoteze h_1 (manji crni kružići) detekcijama odgovarajuće stvarne putanje (veći tamno plavi krugovi) označen je zelenom bojom. Skup $FPA(c)$, označen žutom bojom, sadrži predviđene detekcije hipoteze h_1 koje nisu pridružene niti jednoj stvarnoj putanji ili su pridružene pogrešnoj putanji. Smeđom bojom označen skup $FNA(c)$ obuhvaća detekcije stvarne putanje o_1 kojima je pridružena detekcija pogrešne hipoteze ili im uopće nije pridružena predviđena detekcija. (Slika preuzeta iz [154], uz izmjene.)

IDF1

Za evaluaciju algoritama za praćenje uglavnom se istovremeno koristi više različitih metrika. Uz MOTA i HOTA metrike, često se koristi i $IDF1$ [155] metrika koja se fokusira na točnost identifikacije objekata tokom praćenja. Dok MOTA i HOTA rade pridruživanja na razini detekcija, $IDF1$ to čini na razini putanja. Definiraju se novi skupovi: $IDTP$ (engl. *Identity True Positives*) kao skup parova pridruženih detekcija $(o_i^{(t)}, h_j^{(t)})$ na preklapajućim dijelovima putanja koje su pridružene, $IDFN$ (engl. *Identity False Negatives*) i $IDFP$ (engl. *Identity True Positives*) kao skupovi preostalih stvarnih detekcije iz O_{det} te preostalih predviđenih detekcija iz H_{det} koje se nalaze na putanjama koje nisu uspješno pridružene ili na nepreklapajućim dijelovima pridruženih putanja. Tada je,

$$\textit{ID-Recall} = \frac{|IDTP|}{|IDTP| + |IDFN|}, \quad \textit{ID-Precision} = \frac{|IDTP|}{|IDTP| + |IDFP|}, \tag{2.23}$$

$$IDF1 = \frac{|IDTP|}{|IDTP| + 0.5|IDFN| + 0.5|IDFP|}. \tag{2.24}$$

ID-Recall je udio stvarnih detekcija koje su ispravno identificirane, a *ID-Precision* udio detekcija hipoteza koje su ispravno identificirane. $IDF1$ kombinira *ID-Recall* i *ID-Precision* u jedan broj računajući njihovu harmonijsku sredinu [155]. IDF1 metrika dolazi s nekim nedostacima uključujući prenaglašavanje asocijacija, neintuitivno i nemonotono ponašanje u slučaju detekcija, izostanak evaluacije pogreške lokalizacije te ne razmatranje točnosti asocijacije van preklapajućih dijelova pridruženih putanja [154].

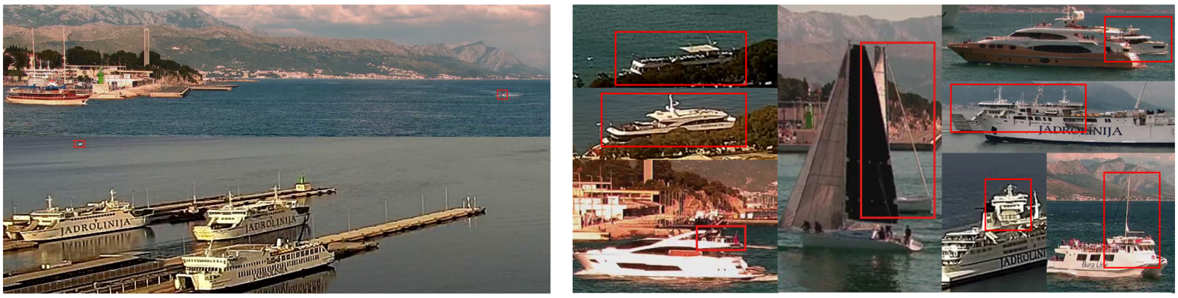
Klasične metrike

Prethodno navedene metrike često se komplementiraju i rezultatima **klasičnih metrika** [142] poput broja putanja stvarnih objekata koji je točno praćen u barem 80% okvira videozapisa (***MT*** - *Mostly Tracked*), broja putanja stvarnih objekata koji je točno praćen u manje od 20% okvira (***ML*** - *Mostly Lost*), broja **fragmenata** (***Frag***), odnosno hipoteza koje pokrivaju manje od 80% stvarne putanje objekta.

2.5. Ključni izazovi algoritama za praćenje više objekata

Algoritmi za praćenje više objekata suočavaju se s nizom izazova koji značajno mogu utjecati na njihovu učinkovitost i performanse. Ti izazovi uglavnom su povezani sa složenošću zadatka detekcije objekata, problemom precizne asocijacije i održavanja konzistentnih identiteta praćenih objekata, osobito u dinamičnim i složenim okruženjima, gdje su česta preklapanja, kolizije i interakcije među objektima.

Budući da se algoritmi za praćenje u velikoj mjeri oslanjaju na ulazne detekcije, njihove performanse mogu značajno biti narušene pogreškama detektora, poput izostanka detekcije objekta, lažno pozitivnih detekcija i nejasnih detekcija preklapajućih objekata [156, 62]. Poseban izazov predstavlja detekcija malih objekata koji na slici zauzimaju malu površinu, bilo da su fizički veliki, ali se nalaze na velikim udaljenostima od kamere, ili su zapravo mali [157, 158], kao i detekcija objekata zaklonjenih drugim objektima ili pozadinom. Slika 2.8 prikazuje nekoliko takvih primjera. Nadalje, velika varijabilnost i nepredvidivost stvarnih okruženja, zajedno s promjenama kuta gledanja, osvjetljenja i pozadinskih uvjeta, dodatno otežava zadatak detekcije i praćenja objekata [159].



(a) (b)

Slika 2.8: Primjeri malih objekata (a) i zaklonjenih objekata (b) koje je potrebno detektirati i pratiti.

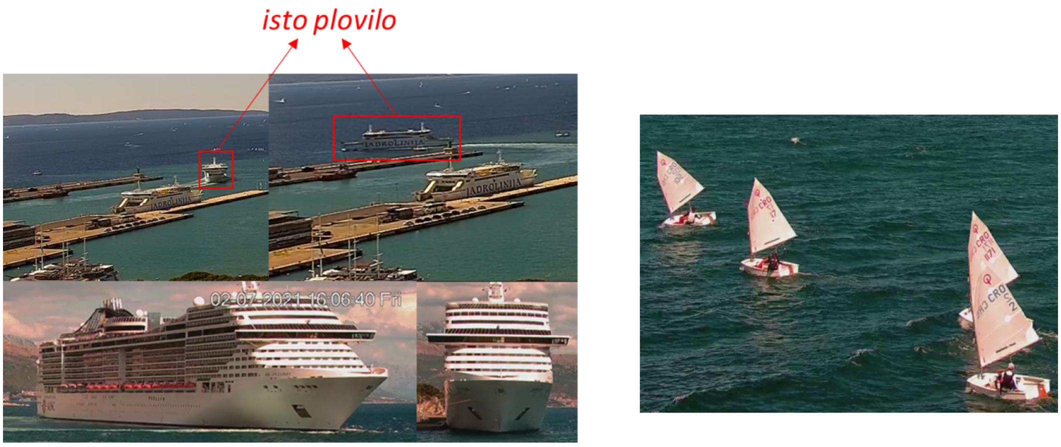
Problem okluzije objekata predstavlja jedan od najvećih izazova u sustavima za praćenje. Već djelomična zaklonjenost uspješno detektiranog objekta otežava njegovo ispravno povezivanje s odgovarajućom putanjom tijekom koraka asocijacije, budući da je kvaliteta dobivenih vektora značajki upitna [22]. U slučaju potpune zaklonjenosti objekta, vizualna informacija u potpunosti nedostaje, te se praćenje zaklonjenog objekta provodi isključivo na temelju predviđanja algoritma poput Kalmanovog filtera. Što je period okluzije dulji, to je procjena algoritma za predviđanje sljedećeg stanja manje pouzdana, te se često događa da objekt nije ispravno povezan s izvornom putanjom nakon završetka okluzije. Vizualne značajke objekta, koje obično imaju ključnu ulogu u reidentifikaciji [59], mogu se znatno razlikovati u trenucima pred potpunu okluziju i nakon nje, kao što je ilustrirano na Slici 2.9. Navedeno otežava ispravnu reidentifikaciju objekta te rezultira povećanim brojem zamjena identiteta tijekom praćenja [160].

Osim zbog okluzija, pogreške u asocijaciji i zamjene identiteta mogu biti uzrokovane: sličnim izgledom različitih objekata (Slika 2.10, b) što otežava razlikovanje njihovih vizualnih značajki [161, 162], značajnim varijacijama u izgledu istog objekta [81] (Slika 2.10, a), te ne-linearnim i nepredvidivim kretanjima objekata [162, 81]. Pored navedenoga, većina praktičnih primjena zahtijeva praćenje objekata u stvarnom vremenu, što implicira potrebu za brzim algoritmima. S druge strane, ograničenja hardverskim resursima mogu značajno

(a) (b)

Slika 2.9: Plovilo netom prije potpune okluzije (a) i prilikom izlaska iz nje (b).

ograničiti kompleksnost i učinkovitost algoritama praćenja.



(a) (b)

Slika 2.10: Varijacije u izgledu istog objekta (a) i sličan izgled različitih objekata (b).

3. PREGLED DOSADAŠNJIH ISTRAŽIVANJA

Ovo poglavlje pruža sustavan pregled dosadašnjih istraživanja u području automatske detekcije i praćenja plovila s posebnim osvrtom na problem okluzije praćenih objekata. U prvom dijelu (3.1) dan je pregled trenutno dostupnih skupova podataka koji se mogu koristiti prilikom treniranja i evaluacije modela za detekciju i praćenje plovila. Drugi dio (3.2) usmjeren je na radove koji se bave implementacijom konkretnih algoritama za detekciju i praćenje plovila. U trećem dijelu (3.3) razmatraju se istraživanja vezana za problem okluzije. Četvrti i posljednji dio (3.4) posvećen je analizi dosadašnjih istraživanja te identifikaciji otvorenih izazova u području automatske detekcije i praćenja plovila. U ovom dijelu razmatraju se ključna ograničenja postojećih pristupa te se ukazuje na smjerove u kojima je potrebno usmjeriti budući istraživački rad.

3.1. Dostupni skupovi podataka

Za razvoj uspješnih i pouzdanih algoritama za detekciju i praćenje plovila nužni su visokokvalitetni, označeni skupovi podataka koji obuhvaćaju svu raznolikost pomorskih okruženja. Takvi podaci trebali bi uključivati različite tipove plovila, različite vremenske uvjete i uvjete na moru, različite razine osvjetljenja, kao i različite kutove snimanja. Visokokvalitetni referentni skupovi podataka ne samo da omogućuju objektivnu usporedbu performansi različitih metoda detekcije i praćenja, već potiču i daljnji napredak ovoga istraživačkog područja.

3.1.1. Općeniti skupovi podataka

Neki općenito poznati, javno dostupni skupovi podataka, poput CIFAR10 [163], Caltech-256 [164] i ImageNet [125] skupova podataka za klasifikaciju, te PASCAL VOC [165] i MS COCO [166] skupova podataka za detekciju objekata, sadrže primjere plovila. No, kategorije plovila koje se javljaju u tim skupovima podataka ne obuhvaćaju svu njihovu raznolikost. Nadalje, plovila u tim skupovima podataka obično zauzimaju veći dio slike te se često nalaze u njenom samom središtu. Za učenje modela temeljenih na dubokom učenju, uz kvalitetu samih primjera, bitna je i njihova kvantiteta. Navedeni skupovi podataka sadrže ograničen broj primjera plovila. Nešto veći broj instanci plovila javlja se u MS COCO skupu podataka, ali oni pripadaju općoj klasi „*boat*“, bez daljnje klasifikacije na potkategorije. U Tablici

3.1 prikazan je broj klasa plovila, slika koje odgovaraju plovilima i odgovarajućih instanci objekata navedenih skupova podataka.

Tablica 3.1: Opći skupovi podataka koji sadrže primjere plovila.

Skup podataka	Klase plovila	Slike	Objekti	Zadatak
CIFAR10 [163]	1	6000	–	klasifikacija
Caltech-256 [164]	4	418	–	klasifikacija
ImageNet [125]	6	525	613	klasifikacija/detekcija
PASCAL VOC [165]	1	363	791	detekcija
MS COCO [166]	1	3025	10759	detekcija

Standardni referentni skupovi podataka za evaluaciju i usporedbu algoritama praćenja uglavnom su usredotočeni na praćenje pješaka. Primjerice, PETS2009 [29] skup podataka i skupovi iz MOT izazova [30, 31, 32, 33]. Afirmirali su se i neki skupovi podataka fokusirani na autonomnu vožnju [39, 34, 36, 38] koji dodatno obuhvaćaju i vozila na cestama. *Međutim, općeprihvaćen skup podataka specifično usmjeren na praćenje plovila još uvijek nedostaje.*

3.1.2. Skupovi podataka iz pomorskih okruženja

Posljednjih godina predstavljeno je nekoliko skupova podataka iz pomorskih okruženja koji sadrže označene slike i/ili videozapise snimljene ***RGB kamerama*** i usmjereni su na zadatke klasifikacije, detekcije i praćenja plovila. Kronološki pregled takvih skupova podataka dan je u Tablici 3.2, dok su u Tablici 3.3 navedene klase objekata koje se javljaju u tim skupovima. U Tablici 3.2, simbol "✓" na kraju retka označava da je skup podataka javno dostupan, dok "×" ukazuje na to da skup podataka nije dostupan javnosti.

VAIS [167], *MARVEL* [169] i *Game of Deep Learning: Ship Dataset* [171] skupovi podataka fokus stavljaju na klasifikaciju različitih vrsta plovila. **VAIS** [167] obuhvaća uparene RGB i infracrvene slike brodova prikupljene tijekom devet dana na šest različitih gatova. Ukupno je 2865 slika (1623 RGB i 1242 infracrvenih), od kojih je 1088 parova. **MARitime VEsseLs** (**MARVEL**) [169] skup podataka sadrži 2 milijuna slika 109 različitih tipova plovila, prikupljenih sa *Shipspotting*[1] web stranice. Korištenjem polu-nadzirane metode grupiranja, konstruirano je 26 superklasa plovila. **Game od Deep Learning: Ship Dataset** [171] je javno dostupan skup podataka predstavljen u sklopu *Game of Deep Learning: Computer Vision Hackathon-a* održanog 2019. godine. Skup podataka sadrži 6252 označene slike za treniranje i 2680 neoznačenih slika za konačnu evaluaciju iz skupa za testiranje.

Iako sadrži isključivo slike iz pomorskih okruženja, **Harbor Surveillance** [3] skup podataka za detekciju ne radi distinkciju između različitih kategorija plovila. Navedeni skup podataka sadrži 48966 slika dobivenih iz videozapisa snimljenih tijekom šestomjesečnog razdoblja s deset različitih pogleda na luku. Za svaki od pogleda, odabrano je i označeno

[1] https://www.shipspotting.com/

Tablica 3.2: Pregled skupova podataka iz pomorskih okruženja.

Skup podataka	Godina	Slike	Objekti	Klase	Rezolucija	Zadaci	
VAIS [167]	2015.	1623	–	6	5056×5056	klasifikacija	✓
SMD [168]	2017.	20367	157668	10	1920×1080	praćenje	✓
MARVEL [169]	2017.	2M (140000)	–	109 (26)	razne rezolucije	klasifikacija	✓
SeaShips [170]	2018.	31455	40077	6	1920×1080	detekcija	✓
Harbor Surveillance [3]	2018.	48966	70513	1	2048×1536	detekcija	×
Game of DL: ship dataset [171]	2019.	8932	–	5	razne rezolucije	klasifikacija	✓
McShips [172]	2020.	14709	26259	13	razne rezolucije	detekcija	✓
ABOships [1]	2021.	9880	41967	11	1920×720	detekcija	✓
GLSD [173]	2021.	152576	212357	13	razne rezolucije	detekcija	×
LMD-TShip [174]	2021.	40240	N/A	5	razne rezolucije	praćenje	✓
MarSyn [175]	2022.	25000	34000	6	1280×720 550×550	detekcija segmentacija	×
SeaSAw [4]	2022.	1.9 M	14.6 M	12	7680×1408, 3840×2056, 3648×2052, 1920×1080	detekcija, praćenje	×
SPSCD [176]	2023.	19337	27849	12	1920×1080	detekcija	✓
FVessel [177]	2023.	7625+	N/A	1	N/A	detekcija, praćenje	✓
MVDD13 [178]	2024.	35474	40839	13	N/A	detekcija	✓

više slika brodova, osiguravajući raznolikost pozadina i orijentacija. **ABOships** [1] skup podataka obuhvaća 9880 slika koje prikazuju čak 41967 označenih objekata. Slike su dobivene iz videozapisa snimljenih kamerom na plovilu za razgledavanje znamenitosti na ruti od grada Turku do grada Rusissalo u Finskoj tijekom 13 dana u lipnju i srpnju 2018. godine. Ovaj skup podataka obuhvaća različite vremenske uvjete tijekom dana, te uključuje slike otvorenog mora, luka i urbanih krajolika. U [175], autori u Blenderu generiraju sintetički **MarSyn** skup podataka koji obuhvaća 25 različitih foto-realističnih videozapisa, pri čemu se svaki sastoji od 1000 okvira. Cilj ovog skupa podataka je simulirati raznolike pomorske scenarije i uvjete, uključujući varijacije vremenskih uvjeta, slike u blizini obale te refleksije na površini vode. Plovila na slikama su različitih vrsta (teretni brodovi, vojni brodovi, ribarski čamci, gliseri, splavi za spašavanje i dr.), duljina (od 3 m do 125 m), oblika i boja.

Tablica 3.3: Klase plovila koje se javljaju u pojedinim skupovima podataka.

Skup podataka	Broj klasa	Klase
VAIS [167]	6	merchant, sailing, medium passenger, medium other, tugboat, small boat
SMD [168]	10	ferry, buoy, vessel/ship, speed boat, boat, kayak, sail boat, swimming person, flying bird/plane, other
MARVEL [169]	26 superklasa	container ship, bulk carrier, passengers ship, ro-ro/passenger ship, ro-ro cargo, tug, vehicles carrier, reefer, yacht, sailing vessel, heavy load carrier, wood chips carrier, fire fighting vessel, patrol vessel, platform, standby safety vassel, combat vessel, training ship, icebreaker, replenishment vessel, tankers, fishing vessels, supply vessels, carrier/floating, dredgers
SeaShips [170]	6	ore carrier, bulk cargo carrier, general cargo ship, container ship, fishing boat, passenger ship
Harbor Surveillance [3]	1	vessel
Game of DL: ship dataset [171]	5	cargo, carrier, cruise, military, tankers
McShips [172]	13	aircraft carrier, submarine, landing ship, auxiliary ship, destroyer, missile boat, speedboat, fishing boat, passenger ship, container ship, tugboat, sailboat, support ship
ABOships [1]	11	boat, cargoship, cruiseship, ferry, militaryship, miscboat, miscellaneous, motorboat, passengership, sailboat, seamark
GLSD [173]	13	sailing boat, fishing boat, warship, passenger ship, general cargo ship, container ship, bulk cargo carrier, barge, ore carrier, speed boat, canoe, oil carrier, tug
LMD-TShip [174]	5	cargo ships, fishing ships, passenger ships, speed boats, unmanned ships
MarSyn [175]	6	cargo ships, military ships, fishing boats, speed boats, rescue rafts, other
SeaSAw [4]	12	ship, recreational vessel, manual craft, sailing vessel, work boat, fishing vessel, towing vessel, dredge, wind turbine, marker, mooring buoy, miscellaneous
SPSCD [176]	12	small craft, small fishing boat, small passenger ship, fishing trawler, large passenger ship, sailing boat, speed craft, motorboat, pleasure yacht, medium ferry, large ferry, high speed craft
FVessel [177]	1	vessel
MVDD13 [178]	13	cargo, passenger, cruise, bulker, tanker, sailingboat, tug, fishing, drill, firefighting, containership, warship, submarine

Neki skupovi, poput *GLSD* [173] i *McShips* [172] skupova podataka, uključuju i primjere prikupljene putem interneta. Većina slika u **GLSD** [173] skupu podataka prikupljena je s interneta te obuhvaća različite svjetske luke, dok manji dio dolazi s nadzornog sustava Zhuhai Hengqin New Area, u Kini. Ovaj skup podataka obuhvaća širok spektar slika koje sadrže male objekte (manje od 32×32 piksela) i objekte srednje veličine (između 32×32 i 96×96 piksela), obuhvaća također i neke neobične primjere i situacije, poput slika plovila u plamenu te mozaike slika plovila. **McShips** [172] skup podataka obuhvaća 14709 slika podijeljenih u skup za treniranje (10297) i skup za testiranje (4412). Slike sadrže šest kategorija ratnih i sedam kategorija civilnih brodova. Podaci su prikupljeni putem različitih izvora poput web-tražilica, foruma, portala te videozapisa i nadzornih kamera, osiguravajući prisutnost različitih pozadina, osvjetljenja te atmosferskih uvjeta.

Najrecentniji skup podataka, **MVDD13** [178], razvijen je kao odgovor na nedostatak javno dostupnih skupova podataka za razvoj modela vizualne percepcije autonomnih besposadnih površinskih vozila (engl. *Unmanned Surface Vehicle*, USV). On uključuje razne vrste plovila te obuhvaća različita osvjetljenja i vremenske prilike, kao i primjere s okluzijama plovila, kako bi se omogućila što bolja generalizacija detektora na stvarne uvjete pomorskih okruženja. Za istraživanje u sklopu doktorskog rada značajan je i **Split Port Ship Classification Dataset** (**SPSCD**) [176] koji obuhvaća slike splitske luke snimljene jednom kamerom u razdoblju od veljače 2020. godine do prosinca 2022. godine, obuhvaćajući različita godišnja doba, dijelove dana te vremenske uvjete. Ovaj skup podataka odražava specifičnosti mediteranskih luka, poput splitske luke, gdje pomorski promet varira od manjih i srednjih plovila koja često nisu praćena uobičajenim sustavima za nadzor pomorskog prometa do velikih putničkih trajekata i kruzera. Na prikupljenim slikama, znanstvenici s Pomorskog fakulteta u Splitu identificirali su i precizno označili dvanaest kategorija brodova. Međutim, ovaj skup podataka prikladan je isključivo za detekciju i klasifikaciju plovila, ali ne omogućava evaluaciju algoritama za praćenje.

Među skupovima podataka navedenim u Tablici 3.2, samo su *SMD* [168], *LMD-Tship* [174], *SeaSAw* [4] i *FVessel* [177] prikladni za evaluaciju algoritama za praćenje plovila. **Singapore Maritime Dataset** (**SMD**) [168] sadrži 51 videozapis visoke rezolucije snimljen na različitim lokacijama i rutama u vodama oko Singapura. Većina videozapisa (njih 40) snimljena je s obale kamerom postavljenom na fiksnom postolju, dok su ostali (njih 11) snimljeni kamerom s broda u pokretu. Od deset klasa objekata, njih polovica ne pripada kategorijama brodova. Nadalje, dostupne anotacije su u .mat formatu te zahtijevaju dodatno procesiranje pri korištenju s popularnim bibliotekama dubokog učenja. **LMD-TShip** [174] skup podataka obuhvaća 191 videozapis snimljen fiksnim kamerama postavljenim na dva plovila te kamerama i mobilnim telefonima s obale. Podaci su podijeljeni u skup za treniranje (152 videozapisa, 31527 okvira) i skup za testiranje (39 videozapisa, 8713 okvira). Uključuje pet klasa plovila: *cargo ships, fishing ships, passenger ships, speed boats, unmanned ships*. Glavni nedostatak ovog skupa podataka je u tome što se uglavnom radi o jednom plovilu po videozapisu. Najopsežniji skup podataka, **Sea Situational Awareness** (**SeaSAw**) [4], sadrži podatke snimane kamerama s brodova u pokretu na nekoliko različitih geografskih lokacija duž Istočne obale SAD-a, u Bostonskoj luci i Europi. Uključuje dvanaest klasa objekata, od kojih njih pet ne odgovara brodovima (*dredge, wind turbine, marker, mooring buoy, miscellaneous*). Iako je predstavljen u sklopu CVPR konferencije 2022. godine, ovaj skup podataka još uvijek nije javno dostupan. **FVessel** [177] je javno dostupan skup podataka koji kombinira videosnimke plovila s pripadajućim AIS zapisima te je namijenjen evaluaciji algoritama za detekciju, praćenje i multimodalnu fuziju podataka. Sastoji se od

26 videozapisa snimljenih fiksnom obalnom kamerom i 7625 dodatnih slika. Skup podataka sadrži samo jednu klasu plovila te obuhvaća raznolike scenarije s primjerima okluzija i promjenama osvjetljenja. Iako videozapisi FVessel skupa podataka traju ukupno više od pet sati, prosječno se pojavljuje tek oko 0.35 plovila po minuti, odnosno jedno plovilo svakih približno tri minute. Primjerice, u videozapisu duljine 8:08 min pojavljuju se samo dva plovila, dok se u snimci trajanja 6:10 min javlja samo jedno.

3.2. Detekcija i praćenje plovila

Prvi dio ovog potpoglavlja donosi pregled radova posvećenih detekciji plovila na RGB slikama. Kvaliteta detekcija izravno utječe na performanse algoritma za praćenje, posebno u slučaju algoritama temeljenih na detekciji. Stoga je važno analizirati postojeće pristupe detekciji kako bi se bolje razumjela njihova ograničenja i prednosti u kontekstu praćenja plovila. U drugom dijelu fokus je na istraživanjima koja se bave samim praćenjem plovila iz RGB videozapisa.

3.2.1. Radovi koji se bave isključivo detekcijom plovila

U razmatranim radovima koji se bave isključivo detekcijom plovila, od dvostupanjskih detektora najzastupljeniji je **Faster R-CNN detektor**. Fu i suradnici [179] koriste model za detekciju plovila baziran na Faster R-CNN detektoru koji upotrebljava ResNet [180] mrežu za ekstrakciju značajki i metodu normalizacije mini-grupe [181]. Ovaj pristup kombinira se s metodom selektivnog učenja na teškim primjerima kako bi se poboljšale performanse detektora u složenim pomorskim okruženjima. Slično, Faster R-CNN s ResNet mrežom i odabirom teških primjera, koristi se i u [182]. Qi i suradnici [183] predlažu modificiranu verziju Faster R-CNN mreže, pri čemu se prije same detekcije provodi postupak smanjenja veličine slike i semantičkog sužavanja scene. Ova tehnika omogućuje isticanje ključnih informacija i usmjeravanje pažnje prema ciljanim područjima gdje bi se plovila mogla nalaziti. Detekcija se zatim provodi Faster R-CNN mreže preoblikovane u hijerarhijsku mrežu sužavanja, čime se smanjuje opseg pretrage detektora i poboljšava brzina detekcije.

Zahvaljujući svojoj sposobnosti brze i precizne detekcije objekata u stvarnom vremenu, jednostupanjski **YOLO detektor** postao je vodeći izbor u primjenama za detekciju plovila, što se jasno može vidjeti iz Tablice 3.4. U radu [184], ispituju se performanse YOLOv2 detektora za detekciju i klasifikaciju plovila, uspoređujući varijantu detektora predtreniranu na PASCAL VOC skupu podataka s onom koja je trenirana na SMD skupu podataka. Za detekciju objekata na površini mora, u radu [185] primjenjuje se unaprijeđena verzija YOLOv3 detektora, koja integrira DenseNet [186] model u Darknet-53 okosnicu, s ciljem poboljšanja prilagodljivosti bespilotnih plovila tijekom dugotrajnih misija. Modificirana varijanta YOLOv5 detektora prilagođena detekciji plovila na snimcima besposadnih površinskih plovila, predstavljena je u [187]. Ova inačica koristi K-means algoritam za optimizaciju referentnih graničnih okvira te uključuje Ghost modul [188] i transformere. Shi i suradnici [19] također predstavljaju model koji poboljšava YOLOv5 i nazivaju ga DrbLSK. Ovaj model integrira modul za dinamički odabir jezgre u okosnici detektora, dilatiranu konvoluciju s velikom jezgrom za smanjenje broja parametara te CIoU [189] funkciju gubitka za ubrzanu konvergenciju. U [190], Cafaro i suradnici koriste YOLOv6-n model za detekciju brodova te dobivene informacije komplementiraju s podacima radara/LIDAR-a i AIS prijemnika. Wu i suradnici [191] predstavljaju poboljšanu varijantu YOLOv7 detektora koja koristi referentne granične okvire koji su bolje prilagođeni različitim veličinama i oblicima brodova, integrira modul za fuziju značajki različitih skala i agregacijsku mrežu za fuziju mapa značajki različitih razina. Unaprijeđena verzija YOLOv7-tiny detektora, nazvana YOLOv7-Ship, predložena je u [192] za detekciju brodova u složenim pomorskim okruženjima. U usporedbi s baznom verzijom, YOLOv7-Ship pokazuje bolju točnost u detekciji objekata različitih veličina, malih i djelomično zaklonjenih objekata. Učinkovitije varijante YOLOv7-tiny detektora, koje pokušavaju riješiti izazov visokih računskih troškova postojećih modela za detekciju brodova uz zadržavanje točnosti detekcije, predložene su i u radovima [193, 194]. Zhao i Song [195] predstavljaju ekstenziju YOLOv8 detektora koja standardnu okosnicu za ekstrakciju značajki zamjenjuje kombinacijom efikasnog MobileViTSF vizualnog transformera i MobileNetv2 [196] mreže, klasične konvolucijske blokove zamjenjuje GSConv blokovima [197] te koristi redizajnirani C2f blok YOLOv8 mreže. YOLO-FE detektor, varijanta YOLOv8-n modela prilagođena za rad na krajnjim uređajima, predstavljena je u [198]. Model integrira FasterNet blok [199] radi smanjenja složenosti i povećanja brzine izračuna, te uključuje EMA mehanizam pažnje [200] za poboljšano razumijevanje globalnog konteksta uz istovremeno smanjenje računskih resursa. U radu [201], razvijena je nova ALF-YOLO arhitektura koja u YOLOv8 integrira: asimptotsku mrežu značajki piramide (engl. *Asymptotic Feature Pyramid Network*, AFPN) [202] kako bi se obogatila reprezentacija značajki korištenjem semantičkih informacija s više razina i LSK (engl. *Large Selective Kernel*) mehanizam pozornosti koji omogućuje detektoru da se više usredotoči na ključne značajke plovila, eliminirajući smetnje složenih okolišnih čimbenika. Učinkovitije varijante YOLOv8 detektora, EL-YOLO i AFF-LightNet, predložene su u [203, 204]. EL-YOLO (*Efficient Lightweight YOLO)* [203] koristi sveobuhvatnu analizu značajki okosnice modela kroz povećanu fuziju informacija, novi AWIoU gubitak i tehniku pohlepnog uklanjanja nepotrebnih filtera s ciljem smanjenja složenosti mreže. Yuan i suradnici [204] u AFF-LightNet modelu kombiniraju optimiziranu operaciju konvolucije GhostConv [188] s iterativnim mehanizmom fuzije značajki temeljenom na mehanizmu pozornosti, te koriste SIoU [205] funkciju gubitka. S druge strane, Liu i Zhu [206] koriste Res-YOLOX model koji u originalni YOLOX uvodi rezidualnu strukturu i CIoU [189] funkciju gubitka.

Iako YOLO detektor prevladava u većini istraživanja, neki autori odlučuju se za **alternativne jednostupanjske detektore**, poput CenterNet-a ili SSD-a. U članku autora Iancu i suradnici [207], evaluiraju performanse CenterNet detektora s različitim ekstraktorima značajki na prilagođenoj varijanti ABOships skupa podataka. Ova varijanta uključuje izbacivanje objekata koji zauzimaju manje od 16^2 piksela, te agregaciju originalnih trinaest klasa ABOship skupa podataka u četiri superklase radi ublažavanja neuravnoteženosti klasa. SSD detektor, prilagođen opažanju ekstremnih varijacija u veličini i obliku brodova, koristi se za detekciju brodova u Harbour Surveillance skupu podataka [3]. Li i suradnici [208] predlažu algoritam za detekciju objekata na vodenoj površini na panoramskim slikama temeljen na poboljšanoj varijanti SSD-a u kojoj je VGG16 [209] zamijenjena s ResNet-50 [180] mrežom te je dodano pet slojeva za ekstrakciju značajki.

Studije koje uspoređuju performanse različitih detektora provedene su u nekoliko istraživačkih radova. U [210] je dana usporedba Faster R-CNN i Mask R-CNN detektora (s ResNet101 okosnicom), koji su prethodno trenirani na skupovima podataka ImageNet i MS COCO, primijenjenih na SMD skupu podataka. Autori rada [170] evaluiraju detektore Fast R-CNN, Faster R-CNN s više različitih okosnica, SSD i YOLOv2 na SeaShips skupu podataka. Usporedba Faster R-CNN, SSD, YOLOv2, YoLOv3 i YOLOv3SPP detektora na McShips skupu podataka dana je u [172]. Detektori Faster R-CNN, SSD, EfficientDet i RFCN su evaluirani na ABOships skupu podataka u [1]. Rad [173] daje usporedbu devet različitih detektora, uključujući Faster R-CNN i RetinaNet detektore, na GLSD skupu podataka. U radu [178], Wang i suradnici uspoređuju performanse osam *state-of-the-art* detektora (Faster R-CNN, SSD, YOLOv3, RetinaNet, YOLOv5s, FCOS, YOLOX, DETR) na MVDD13 skupu podataka. U svom sljedećem istraživanju [211] razvijaju AodeMar, novi detektor optimiziran za bolju detekciju zaklonjenih brodova. Njegova robusnost složenim pomorskim uvjetima također je testirana na MVDD13 skupu podataka, uz usporedbu s detektorima Faster R-CNN, SSD, YOLOv3-v5, YOLOv7, YOLOv8 i DETR. Zhao i suradnici [212] uspoređuju dvanaest različitih detektora, među kojima su Faster R-CNN, SSD, RetinaNet i razne varijante YOLO detektora, na privatnom skupu podataka koji sadrži slike snimljene bespilotnim letjelicama. U radu [213], analiziraju se performanse tri različite varijante YOLO detektora (YOLOv5-s, YOLOv7, YOLOv8-s) na zadatku detekcije brodova na slikama snimljenim pri ulazu u marinu u Biogradu na Moru te na ulazu u Kanal sv. Ante u Šibeniku. S druge strane, Heller i suradnici [214] istražuju primjenu dviju verzija YOLOv4 detektora za detekciju objekata na moru, uključujući i različite vrste plovila, na krajnjim uređajima ugradbenih sustava. U svom radu koriste kombinaciju SMD i SeaShips skupova podataka, zajedno s MODD skupom podataka [215, 216] koji je posebno dizajniran za zadatak segmentacije slike u tri prirodna područja (nebo, obalu i more), te detekcije velikih i malih prepreka na morskoj površini.

Tablica 3.4: Dosadašnja istraživanja vezana za detekciju plovila.

Rad	Godina	Detektor	Podaci
Zwemer i sur. [3]	2018.	modificirani SSD	Harbour Surveillance
Lee i sur. [184]	2018.	YOLOv2	PASCAL VOC, SMD
Fu i sur. [179]	2018.	modificirani Faster R-CNN	privatni podaci
Shao i sur. [170]	2018.	Fast R-CNN i Faster R-CNN s više okosnica, SSD, YOLOv2	SeaShips
Zou i sur. [182]	2019.	modificirani Faster R-CNN	SMD
Qi i sur. [183]	2019.	modificirani Faster R-CNN	privatni podaci
Moosbauer i sur. [210]	2019.	Faster R-CNN, Mask R-CNN	SMD
Li i sur. [185]	2020.	modificirani YOLOv3	USV privatni podaci
Zheng i sur. [172]	2020.	Faster R-CNN, SSD, YOLOv2, YoLOv3 i YOLOv3SPP	McShips
Li i sur. [208]	2021.	modificirani SSD	slike iz različitih izvora, panoramske slike
Iancu i sur. [1]	2021.	Faster R-CNN, SSD, EfficientDet, RFCN	ABOships
Shao i sur. [173]	2021.	devet različitih detektora (Faster R-CNN, RetinaNet i dr.)	GLSD
Heller i sur. [214]	2022.	YOLOv4, YOLOv4-tiny	kombinacija SMD, SeaShips i MODD
Liu i Zhu [206]	2022.	Res-YOLOX	privatni podaci
Zhang i sur. [187]	2023.	modificirani YOLOv5	Game of DL: ship dataset, privatni USV podaci
Cafaro i sur. [190]	2023.	YOLOv6-n	COCO, ABOships privatni podaci
Wu i sur. [191]	2023.	modificirani YOLOv7	SeaShips
Zhao i Song [195]	2023.	modificirani YOLOv8	SeaShips
Iancu i sur. [207]	2023.	CenterNet	poboljšani ABOships
Cheng i sur. [198]	2023.	YOLO-FE baziran na YOLOv8-n	privatni podaci

Nastavlja se na sljedećoj stranici

Tablica 3.4 – Nastavak

Zhao i sur. [212]	2024.	dvanaest različitih detektora (Faster R-CNN, SSD, RetinaNet, razne varijante YOLO-a i dr.)	UAV privatni podaci
Wang i sur. [201]	2024.	ALF-YOLO baziran na YOLOv8	SeaShips, McShips
Jiang i sur. [193]	2024.	YOLOSeaShip baziran na YOLOv7-tiny	SeaShips
Correira i sur. [213]	2024.	YOLOv5-s, YOLOv7 YOLOv8-s	privatni podaci
Yang i sur. [203]	2024.	EL-YOLO baziran na YOLOv8	ABOships, SeaShips
Wang i sur. [178]	2024.	Faster R-CNN, SSD, YOLOv3, RetinaNet, YOLOv5s, FCOS, YOLOX, DETR	MVDD13
Wang i sur. [211]	2024.	novi *AodeMar* detektor	MVDD13
Jin i sur. [194]	2025.	YOLO-GCV baziran na YOLOv7-tiny	SeaShips + dodatni privatni i javni pdoaci
Yuan i sur. [204]	2025.	AFF-LightNet baziran naYOLOv8	SeaShips
Shi i sur. [19]	2025.	DrbLSK baziran naYOLOv5	ABOships

3.2.2. Praćenje plovila

Iako su detekcija i praćenje plovila usko povezani zadaci, istraživanja u području praćenja plovila znatno zaostaju za onima koja se bave detekcijom. To potvrđuju bibliometrijski podaci - od 2020. do danas (studeni, 2025.), Web of Science (WoS) baza podataka bilježi 2114 radova vezanih uz pojam *"ship detection"*, dok je radova povezanih s pojmom *"ship tracking"* svega 174. Sličan omjer prisutan je i u IEEE Xplore bazi (10257 prema 91). Kada se dodatno filtriraju istraživanja koja ne koriste standardne RGB videozapise u kombinaciji s metodama dubokog učenja za automatizirano praćenje, ionako malen broj relevantnih radova dodatno se smanjuje.

Sažeti prikaz relevantnih istraživanja koja se bave praćenjem plovila na RGB videozapisima dan je u Tablici 4.1. Zbog preglednosti, u stupcu *"Praćenje"* koristi se oznaka "*" kada se u istraživanju koristi modifikacija ili proširenje navedenog algoritma praćenja. Primjerice, *StrongSORT** označava da se u znanstvenom radu koristi modificirana verzija StrongSORT-a, dok *StrongSORT* predstavlja originalnu implementaciju tog algoritma.

Iz Tablice 4.1 može se uočiti da recentni radovi iz područja praćenja plovila najčešće koriste kombinaciju YOLO detektora i algoritama praćenja iz SORT familije, pri čemu je najzastupljeniji **DeepSORT algoritam**. Modificiranu verziju YOLOv3 algoritma u kombinaciji s DeepSORT algoritmom koriste u [217] za detekciju i praćenje brodova u unutarnjim plovnim putevima - na rijeci Yangtze u Kini. Uz optimizaciju referentnih graničnih okvira, autori zamjenjuju sigmoidnu aktivacijsku funkciju klasifikatora softmax aktivacijom, te predlažu korištenje poboljšane verzije NMS algoritma za učinkovitije uklanjanje redundantnih graničnih okvira. U [218], fokus je na praćenju brodova za maglovita vremena. Prvo se pomoću konvolucijske neuronske mreže iz svakog okvira videozapisa ukloni magla, a zatim se za detekciju i praćenje koriste YOLOv5 i DeepSORT. Poboljšane verzije YOLOv5 i YOLOX algoritma u kombinaciji s originalnim DeepSORT algoritmom koriste se za praćenje plovila i u radovima [219, 220]. S druge strane, **modifikacije DeepSORT algoritma** implementirane su u [221, 222, 223]. Qi i suradnici [221] produljuju "vrijeme čekanja" prije brisanja izgubljene putanje sa 30 na 300 uzastopnih okvira. Dodatno se, u svakom koraku čekanja, površina posljednjeg graničnog okvira te putanje povećava za 10%, sve do maksimalnih 2000 piksela, nakon čega se putanja briše. U [222], predlaže se kombinacija poly-YOLO modela za detekciju i poboljšanog DeepSORT modula (engl. *Enhanced DeepSORT, EDS*) za praćenje plovila. EDS modul koristi Gaussov filter za smanjenje šuma i pozadinske interferencije, te primjenjuje normalizaciju histograma kako bi se uravnotežila raspodjela kontrasta slike, čime se poboljšava praćenje plovila pri slaboj vidljivosti. Zhang i suradnici [223] koriste optimizirani YOLOv7 detektor zajedno s varijantom DeepSORT algoritma koja u finalnoj fazi pridruživanja umjesto IoU koristi DIoU [189] mjeru sličnosti.

U nekoliko radova koriste se i drugi popularni algoritmi temeljeni na detekciji, preciznije **ByteTrack** i **StrongSORT**. Wu i suradnici [224] koriste poboljšani YOLOv8 u kombinaciji sa StrongSort algoritmom za praćenje brodova na videozapisima snimljenim bespilotnom letjelicom. U radu [225], Han i Jung predlažu varijantu StrongSORT algoritma nazvanu StableSORT, koja učinkovito rješava izazove nepravilnih kretanja besposadnih plovila na kojima je postavljena kamera, kao i nestabilnosti uzrokovane valovima i vjetrovima. StableSORT koristi B-IoU [226] kako bi prevladao ograničenja standardne IoU metrike u dinamičnim uvjetima, te prilagođeni NSA Kalmanov filter koji c u $\widetilde{R}_t = (1-c)R_t$ postavlja na jedan kada pouzdanost detekcije premašuje zadanu granicu. Navedenim se uklanja šum mjerenja i maksimizira utjecaj detekcije visoke pouzdanosti prilikom ažuriranja stanja sustava u Kalmanovom filteru. Hao i suradnici [227] koriste poboljšanu verziju YOLOv7-tiny detektora, koja pokazuje bolje performanse u detekciji brodova na složenim pozadinama, različitim skalama te u slučajevima djelomične zaklonjenosti. Detektor se kombinira s unaprijeđenom implementacijom StrongSORT algoritma, koji koristi OSNet [228] mrežu za ekstrakciju vizualnih značajki, čime se smanjuje učestalost zamjena identiteta praćenih plovila. Za praćenje plovila u stvarnom vremenu, Xing *et al.* [229] sugeriraju korištenje YOLOv8-FAS algoritma, poboljšane verzije YOLOv8n algoritma dizajnirane za primjenu na uređajima s ograničenom memorijom i računalnim resursima, zajedno s ByteTrack algoritmom. AdapTrack algoritam za praćenje plovila, koji se oslanja na FairMOT algoritam, opisan je u radu [230]. AdapTrack algoritam implementira strategiju asocijacije ByteTrack algoritma. U prvoj fazi, detekcije visoke pouzdanosti se pridružuju putanjama koristeći vizualne karakteristike. Zatim, u drugoj fazi, preostale putanje se povezuju s detekcijama niske pouzdanosti temeljem IoU vrijednosti. S druge strane, Zou i sur. [231] u obje faze asocijacije, pri povezivanju s detekcijama visoke pouzdanosti i s onima niske pouzdanosti, primjenjuju jedinstvenu mjeru sličnosti koja, uz IoU, uključuje i smjer kretanja plovila. U svom sljedećem radu, Zou i sur. [232] također kombiniraju IoU sa smjerom kretanja plovila u koraku asocijacije, ali ovog puta koriste VFE-YOLO detektor. Ova poboljšana varijanta YOLOX detektora koristi višeskalne veze u arhitekturi izdvajanja značajki koje poboljšavaju prepoznavanje plovila. Autori uvode i strukturni pozornostni modul koji smanjuje utjecaj redundantnih informacija te povećava reprezentacijsku sposobnost modela.

Za razliku od pristupa koji se oslanjaju na popularne algoritme praćenja, u radovima [233, 234, 235, 236] predstavljeni su **alternativni pristupi praćenju plavila temeljeni na detekciji**. Wu i suradnici [233] koriste YOLOv3 i YOLOv5 detektore u kombinaciji s RoDAN (*Robust Deep Affinity Network*) algoritmom praćenja koji, kroz tri različita modula, objedinjuje informacije o tri ključna aspekta objekta: njegovoj veličini, regiji koju obuhvaća i njegovom kretanju. Prvi modul spaja značajke različitih skala, čime omogućuje bogatiju i semantički potpuniju reprezentaciju svakog plovila, smanjujući njegovu osjetljivost na varijacije u veličini. Drugi modul poboljšava reprezentaciju značajki i rješava problem velikih graničnih okvira koji obuhvaćaju pozadinu, osobito u slučaju plovila s dodacima poput jarbola ili drugih nadogradnji. Treći modul dodatno usmjerava i fino podešava preliminarne rezultate praćenja, provjeravajući uspješno povezane, nove i izgubljene putanje, čime algoritam postaje manje ovisan o detektoru i otporniji na okluzije. Autori u [234] evaluiraju četiri varijante YOLO algoritma za detekciju brodova na SMD skupu podataka te predlažu algoritam za praćenje plovila na videozapisima snimljenim kamerom na brodu u pokretu koji koristi pridruživanje temeljem IoU vrijednosti detekcija iz trenutnog okvira i praćenih objekata i pridruživanje zasnovano na sličnosti ORB (*oriented FAST and rotated BRIEF*) značajki i veličini graničnih okvira. Wang i suradnici [235] koriste YOLOv5 detektor u kombinaciji s algoritmom praćenja koji integrira znanje da "objekti ne mogu iznenada nestati". U tu svrhu, algoritam uključuje dodatnu granu kojoj je cilj locirati sve objekte iz prethodnog kadra u trenutnom okviru videozapisa. Za to se koristi algoritam praćenja pojedinačnih objekata temeljen na sijamskim neuronskim mrežama, koji računa sličnost između objekta i potencijalnih regija u trenutnom okviru gdje bi se objekt mogao nalaziti. Ako značajke iz novog okvira značajno odstupaju od onih u posljednjih N okvira, objekt se smatra nestalim i prekida se potraga za njim. U radu [236], koristi se jednostavan model praćenja nalik SORT algoritmu uz dodatak kaskadnog pridruživanja. Algoritam praćenja temelji se na predviđanjima

Kalmanovog filtera, te kaskadnom i IoU pridruživanju mađarskim algoritmom.

Nešto drugačiji pristupi praćenju, koji nadilaze klasične okvire popularnih algoritama, predstavljeni su u [237, 238, 239]. Shan i suradnici [237] koriste SiamFPN model za praćenje brodova koji se sastoji od sijamske mreže s FPN podmrežama i tri mreže za predlaganje regija od interesa. Autori u [239] predstavljaju novi algoritam praćenja zasnovan na mehanizmu dinamičke memorije i hijerarhijskom modelu koji je svjestan konteksta. Mehanizam dinamičke memorije pohranjuje značajke prethodnih okvira te ih dinamički integrira s trenutnim značajkama kako bi se u model inkorporirao vremenski kontekst i međusobna koreliranost okvira videozapisa. Hijerarhijski model koristi se za ekstrakciju kontekstualne informacije na različitim skalama te globalnih i lokalnih informacija pomoću slojeva sažimanja i konvolucije s dilatacijom. Luo i suradnici [238] primjenjuju CO-Tracker [240], model zasnovan na transformerima, u kombinaciji s LSTM i graf neuronskim mrežama s mehanizmom pozornosti.

Tablica 3.5: Dosadašnja istraživanja vezana za praćenje plovila.

Rad	**Godina**	**Detekcija**	**Praćenje**	**Podaci**
Shan i sur. [237]	2020.	SiamFPN model zasnovan na sijamskim neuronskim mrežama		privatni podaci
Wu i sur. [233]	2021.	**YOLOv3** i **YOLOv5**	RoDAN	SMD i privatni HSD skup podataka
Jie i sur. [217]	2021.	poboljšani **YOLOv3**	DeepSORT	privatni podaci
Li i sur. [219]	2022.	poboljšani **YOLOv5**	DeepSORT	privatni podaci
Park i sur. [234]	2022.	četiri varijante **YOLO**-a	algoritram temeljen na IoU i ORB_vel pridruživanju	SMD
Xing i sur. [229]	2023.	**YOLOv8-FAS** zasnovan na **YOLOv8-n**	ByteTrack	privatni podaci
Chen i sur. [230]	2023.	AdapTrack: algortiam zasnovan na FairMOT-u s ByteTrack asocijacijom		SMD
Yang i sur. [239]	2023.	Algoritam zasnovan na mehanizmu dinamičke memorije i hijerarhijskom modelu		LMD-Tship
Zhou i sur. [218]	2023.	**YOLOv5**	DeepSORT	privatni podaci

Nastavlja se na sljedećoj stranici

Tablica 3.5 – Nastavak

Liu i Li [220]	2023.	poboljšani **YOLOX**	DeepSORT	LMD-TShip
Wang i sur. [235]	2023.	**YOLOv5**	algoritam temeljen na znanju	SeaShips i privatni podaci
Wu i sur. [224]	2024.	poboljšani **YOLOv8**	StrongSORT	PASCAL VOC i privatni podaci
Luo i sur. [238]	2024.	CO-Tracker zasnovan na transformerima s LSTM graf neuronskim mrežama		privatni podaci
Qi i sur. [221]	2024.	poboljšani **YOLOv5**	DeepSORT*	privatni podaci
Han i Jung [225]	2024.	**YOLOv5**	StrongSORT*	privatni podaci
Cen i sur. [236]	2024.	poboljšani **YOLOv7**	praćenje temeljeno na Kalmanovom filteru	SeaShips i privatni podaci
Chen i sur. [222]	2024.	**poly-YOLO**	DeepSORT*	SMD
Zhang i sur. [223]	2025.	poboljšani **YOLOv7**	DeepSORT*	privatni podaci
Hao i sur. [227]	2025.	poboljšani **YOLOv7-tiny**	StrongSORT*	SMD
Zou i sur. [231]	2025.	**YOLOX**	asocijacija u dvije faze (IoU + smjer kretanja)	FVessel
Zou i sur. [232]	2025.	**VFE-YOLO**	VSATrack	FVessel

3.3. Problem okluzije objekata

U posljednje vrijeme, objavljena je nekolicina radova koji istražuju kako poboljšati performanse algoritama praćenja tijekom okluzija objekata. Većina dosadašnjih istraživanja o okluzijama fokusira se na praćenje pješaka i vozila, dok je područje praćenja plovila znatno manje istraženo.

3.3.1. Istraživanja problema okluzije općenito

Cao i suradnici [27] predstavljaju OC-SORT algoritam, unaprijeđenu verziju klasičnog SORT [58] algoritma koja je usmjerena na prevladavanje ograničenja vezanih za okluzije i ne-linearna kretanja objekata. Njihov pristup uključuje dodatno ažuriranje parametara Kalmanovog filtera kada se objekt ponovno pojavi nakon što je bio zaklonjen, čime se smanjuje akumulirana pogreška. Ovo ažuriranje temelji se na virtualnoj putanji objekta generiranoj korištenjem podataka o posljednjoj detekciji prije okluzije i prvoj nakon nje. Nadalje, mjera sličnosti za asocijaciju putanja i detekcija uključuje i analizu konzistentnosti brzine kretanja objekata, što dodatno doprinosi preciznosti praćenja. U [52], autori sugeriraju zamjenu Kalmanovog filtera mrežom konvolucijskih propusnih povratnih ćelija (ConvGRU) kako bi se riješio problem dugotrajnih okluzija u algoritmu zajedničke detekcije i praćenja. Ova mreža omogućuje korištenje povijesnih prostorno-vremenskih informacija više objekata i učenje dugoročnih ovisnosti.

Zhang i suradnici [160] predlažu strategiju asocijacije koja je robusnija na okluzije objekata. Za razliku od DeepSORT [59] algoritma, zadržavaju granične okvire probnih putanja čak i kada im tri uzastopna okvira nije pridružena detekcija, te ne brišu granične okvire putanja maksimalne starosti jer bi upravo ti okviri mogli odgovarati objektima koji su dulje vrijeme zaklonjeni. Navedene granične okvire nazivaju graničnim okvirima visoke vrijednosti. Nepridružene detekcije u koraku asocijacije pridružuju se graničnim okvirima visoke vrijednosti i predviđenim graničnim okvirima, te se odabire ono pridruživanje koje ima manju kosinusnu udaljenost. Za predviđanje se koristi metoda najmanjih kvadrata ako je broj okvira od kada je putanji pridružena detekcija manji od zadane granične vrijednosti, inače se koristi Kalmanov filter.

S druge strane, ByteTrack [60] i BoT-SORT [61] algoritmi zadržavaju gotovo sve detektirane granične okvire, uključujući i one s niskom pouzdanošću koji se obično filtriraju, jer bi upravo ti granični okviri mogli sadržavati objekte koji su *djelomično* zaklonjeni. Pridruživanje detekcija i putanja provodi se u dvije faze. U prvoj fazi, detekcije visoke pouzdanosti pokušavaju se pridružiti postojećim putanjama korištenjem IoU i/ili vizualne sličnosti, dok se u drugoj fazi asocijacije neuparenim putanjama pridružuju detektirani granični okviri niske pouzdanosti koristeći isključivo IoU kao mjeru sličnosti, budući da vizualne značajke nisu pouzdane u slučaju zaklonjenih objekata. Autori u [46] primjenjuju sličan pristup. Međutim, oni u drugoj fazi asocijacije, u kojoj se detekcije niske pouzdanosti pridružuju preostalim putanjama, koriste prostorno-vremenske značajke koje u obzir uzimaju odnos prostornog položaja objekta i njegovog susjednog okruženja.

U radu [241], predlaže se unaprijeđeni algoritam praćenja temeljen na regresiji koji koristi regresijsku glavu detektora u dvije faze za predviđanje sljedeće pozicije praćenog objekta. Osim regresije aktivnih putanja, provodi se i regresija neaktivnih putanja kojima je regresijska vrijednost pala ispod zadane minimalne vrijednosti s_{min} ili im je regresijom dobiveni granični okvir filtriran algoritmom ne-maksimalnog potiskivanja, što može biti posljedica okluzije objekta. Pojam zaklonjena putanja odnosi se na aktivnu putanju koja je potisnuta NMS algoritmom zbog preklapanja s drugom aktivnom zaklanjajućom putanjom. Za nove detekcije koje nisu filtrirane aktivnim putanjama i NMS algoritmom, provjerava se pripadaju li neaktivnim zaklonjenim putanjama koristeći udaljenost centra graničnog okvira detekcije i predviđene pozicije neaktivne putanje. Ako je ta udaljenost manja od zadane granične vrijednosti, neaktivna putanja se ponovno aktivira. Granična vrijednost koja se koristi za aktivaciju putanje raste s povećanjem brzine putanje i broja okvira u kojima je putanja bila neaktivna, budući da oba termina utječu na (ne)pouzdanost predviđene pozicije.

Han i suradnici [242] predlažu inovativni LSTP modul za predviđanje sljedeće pozicije objekta koji koristi kombinaciju prostornog i vremenskog Transformera, uz metodu asocijacije koja prioritet daje aktivnim i zaklonjenim putanjama. Integrirajući prostorni koder Transformera za česte interakcije objekata i vremenski koder Transformera za praćenje vremenskog kontinuiteta putanja, LSTP modul daje dugoročno bolja predviđanja za neaktivne putanje. Osim predviđanja sljedeće pozicije objekta, LSTP modul također procjenjuje vidljivost predviđenog graničnog okvira. Na temelju procijenjene vidljivosti, putanje se kategoriziraju kao zaklonjene (vidljivost $<$ granične vrijednosti) i zaklanjajuće (vidljivost $\geq$ granične vrijednosti). U koraku asocijacije, prioritet se daje pridruživanju aktivnih putanja nad neaktivnima te zaklanjajućih putanja nad zaklonjenima. Nadalje, detektirani granični okviri se razvrstavaju prema pouzdanosti, sprječavajući pridruživanje putanja visoke vidljivosti s detektiranim graničnim okvirima niske pouzdanosti, koji obično sadrže zaklonjene objekte.

Zhou i suradnici [22] u svom algoritmu praćenja implementiraju modul zadužen za detekciju okluzija, pri čemu objekt smatraju zaklonjenim ako se središte njegovog graničnog okvira nalazi unutar nekog drugog graničnog okvira. Vizualne značajke objekata za koje je detektirana okluzija se ne ažuriraju kako bi se izbjeglo njihovo „zagađenje" tijekom kratkotrajnih okluzija. U radu [243], autori predlažu adaptivni model koji kombinira predviđanja dobivena modelima kratkoročnog i dugoročnog predviđanja kako bi se poboljšale performanse tijekom okluzija različitih trajanja. Model kratkoročnog predviđanja kombinira predviđanja dobivena temeljem vizualnih informacija i informacija o kretanju, dok se model dugoročnog predviđanja oslanja na kubičnu splajn interpolaciju. Dendorfer i suradnici [23] pokušavaju riješiti problem dugotrajnih okluzija predviđajući buduće pozicije objekata koristeći simulaciju scene iz ptičje perspektive i generativne suparničke neuronske mreže. U [244], korak asocijacije koristi sekvencijalne značajke putanja koje integriraju informacije iz više okvira videozapisa, što ih čini otpornijima na određene anomalije. Za ekstrakciju sekvencijalnih značajki, koje istovremeno obuhvaćaju i prostornu i vremensku informaciju o objektu, koristi se AP3D mreža. U radu [245], autori koriste rezultate dobivene modelom segmentacije prepreka u sceni kako bi identificirali situacije u kojima je objekt zaklonjen preprekom/pozadinom, dok se vizualni model temeljen na pažnji koristi za rješavanje problema okluzije objekta drugim objektom. PSMOT [246] algoritam koristi model zajedničke detekcije i ekstrakcije značajki s R-FCN detektorom u dvije faze. Za rješavanje problema okluzije, PSMOT koristi model osjetljivosti na položaj. Dobivene klasifikacijske mape osjetljive na položaj transformiraju se u binarne mape koje ukazuju na dijelove objekta koji su zaklonjeni, omogućujući tako ignoriranje tih dijelova prilikom agregacije značajki. Na taj način se efikasno smanjuju smetnje uzrokovane okluzijama.

3.3.2. Istraživanja vezana za okluzije plovila

Wang i suradnici, u svom radu [211], bave se problemom **detekcije** zaklonjenih plovila, te predlažu novi model AodeMar, koji unutar YOLO okvira implementira mehanizam pažnje s ciljem preciznijeg prepoznavanja zaklonjenih brodova. Dani model pomoću RCAC3 modula kombinira značajke niskog i visokog semantičkog značaja koristeći rezidualne veze, čime omogućuje preciznije određivanje graničnih okvira zaklonjenih brodova. Nadalje, koriste SP-STR modul za korelaciju semantike značajki na više razina koji kombinira SPP [247] (*Spatial Pyramid Pooling*) i STR [248] blok s mehanizmom samopozornosti kako bi se poboljšala sposobnost klasifikacije zaklonjenih brodova. No, ovaj rad je fokusiran isključivo na **poboljšanje detekcije zaklonjenih plovila**, ne i na druge aspekte praćenja objekta tijekom okluzije. Hao i suradnici [227] koriste poboljšanu verziju YOLOv7-tiny detektora, koja pokazuje bolje performanse u detekciji djelomično zaklonjenih plovila. Dodatno, integracijom OSNet [228] mreže za ekstrakciju vizualnih značajki u StrongSORT algoritam, reducira se broj zamjena identiteta koje mogu biti uzrokovane okluzijama objekata.

Chen i suradnici u svom radu [249] iz 2020. godine obrađuju problem praćenja **jednog, unaprijed definiranog plovila** za vrijeme okluzije, oslanjajući se pritom na klasične metode kerneliziranih korelacijskih filtera. U recentnom radu iz 2025. [250] koristi se fuzija informacija iz različitih izvora te OC-SORT algoritam praćenja kako bi se umanjio izazov praćenja plovila za vrijeme okluzija i nelinearnih kretanja. Wang i sur. [235] u svoj algoritam praćenja integriraju spoznaju da objekti ne mogu iznenada nestati, te nastoje locirati sve objekte koji su bili prisutni u prethodnom okviru. Ova metoda prepoznaje nestanak objekta, odnosno **detektira okluziju**, kada se značajke u novom okviru značajno razlikuju od značajki objekta u prethodnih N okvira. Međutim, **ne nudi rješenja za praćenje zaklonjenih objekata**, već jednostavno prekida pretragu za objektima koji su nestali.

U radovima [233, 221] problem lakšeg uparivanja zaklonjenih objekata s novim detekcijama pokušava se riješiti postupnim povećavanjem posljednjeg graničnog okvira izgubljenog objekta. Wu i suradnici [233] dodatno provjeravaju inicijalne rezultate asocijacije kako bi identificirali putanje koje su stvarno izgubljene. Tim se putanja posljednji granični okvir povećava kroz maksimalno 50 uzastopnih okvira, a ako nakon tog razdoblja ostanu neuparene, onda se uklanjaju. S druge strane, u [221], neuparenim putanjama koje čekaju na pridruživanje u svakom koraku se granični okvir proširuje za 10% (s centrom kao središtem) do maksimalnih 2000 piksela. Ako se neuparena putanja, koja je dosegla maksimalnu površinu graničnog okvira, ne bude povezana s detekcijom, ona se briše. Također, maksimalno vrijeme čekanja prije brisanja putanje produljuje se na 300 uzastopnih okvira. Zou i sur. u radovima [231, 232] za rješavanje problema reidentifikacije plovila nakon okluzije predlažu mjeru sličnosti koja, uz IoU, dodatno integrira i smjer gibanja plovila. Ovakav pristup rezultira robusnijim praćenjem te smanjenim brojem zamjena identiteta uzrokovanih okluzijama u scenarijima plovidbe unutarnjim vodnim putovima.

3.4. Analiza dosadašnjih istraživanja i otvoreni izazovi

Analiza do sada predstavljenih skupova podataka iz pomorskih okruženja ukazuje na značajan nedostatak javno dostupnih, označenih skupova podataka koji su prikladni za razvoj i evaluaciju algoritama za praćenje plovila, a koji istovremeno adekvatno obuhvaćaju specifičnosti pomorskih okruženja i raznolikosti plovila. Kod dostupnih skupova podataka ograničenje predstavljaju scene male gustoće plovila, koje nedovoljno odražavaju stvarne uvjete. Takvi skupovi često sadrže tek nekoliko plovila u vidnom polju što otežava procjenu performansi algoritama u dinamičnim scenarijama s visokom gustoćom prometa. Također, za ispitivanje robusnosti algoritma na dugotrajne okluzije plovila, ključno je da skup podataka sadrži različite primjere plovila koja su dugo vremena zaklonjena. Zbog navedenog, istraživanje u domeni praćenja plovila, osobito uz primjenu najnovijih modela dubokog učenja, drastično zaostaje za mnogobrojnim istraživanjima u području praćenja pješaka, a situacija je još nepovoljnija kada je riječ o praćenju plovila za vrijeme okluzija. Nadalje, većina istraživanja vezanih za praćenje plovila oslanja se na privatne podatke, što otežava objektivnu usporedbu tih metoda. Stoga je neophodno razviti javno dostupne i reprezentativne skupove podataka za praćenje plovila kako bi se potakao daljnji napredak ovog istraživačkog područja i omogućila objektivna usporedba postojećih metoda.

Iako su detekcija i praćenje plovila komplementarni zadaci, istraživanja u području praćenja plovila znatno zaostaju za istraživanjima usmjerenima na njihovu detekciju. Neki od mogućih razloga uključuju veći broj javno dostupnih skupova podataka za zadatak detekcije, kao i jednostavniju implementaciju detekcije u odnosu na praćenje. Naime, praćenje plovila zahtijeva ne samo detekciju plovila, već i dodatne zadatke poput predviđanja budućih pozicija praćenih plovila, ekstrakciju vizualnih značajki, preciznu identifikaciju plovila te održavanje konzistentnog identiteta tokom vremena. Mnogi radovi, poput [217, 229, 224, 219, 220], usredotočeni su na poboljšanje detektora, dok za praćenje koriste popularne algoritme poput DeepSORT-a, ByteTrack-a i StrongSORT-a u osnovnom obliku, bez dodatnih prilagodbi ili poboljšanja. Iako kvaliteta detekcije značajno utječe na kvalitetu praćenja, samo poboljšanje detektora ne rješava problem dugotrajnih okluzija. Stoga se potrebno fokusirati na moguća poboljšanja u ostalim koracima praćenja kako bi se potencijalno riješio problem okluzija.

U nekoliko nedavno objavljenih radova razmatraju se i modifikacije popularnih algoritama ili se predlažu neki novi algoritmi praćenja. Međutim, ovi radovi problem okluzije plovila ili u potpunosti zanemaruju ili ga samo djelomično rješavaju. U radovima [211, 227] fokus je na detekciji djelomično zaklonjenih objekata. Dok se [211] uopće ne razmatra problema praćenja za vrijeme okluzije niti reidentifikacije izgubljenih objekata, [227] uvodi poboljšani model za ekstrakciju vizualnih značajki koji može smanjiti broj pogrešno dodijeljenih identiteta nakon perioda okluzije. Ipak, pritom treba uzeti u obzir da vizualne značajke objekta koji tek izlazi iz okluzije često nisu reprezentativne, te ovaj pristup sam po sebi ne rješava u potpunosti problem reidentifikacije objekta nakon okluzije. Wang i suradnici [235] koriste metodu koja identificira nestanak objekta, ali umjesto pokušaja ponovnog pronalaska, putanja nestalog objekta se jednostavno briše. Pristup postepenog povećanja graničnog okvira izgubljenog objekta s ciljem lakšeg budućeg uparivanja predstavljen je u [233, 221]. Ovaj pristup poprilično je ograničen. Pretjerano proširenje graničnog okvira može dovesti do značajnih preklapanja s detekcijama drugih objekata, što otežava precizno pridruživanje i može negativno utjecati na pouzdanost praćenja. Chen i suradnici [250] koriste OC-SORT algoritam, koji uključuje dodatni korak ažuriranja kako bi smanjio pogreške Kalmanovog filtera akumulirane tijekom okluzije. Međutim, ovaj algoritam pretpostavlja da će izgubljeni objekt biti ponovno detektiran i ispravno reidentificiran kako bi se dodatno ažuriranje izvršilo, što često nije slučaj, osobito pri duljim periodima okluzije.

Od istraživanja vezanih uz praćenje plovila, radovi [249, 232, 231] detaljnije se bave problemom praćenja plovila tijekom okluzije. U radu [249] autori pritom koriste klasične metode kerneliziranih korelacijskih filtera. Nadalje, fokus je na praćenju samo jednog, unaprijed definiranog plovila, što predstavlja znatno uži obuhvat u odnosu na problem okluzija koji se javlja pri praćenju većeg broja objekata.

Zou i sur. [232, 231] u svojim se radovima prvenstveno oslanjaju na prostorne i dinamičke informacije o kretanju plovila, koje u razdobljima duljih okluzija postaju sve nepouzdanije. Budući da se pritom vizualne značajke plovila u potpunosti zanemaruju, model ostaje osjetljiv na odstupanja u predviđenim gibanjima te na scenarije u kojima više plovila pokazuje slične obrasce kretanja. Takav pristup ograničava pouzdanost reidentifikacije nakon okluzije i dodatno motivira integraciju komplementarnih vizualnih informacija kako bi se postiglo stabilnije praćenje u složenim uvjetima.

Iz navedenog razmatranja evidentni su nedostaci u postojećim istraživanjima vezanim za praćenje plovila, osobito u kontekstu okluzija. Radovi koji se konkretno bave okluzijama objekata uglavnom su usmjereni na rješavanje problema okluzija u zadacima praćenja pješaka, gdje je karakteristična česta interakcija među objektima, nagli i nepredvidljivi obrasci kretanja te češće kratkotrajne zaklonjenosti. Ideje koje su iznesene u radovima koji istražuju okluzije pješaka mogu biti korisne i za rješavanje problema okluzija kod plovila. Međutim, važno je uzeti u obzir specifične karakteristike pomorskih okruženja, plovila i njihovog kretanja kako bi se razvili efikasniji algoritmi praćenja koji će općenito imati bolje performanse i biti otporniji na dugotrajne okluzije. Dakle, postoji značajna potreba za daljnjim istraživanjima koja bi mogla rezultirati naprednijim tehnikama i algoritmima, posebno usmjerenih na rješavanje problema dugotrajnih okluzija plovila, čime bi se unaprijedila robusnost i učinkovitost sustava za praćenje plovila u složenim uvjetima.

4. DIZAJN I IMPLEMENTACIJA *SPLIT SHIP MOT* SKUPA PODATAKA

Pouzdanost i robusnost metoda za praćenje plovila uvelike ovise o kvaliteti i raznovrsnosti podataka korištenih tijekom njihovog razvoja. Pregled ranije predstavljenih skupova podataka iz pomorskih okruženja pokazuje značajan nedostatak javno dostupnih označenih skupova podataka koji su pogodni za razvoj i evaluaciju algoritama praćenja plovila, a istovremeno adekvatno obuhvaćaju sve specifičnosti pomorskih okruženja i raznolikosti plovila. Dodatan problem predstavlja nedovoljna zastupljenost složenih scena koje uključuju istovremeno praćenje većeg broja plovila, njihovo međusobno mimoilaženje te situacije u kojima dolazi do djelomičnih ili potpunih okluzija među plovilima. Nadalje, za ispitivanje robusnosti algoritama u uvjetima dugotrajnih okluzija nužno je da skup podataka sadrži i primjere plovila koja su dulje vrijeme zaklonjena. Bez takvih primjera nije moguće sveobuhvatno procijeniti stabilnost sustava praćenja u realnim, dinamičnim uvjetima.

U okviru doktorske disertacije kreiran je novi ***Split Ship MOT (SSMOT)*** skup podataka koji se sastoji od tri komplementarne komponente:

- **Skupa za detekciju**, koji omogućava razvoj i evaluaciju modela sposobnih za preciznu lokalizaciju i klasifikaciju plovila u pojedinačnim okvirima videozapisa.
- **Skupa za reidentifikaciju (ReID)**, koji sadrži slike istog plovila snimljenog u različitim položajima i uvjetima, čime se omogućuje učenje diskriminativnih značajki nužnih za ponovno prepoznavanje istog objekta nakon prekida praćenja zbog ograničenja detektora ili uslijed okluzije.
- **Skupa za praćenje**, koji obuhvaća označene videozapise sa scenarijima dugotrajnih okluzija i mimoilaženjima plovila, pružajući tako sveobuhvatan i realističan okvir za procjenu performansi algoritama praćenja.

Ovakav višekomponentni dizajn osigurava veću fleksibilnost i raznolikost u odnosu na standardne skupove podataka koji se temelje isključivo na označenim videozapisima. U takvim skupovima, za treniranje detektora najčešće se koriste uzastopni okviri iz ograničenog broja videozapisa koje karakterizira visoka redundancija [251, 252]. U kratkim vremenskim razmacima javljaju se gotovo identični prikazi istih objekata, dok su uvjeti snimanja i osvjetljenja unutar jednog videozapisa uglavnom konstantni. Kao rezultat, detektori trenirani na uzastopnim okvirima videozapisa imaju smanjenu sposobnost generalizacije na nove scene i uvjete. Sličan problem javlja se i u kontekstu reidentifikacije [253, 254]. Kada se slike istog plovila uzimaju iz susjednih okvira, dobivaju se gotovo identični vizualni prikazi koji ne doprinose učenju diskriminativnih značajki potrebnih za prepoznavanje plovila u različitim položajima, kutovima gledanja te pri promjenama vremenskih uvjeta i osvjetljenja.

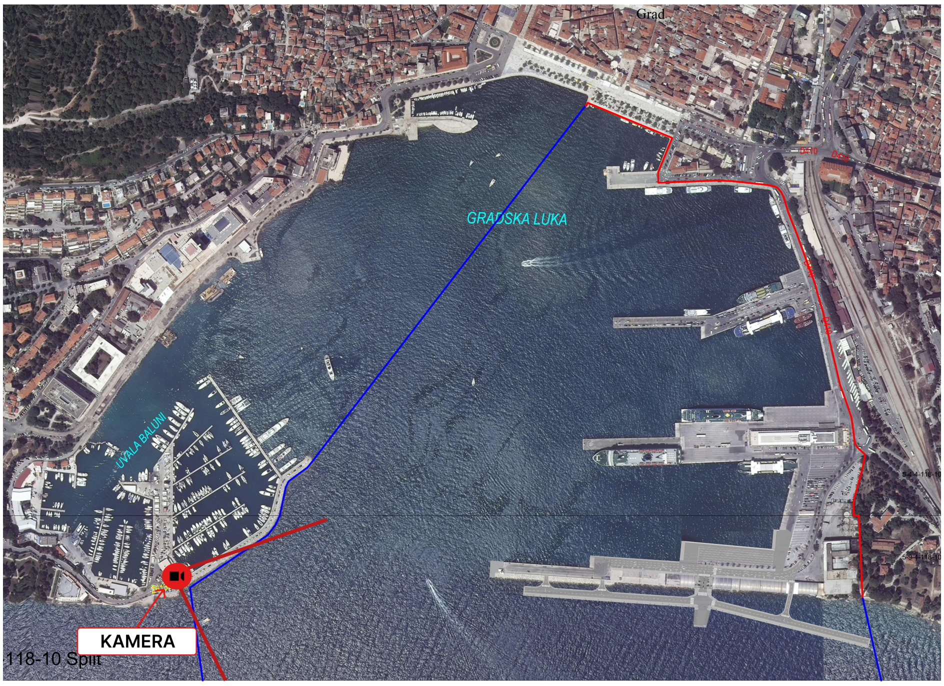
Alternativni pristup u literaturi temelji se na evaluaciji praćenja na danim skupovima videozapisa, dok se detektor i ReID model treniraju i evaluiraju na zasebnim skupovima podataka, posebno dizajniranim upravo za te zadatke [235, 224]. Ključni nedostatak ovog pristupa jest neusklađenost distribucija podataka različitih podskupova, budući da vizualne karakteristike objekata, uvjeti snimanja ili raspodjela scena u skupu za detekciju i ReID skupu često ne odgovaraju onima u videozapisima korištenima za evaluaciju praćenja. Dodatno, klase objekata u tim skupovima ne moraju nužno odgovarati klasama koje se javljaju u videozapisima za evaluaciju praćenja. Zbog toga, u procesu integracije može doći do pada performansi – detektor i ReID model uče reprezentacije koje nisu u potpunosti prilagođene specifičnostima scena u kojima se od njih očekuje zajedničko djelovanje. Takva neusklađenost otežava i realnu procjenu sustava praćenja, jer rezultati evaluacije djelomično odražavaju različite distribucije i klase podataka koji su se koristili za treniranje pojedinih komponenti.

Upravo navedene probleme adresira SSMOT skup podataka, u kojem su sve tri komponente – detekcija, reidentifikacija i praćenje – izvedene iz istog izvora i oblikovane prema konzistentnim kriterijima. Na taj se način osigurava da vizualne karakteristike, uvjeti snimanja i distribucija scena budu usklađeni kroz sve zadatke, čime se uklanja problem neusklađenosti odvojenih podskupova. Istodobno, način konstrukcije detekcijskog i ReID skupa sprječava pojavu redundancije uzrokovane uzimanjem uzastopnih, gotovo identičnih kadrova, te osigurava dovoljnu varijabilnost u prikazima istih objekata. Time detektor i ReID model uče reprezentacije prilagođene stvarnim izazovima praćenja, ali i dovoljno robusne da generaliziraju na različite uvjete snimanja. Evaluacija praćenja je pritom pouzdanija, budući da se temelji na jedinstvenoj, konzistentnoj i raznolikoj podatkovnoj osnovi. Takav dizajn otvara prostor za razvoj robusnijih sustava, što SSMOT skup podataka čini značajnim doprinosom u području praćenja plovila.

4.1. Opis lokacije i tehničkih specifikacija kamere

Slike i videozapisi korišteni u SSMOT skupu podataka prikupljeni su u splitskoj luci, najvećoj putničkoj luci u Hrvatskoj i jednoj od najprometnijih luka na Mediteranu [176]. Splitsku luku karakterizira intenzivan i raznolik promet koji obuhvaća trajekte, katamarane, kruzere te velik broj manjih brodica i jedrilica. Upravo zbog prisutnosti brojnih malih i srednjih plovila koja nisu obuhvaćena standardnim sustavima za nadzor i upravljanje [255, 256], splitska luka predstavlja izazovno okruženje za automatsko praćenje i analizu pomorskog prometa.

Za snimanje je korištena nadzorna kamera *Dahua DH-TPC-PT8620A-T* [257] opremljena $1/1.9''$ Sony CMOS senzorom rezolucije 1944×1092 piksela (≈ 2 MP). Kamera omogućuje snimanje u slabim svjetlosnim uvjetima (0.001 lux u boji) te posjeduje objektiv sa žarišnom duljinom $6-180$ mm, kutom vidnog polja $59^\circ - 2.4^\circ$ i $30\times$ optičkim zumom. Kamera je postavljena na zgradi na ulazu u luku, čime je omogućeno sustavno praćenje cjelokupnog dolaznog i odlaznog pomorskog prometa u luci Split. Preciznije, kamera se nalazi na koordinatama $43^\circ 30' 04''$ N i $16^\circ 25' 48''$ E, te je montirana na visini od približno 9 m iznad razine mora. Pozicija kamere u luci prikazana je na Slici 4.1 crvenim krugom s ikonom kamere. Iz kruga se protežu crvene linije koje približno označavaju stvarno vidno polje kamere.



Slika 4.1: Lokacija kamere u splitskoj luci. (Slika preuzeta iz [258], uz izmjene.)

4.2. Skup podataka za detekciju

Točna i precizna detekcija plovila ključna je za njihovo uspješno praćenje, budući da svaki propust ili pogreška u ovom koraku izravno otežava kasniju reidentifikaciju i narušava konzistentnost dodijeljenih identiteta kroz vrijeme. Kako bi se omogućilo učenje robusnih i univerzalnih značajki, SSMOT detekcijski podskup obuhvaća raznolike primjere plovila snimljene u različitim uvjetima, od različitih doba dana i razina osvjetljenja do promjenjivih okolišnih uvjeta. Pri njegovoj izradi posebna je pažnja posvećena izbjegavanju redundantnih i uniformnih vizualnih prikaza iz susjednih okvira videozapisa. Umjesto toga, odabrani su reprezentativni i međusobno raznoliki primjeri, čime se povećava varijabilnost skupa podataka i omogućuje razvoj detektora s boljim generalizacijskim sposobnostima, koji su prikladni za primjenu u stvarnim, dinamičnim uvjetima.

4.2.1. Klasifikacija plovila

Prilikom definiranja kategorija plovila u SSMOT skupu podataka kao polazište je korištena klasifikacija iz SPSCD skupa za detekciju koji je razvijen na Pomorskom fakultetu Sveučilišta u Splitu uz sudjelovanje stručnjaka sa Zavoda za nautiku [176]. U tom skupu podataka kategorije su određene prema tipu i duljini plovila, no u praksi duljinu plovila nije moguće pouzdano razaznati iz samih snimaka. Stoga je u okviru SSMOT skupa naglasak stavljen na vizualna obilježja plovila koja su lako prepoznatljiva iz videozapisa.

Na temelju ciljeva i svrhe SSMOT skupa podataka, kao i specifičnosti prometa u splitskoj luci, definirano je ukupno 11 kategorija plovila. Kategorije *Medium Ferry* i *Large Ferry* iz izvornog SPSCD skupa podataka objedinjene su u jedinstvenu kategoriju *Ferry* zbog jako sličnih vizualnih obilježja. Detaljan pregled kategorija s pripadajućim opisima i karakterističnim vizualnim obilježjima dan je u Tablici 4.1, a na Slici 4.2 prikazani su reprezentativni primjeri objekata iz svake kategorije.

Tablica 4.1: Pregled kategorija plovila korištenih u SSMOT detekcijskom podskupu.

ID	Klasa	Opis	Vizualna obilježja	Tipična namjena
0	*Small Craft*	Vrlo mala plovila namijenjena rekreaciji, sportu ili osnovnoj obuci	Vrlo malih dimenzija, izloženo sjedenje, npr. *male trening jedrilice, jet ski, daske i kanui*	Rekreativno veslanje, jedriličarska obuka, sportovi na vodi
1	*Small Fishing Boat*	Manje privatne i ribarske brodice skromnih performansi	Jednostavno plovilo tradicionalnog izgleda, otvoren trup, skromna kabina i/ili tenda, vanbrodski motor	Rekreativni ili poluprofesionalni ribolov, osobni prijevoz lokalaca
2	*Small Passenger Ship*	Prijevoz ograničenog broja putnika na kraćim relacijama ili turističkim vožnjama	Obično više paluba, kabine s prozorima, moguće otvorena paluba, ponekad jedra (*turistički jedrenjaci*)	Turistički izleti, linijski prijevoz, kratke kružne ture
3	*Fishing Trawler*	Veći ribarski i radni brodovi	Masivan radni trup, oprema za koćarenje, visoki pramac	Industrijski ribolov, višednevne ribarske ekspedicije

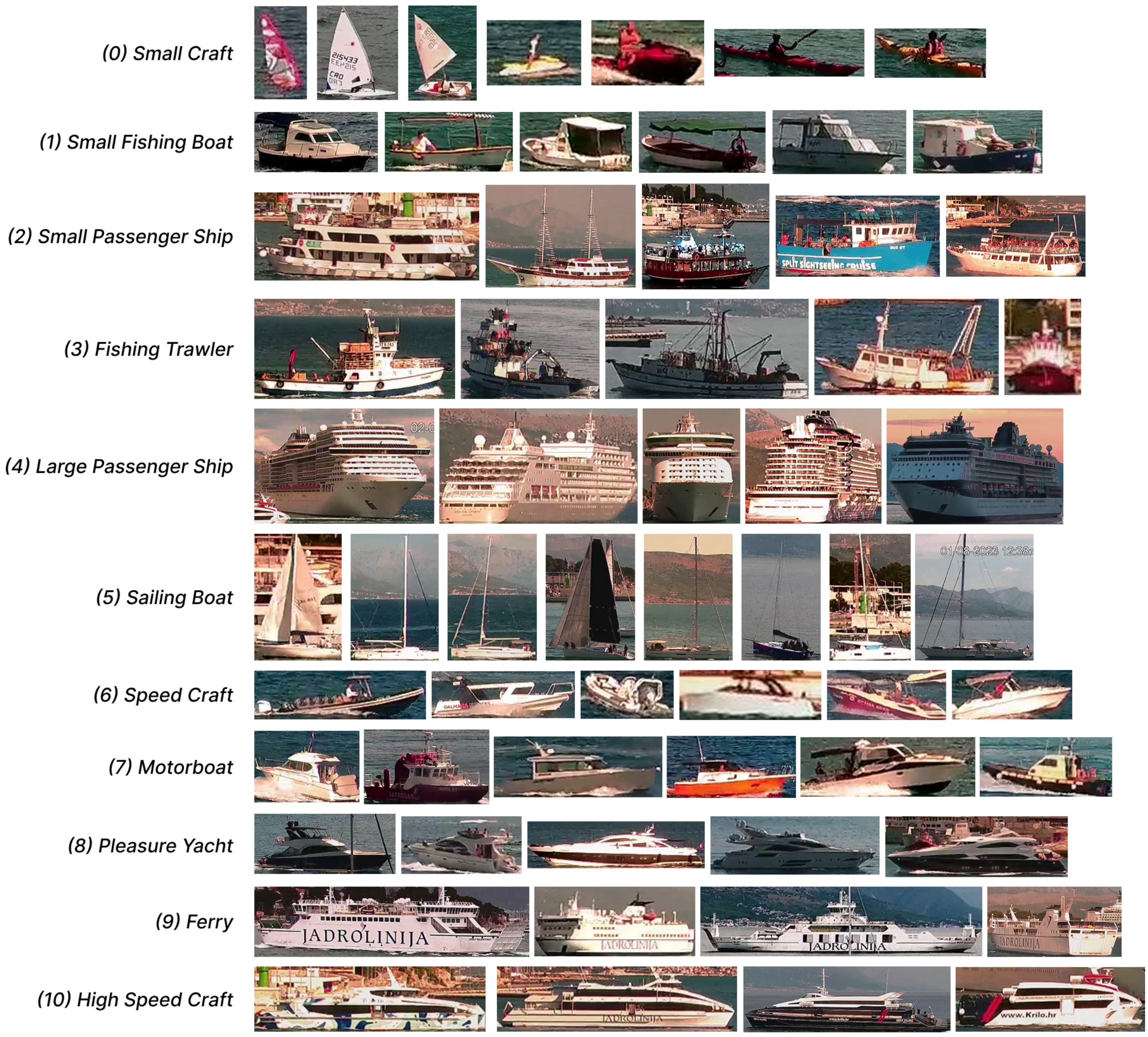
Nastavlja se na sljedećoj stranici

Tablica 4.1 – Nastavak

4	*Sailing Boat*	Plovila na jedra koja koriste vjetar za pogon uz mogući pomoćni motor	Jarbol(i), jedra, vitak trup srednje do velike veličine	Rekreacijsko jedrenje, regate, charter, luksuzna plovidba
5	*Speed Craft*	Brza, sportska motorna plovila	Nizak, izdužen i aerodinamičan trup, otvorena ili poluzatvorena kabina, izražen sportski dizajn	Brza vožnja, glisiranje, adrenalinske aktivnosti, hitne intervencije, brzi transferi
6	*Motorboat*	Višenamjenska motorna plovila srednje veličine	Stabilan, robustan trup, zatvorena/poluzatvorena kabina, udoban i funkcionalan dizajn	Obiteljski izleti, svakodnevna plovidba, službene dužnosti poput ophodnje
7	*Pleasure Yacht*	Luksuzne privatne, sportske i charter jahte	Elegantna i moderna silueta, višekatno nadgrađe ili otvoreni sportski dizajn	Privatna i VIP charter plovidba, društveni događaji, rekreacija
8	*Ferry*	Trajekti za putnike i vozila	Višepalubno plovilo s rampama	Redovni linijski promet kopno-otok i međunarodne linije preko mora
9	*High Speed Craft*	Brzi katamarani i slična plovila	Moderan, aerodinamičan dizajn, uska silueta, niski trup, često katamaran	Brzi prijevoz putnika na udaljene relacije
10	*Large Passenger Ship*	Veliki brodovi za prijevoz većeg broja putnika, opremljeni za duža putovanja	Višekatna paluba, velik trup, brojni prozori/kabine	Kružna turistička putovanja (kruzeri), prekooceanske linije

4.2.2. Proces prikupljanja i anotacije slika

Za potrebe izrade detekcijskog podskupa korišteni su videozapisi snimljeni u razdoblju od 31. srpnja do 17. listopada 2023. godine, koje obuhvaća ljetne mjesece s povećanom prometnom aktivnošću u splitskoj luci, te dodatni zapisi iz svibnja 2024. godine. Iz njih je izdvojeno ukupno **1040 slika**. Dodatno je uključena i **2091 slika** iz postojećeg SPSCD skupa podataka, koji sadrži ukupno 19337 slika [176]. Pri odabiru slika iz SPSCD-a posebna je pažnja posvećena osiguravanju raznolikosti, budući da skup sadrži velik broj međusobno vrlo sličnih



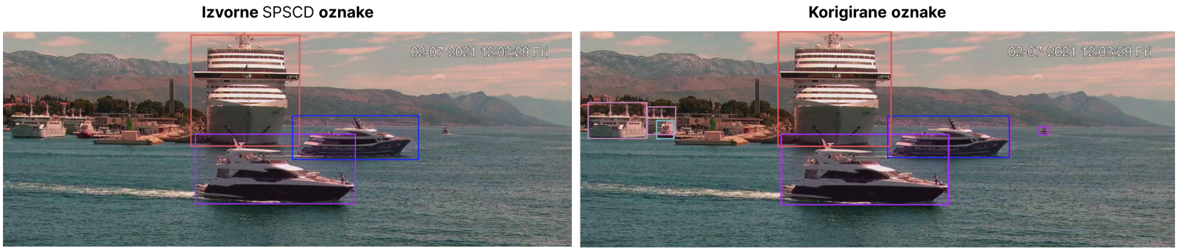
Slika 4.2: Reprezentativni primjeri objekata pojedinih klasa plovila.

kadrova dobivenih iz uzastopnih okvira videozapisa. Uključivanjem SPSCD slika dobivena je i dodatna sezonska raznolikost, jer one obuhvaćaju i zimske mjesece koji nisu zastupljeni u vlastitim snimkama. Postojeće anotacije odabranih SPSCD slika su prilagođene i korigirane, budući da su originalni granični okviri bili nedovoljno precizni (Slika 4.3), a sva plovila od interesa nisu uvijek bila označena (Slika 4.4). Također, ispravljene su uočene nekonzistentnosti u oznakama istog plovila na različitim slikama.

Sve slike anotirane su koristeći *Computer Vision Annotation Tool* (*CVAT*) [259], standardni alat za označavanje objekata na slikama i u videozapisima. Korišten je *YOLO* format oznaka [88], u kojem je svaki objekt na slici reprezentiran uređenom petorkom:

$$(ID_{class}, \frac{x_c}{w_{img}}, \frac{y_c}{h_{img}}, \frac{w}{w_{img}}, \frac{h}{h_{img}}),$$

Slika 4.3: Primjeri graničnih okvira koji obuhvaćaju nepotrebne dijelove pozadine.

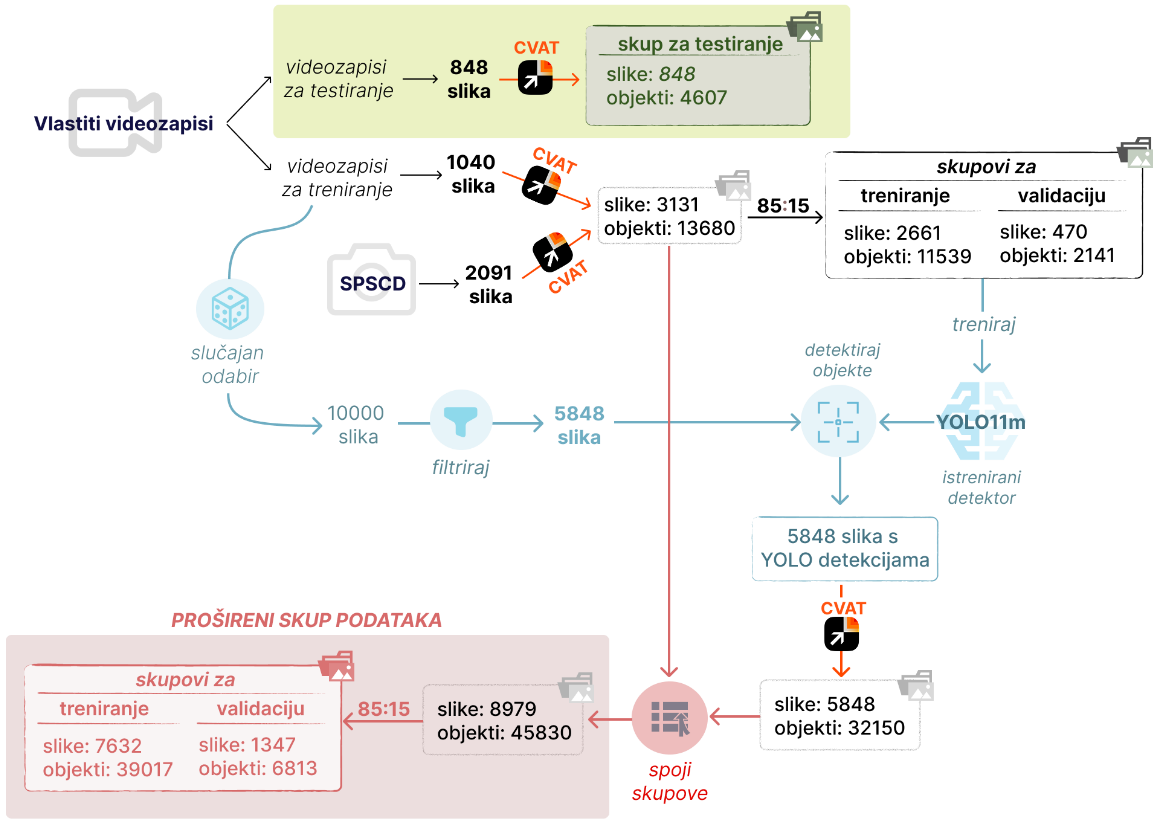


Slika 4.4: Razlika u broju označenih plovila u izvornim SPSCD anotacijama i korigiranoj verziji oznaka.

pri čemu ID_{class} predstavlja numeričku oznaku klase kojoj objekt pripada, x_c i y_c koordinate središta graničnog okvira, w i h njegovu širinu i visinu, a w_{img} i h_{img} širinu i visinu slike. Označene slike podijeljene su na skup za treniranje i validaciju u omjeru 85:15. Od ukupno 3131 slike s 13680 objekata, na skup za treniranje otpada 2661 slika s 11539 objekata, dok skup za validaciju čini 470 slika s 2141 objektom.

Kako bi se dodatno povećala raznolikost i proširio skup podataka za detekciju, primijenjen je poluautomatizirani postupak. Na pripremljenim podacima treniran je YOLO11m detektor tijekom 100 epoha, s veličinom miniserije (engl. *mini-batch*) 8. Iz videozapisa namijenjenih za treniranje izdvojeno je 10000 slučajno odabranih okvira, na koje je primijenjen prethodno istrenirani YOLO11m detektor iz epohe s najboljim performansama na skupu za validaciju. Slike bez detektiranih objekata, kao i one s isključivo privezanim plovilima, uklonjene su. Detekcije dobivene YOLO detektorom preostalih slika korištene su kao početna točka za njihovu bržu i učinkovitiju anotaciju u CVAT-u. Naposljetku je dobiven **prošireni skup podataka**, znatno veće raznolikosti, s **8981 slikom i 45830 objekata**, koji je također podijeljen na skup za treniranje i validaciju u omjeru 85:15. Detaljan pregled broja slika i instanci objekata po klasama, u inicijalnoj i u proširenoj verziji skupa podataka, prikazan je u Dodatku A.

Za potrebe konačne evaluacije, iz posebnih videozapisa koji nisu korišteni za generiranje skupa za treniranje, izdvojeno je i označeno **848 slika** koje čine nezavisni **testni skup**. Cjelokupni proces prikupljanja, filtriranja, anotacije i proširenja skupa za detekciju prikazan je na Slici 4.5, koja ilustrira ključne korake u stvaranju SSMOT detekcijskog podskupa.



Slika 4.5: Vizualni prikaz procesa stvaranja SSMOT podskupa za detekciju.

4.2.3. Karakteristike detekcijskog podskupa

Radi boljeg razumijevanja SSMOT detekcijskog skupa podataka, u nastavku je provedena analiza njegovih osnovnih karakteristika poput broja instanci objekata po klasama, broja objekata po slici te dimenzija i površina graničnih okvira. Osim numeričkih pokazatelja, uključeni su i ilustrativni primjeri slika koji prikazuju različite uvjete snimanja, promjene osvjetljenja, djelomične okluzije objekata i druge izazovne situacije.

Struktura SSMOT skupa podataka za detekciju

SSMOT skup podataka za detekciju sastoji se od ukupno 9827 slika u dvije rezolucije: 1920×1080 (9601 slika) i 1280×720 (226 slika). Skup je podijeljen na **podskup za treniranje** (7632 slika), **validaciju** (1347 slika) i **testiranje** (848 slika), pri čemu testni podskup čini nezavisnu cjelinu jer slike potječu iz videozapisa različitih od onih korištenih za treniranje i validaciju. Struktura SSMOT skupa za detekciju, s detaljnim prikazom broja slika na kojima se javljaju objekti pojedine klase i broja instanci po klasama u pojedinim podskupovima i sveukupno, dana je u Tablici 4.2. U retku "Ukupno" Tablice 4.2 u stupcima

"Slike" naveden je ukupan broj slika u pojedinom skupu podataka [1]. Slika 4.6 dodatno ilustrira broj instanci po klasama u SSMOT skupu za detekciju, s naznačenim udjelom pojedinih podskupova.

Tablica 4.2: Broj slika i instanci objekata po klasama u SSMOT detekcijskom podskupu.

Klasa	**Train**		**Val**		**Test**		**Sveukupno**	
	Slike	Inst.	Slike	Inst.	Slike	Inst.	Slike	Inst.
Small craft	517	1417	90	199	92	269	699	1885
Small Fishing Boat	1023	1101	168	186	83	85	1274	1372
Small Passenger Ship	2755	3496	483	613	257	301	3495	4410
Fishing Trawler	3751	4098	652	714	572	589	4975	5401
Large Passenger Ship	3360	4150	604	741	497	607	4461	5498
Sailing Boat	2139	3261	390	612	262	333	2791	4206
Speed Craft	3203	4501	540	769	410	567	4153	5837
Motorboat	1780	1902	320	343	186	210	2286	2455
Pleasure Yacht	1282	1383	225	243	136	151	1643	1777
Ferry	7088	12351	1242	2133	793	1315	9123	15799
High-speed craft	1274	1357	233	260	170	180	1677	1797
Ukupno	**7632**	**39017**	**1347**	**6813**	**848**	**4607**	**9827**	**50437**

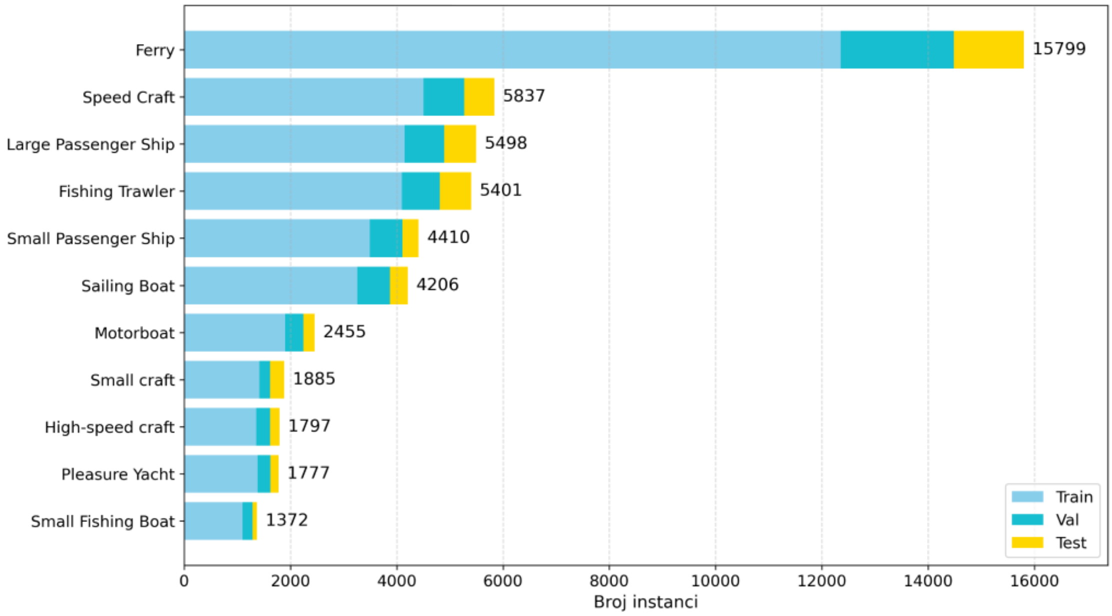
Vidljivo je da klasa *Ferry* značajno dominira po broju instanci, što se može objasniti činjenicom da se trajekti često pojavljuju privezani u luci te su stoga prisutni na velikom broju slika. Sličan razlog djelomično objašnjava i povećan broj instanci klasa *Large Passenger Ship* i *Fishing Trawler*, budući da se i oni nerijetko pojavljuju u kadru dok su privezani. Klasa *Speed Craft* obuhvaća objekte koji, za razliku od prethodno navedenih, nisu privezani već se uglavnom pojavljuju u pokretu, no unatoč tome bilježe velik broj instanci. To se može objasniti njihovom čestom prisutnošću u stvarnim scenama te izraženom intra-klasnom varijabilnošću: ova klasa obuhvaća širok raspon modela koji se razlikuju po obliku, boji, vrsti pogona i po materijalu izrade (npr. guma, plastika, aluminij).

Broj objekata po slici

Kako bi se dobio sveobuhvatan uvid u strukturu podataka, analize koje slijede provode se nad cjelokupnim detekcijskim skupom, objedinjavanjem podskupova za treniranje, validaciju i testiranje.

Na Slici 4.7 prikazana je **distribucija broja objekata po slici** koristeći dvije komplementarne vizualizacije. Gornji stupčasti dijagram (a) prikazuje apsolutan broj slika koje sadrže određenu vrijednost broja objekata, uz dodatno naznačene relativne udjele. Donji prikaz (b) predstavlja boxplot dijagram koji naglašava centralne tendencije i raspon vrijednosti

[1] Vrijednost u retku "Ukupno" za stupce "Slike" ne mora odgovarati zbroju vrijednosti tog stupca po klasama, budući da se na istoj slici može nalaziti više različitih objekata.

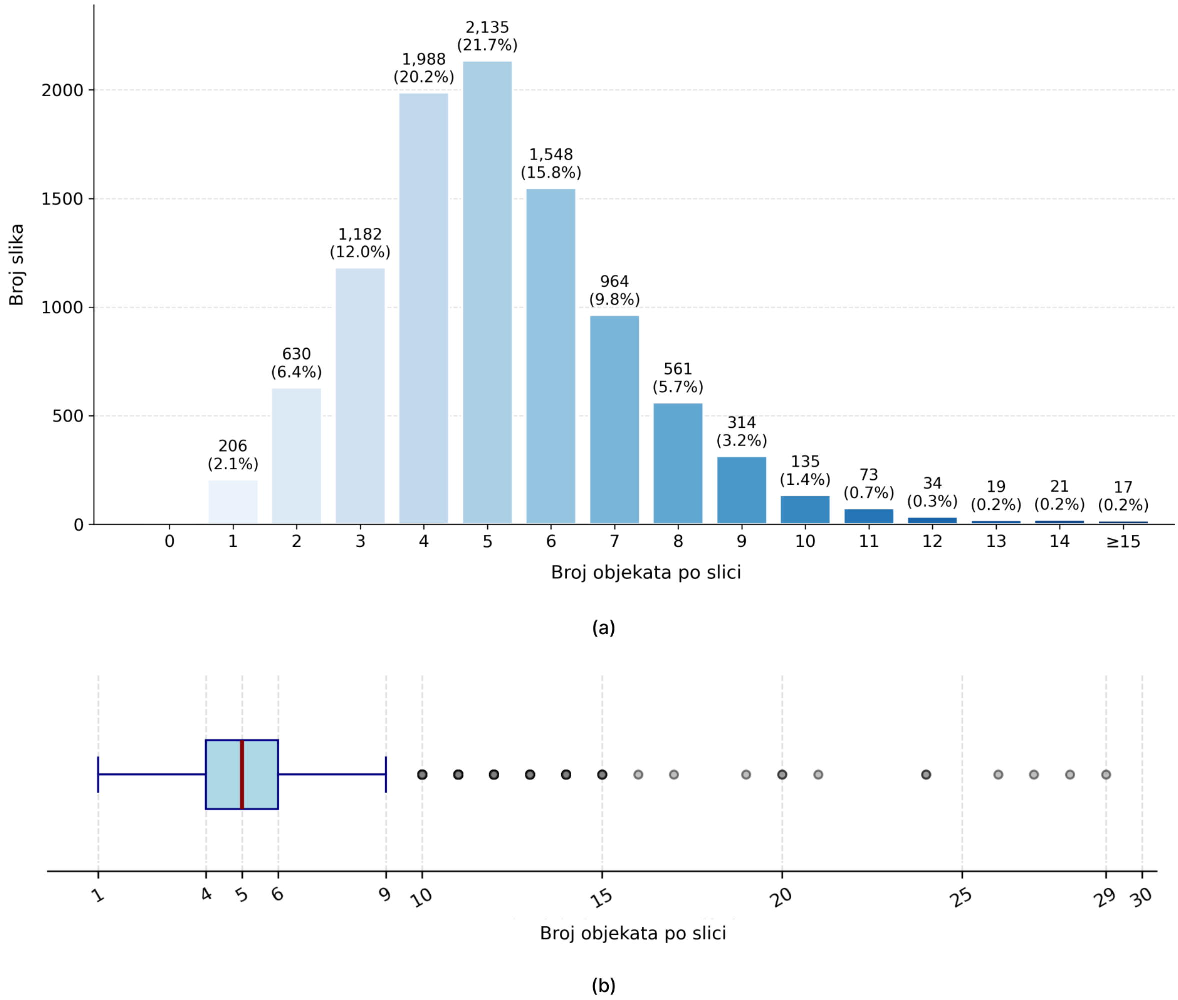


Slika 4.6: Distribucija broja objekata pojedinih klasa u SSMOT skupu za detekciju.

u danim podacima. Analiza distribucije pokazuje da se najveći udio slika u SSMOT detekcijskom skupu podataka (57.7%) nalazi u rasponu od četiri do šest objekata po slici. Nasuprot tome, slike s manje od dva objekta te one s više od deset objekata javljaju se znatno rjeđe, pri čemu njihov udio ne prelazi 5% ukupnog skupa. Slike s 10 objekata i više, njih 299 (3%), detektirane su kao skup *outliera* što ukazuje na to da se takvi primjeri pojavljuju rjeđe i odstupaju od dominantnog obrasca distribucije. Čak 59.8% slika sadrži 5 ili više objekata, a njih 22.3% 7 objekata ili više. U skupu podataka nema slika bez objekata; na svakoj slici prisutan je barem jedan brod, bilo privezan u luci ili u pokretu.

Usporedba distribucije broja objekata po slici u SSMOT i SPSCD skupovima podataka pokazuje na njihove značajne razlike. Više od polovice slika (55.2%) SPSCD skupa podataka sadrži samo jedan objekt, dok dodatnih 11.7% ne sadrži niti jedan objekt, što znači da gotovo dvije trećine skupa (66.9%) obuhvaća slike s najviše jednim objektom. Slike s dva broda čine 17.7%, a s tri broda 10.7% ukupnog skupa. Udio slika sa složenijim scenama, tj. s četiri ili više brodova, iznosi svega 4.6%, a tek je 26 slika s maksimalnim brojem objekata 7 (0.1%). Važno je istaknuti da u SPSCD skupu podataka nisu označavana privezana plovila unutar luke, već su uglavnom anotirana plovila u pokretu, a ona malo udaljenija često su izostavljena. Međutim, broj takvih neoznačenih plovila u SPSCD-u nije toliko velik da bi bio jedini uzrok velike razlike u distribuciji broja objekata po slici SPSCD i SSMOT skupa podataka. SSMOT se ističe znatno većim brojem objekata po slici te složenijim scenama, što ga čini pogodnijim za razvoj i evaluaciju detekcijskih modela u realnim uvjetima.

Za razliku od SPSCD skupa podataka, SSMOT skup obuhvaća i sva vidljiva plovila – ona u pokretu, kao i ona koja su privezana. Budući da vizualna obilježja plovila ostaju slična neovisno o tome kreću li se ili miruju, njihovo selektivno označavanje ne utječe samo



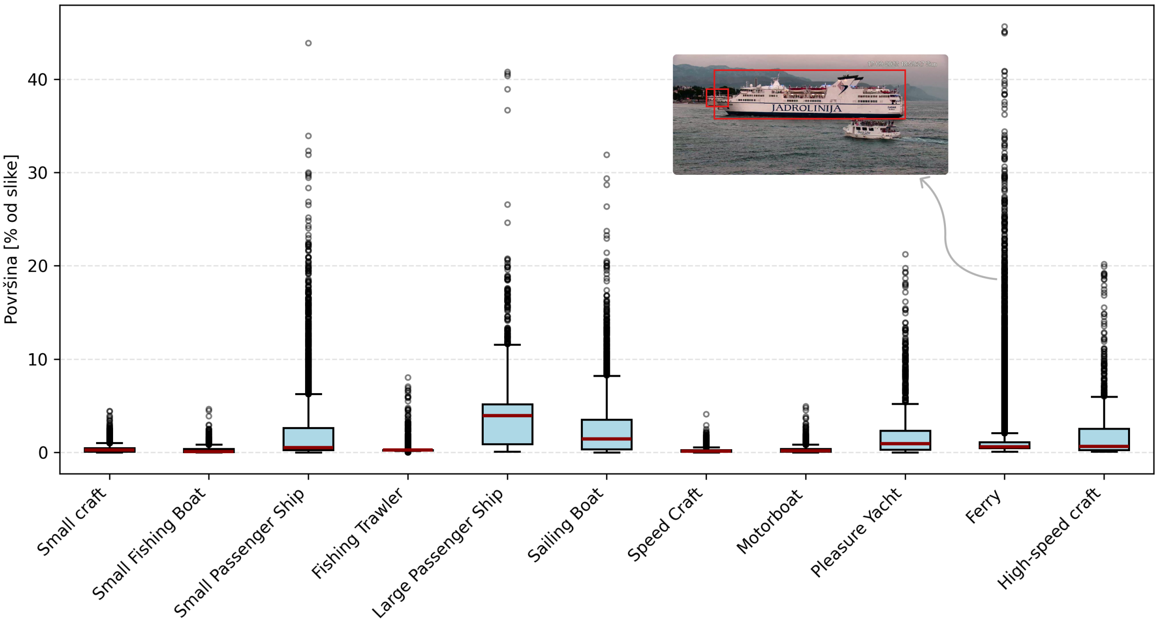
Slika 4.7: Distribucija broja objekata po slici u SSMOT detekcijskom skupu podataka prikazana stupčastim dijagramom (a) i boxplotom (b).

na smanjenje prosječnog broja objekata po slici, već može i zbunjivati detekcijske modele. Naime, kada je plovilo sa sličnim vizualnim karakteristikama u jednoj slici označeno, a u drugoj ne, otežava se proces učenja jasnih i konzistentnih značajki. Prednost SSMOT skupa podataka je u dosljednom označavanju svih plovila, koje čini osnovu za treniranje stabilnih modela sposobnih za pouzdanu detekciju u praktičnim uvjetima.

Veličine i proporcije objekata

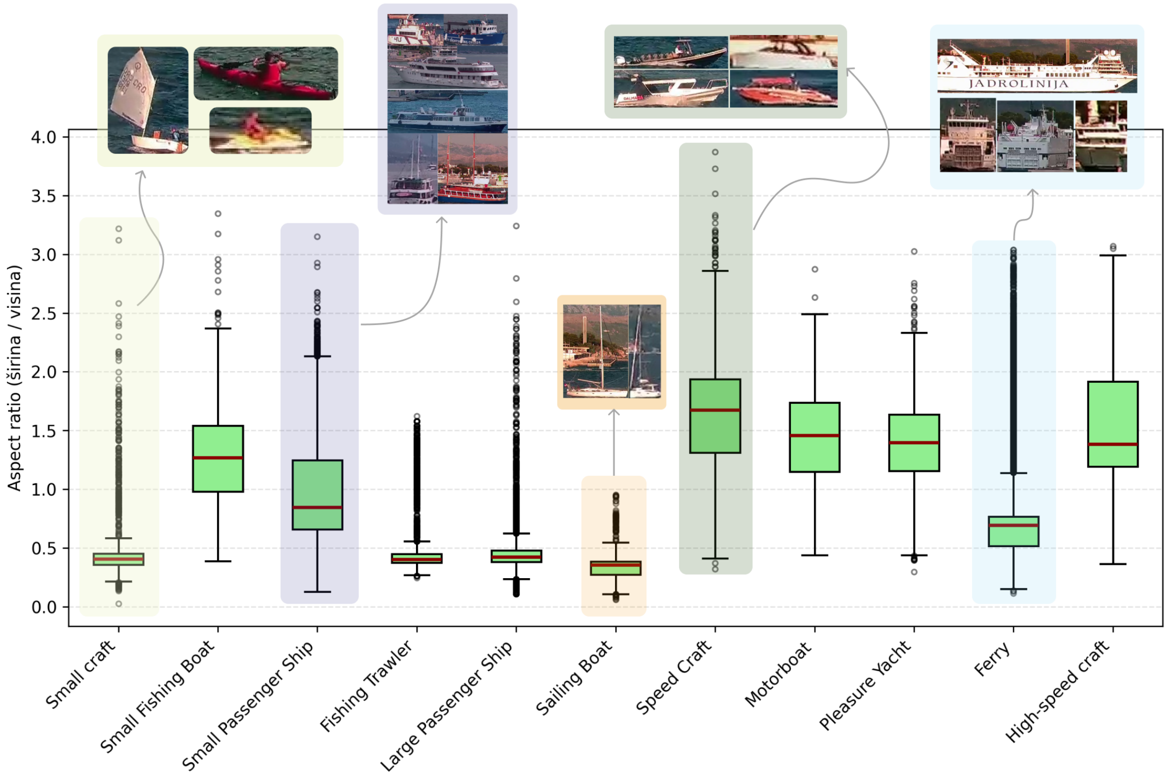
Analiza veličina i proporcija označenih objekata pruža važan uvid u karakteristike samog skupa podataka. Površine graničnih okvira otkrivaju koliki dio slike plovila obično zauzimaju, dok omjer stranica graničnog okvira (engl. *aspect ratio*) daje uvid u karakteristične oblike plovila koja se javljaju. Analiza je provedena po klasama kako bi se lakše uočili specifični obrasci karakteristični za pojedine kategorije plovila.

Na Slici 4.8 prikazana je distribucija **površina graničnih okvira** po klasama brodova, izraženih kao postotak ukupne površine slike. Klase poput *Small craft*, *Small Fishing Boat*, *Speed Craft* i *Motorboat* obuhvaćaju male objekte koji obično zauzimaju vrlo mali udio slike (treći kvartil Q_3 za sve klase manji od 0.45%), što je i očekivano budući da se tu generalno radi o manjim plovilima. S druge strane, kategorije koje sadrže veća plovila poput velikih trajekata (*Ferry*) i kruzera (*Large Passenger Ship*) imaju jako širok raspon vrijednosti, od primjera kada se javljaju kao dominantni objekti u kadru zauzimajući iznimno velik dio slike (preko 30%) do situacija kada su privezani ili se nalaze u daljini pa su im površine znatno manje. Navedeno je ilustrirano primjerom slike (na Slici 4.8) na kojoj se nalaze dva granična okvira klase *Ferry* bitno različitih površina.



Slika 4.8: Box-plot dijagrami površina graničnih okvira po klasama.

S druge strane, Slika 4.9 pruža pregled raspodjele **omjera stranica** (širine i visine) **graničnih okvira** po klasama plovila. Može se uočiti da pojedine klase imaju karakteristične proporcije. Primjerice, granični okviri plovila klase *Sailing Boat* su uvijek vertikalno izduženi zbog karakterističnog jarbola ($Q_3 = 0.39$), dok kod klase *Speed Craft* prevladavaju horizontalno izduženi granični okviri ($Q_1 = 1.31$). Vidljiva je i varijabilnost oblika graničnih okvira unutar pojedinih klasa, koja proizlazi iz raznolikosti izgleda plovila u različitim položajima, kao i heterogenosti plovila unutar same kategorije. Primjerice, unutar klase *Small Craft* prevladavaju male trening jedrilice čiji jarboli uzrokuju okomito izdužene okvire, dok se istodobno pojavljuju i drugi tipovi plovila, poput kanua, kod kojih su granični okviri izrazito horizontalno izduženi. Slična raznolikost ($Q_1 = 0.66$, $Q_3 = 1.25$) prisutna je i u klasi *Small Passenger Ship*, koja obuhvaća širok spektar putničkih brodova, od manjih plovila do većih turističkih jedrenjaka i brodova namijenjenih kraćim kružnim turama. Kod klasa poput *Ferry*, *Large Passenger Ship* i *Fishing Trawler* učestalo se pojavljuju primjerci privezani uz obalu, zbog čega je prisutan znatan broj graničnih okvira s omjerom stranica manjim od 1.



Slika 4.9: Box-plot dijagram omjera širine i visine graničnih okvira po klasama.

Velika raznolikost veličina objekata, kako među klasama tako i unutar iste klase, zajedno s varijabilnošću njihovih proporcija, čini skup podataka reprezentativnijim, ali istodobno predstavlja i veći izazov za detekcijske modele koji moraju jednako pouzdano prepoznavati vrlo male i izrazito velike objekte različitih oblika.

Vizualna raznolikost i izazovni scenariji

Kako bi se dodatno prikazala raznolikost i složenost SSMOT skupa podataka za detekciju, u nastavku su prikazani reprezentativni primjeri slika koje obuhvaćaju različite vremenske i morske uvjete, razine osvijetljenosti te situacije djelomične zaklonjenosti i međusobnog mimoilaženja brodova. Uvrštavanjem ovakvih primjera osigurava se da detekcijski model ne uči isključivo iz idealnih situacija, nego i iz složenih scenarija. Time se razvija sposobnost generalizacije detekcijskog modela i njegova otpornost na različite uvjete i okolnosti koje odstupaju od uobičajenih, poput maglovita vremena ili djelomične zaklonjenosti plovila koje je potrebno detektirati.

Primjeri prikazani na Slici 4.10 ukazuju na to da skup podataka SSMOT obuhvaća raznolike vremenske i okolišne uvjete. Uključene su scene snimljene pri sunčanom i oblačnom vremenu, kao i u uvjetima magle i kiše, uz različite razine osvijetljenosti – od ranih jutarnjih sati, preko dnevnih i sumračnih scena, do večernjih uvjeta. Na pojedinim primjerima uočljivi su i valovi te odsjaji sunčeve svjetlosti na površini mora.

Slika 4.10: Primjeri različitih uvjeta i razina osvijetljenosti u SSMOT skupu podataka za detekciju.

Na Slici 4.11 prikazane su situacije djelomične zaklonjenosti i međusobnog mimoilaženja plovila, koje dodatno doprinose složenosti skupa podataka. Ove raznolikosti čine skup podataka zahtjevnijim, ali istovremeno omogućuju treniranje detekcijskih modela veće robusnosti koji će biti pouzdani i u stvarnim uvjetima primjene.

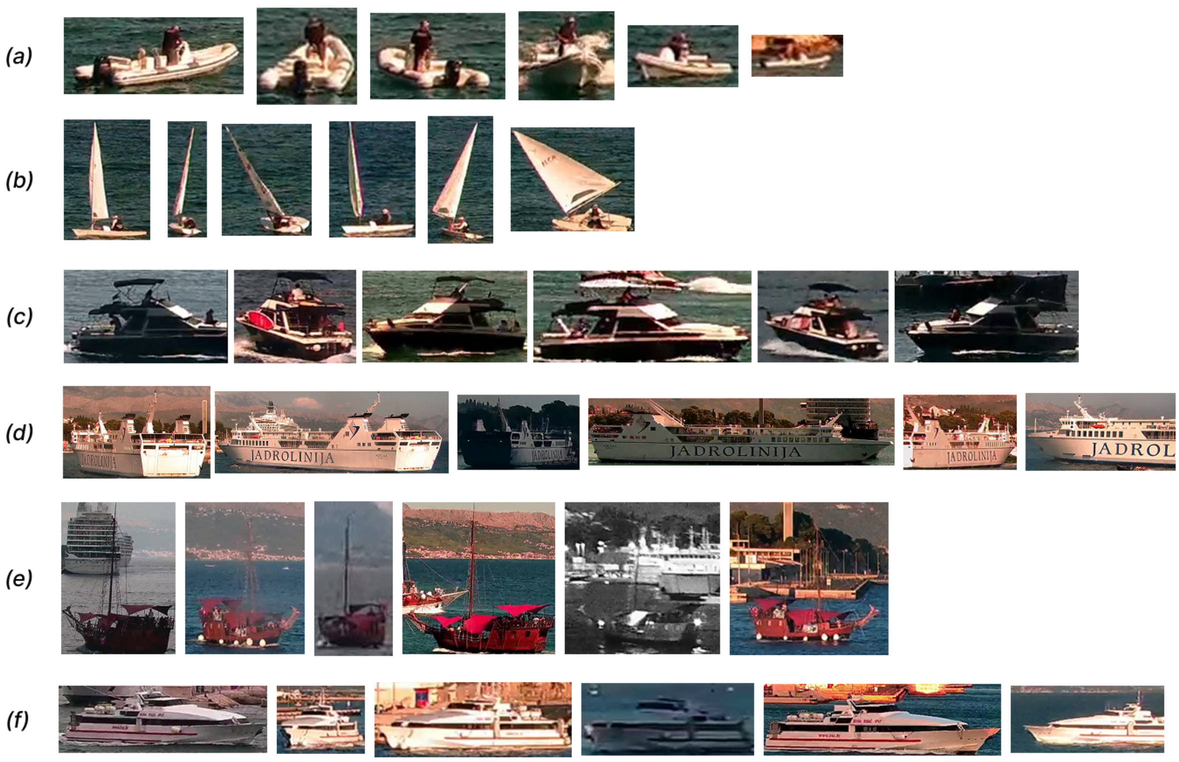
Slika 4.11: Primjeri složenijih scena s djelomično zaklonjenim plovilima koja se mimoilaze.

4.3. ReID skup podataka

Reidentifikacija plovila čini osnovu naprednih sustava za praćenje, budući da omogućava prepoznavanje i povezivanje istog objekta kroz vremenski razdvojene okvire videozapisa, čak i u situacijama privremenog izostanka detekcije uslijed pogreške detektora ili okluzije. Za razliku od skupa za detekciju, koji se sastoji od pojedinačnih okvira videozapisa s označenim lokacijama i klasama plovila, skup podataka za reidentifikaciju sadrži slike istih plovila prikazanih u različitim okolnostima - pri promjenama položaja, udaljenosti od kamere, uvjeta osvjetljenja i snimanja. Ovakva struktura podataka omogućuje učenje diskriminativnih vizualnih obilježja plovila, koja ostaju relativno nepromijenjena unatoč varijacijama u njihovom izgledu i vanjskim okolnostima snimanja.

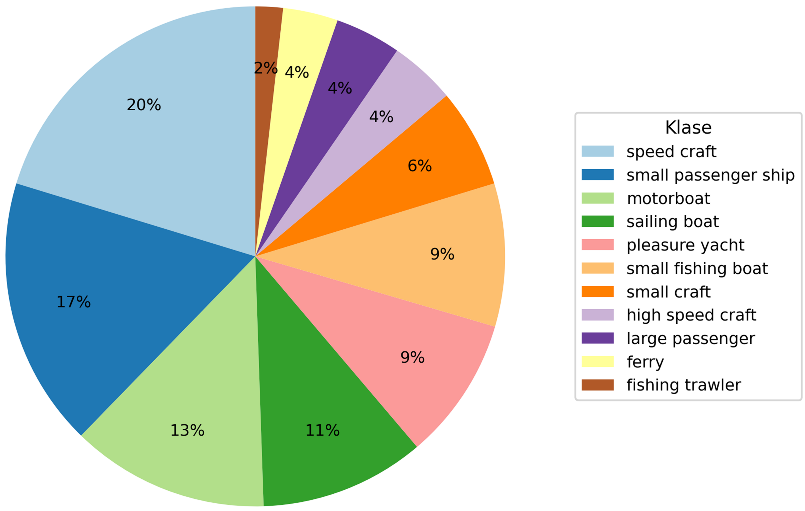
4.3.1. Osnovne karakteristike

SSMOT skup podataka za reidentifikaciju, razvijen za specifičan problem praćenja plovila u splitskoj luci, obuhvaća ukupno **281 različito plovilo** (identitet). Za svako plovilo prikupljeno je najmanje osam slika koje ga prikazuju u različitim položajima i/ili su snimljene u različitim vremenskim razdobljima. Kod nekih plovila, poput onih na redovnim trajektnim i turističkim linijama, bilo je moguće prikupiti uzorke snimljene u različitim vremenskim razdobljima i pod različitim uvjetima. Nasuprot tome, kod manjih privatnih brodica takav pristup nije bio uvijek izvediv, pa su varijacije uglavnom ostvarene primjerima plovila u različitim položajima. Slika 4.12 prikazuje uzorke slika šest različitih identiteta iz SSMOT skupa za reidentifikaciju.



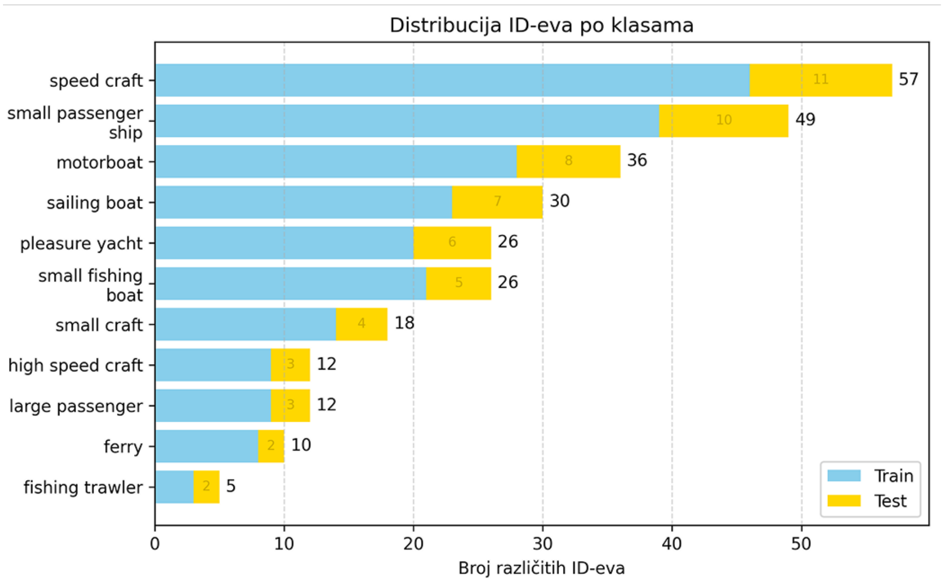
Slika 4.12: Primjeri slika šest identiteta iz SSMOT ReID skupa podataka.

Na Slici 4.13 prikazan je kružni dijagram koji ilustrira raspodjelu pojedinih klasa plovila u SSMOT skupu podataka za reidentifikaciju. Nasuprot detekcijskom skupu podataka, u kojem prevladavaju klase poput *Ferry*, čija je visoka zastupljenost rezultat same strukture prometa u splitskoj luci - učestalog prometovanja redovnih trajektnih linija te čestog boravka trajekata privezanih u luci - u ReID skupu podataka zastupljenije su klase poput *Speed Craft*, *Small Passenger Ship* i *Motorboat*, koje karakterizira izraženija unutar-klasna varijabilnost. Veća zastupljenost takvih klasa u ReID skupu podataka je iznimno važna jer omogućuje modelu da razlikuje pojedinačna plovila iste kategorije koja dijele slična vizualna obilježja, što predstavlja temelj za pouzdano očuvanje identiteta plovila tijekom praćenja.



Slika 4.13: Kružni dijagram udjela pojedinih klasa u identitetima SSMOT ReID skupa podataka.

Od ukupno 281 identiteta u ReID skupu podataka, njih 61 je izdvojeno za testiranje, dok je preostalih 220 namijenjeno treniranju reidentifikacijskih modela. U oba podskupa zastupljene su sve klase plovila, što je ilustrirano na Slici 4.14.



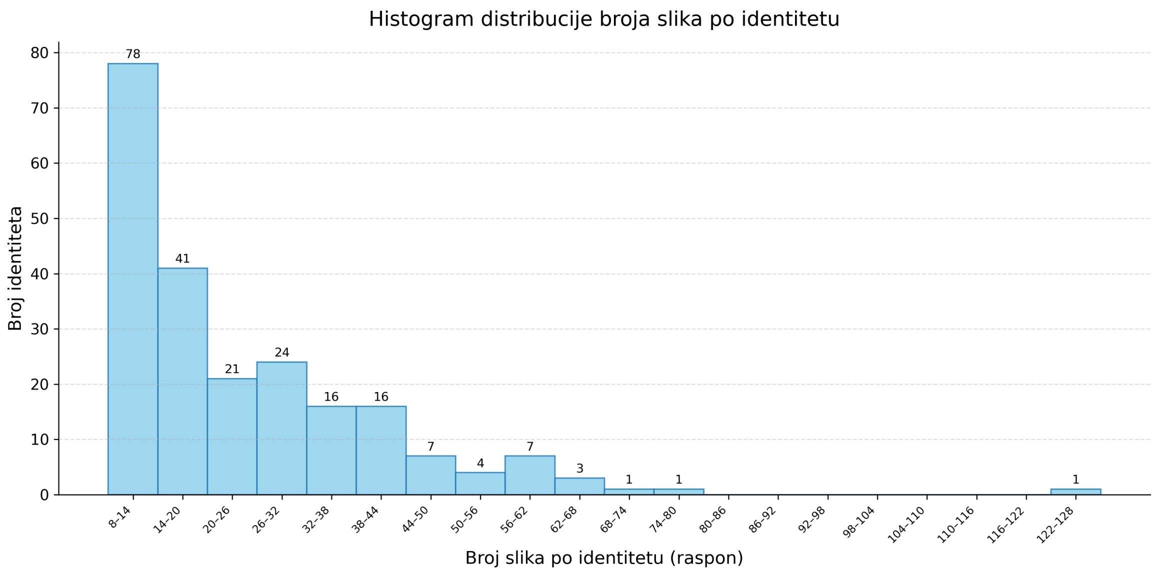
Slika 4.14: Distribucija broja identiteta SSMOT ReID skupa podataka po klasama s podjelom na skup za treniranje i testiranje.

4.3.2. Struktura ReID skupa podataka

SSMOT skup podataka za reidentifikaciju podijeljen je na tri particije standardne za ReID skupove podataka [260, 261, 262, 263], a to su: skup za učenje, skup upita i galerija.

Skup za učenje (engl. *train set*) koristi se za treniranje ReID modela i sadrži slike identiteta namijenjene isključivo procesu učenja. S druge strane, za evaluaciju performansi dobivenih modela, koriste se skup upita i galerija. **Skup upita** (engl. *query set*) najčešće se sastoji od jedne do dvije slike za svaki testni identitet. Cilj ReID modela je za svaku sliku iz skupa upita pronaći odgovarajuće slike istog identiteta unutar galerije. **Galerija** (engl. *gallery*) predstavlja pretraživački prostor slika u kojem se traže pozitivni primjeri za zadane upite. Ona sadrži veći broj slika svakog testnog identiteta, kao i distraktore - vizualno slične primjere identiteta koji se ne nalaze među upitima. Dodavanjem distraktora u galeriju modelima se dodatno otežava zadatak reidentifikacije, a sama procjena performansi modela postaje realističnija i pouzdanija jer je bliža stvarnim uvjetima primjene.

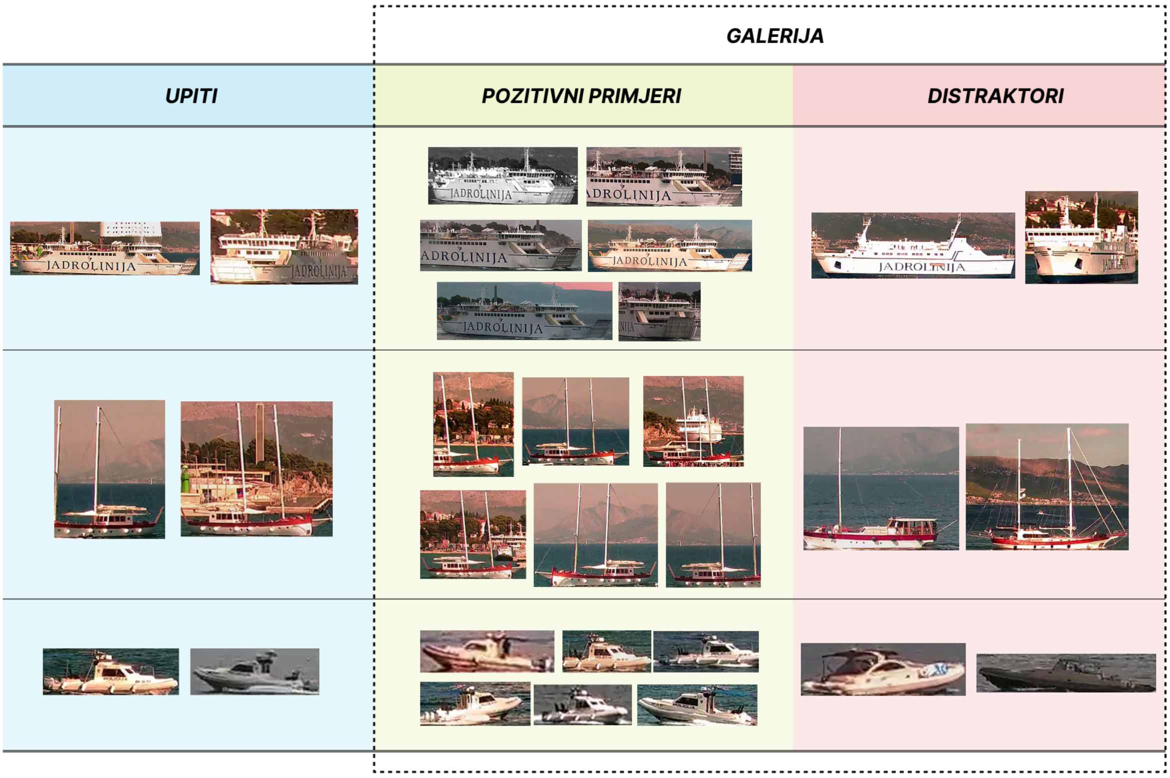
SSMOT ReID skup podataka konkretno sadrži: 1) **5227 slika** u skupu za treniranje - minimalno osam primjera za svaki od identiteta namijenjenih za učenje ReID modela; 2) **122 slike** u skupu upita - po dvije slike za svaki testni identitet; 3) **488 slika** u galeriji - šest pozitivnih primjera i dva distraktora za svaki testni identitet. U skupu za treniranje broj primjera po identitetu varira: od minimalnih 8 do maksimalnih 128, pri čemu većina identiteta (54%) ima između 8 i 20 slika. Distribucija broja primjera po identitetu u ReID skupu za treniranje ilustrirana je na Slici 4.15.



Slika 4.15: Histogram distribucije broja slika po identitetu u ReID skupu za treniranje.

Korištenje većeg broja upita po identitetu u evaluaciji reidentifikacijskog modela može doprinijeti pouzdanijoj procjeni njegovih performansi. Odabirom dviju slika po testnom identitetu u skupu upita evaluacija postaje pouzdanija nego kad se koristi samo jedno, a pritom se ne uvodi značajno dodatno opterećenje. Na taj je način moguće provjeriti koliko je model stabilan u pronalasku istog identiteta kada se koriste njegovi različiti vizualni prikazi. Upravo takav pristup, s dvjema slikama po identitetu, koristi se i u [263, 264]. Kako bi se ReID modelima dodatno otežao postupak pretrage galerije s ciljem pronalaska pozitivnih primjera, za svaki testni identitet u galeriju su uključena i dva distraktora koja vizualno nalikuju slikama upita. Na Slici 4.16 prikazani su primjeri distraktora za tri testna identiteta.

Uz osnovnu galeriju, kao "teži" scenarij, promatrana je i proširena galerija s 1138 slika, koja je zbog ograničenog broja dostupnih identiteta proširena dodavanjem slika iz skupa za učenje te dodatnih distraktora. Iako idealno rješenje podrazumijeva korištenje primjera identiteta koji nisu prisutni ni u skupu za učenje ni u skupu za testiranje, uključivanje uzoraka iz skupa za učenje ipak je prihvatljivo jer ne dolazi do preklapanja s testnim identitetima, a istovremeno se povećava raznolikost i veličina galerije. Na taj način evaluacija postaje zahtjevnija i realističnija, budući da bolje odražava uvjete u kojima je potrebno pronaći traženi identitet u znatno većem prostoru pretraživanja.



Slika 4.16: Primjeri: upit - pozitivni primjeri + distraktori.

4.4. Skup videozapisa za praćenje plovila

Skup videozapisa završna je i ključna komponenta SSMOT skupa podataka, namijenjena istraživanju i evaluaciji metoda za praćenje plovila u stvarnim pomorskim uvjetima. Za razliku od statičnih slika koje su korištene u detekcijskom podskupu, videozapisi omogućuju analizu vremenske povezanosti između uzastopnih okvira te praćenje dinamičkih promjena u kretanju plovila kroz vrijeme. Osim samog kretanja, videozapisima se mogu proučavati i međusobni prostorni odnosi plovila u vidnom polju, poput njihovog približavanja, preklapanja i privremene zaklonjenosti, odnosno pojave okluzije.

4.4.1. Način označavanja podataka

Svi videozapisi unutar SSMOT skupa podataka anotirani su korištenjem alata *Computer Vision Annotation Tool* (*CVAT*) [259]. Proces anotiranja obavljen je ručno, uz naknadnu provjeru i korekciju, kako bi se osigurala točnost i konzistentnost anotacija. Korišten je *MOT 1.1.* (*Multiple Object Tracking*) format oznaka, koji omogućuje praćenje svih instanci plovila kroz uzastopne okvire videozapisa.

U odnosu na detekcijski podskup, u kojem se svaka slika označava zasebno, MOT anotacije zahtijevaju očuvanje identiteta svih objekata kroz okvire videozapisa. Svakom plovilu pri prvom pojavljivanju dodjeljuje se jedinstveni identifikator (*ID*) koji ostaje nepromijenjen tijekom cijelog razdoblja njegova praćenja, uključujući i vrijeme djelomične ili potpune okluzije. Takav pristup omogućuje kontinuirano i dosljedno praćenje objekata kroz cijeli videozapis. Kategorije plovila u MOT skupu usklađene su s onima korištenima u detekcijskom podskupu, kako bi se osigurala konzistentnost između različitih dijelova SSMOT skupa podataka. Sve klase koriste iste nazive i oznake kao u detekcijskom dijelu (0 - *Small Craft*, 1 - *Small Fishing Boat*, 2 - *Small Passenger Ship*, itd.), čime je omogućena njihova zajednička i međusobno kompatibilna uporaba.

Svaki videozapis ima odgovarajuću tekstualnu datoteku s anotacijama u sljedećem formatu:

```
frame, id, x, y, width, height, confidence, class, visibility
```

gdje svaki redak predstavlja jednu instancu plovila u određenom okviru videozapisa. Parametar `frame` označava redni broj okvira u kojem se plovilo pojavljuje, dok `id` predstavlja njegov jedinstveni identifikator koji ostaje nepromijenjen kroz sve okvire u kojima je plovilo prisutno. Koordinate `x` i `y` definiraju položaj gornjeg lijevog kuta graničnog okvira plovila, a `width` i `height` njegovu širinu i visinu izraženu u pikselima. Vrijednost `confidence` označava razinu pouzdanosti detekcije (u *ground-truth* oznakama obično postavljena na 1), dok se vrijednost `class` odnosi na pripadajuću kategoriju plovila. Zadnji parametar `visibility` koristi se kao binarna oznaka vidljivosti, pri čemu vrijednost `0` označava potpuno zaklonjen, a vrijednost 1 vidljiv objekt.

Primjer isječka iz anotacijske datoteke prikazan je na Slici 4.17. Prikazani redci odgovaraju zapisima o tri različita plovila unutar prva dva okvira videozapisa, gdje svaka linija sadrži njihove prostorne koordinate, dimenzije graničnih okvira, te informacije o pripadajućoj klasi i trenutnoj vidljivosti objekta.

```
1,1,160.6,165.15,115.70,84.13,1,10,1
1,2,598.62,145.75,276.77,124.46,1,10,1
1,3,1052.76,223.45,10.60,6.30,1,8,1
2,1,160.6,165.15,115.70,84.13,1,10,1
2,2,599.15,145.77,276.61,124.45,1,10,1
2,3,1052.84,223.45,10.60,6.30,1,8,0
```

Slika 4.17: Primjer formata MOT 1.1 oznaka preuzetih iz CVAT-a.

4.4.2. Opće karakteristike SSMOT videozapisa

SSMOT skup za praćenje obuhvaća videozapise snimljene u različitim uvjetima osvjetljenja i vidljivosti, tijekom različitih doba dana. U njemu su zastupljene jednostavnije scene s manjim brojem plovila u vidnom polju, kao i složenije sekvence s pojačanim prometom te međusobnim preklapanjima i mimoilaženjima plovila. Primjeri reprezentativnih kadrova SSMOT videozapisa prikazani su na Slici 4.18. Kadrovi različitih razina osvjetljenja i broja objekata prikazani su u prvom retku na Slici 4.18, dok su u drugom retku ilustrirane situacije međusobnog mimoilaženja i preklapanja plovila. Skup također obuhvaća izazovne slučajeve u kojima je potrebno istodobno pratiti više vizualno sličnih objekata (primjer na Slici 4.19), kao i situacije djelomičnih i potpunih okluzija te ulazaka i izlazaka objekata iz vidnog polja.

Slika 4.18: Primjeri scena iz SSMOT skupa videozapisa.

Skup uključuje ukupno 18 videozapisa različitog trajanja, snimljenih u stvarnim obalnim uvjetima u kolovozu i rujnu 2023. godine pomoću nadzorne kamere opisane u potpoglavlju 4.1. Većina videozapisa u skupu ima frekvenciju od 25 okvira u sekundi (FPS), dok se manji broj bilježi pri 10 ili 30 FPS. Ukupno trajanje videozapisa iznosi 9 minuta i 34 sekunde, što odgovara 12946 pojedinačnih okvira. U njima se javlja 149 različitih plovila, s

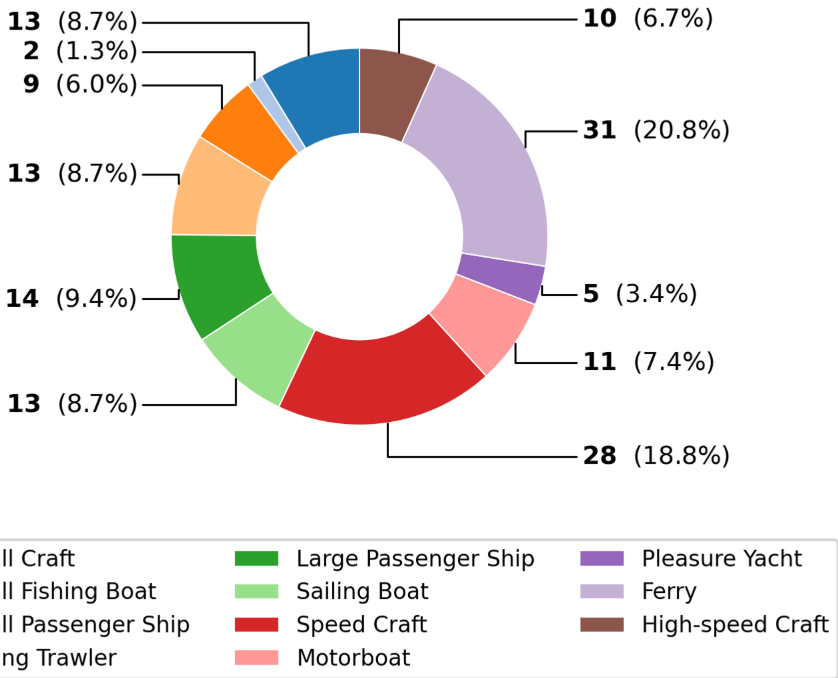
Slika 4.19: Primjer videozapisa (SSMOT_12) u kojem je potrebno istovremeno pratiti veći broj vizualno sličnih plovila koja se međusobno preklapaju i mimoilaze.

ukupno 91006 graničnih okvira. Videozapisi su pohranjeni u `.mp4` formatu s razlučivošću 1280×720 piksela. Detaljan pregled karakteristika pojedinih videozapisa prikazan je u Tablici 4.3. Tablica prikazuje osnovne tehničke specifikacije svakog videozapisa — njegovu duljinu, broj okvira i anotiranih putanja plovila, kao i ukupan broj graničnih okvira. Stupac *Gustoća* prikazuje prosječan broj označenih plovila po okviru videozapisa, čime se procjenjuje razina složenosti pojedine scene. Vrijednost u retku *Ukupno* predstavlja prosječnu gustoću izračunatu za čitav skup, odnosno omjer ukupnog broja graničnih okvira i ukupnog broja okvira svih videozapisa.

Tablica 4.3: Osnovne karakteristike videozapisa u SSMOT skupu za praćenje.

Naziv	**FPS**	**Broj okvira**	**Duljina**	**Putanje**	**Granični okviri**	**Gustoća**
SSMOT_1	25	287	0:11	5	1435	5.0
SSMOT_2	30	604	0:20	9	4415	7.3
SSMOT_3	25	1037	0:41	7	7049	6.8
SSMOT_4	25	370	0:14	8	2850	7.7
SSMOT_5	25	433	0:17	7	2901	6.7
SSMOT_6	25	856	0:34	11	7990	9.3
SSMOT_7	25	451	0:18	4	1804	4.0
SSMOT_8	25	1697	1:07	11	10494	6.2
SSMOT_9	25	922	0:36	8	6357	6.9
SSMOT_10	25	836	0:33	12	6571	7.9
SSMOT_11	25	277	0:11	12	3185	11.5
SSMOT_12	25	928	0:37	10	6930	7.5
SSMOT_13	10	646	1:04	7	3647	5.6
SSMOT_14	10	598	0:59	9	5072	8.5
SSMOT_15	25	585	0:23	7	4095	7.0
SSMOT_16	25	940	0:37	10	8742	9.3
SSMOT_17	30	941	0:31	4	3764	4.0
SSMOT_18	25	538	0:21	8	3705	6.9
Ukupno		12946	9:34	149	91006	7.0

Videozapisi obuhvaćaju svih jedanaest kategorija plovila, pri čemu je raspodjela putanja po kategorijama prikazana na Slici 4.20. Detaljni podaci o broju plovila svake kategorije u pojedinim videozapisima dostupni su u Dodatku B. Iako su pojedine klase, poput *Small Fishing Boat*, slabije zastupljene, to ne predstavlja značajno ograničenje ovog skupa podataka, čija je osnovna svrha ispitati konzistentnost praćenja i zadržavanja identiteta objekata kroz videozapis. Sama klasifikacija plovila temelji se na predviđanjima detektora, a evaluacija performansi detektora provodi se na zasebno pripremljenom testnom podskupu SSMOT skupa za detekciju.



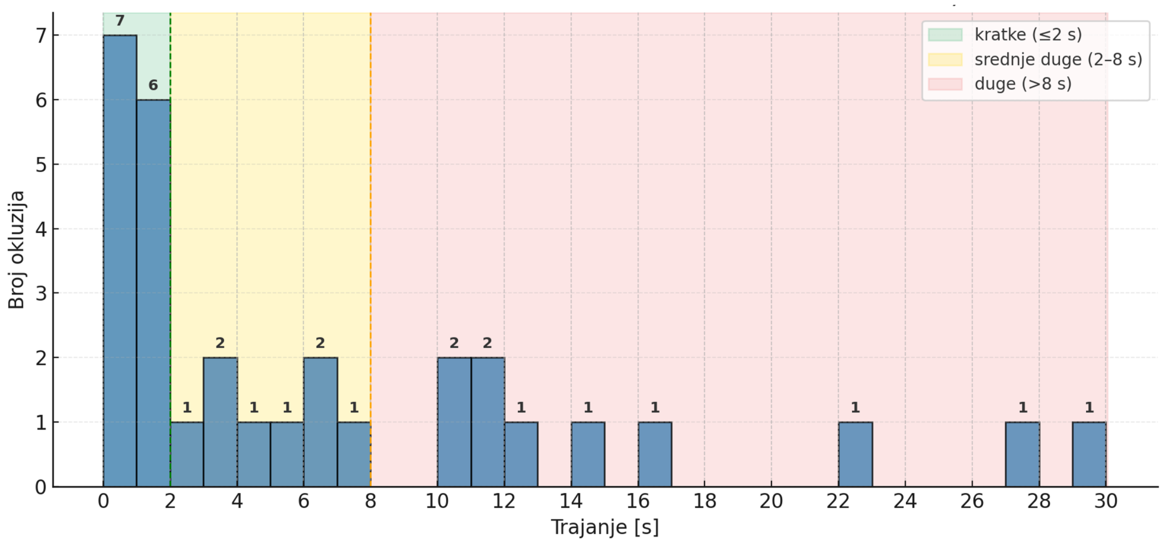
Slika 4.20: Zastupljenost pojedinih kategorija plovila u putanjama SSMOT videozapisa.

4.4.3. Okluzije u videozapisima

Okluzije plovila u videozapisima mogu biti djelomične ili potpune. Djelomične okluzije javljaju se kada je plovilo samo dijelom prekriveno drugim plovilom ili elementom scene, dok potpune okluzije označavaju situacije u kojima je plovilo u potpunosti zaklonjeno određeni broj uzastopnih okvira. U daljnjem tekstu pojam *okluzije* odnosi se upravo na situacije potpune zaklonjenosti u kojima plovilo privremeno nije vidljivo u kadru te se potom se ponovno pojavljuje. Takvi su primjeri osobito važni za procjenu sposobnosti algoritama praćenja da zadrže identitet objekta unatoč privremenom prekidu vidljivosti. Stoga je pri konstrukciji SSMOT skupa videozapisa posebna pažnja posvećena uključivanju dovoljnog broja primjera potpunih okluzija, kako bi se osigurali uvjeti za realniju procjenu robusnosti algoritma u složenim scenarijama zaklonjenosti.

Okluzije se dodatno mogu razvrstati na kratkotrajne i dugotrajne, ovisno o trajanju razdoblja u kojem je objekt zaklonjen. Kratkotrajne okluzije obično traju svega nekoliko uzastopnih okvira i predstavljaju manji izazov za algoritme praćenja, dok dugotrajne okluzije, koje mogu potrajati i nekoliko sekundi, značajno povećavaju rizik od gubitka identiteta objekta i pogrešne reasocijacije nakon ponovnog pojavljivanja U postojećoj literaturi ne postoji kriterij koji jednoznačno definira granicu između kratkotrajnih i dugotrajnih okluzija. Predloženi pragovi značajno se razlikuju među autorima - od pragova od 20 [25] i 30 [265] okvira, do vremenskih pragova od 2 s [23] i 5 s [24].

U okviru SSMOT skupa videozapisa evidentirana je ukupno 31 okluzija. Histogram njihovog trajanja prikazan je na Slici 4.21. Budući da skup obuhvaća videozapise različitih frekvencija okvira u sekundi (FPS), trajanje okluzija izraženo je u sekundama. Od ukupnog broja, sedam okluzija traje do 1 s, pri čemu najkraća traje 0.4 s (10 okvira). Četrnaest okluzija (45.16%) traje dulje od 5 s, dok su tri dulje od 20 s, pri čemu najdulja traje 29.05 s. Detaljan popis svih zabilježenih okluzija, s pripadajućim informacijama o videozapisu, ID-u objekta, kategoriji, početnom i završnom okviru okluzije te njenom trajanju u okvirima i sekundama, nalazi se u Dodatku B.



Slika 4.21: Histogram trajanja potpunih okluzija u SSMOT videozapisima.

Okluzije se u ovom radu, s obzirom na trajanje izraženo u sekundama, dijele na **kratke**, **srednje duge** i **duge**. Kratkima se smatraju okluzije koje traju najviše 2 s ($\leq$ 2 s), srednje dugima one čije je trajanje dulje od 2 s, ali ne prelazi 8 s ($>$ 2 s, $\leq$ 8 s), dok se dugima smatraju okluzije dulje od 8 s ($>$ 8 s). U skladu s definiranom klasifikacijom, u SSMOT skupu videozapisa zabilježeno je **13** (41.9%) kratkih, **8** (25.8%) srednje dugih i **10** (32.3%) dugih okluzija. Slika 4.22 prikazuje po jedan primjer svake kategorije okluzije: na vrhu je prikazana kratka okluzija iz *SSMOT_16* videozapisa koja traje tek 0.4 s, u sredini se nalazi srednje duga okluzija u trajanju od 2.08 s iz *SSMOT_5* videozapisa, a na dnu duga okluzija iz *SSMOT_14* videozapisa u kojoj trajekt zaklanja jedrilicu 29.05 s.

Slika 4.22: Primjeri okluzija različitih trajanja iz SSMOT skupa podataka.

5. MODEL ZA DETEKCIJU PLOVILA

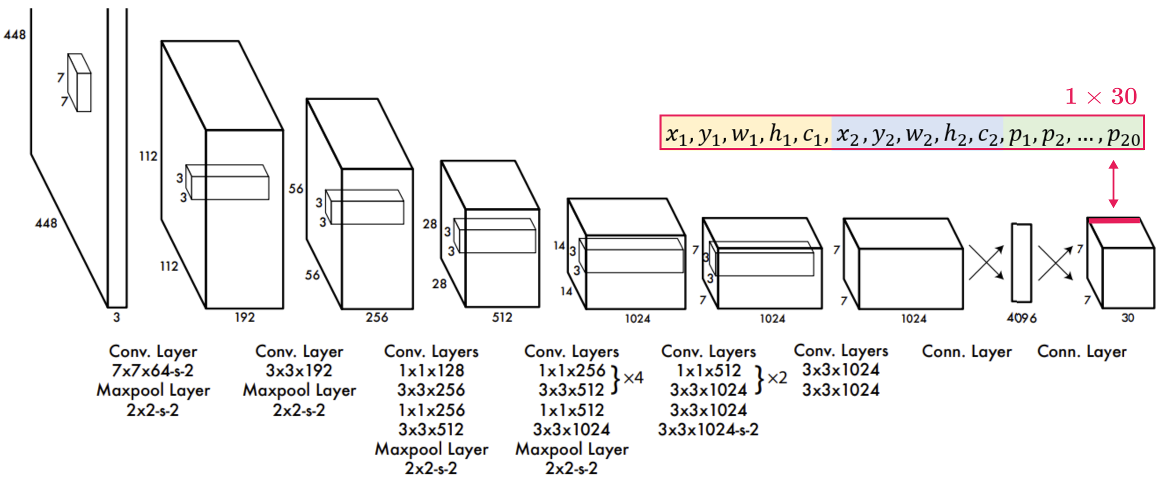
Detekcija plovila ključan je korak modernih algoritama za praćenje i nadzor pomorskog prometa, koji značajno utječe na njihove konačne performanse. Zbog svojih ograničenja, tradicionalni pristupi detekciji objekata temeljeni na pomičnim prozorima i ručno dizajniranim značajkama poput HOG [266], SIFT [267], SURF [268] i Haarovih značajki [269], gotovo su u potpunosti zamijenjeni detektorima dubokog učenja, koji zahvaljujući automatskom učenju robusnih značajki i superiornim performansama danas predstavljaju standard [270, 271, 73]. Među detektorima temeljenima na dubokom učenju posebno se ističu detektori iz YOLO familije [88, 272, 273, 274, 275, 276, 277, 278, 279, 280], koji su zahvaljujući kombinaciji brzine i točnosti postali vodećim odabirom u mnogim aplikacijama za detekciju objekata u stvarnom vremenu.

U ovom poglavlju opisan je model razvijen za automatsku detekciju plovila, koji se temelji na **YOLO11** arhitekturi. U nastavku je prikazan postupak razvoja, treniranja i evaluacije YOLO11 detektora za prepoznavanje plovila. Prvo je opisana sama arhitektura YOLO modela, s naglaskom na prednosti korištene verzije. Zatim je detaljno opisan proces treniranja modela, uključujući pripremu podataka, specifikacije okruženja na kojem se model trenirao, kao i vrijednosti korištenih hiperparametara. Na kraju su prikazani i analizirani rezultati evaluacije dobivenog modela, pri čemu su korištene standardne metrike poput preciznosti, odziva i srednje prosječne preciznosti (mAP).

5.1. YOLO detektor

Joseph Redmon i suradnici [88] predstavili su 2016. godine prvu verziju ***"You Only Look Once" (YOLO)*** algoritma, koja je revolucionirala pristup detekciji objekata koristeći samo jedan prolaz kroz neuronsku mrežu za cijeli proces detekcije. YOLO dijeli ulaznu sliku na mrežu od $S \times S$ ćelija, pri čemu se za svaku ćeliju predviđa B graničnih okvira i C vjerojatnosti klasa. Za svaki granični okvir, model predviđa koordinate središta (x, y), širinu i visinu (w, h) te sigurnost (p_c) da okvir sadrži objekt. Izlaz mreže je $S \times S \times (B \cdot 5 + C)$ volumen, a arhitektura, inspirirana GoogleLeNet-om [281], prikazana je na Slici 5.1.

Prednosti predloženog YOLO detektora uključuju jednostavnu implementaciju, brzu detekciju u stvarnom vremenu, implicitno kodiranje kontekstualnih i vizualnih informacija analizom cijele slike te *end-to-end* optimizaciju. Međutim, ovakav YOLO detektor ima odre-



Slika 5.1: Arhitektura originalnog YOLO detektora: $S = 7$, $B = 2$, $C = 20$. (Slika preuzeta iz [88], uz izmjene.)

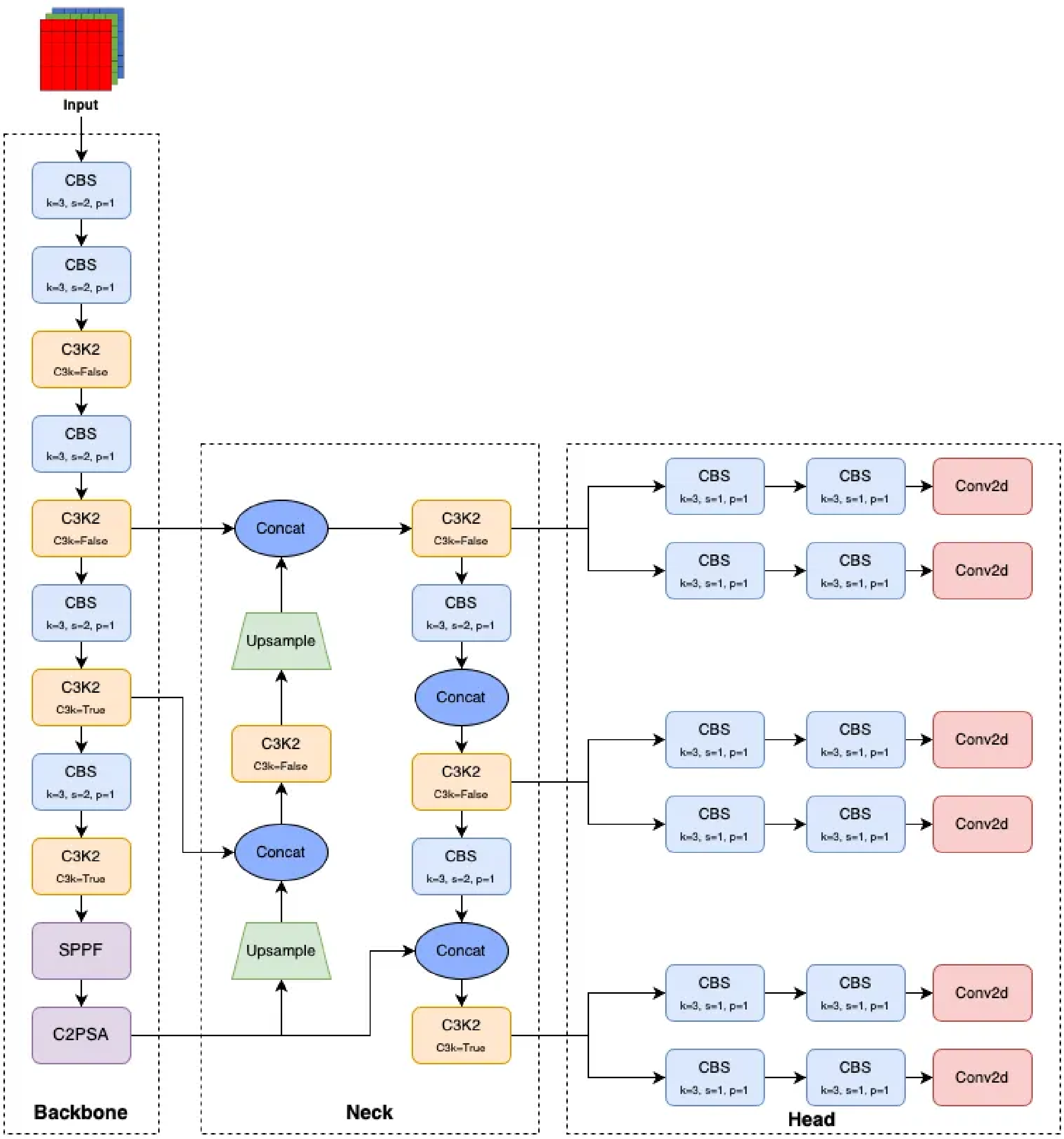
đena ograničenja u odnosu na druge *state-of-the-art* detektore: manje je precizan u lokalizaciji objekata, ima poteškoće s detekcijom manjih i blisko grupiranih objekata, a svaka ćelija može predvidjeti samo dva granična okvira ($B = 2$) i jednu klasu.

YOLOv2 [272], također poznat kao YOLO9000, proširuje mogućnosti detekcije objekata na preko 9000 različitih klasa uz povećanu rezoluciju ulaznih slika. Model uvodi referentne okvire (engl. *anchor boxes*), unaprijed definirane pravokutne regije različitih veličina i omjera, koje služe kao polazna točka za predviđanje graničnih okvira, a njihov se odabir optimizira primjenom K-Means [282] algoritma. Također primjenjuje normalizaciju mini-serija (engl. *batch normalization*) [181] i Darknet-19 okosnicu baziranu na VGG [209] arhitekturi. Treća verzija YOLO detektora, **YOLOv3** [273], unaprjeđuje model korištenjem Darknet-53 okosnice s rezidualnim blokovima [180] te uvodi koncept sličan piramidalnoj mreži značajki [283] za poboljšanu detekciju objekata različitih veličina.

YOLOv3 posljednja je verzija YOLO detektora koju je razvio originalni autor YOLO-a, Joseph Redmon. U međuvremenu, predstavljene su verzije YOLO detektora od **YOLOv4** [274] do najrecentnije verzije **YOLOv12** [284], te verzije poput **PP-YOLO** [278], **YOLOR** [279], **YOLOX** [280] i **YOLO-NAS** [285] detektora. Verzije YOLO algoritma, od YOLOv1 do YOLOv4, koriste DarkNet okvir otvorenog koda, koji je razvijen u programskom jeziku C i CUDA-i. Verzija **YOLOv5** [275], koju je razvio tim iz tvrtke Ultralytics, prva je verzija YOLO-a koja je umjesto u DarkNet okviru implementirana u PyTorch-u [286]. Budući da je YOLOv5 objavljen samo kao GitHub repozitorij, a ne kao recenzirano istraživanje, postojale su sumnje u autentičnost i učinkovitost tog modela. Iako odgovarajući istraživački rad nije bio dostupan, činjenica da je YOLOv5 kasnije primijenjen u brojnim aplikacijama s učinkovitim rezultatima počela je jačati kredibilitet modela [287]. U 2023. godini, isti tim je izdao poboljšani YOLO detektor nazvan **YOLOv8** [276], a 2024. godine i verziju **YOLO11** [288].

5.1.1. YOLO11

Svaka verzija YOLO serije postavila je nove standarde u detekciji objekata, uvodeći ključne arhitektonske inovacije koje unaprjeđuju performanse detektora u različitim zadacima računalnog vida [159]. Posljednja verzija razvijena od strane Ultralytics tima, YOLO11 [288], donosi poboljšanja u vidu ekstrakcije značajki i balansa između točnosti i brzine izvođenja. Detaljna arhitektura YOLO11 detektora prikazana je na Slici 5.2.



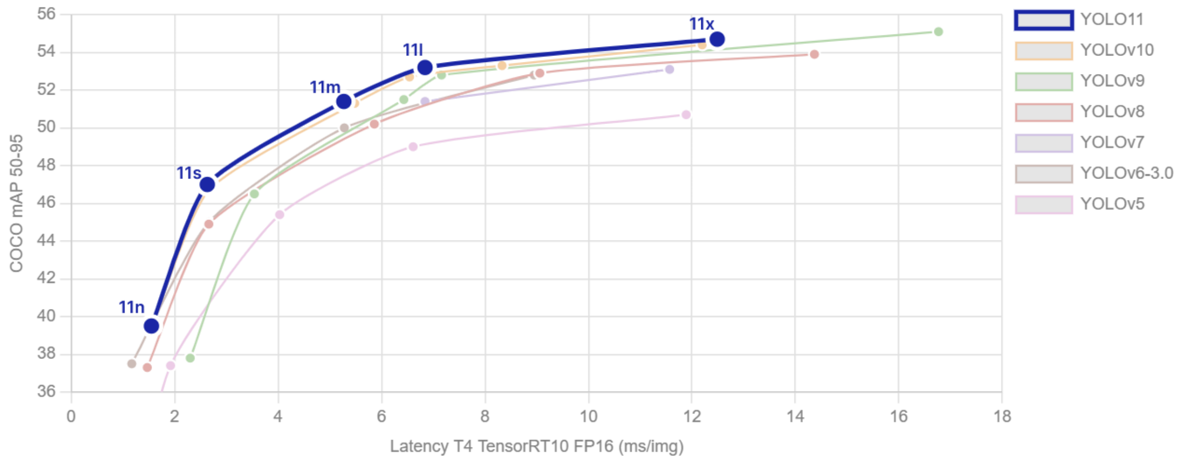
Slika 5.2: Arhitektura YOLO11 detektora. (Slika preuzeta iz [289].)

YOLO11 detektor sastoji se od tri osnovne komponente: okosnice (engl. *backbone*), vrata (engl. *neck*) i glave (engl. *head*). Okosnica detektora zadužena je za izdvajanje relevantnih značajki različitih skala iz dane ulazne slike. Tako dobivene značajke se potom agregiraju i obogaćuju u međusloju, odnosno vratu detektora, kako bi se povećala sposobnost modela da prepoznaje objekte različitih veličina. Na temelju agregiranih značajki glava detektora, putem triju paralelnih grana, generira predviđanja koordinata graničnih okvira te vjerojatnosti njihove pripadnosti pojedinim klasama, zajedno s odgovarajućim pouzdanostima (engl. *confidence score*).

Temelj arhitekture YOLO11 detektora čine CBS (*Conv-BN-SiLU*) blokovi koji objedinjuju konvoluciju (*Conv*), normalizaciju mini-serija (engl. *Batch Normalization*, *BN*) i SiLU (engl. *Sigmoid Linear Unit*) aktivacijsku funkciju [290]. Ovi blokovi osiguravaju izdvajanje relevantnih značajki, stabilizaciju protoka podataka i poboljšane performanse modela, generirajući rafinirane mape značajki potrebne za predviđanje graničnih okvira i klasifikaciju objekata [291]. C2f blok, korišten u YOLOv8 detektoru, u YOLO11 je zamijenjen C3k2 blokom. Ovaj blok koristi dvije manje konvolucije (s jezgrom veličine 2) umjesto jedne velike, čime se postiže brža obrada uz zadržavanje visoke točnosti modela [289]. Nadalje, YOLO11 zadržava SPPF komponentu (*Spatial Pyramid Pooling - Fast*) prethodnih verzija, uz dodatak modula prostorne pažnje C2PSA (*Convolutional block with Parallel Spatial Attention*), koji modelu omogućuje preciznije usmjeravanje pažnje na ključne regije slike, čime se dodatno poboljšava ekstrakcija značajki [291]. Ove arhitektonske nadogradnje omogućuju YOLO11 precizniju detekciju složenih detalja na slikama, posebno u zahtjevnim scenarijima s malim ili djelomično zaklonjenim objektima [159].

Dostupno je više različitih inačica YOLO11 detektora koje su optimizirane za različite potrebe u pogledu brzine, veličine modela i točnosti. Najmanja varijanta, u pogledu broja parametara i računalne složenosti, **YOLO11n**, dizajnirana je za rad u stvarnom vremenu na uređajima ograničenih resursa. S druge strane, **YOLO11x** je najveći model koji karakteriziraju najbolje performanse detekcije, ali i znatno veći zahtjevi za računalnim resursima i dulje vrijeme izvođenja. Između navedenih krajnosti nalaze se inačice, **YOLO11s**, **YOLO11m** i **YOLO11l**, koje nude dobar balans između točnosti detekcije i računske složenosti. Na slici 5.3 prikazani su rezultati evaluacije na MS COCO skupu podataka [166], gdje su uspoređene različite inačice YOLO11 modela s prethodnim verzijama YOLO detektora. Može se uočiti da YOLO11 konzistentno ostvaruje povoljniji kompromis između brzine izvođenja i točnosti detekcije, pri čemu manji modeli (n, s) omogućuju bržu inferenciju, dok veći (l, x) postižu višu razinu preciznosti.

Iako je YOLOv12 najrecentniji model iz YOLO obitelji detektora, rezultati komparativnih studija pokazuju da YOLO11 ostvaruje bolje performanse, nadmašujući pritom svog nasljednika [289, 293, 294]. Arhitektonske nadogradnje poput mehanizma pažnje po područjima i R-ELAN modula u YOLOv12 povećale su složenost modela, ali nisu donijele očekivana poboljšanja, čime su dodatno istaknuti izazovi učinkovite integracije naprednih mehanizama pažnje u YOLO okvir [289]. U odnosu na YOLO11, YOLOv12 postiže slabiju brzinu izvođenja i lošije rezultate, uz veće računalne zahtjeve [289, 293, 294]. Za primjene u stvarnom vremenu koje zahtijevaju visoku brzinu i preciznost, optimalnim izborom se pokazala inačica YOLO11 serije, YOLO11n [293].



Slika 5.3: Usporedba performansi inačica YOLO11 detektora s prethodnim YOLO modelima (YOLOv5–YOLOv10). (Slika preuzeta iz [292].)

5.2. Treniranje modela

5.2.1. Tehničke specifikacije računalnog okruženja

U okviru istraživanja korištena je usluga **Napredno računanje** Sveučilišta u Zagrebu Sveučilišnog računskog centra (Srce) [295]. Preciznije, za potrebe treniranja modela korištena je **platforma Jupyter** s operacijskim sustavom *Ubuntu 22.04.4 LTS (Jammy Jellyfish)*. Korisničkom okruženju unutar SRCE Jupyter platforme dodijeljena je MIG instanca GPU-a *NVIDIA A100 s 4.9 GiB VRAM-a*, dok su CPU i RAM resursi bili ograničeni unutar zajedničkog fizičkog čvora (*AMD EPYC 7713P, 32 jezgre, 239 GiB RAM-a*). Za pripremu, treniranje i evaluaciju modela korišteno je okruženje *JupyterLab*, uz dodatnu uporabu Python skripti i Ultralytics CLI-a. Softversko okruženje uključuje programski jezik *Python 3.11.9*, uz korištenje paketa *Ultralytics 8.3.178* i okvira *PyTorch 2.8.0*, dok su za obradu i vizualizaciju podataka korištene standardne biblioteke poput *NumPy-a*, *Pandas-a*, *OpenCV-a*, *Albumentations-a* i *Matplotlib-a*.

5.2.2. Odabrani modeli i vrijednosti hiperparametara

U radu su korištene tri inačice YOLO11 detektora: *YOLO11n*, *YOLO11s* i *YOLO11m*. Veće inačice *YOLOv11l* i *YOLO11x* nisu korištene zbog znatno veće računalne složenosti i memorijskih zahtjeva. Odabrane manje inačice predstavljaju balans između točnosti i učinkovitosti te su pogodnije za primjenu u stvarnom vremenu. Odabrani modeli trenirani su kroz ukupno 250 epoha s mini-serijama veličine 8, koje predstavljaju kompromis između potrošnje memorije i brzine izvođenja [293]. Sve ulazne slike svedene su na rezoluciju 640×640, koja se obično koristi pri treniranju YOLO modela [284, 293]. Nastavno na praksu ranijih istraživanja [277, 284, 296, 293], za optimizaciju težina koristi se SGD (*Stochastic Gradient Descent*) optimizator s momentom 0.937 i koeficijentom L_2 regularizacije (engl. *weight decay*) $5 \cdot 10^{-4}$, dok je vrijednost stope učenja postavljena na 0.01. Dodavanjem momenta osigurava se stabilnija konvergencija tijekom procesa treniranja, a regularizacija težina sprječava prenaučenost modela [297]. Treniranje je započelo fazom zagrijavanja (engl. *warmup*) u trajanju od tri epohe, tijekom kojih se stopa učenja i moment postupno povećavaju do zadanih vrijednosti kako bi se stabilizirao proces učenja. Funkcija gubitka definirana je kao linearna kombinacija lokalizacijskog ($\mathcal{L}_{box}$), klasifikacijskog ($\mathcal{L}_{cls}$) i distribucijskog (engl. *Distribution Focal Loss*, $\mathcal{L}_{dfl}$) gubitka: $\mathcal{L}_{total} = \lambda_{box} \cdot \mathcal{L}_{box} + \lambda_{cls} \cdot \mathcal{L}_{cls} + \lambda_{dfl} \cdot \mathcal{L}_{dfl}$, pri čemu su težine postavljene na $\lambda_{box} = 7.5$, $\lambda_{cls} = 0.5$, $\lambda_{dfl} = 1.5$ [277, 296, 288, 284], čime je naglašena važnost precizne lokalizacije objekata. Prag pouzdanosti detekcije postavljen je na 0.25, a prag preklapanja graničnih okvira (IoU) na 0.7. Također je korištena i metoda ranog zaustavljanja s pragom od 15 epoha kako bi se spriječilo nepotrebno produžavanje treniranja detektora u slučaju stagnacije performansi modela na skupu za validaciju.

Kako bi se povećala raznolikost skupa podataka, smanjila prenaučenost i osigurala veća robusnost modela u stvarnim uvjetima primjene, tijekom treniranja primijenjene su različite **augmentacije podataka** [297, 298]. Korištene su *fotometrijske augmentacije* (promjene nijanse, zasićenosti i svjetline), *geometrijske transformacije* (translacije, skaliranje, horizontalno zrcaljenje), *nasumično brisanje dijela slike* te kompozitna *augmentacija mozaika* [274] koja kombinira više slika iz skupa za treniranje u jednu. Uz prethodno navedene, ručno definirane augmentacije korišten je i *RandAugment* [299], koji dodatno proširuje varijabilnost podataka nasumičnim odabirom transformacija iz unaprijed definiranog skupa. Detaljne vrijednosti hiperparametara prikazane su u Tablici 5.1.

5.2.3. Proces treniranja

Sve inačice detektora, YOLO11n/s/m, prvo su trenirane 100 epoha na manjoj verziji SSMOT podskupa za detekciju, koja sadrži 3131 sliku s ukupno 13 680 objekata. Podaci su podijeljeni u omjeru 85:15 na skup za treniranje (2661 slika, 11 539 objekata), korišten za optimizaciju parametara modela, i na skup za validaciju (470 slika, 2141 objekt), korišten za praćenje performansi na neviđenim podacima. Kao početne točke, koristile su se težine YOLO11n/s/m modela predtrenirane na MS COCO skupu podataka.

U svrhu osiguravanja veće raznolikosti objekata i scena te poboljšanja sposobnosti generalizacije detekcijskog modela, provedeno je poluautomatizirano proširenje skupa podataka. YOLO11m model s najboljim rezultatima na skupu za validaciju korišten je za generiranje inicijalnih anotacija koje su potom ručno korigirane. Na taj je način dobiven prošireni SSMOT skup podataka za detekciju koji obuhvaća 8979 slika s ukupno 45 830 instanci objekata, podijeljen u omjeru 85:15 na skup za treniranje i skup za validaciju. Više o samoj strukturi i karakteristikama SSMOT skupa podataka za detekciju može se pronaći u odjeljku 4.2.

Tablica 5.1: Hiperparametri korišteni prilikom treniranja YOLO11n/s/m modela.

	Hiperparametar	Vrijednost
Konfiguracija: treniranje	Ukupan broj epoha	250
	Veličina mini-serije	8
	Veličina ulazne slike	640×640
	Optimizator	SGD
	Stopa učenja	0.01
	Moment	0.937
	L_2 regularizacija	0.0005
	Prag za rano zaustavljanje	15
	Početno zagrijavanje	
	epohe	3
	moment	0.8
	stopa učenja pomaka	0.1
Gubitak	λ_{box}	7.5
	λ_{cls}	0.5
	λ_{dfl}	1.5
Augmentacije	**Fotometrijske**	
	nijansa	0.015
	zasićenost	0.7
	svjetlina	0.4
	Geometrijske	
	translacija	0.1
	skaliranje	0.5
	horiz. zrcaljenje	0.5
	Slučajno brisanje	0.4
	Mozaik	1.0
	Auto augmentacije	*RandAugment*

Nakon toga, svi modeli su se nastavili trenirati dodatnih 150 epoha na proširenom skupu podataka za detekciju, pri čemu su upotrijebljene iste vrijednosti hiperparametara kao i u prvih 100 epoha. Naposlijetku, YOLO11n/s/m modeli s najboljim performansama na validacijskom skupu evaluirani su na zasebnom testnom podskupu SSMOT skupa podataka. Rezultati evaluacije prikazani su i raspravljeni u sljedećem odjeljku.

5.3. Evaluacija modela

Kako bi se procijenila učinkovitost treniranih detektora, provedena je njihova evaluacija na zasebnom testnom podskupu SSMOT skupa podataka za detekciju. Pri tom su se koristile standardne metrike koje objektivno kvantificiraju sposobnost modela u detekciji objekata različitih klasa. U nastavku se prvo definiraju metrike koje su korištene (5.3.1), a potom se prikazuju i analiziraju dobiveni rezultati (5.3.2).

5.3.1. Korištene metrike

Standardne metrike koje su korištene za evaluaciju detekcijskih modela su redom preciznost, odziv, mAP_{50} i $mAP_{50:95}$.

Detekcija modela se smatra točnom (engl. *True Positive*) ako preklapanje predviđenog B_p i stvarnog B_{gt} graničnog okvira prelazi zadani prag μ, odnosno ako vrijedi $IoU(B_p, B_{gt}) =$ površina$(B_p \cap B_{gt})$/površina$(B_p \cup B_{gt}) \geq \mu$. U suprotnom, detekcija se smatra lažnom (engl. *False Negative*). Stvarni objekti koji nisu detektirani predstavljaju lažno negativne (engl. *False Negative*) primjere. Temeljem navedenih klasifikacija definiraju se **preciznost** (engl. *Precision*, **P**) i **odziv** (engl. *Recall*, **R**):

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \tag{5.1}$$

gdje TP, FP i FN redom označavaju broj točnih detekcija, lažnih detekcija i lažno negativnih primjera pri zadanom IoU pragu μ. Navedene metrike najprije se računaju za svaku klasu zasebno, a ukupna vrijednost se potom izračuna kao aritmetička sredina vrijednosti po klasama.

Prosječna preciznost (engl. *Average Precision*, **AP**) sažima odnos preciznosti i odziva u jednu vrijednost. Može se interpretirati kao površina ispod *preciznost-odziv* krivulje:

$$AP_\mu = \int_0^1 p(r)dr, \tag{5.2}$$

gdje $p(r)$ označava vrijednost preciznosti za zadanu vrijednost odziva r. U praksi se računa prema standardu MS COCO [166] korištenjem 101 točke odziva u intervalu od 0 do 1. AP_{50} označava prosječnu preciznost pri fiksnom IoU pragu od 0.5, dok $AP_{50:95}$ predstavlja prosjek vrijednosti AP-a kroz pragove od 0.5 do 0.95, u koracima od 0.05, čime se dobiva stroži i sveobuhvatniji pokazatelj uspješnosti modela. AP vrijednost se računa za svaku klasu zasebno, a zatim se **srednja prosječna preciznost** (engl. *mean Average Precision*, **mAP**) dobije kao aritmetička sredina vrijednosti po klasama.

5.3.2. Rezultati evaluacije

Pri izračunu metrika razmatrane su detekcije čija je pouzdanost veća od 0.5, dok je prag preklapanja (IoU) postavljen na 0.5. Tablica 5.2 prikazuje rezultate evaluacije odabranih inačica YOLO11 detektora ***na testnom podskupu*** SSMOT skupa za detekciju, nakon treniranja na manjoj verziji SSMOT skupa za detekciju te nakon treniranja na njegovoj proširenoj verziji. Zajedno s vrijednostima evaluacijskih metrika navedene su i arhitekturne karakteristike detektora, broj slojeva i parametara modela, te računalna složenost izražena GFLOP (*Giga FLoating point OPeration*)[1] vrijednostima za veličinu ulazne slike 640×640.

Tablica 5.2: Usporedba arhitekturnih karakteristika i evaluacijskih rezultata odabranih inačica YOLO11 detektora na SSMOT testnom podskupu.

Model	Slojevi	Parametri	GFLOPs	Metrika	Prije proširivanja	Poslije proširivanja
YOLO11n	181	2 591 985	6.5	P	0.834	0.908
				R	0.730	0.812
				mAP_{50}	0.796	0.875
				$mAP_{50:95}$	0.660	**0.764**
YOLO11s	181	9 432 049	21.6	P	0.869	0.915
				R	0.759	0.837
				mAP_{50}	0.818	0.891
				$mAP_{50:95}$	0.707	**0.799**
YOLO11m	231	20 061 489	68.2	P	0.885	0.924
				R	0.780	0.862
				mAP_{50}	0.844	0.908
				$mAP_{50:95}$	0.752	**0.829**

Iz rezultata evaluacije prikazanih u Tablici 5.2 vidljiv je pozitivan učinak treniranja modela na proširenom skupu podataka za detekciju. Zabilježen je porast vrijednosti svih metrika kod svih detektora, što upućuje na to da treniranje na većem i raznolikijem skupu podataka značajno doprinosi povećanju robusnosti i pouzdanosti modela u detekciji objekata. Najbolje performanse, očekivano, ima najveći razmatrani model YOLO11m s vrijednošću $mAP_{50:95} = 0.829$. Međutim, performanse manjih modela ne zaostaju značajno: YOLO11s postiže $mAP_{50:95} = 0.799$, dok YOLO11n ostvaruje $mAP_{50:95} = 0.764$ s višestruko manjom računalnom složenošću i brojem parametara, što ga čini dobrim kompromisom između točnosti i učinkovitosti, osobito za primjene u stvarnom vremenu ili na uređajima s ograničenim resursima. Nakon dodatnog treniranja na proširenom skupu podataka, YOLO11n postiže $mAP_{50:95} = 0.764$, što nadmašuje performanse najvećeg modela YOLO11m treniranog isključivo na manjem skupu podataka s vrijednosti $mAP_{50:95} = 0.752$. Ovakav rezultat naglašava važnost dodatnog treniranja modela na opsežnijem i reprezentativnom skupu podataka, koji u pojedinim slučajevima može imati veći utjecaj na performanse modela od njegove same strukturne složenosti. Cjelovit prikaz rezultata svih metrika po klasama i modelima, treniranima prije i nakon proširenja detekcijskog podskupa, nalazi se u Dodatku C (Tablica C.1), dok su rezultati po klasama finalnih modela izdvojeno prikazani u Tablici 5.3.

[1] GFLOP = broj milijardi operacija s pomičnim zarezom potrebnih za jedan prolaz kroz mrežu s određenom veličinom ulazne slike

Tablica 5.3: Rezultati evaluacije YOLO11n/s/m detektora na testnom SSMOT podskupu, prikazani po klasama.

Klasa	YOLO11n				YOLO11s				YOLO11m			
	P	R	mAP_{50}	$mAP_{50:95}$	P	R	mAP_{50}	$mAP_{50:95}$	P	R	mAP_{50}	$mAP_{50:95}$
Small craft	0.974	0.848	0.921	0.747	0.979	0.874	0.935	0.768	0.983	0.877	0.937	0.801
Small Fishing Boat	0.767	0.388	0.599	0.485	0.692	0.424	0.600	0.530	0.688	0.518	0.640	0.573
Small Passenger Ship	0.910	0.874	0.917	0.798	0.907	0.910	0.937	0.830	0.961	0.904	0.946	0.860
Fishing Trawler	0.982	0.951	0.975	0.816	0.983	0.959	0.979	0.845	0.986	0.973	0.985	0.877
Large Passenger Ship	0.995	0.993	0.995	0.975	0.998	0.993	0.994	0.974	0.998	0.992	0.992	0.977
Sailing Boat	0.958	0.895	0.943	0.866	0.974	0.916	0.955	0.903	0.969	0.931	0.964	0.917
Speed Craft	0.835	0.801	0.845	0.680	0.861	0.850	0.883	0.740	0.901	0.869	0.904	0.786
Motorboat	0.748	0.467	0.606	0.500	0.780	0.557	0.668	0.580	0.842	0.633	0.745	0.647
Pleasure Yacht	0.878	0.762	0.852	0.743	0.928	0.768	0.871	0.798	0.904	0.815	0.893	0.829
Ferry	0.985	0.983	0.990	0.928	0.984	0.983	0.989	0.932	0.983	0.990	0.994	0.949
High-speed craft	0.956	0.967	0.980	0.866	0.978	0.972	0.985	0.888	0.951	0.978	0.988	0.907
Zajedno	0.908	0.812	**0.875**	**0.764**	0.915	0.837	**0.891**	**0.799**	0.924	0.862	**0.908**	**0.829**

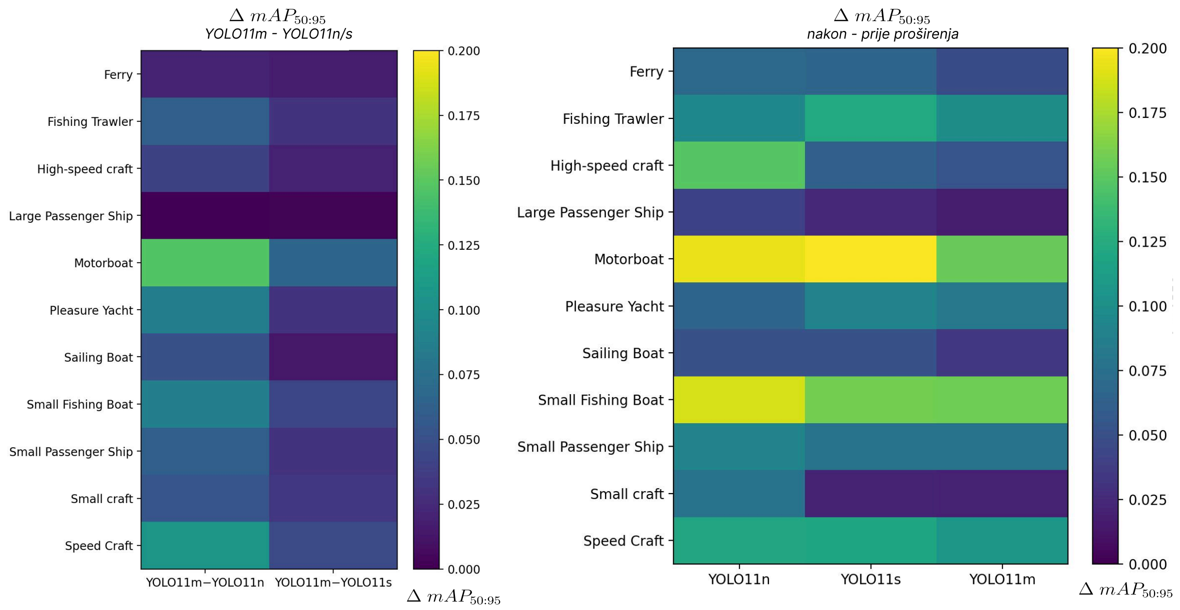
Modeli najviše poteškoća imaju u ispravnoj detekciji objekata klase *Small Fishing Boat*, koja je ujedno i najmanje zastupljena u SSMOT skupu podataka za detekciju. Odziv modela YOLO11n na toj klasi iznosi svega 0.338, dok je kod modela YOLO11m nešto viši i doseže 0.518. Međutim, problem ne proizlazi samo iz nedovoljne zastupljenosti primjera, nego i iz visoke vizualne sličnosti objekata klase *Small Fishing Boat* s objektima drugih klasa, poput klase *Speed Craft* (otvorene brodice bez nadgrađa imaju slične vizualne karakteristike kao manja otvorena brza plovila). Ta česta zabuna jasno se očituje u matricama konfuzije na Slici 5.5. Prikazane matrice konfuzije su normalizirane po stupcima, prikazujući tako relativne udjele umjesto apsolutnih vrijednosti čime se omogućuje jasniji uvid u performanse modela po klasama, neovisno o njihovoj zastupljenosti u danom skupu podataka [2].

Mali broj primjera u skupu podataka za treniranje, ne mora nužno predstavljati ograničenje ako je riječ o klasi s jasno prepoznatljivim i distinktivnim vizualnim obilježjima. To potvrđuju vrlo dobri rezultati na klasi *High Speed Craft*, koja je također slabo zastupljena u danom skupu (1357 od 39017 instanci u skupu za treniranje, naspram 1101 za *Small Fishing Boat*; Poglavlje 4, odjeljak 4.2, Tablica 4.2). Uz *Small Fishing Boat*, detektori se često „muče” i s klasom *Motorboat*, čije se instance nerijetko zamjenjuju s onima klase *Speed Craft* ili ostaju nedetektirane pri zadanoj granici pouzdanosti. Srednje vrijednosti postiže i sama klasa *Speed Craft*, što se može pripisati intraklasnoj varijabilnosti te djelomičnoj vizualnoj sličnosti s drugim klasama, primjerice sa spomenutom *Small Fishing Boat*.

Upravo su na tim zahtjevnijim klasama ostvareni najznačajniji pomaci promjenom arhitekture detektora s jednostavnije YOLO11n na složeniju YOLO11m, kao i dodatnim treniranjem na proširenom skupu podataka s većim brojem i raznolikijim primjerima tih klasa. Navedeno je sažeto ilustrirano na Slici 5.4, koja prikazuje toplinske karte (engl. *heatmap*) razlika $mAP_{50:95}$ vrijednosti: (a) finalnog detektora YOLO11m i jednostavnijih varijanti

[2]Vrijednosti na dijagonali normalizirane matrice konfuzije u pravilu prikazuju odziv pojedine klase, jer prikazuju udio ispravno detektiranih primjera u odnosu na sve uzorke te klase. Ipak, javljaju se manja odstupanja u odnosu na formalno izračunate vrijednosti odziva, ponajprije zbog načina na koji YOLO implementira evaluaciju (izračunu preciznosti i odziva s dodatnim koracima poput NMS filtriranja i obrade višestrukih detekcija), kao i zbog numeričkog zaokruživanja prilikom prikaza.

YOLO11s/m; (b) modela treniranih isključivo na prvoj, manjoj varijanti skupa za detekciju i onih treniranih na proširenom SSMOT skupu za detekciju. Naglasak je stavljen na $mAP_{50:95}$ metriku jer ona integrira više različitih IoU pragova i preciznost-odziv krivulju, pružajući tako robusniju usporedbu od jednopragovnog *mAP* te pojedinačnih mjera preciznosti i odziva.

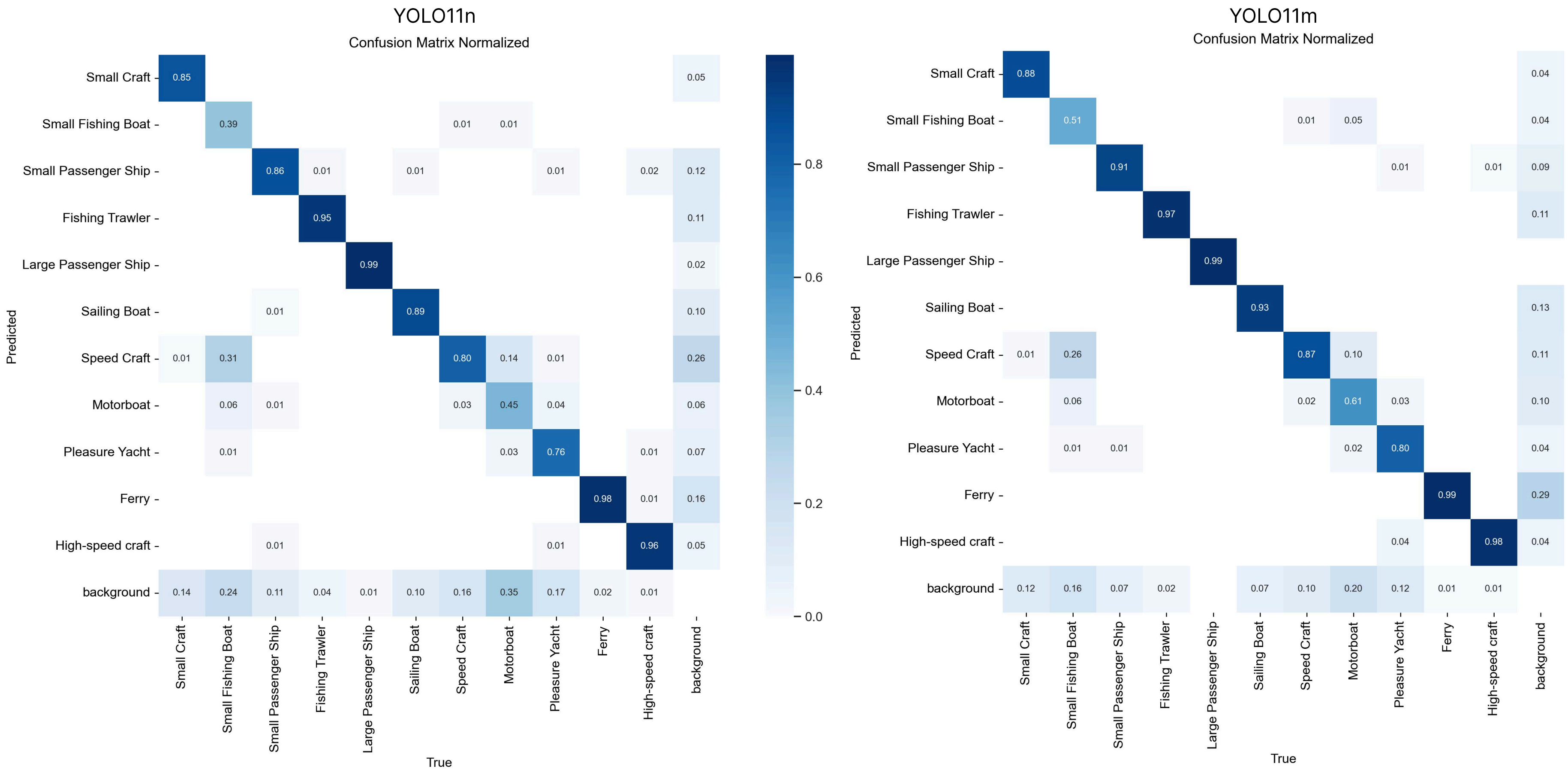


(a) (b)

Slika 5.4: Toplinske karte razlika $mAP_{50:95}$ na testnom skupu: (a) arhitekturni učinak: YOLO11m u odnosu na YOLO11n/s; (b) učinak treniranja na proširenom skupu za detekciju: modeli trenirani na proširenoj naspram modela treniranih samo na početnoj verziji SSMOT skupa za detekciju.

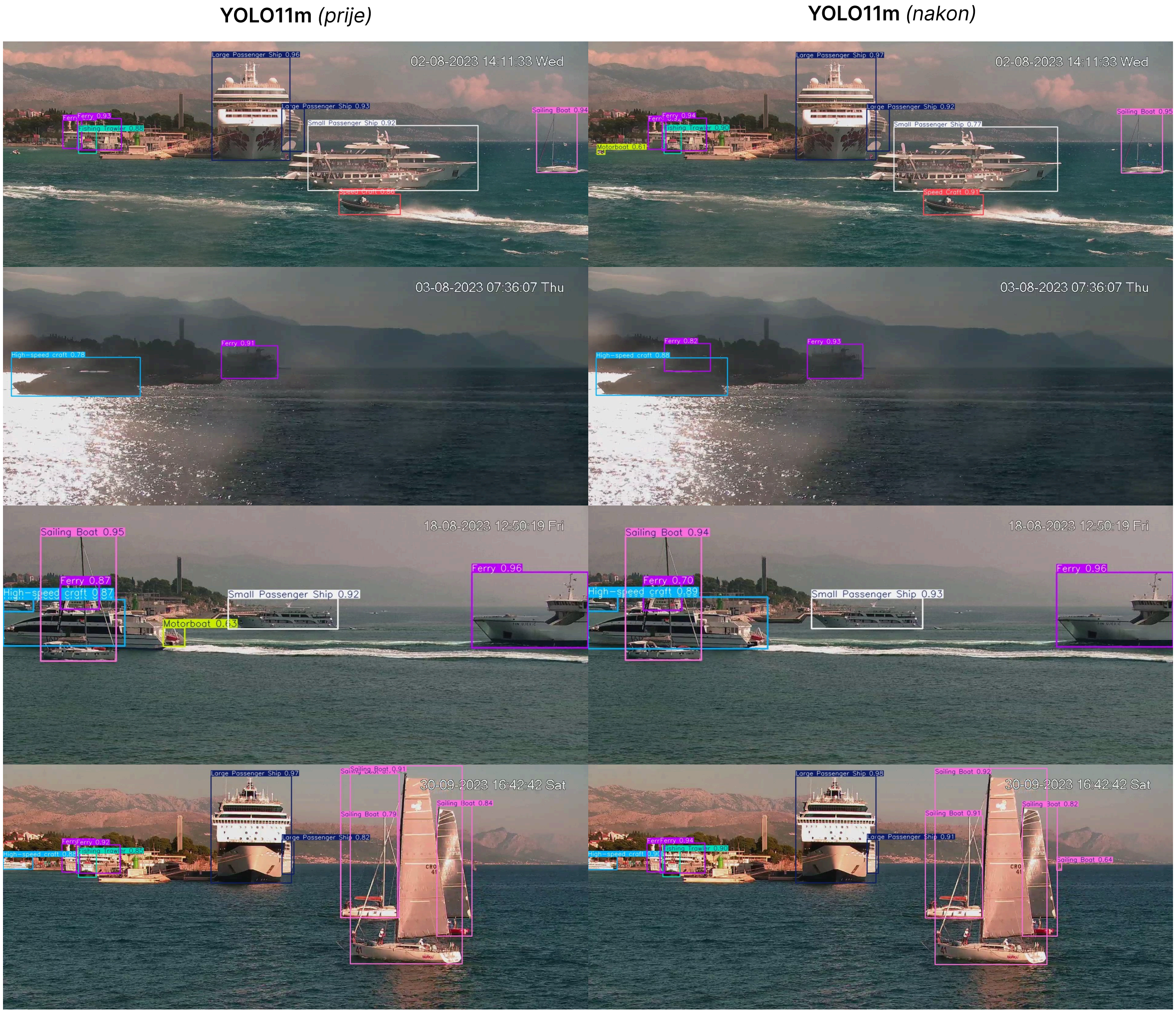
Iz Slike 5.4 (a) je jasno da složenija arhitektura ima konzistentan dobitak nad manjima pri čemu je dobitak YOLO11m detektora nad manjim modelom YOLO11n izraženiji, što je bilo i očekivano. Nadalje, najveći dobitak dobiven je na klasi *Motorboat* (0.147), a slijede *Speed Craft* (0.106) i *Small Fishing Boat* (0.088). Na Slici 5.4 (b) može se uočiti značajan pozitivan učinak treniranja modela na proširenom skupu podataka, osobito kada je riječ o problematičnim klasama *Motorboat* (dobitak: YOLO11n - 0.194, YOLO11m - 0.154) *Speed Craft* (dobitak: YOLO11n - 0.121, YOLO11m - 0.106) i *Small Fishing Boat* (dobitak: YOLO11n - 0.188, YOLO11m - 0.157). Veće, jasno odijeljene klase poput *Ferry* i *Large Passenger Ship* bilježe pozitivne promjene, ali nešto manjeg intenziteta. Pozitivan učinak proširenja izraženiji je kod manjih modela, dok je kod najvećeg modela YOLO11m prisutan, ali skromniji.

Kako bi se dobio uvid u performanse različitih varijanti detektora, u nastavku su prikazani konkretni primjeri njihovih detekcija na testnim slikama.



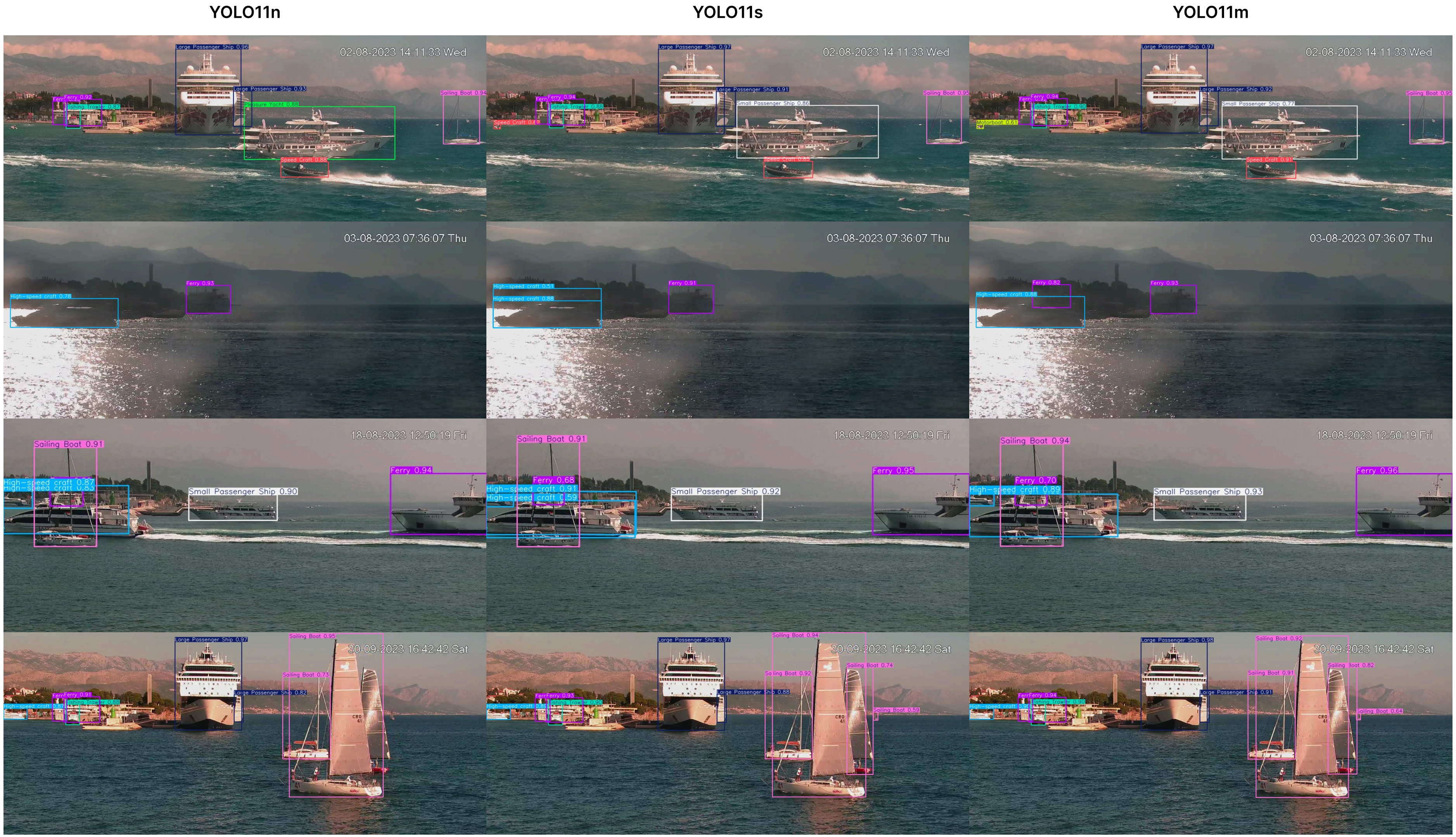
Slika 5.5: Normalizirane matrice konfuzije YOLO11n/m detektora na testnom podskupu.

Primjeri detekcija YOLO11m detektora treniranog na manjem i na proširenom skupu podataka, prikazani su na Slici 5.6. Mogu se uočiti poboljšanja u detekciji objekata pri smanjenoj vidljivosti (druga slika), u prepoznavanju manjih plovila u daljini (prva i četvrta slika) te u detekciji djelomično zaklonjenih objekata i objekata s preklapajućim graničnim okvirima (treća i četvrta slika). Istodobno, zabilježeni su i izazovi finalnog modela, primjerice pri detekciji dvaju plovila koja plove paralelno, a model ih spaja u jednu detekciju, kao i pri ispravnoj klasifikaciji manjih objekata u daljini (prva slika).



Slika 5.6: Detekcije dobivene YOLO11m detektorom treniranom na manjem skupu podataka (lijevo) i nakon treniranja na proširenom skupu za detekciju (desno).

S druge strane, Slika 5.7 na istim primjerima usporedno prikazuje detekcije dobivene finalnim varijantama YOLO11n/s/m detektora. YOLO11m postiže najbolje rezultate u uvjetima smanjene vidljivosti, detektirajući sva tri prisutna plovila (druga slika). U slučaju preklapanja/zaklonjenosti jedrilica na četvrtoj slici, YOLO11m i YOLO11s ostvaruju bolje rezultate od YOLO11n, dok se u scenariju preklapanja objekata na trećoj slici najuspješniji pokazuje YOLO11m: YOLO11n generira neprecizan granični okvir, YOLO11s daje dvije odvojene detekcije umjesto jedne, dok YOLO11m pravilno prepoznaje objekt.



Slika 5.7: Primjeri detkecija dobivenih finalnim verzijama detektora YOLO11n/s/m.

6. REIDENTIFIKACIJSKI MODEL

Zadatak reidentifikacije plovila od iznimne je važnosti za sustave automatskog praćenja, jer omogućuje prepoznavanje istog plovila u različitim vremenskim trenutcima i iz različitih perspektiva. U ovom su poglavlju najprije predstavljena dva modela korištena za reidentifikaciju plovila - ResNet50 i OSNet, zatim je detaljno opisan postupak njihove implementacije i treniranja, a na kraju su opisane metrike korištene za evaluaciju te su prikazani i analizirani dobiveni rezultati.

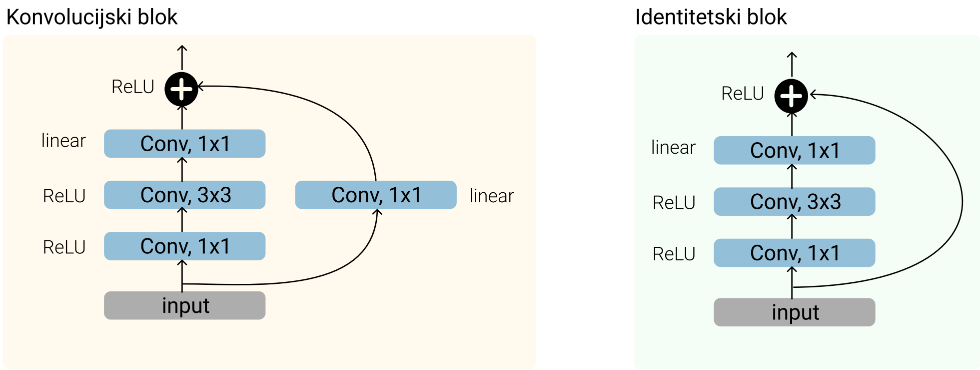
6.1. Korišteni modeli

U svrhu rješavanja problema reidentifikacije odbrane su dvije različite arhitekture dubokih konvolucijskih mreža - ResNet50 i OSNet. ResNet50 je jedna od najčešće korištenih arhitektura u području računalnog vida koja se, zahvaljujući svojoj robusnosti i fleksibilnosti, koristi u širokom spektru zadataka, od klasifikacije slika [116, 180, 300] i semantičke segmentacije [301, 302, 303] pa sve do reidentikacije objekata [304, 305, 306]. S druge strane, OSNet arhitektura je specijalno razvijena za potrebe zadatka reidentifkacije. Paralelnom primjenom i evaluacijom ovih dvaju modela omogućeno je vrednovanje performansi općeg, široko primjenjivog modela i njegova usporedba s modelom dizajniranim specifično za zadatak reidentifikacije.

6.1.1. ResNet50

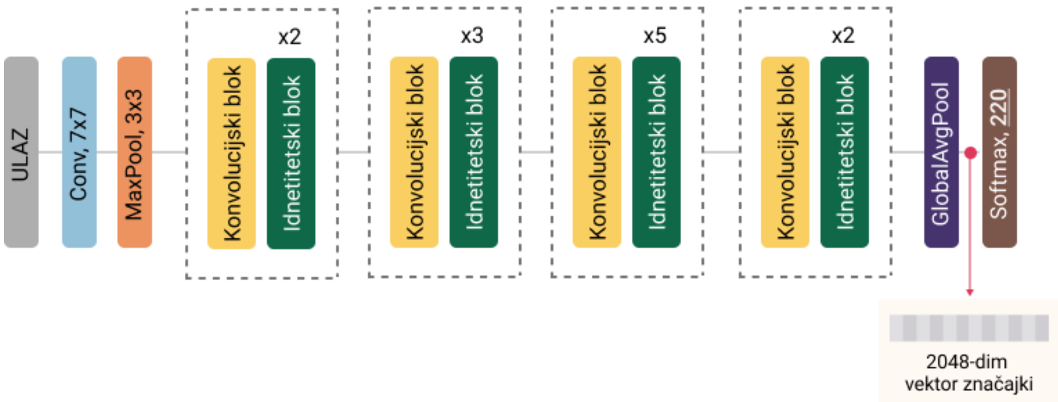
ResNet50 arhitektura dio je obitelji rezidualnih mreža koje su predstavili He i sur. [180] s ciljem prevladavanja problema degradacije točnosti kod vrlo dubokih neuronskih mreža. Temeljna ideja počiva na uvođenju tzv. prečac (engl. *shortcut*) veza koje izravno povezuju ulaz i izlaz određenog sloja mreže tako da se izlaz ne računa samostalno već se zbraja s izvornim ulaznim podatkom. Na taj način mreža ne uči direktno ciljno preslikavanje $f(x)$, već rezidualnu funkciju $f_R(x) = f(x) - x$, što u praksi olakšava optimizaciju i omogućuje treniranje znatno dubljih arhitektura.

ResNet50 sastoji se od 50 slojeva organiziranih kroz dvije vrste rezidualnih modula: identitetskih (engl. *identity*) i konvolucijskih (engl. *convolution*) blokova. Identitetski blokovi primjenjuju izravne prečac veze te se stoga koriste kada su dimenzije ulaza i odabranog izlaza jednake. S druge strane, u slučajevima različitih dimenzija, koriste se konvolucijski blokovi s projekcijskim prečac vezama koje pomoću 1×1 konvolucije usklađuju dimenzije ulaza i izlaza. Ilustracija dviju vrsta rezidualnih blokova prikazana je na Slici 6.1.



Slika 6.1: Dvije vrste rezidualnih blokova. Svaki blok računa funkciju $f_R(x) + g(x)$, gdje x označava ulaz modula, a $f_R(x)$ izlaz triju uzastopnih konvolucijskih slojeva. Funkcija g u konvolucijskom bloku (lijevo) djeluje kao projekcija kojom se usklađuju dimenzije x i $f_R(x)$, dok je u identitetskom bloku (desno) definirana kao $g(x) = x$. (Slika preuzeta iz [116].)

Osim prečac veza, važnu ulogu u ResNet arhitekturi ima i normalizacija mini-serija, koja se primjenjuje neposredno iza konvolucije, a prije nelinearne aktivacijske funkcije. Na taj se način dodatno stabilizira i ubrzava proces treniranja mreže. Detaljna arhitektura ResNet50 mreže korištene u ovom radu, prikazana je na Slici 6.2. U odnosu na izvorni model, izmijenjen je broj neurona u završnom softmax sloju s 1000 (koliko ih ima u ImageNet skupu podataka) na 220, što odgovara broju različitih identiteta u SSMOT ReID skupu podataka. Za svaku ulaznu sliku, model izdvaja 2048-dimenzionalni vektor značajki, dobiven na izlazu sloja globalnog prosječnog sažimanja (engl. *Global Average Pooling*) [307], koji se potom koristi kao reprezentacija plovila u zadatku reidentifikacije. Informacije sadržane u 2048 dvodimenzionalnih mapa značajki posljednjeg identitetskog bloka sažimaju se u vektor iste dimenzionalnosti, pri čemu je i-ta komponenta vektora dobivena kao aritmetička sredina vrijednosti i-te mape značajki, $i \in \{1, \ldots, 2048\}$.

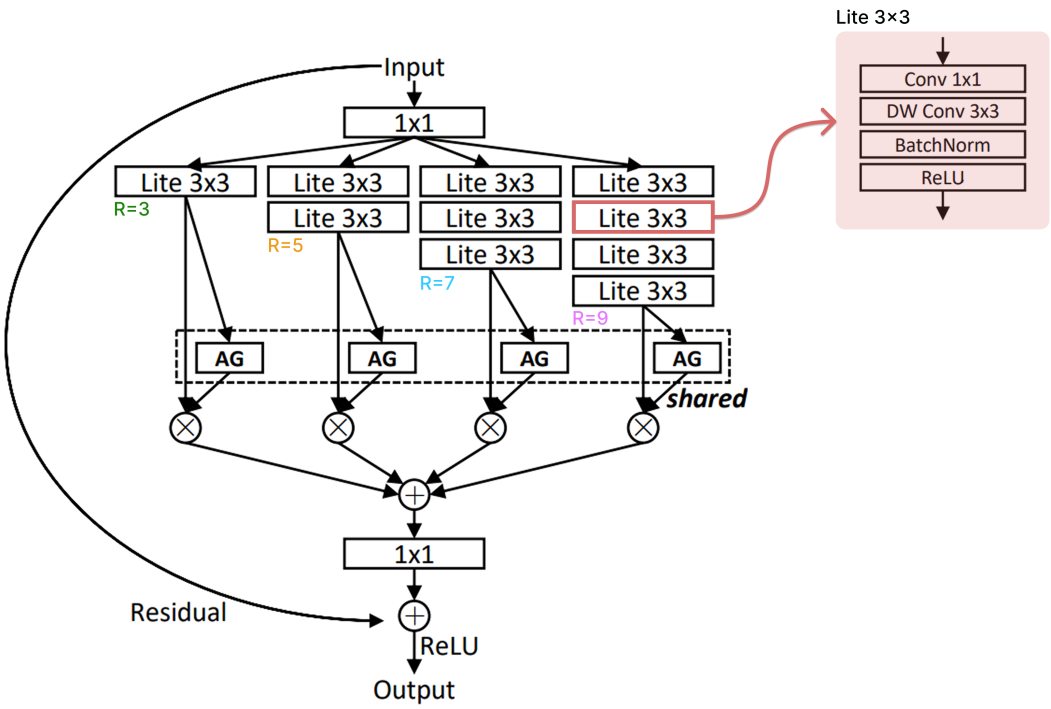


Slika 6.2: Arhitektura ResNet50 modela. (Slika preuzeta iz [116], uz izmjene.)

6.1.2. OSNet

Kako bi proces reidentifikacije bio što uspješniji, nužno je razviti značajke koje su dovoljno diskriminativne da jasno razlikuju vrlo slične objekte, a istodobno dovoljno robusne da ostanu primjenjive u različitim uvjetima i okruženjima. U tom kontekstu, Zhou i sur. [228] predlažu **OSNet** (*Omni-Scale Network*) model, koji uvodi koncept ***učenja značajki na više skala*** (engl. *omni-scale feature learning*). Osnovna ideja pritom je istovremeno obuhvatiti značajke različitih razmjera te ih dinamički kombinirati u jedinstvenu reprezentaciju koja najbolje odgovara zadanom ulazu.

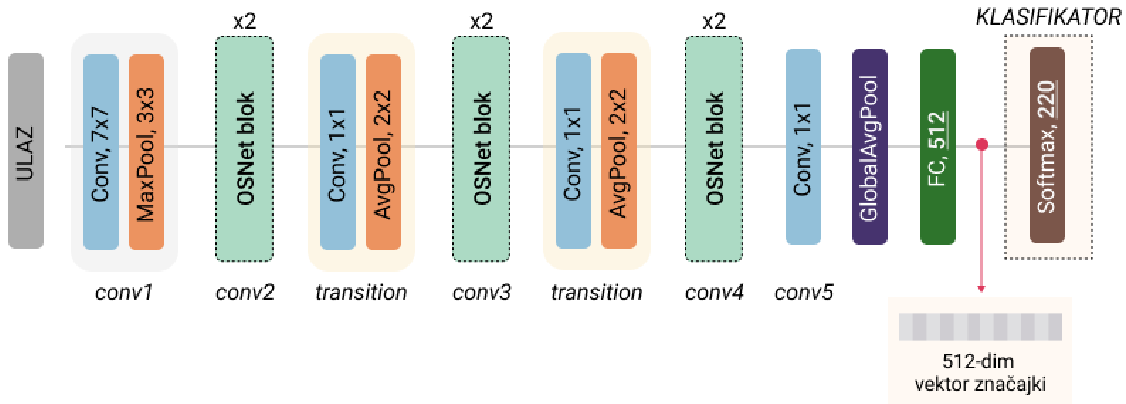
Temelj OSNet arhitekture čine *omni-scale* rezidualni blokovi, prikazani na Slici 6.3, koji se sastoje od više konvolucijskih grana s receptivnim poljima različitih veličina. Receptivno polje dimenzije $(2t+1)\times(2t+1)$ dobije se sukcesivnim slaganjem t "laganih" (engl. *lite*) 3×3 konvolucijskih blokova. Ovi blokovi koriste 3×3 dubinski-razdvojene konvolucije (engl. *depth-wise separable convolutions*), koje zahtijevaju približno 8 do 9 puta manje operacija u odnosu na standardne konvolucije [308]. Za razliku od izvorne varijante dubinski-razdvojene konvolucije, u OSNet-u se najprije provodi 1×1 konvolucija po elementima (engl. *point-wise convolution*), kojom se vrijednosti različitih kanala ulaza na istoj prostornoj poziciji linearno kombiniraju u novu informaciju. Tek potom se primjenjuje 3×3 dubinska konvolucija (engl. *depth-wise convolution*), kojom se ulazne značajke filtriraju korištenjem zasebne konvolucijske jezgre za svaki kanal.



Slika 6.3: Osnovni građevni blokovi OSNet modela. Oznake: AG - agregacijaska vrata; R - veličina receptivnog polja. (Slika preuzeta iz [228], uz izmjene.)

Svaka konvolucijska grana generira homogene značajke jedne prostorne skale. Kako bi se istodobno obuhvatile značajke više različitih skala, OSNet uvodi mehanizam agregacijskih vrata (engl. *Agregation Gate*, AG), koji dinamički kombinira izlaze pojedinih grana dodjeljujući im težine, ovisno o ulazu. Time se modelu omogućuje da, za svaku ulaznu sliku, naglasi one skale ili njihove kombinacije koje su najrelevantnije. Agregacijska vrata najprije trodimenzionalne ulazne značajke komprimiraju u dvodimenzionalni vektor koristeći sloj globalnog prosječnog sažimanja. Na dobiveni vektor zatim se primjenjuje mala neuronska mreža sastavljena od potpuno povezanih slojeva, koja generira koeficijente što služe kao dinamičke težine za kombiniranje izlaza konvolucijskih grana. Pri tome se koristi jedinstvena mreža čije su težine zajedničke za sve grane.

Detaljna arhitektura OSNet modela prikazana je na Slici 6.4. Model je dostupan u više varijanti koje se razlikuju prema širini mreže. Oznake poput 0.5, 0.75 ili 1.0 označavaju omjer broja kanala u konvolucijskim slojevima u odnosu na osnovnu varijantu, *OSNet 1.0*. Manje varijante, poput *OSNet 0.75*, imaju smanjen broj parametara, zahtijevaju manje računalnih resursa te omogućuju bržu obradu i manju potrošnju memorije, uz blago smanjenje performansi.



Slika 6.4: Arhitektura implementiranog OSNet modela.

6.2. Implementacija i treniranje ReID modela

Za implementaciju modela reidentifikacije korištena je *Torchreid* [309] biblioteka, u kojoj modeli dolaze s težinama predtreniranim na ImageNet [125] skupu podataka. Pritom zadane vrijednosti dimenzija ulaznih slika iznose visina = 256 i širina = 128, što je standard u reidentifikaciji osoba [228, 310, 311] jer odgovara specifičnom obliku graničnih okvira za ljude. Budući da takve dimenzije nisu karakteristične za plovila, provedena je prilagodba: prvo je na podacima za učenje iz SSMOT ReID skupa podataka određena medijalna vrijednost omjera širine i visine (≈ 2.3), a zatim je odabrana kombinacija visina = 160 i širina = 368 koja zadovoljava taj omjer uz razuman kompromis između očuvanja detalja slike i potrošnje memorije.

6.2.1. Konstrukcija mini-serija

Tijekom treniranja koriste se mini-serije veličine 32 ($P \times K$), pri čemu svaka sadrži $P = 8$ različitih identiteta i $K = 4$ slika za svaki identitet. Od tih osam identiteta četiri pripadaju istoj fokus-klasi, dok se preostala četiri biraju iz ostalih klasa. Na taj način model istodobno uči izražene međuklasne razlike (npr. između trajekta i jahte) i suptilne razlike unutar iste klase (npr. između dva trajekta). Konstrukcija mini-serije započinje odabirom fokus-klase uz vjerojatnosti proporcionalne potenciji N_i^{γ}, gdje je N_i broj različitih identiteta klase i, dok eksponent $\gamma = 0.5$ ublažava dominaciju velikih klasa. Pri tom se koristi dodatno pravilo kojim se s vjerojatnošću 0.3 potiče odabir manjinskih klasa (onih s najviše 10 identiteta). Iz odabrane fokus-klase se nasumično odabiru četiri različita identiteta, a preostala četiri iz ostalih klasa. Za svaki identitet uzimaju se četiri nasumično odabrane slike.

Kako bi se povećala robusnost modela i njegova sposobnost generalizacije na različite položaje plovila i uvjete snimanja, tijekom treniranja primijenjene su sljedeće augmentacije podataka: *(i) nasumično horizontalno zrcaljenje*, koje smanjuje osjetljivost modela na lijevo-desnu orijentaciju plovila; *(ii) nasumično izrezivanje* kojim se simuliraju pomaci i neprecizni granični okviri; *(iii) nasumične promjene svjetline i kontrasta* radi veće otpornosti na promjene osvjetljenja i vremenskih prilika; te *(iv) nasumično brisanje* pravokutnog područja kojim se imitira djelomična okluzija plovila.

Prilikom učitavanja u mini seriju, dimenzije odabranih slika se najprije prilagođava zadanim dimenzijama (visina = 160, širina = 368) primjenom bilinearne interpolacije. Zatim se, s vjerojatnošću 0.5, primjenjuje horizontalno zrcaljenje. U sljedećem koraku slika se bilinearnom interpolacijom privremeno povećava na dimenzije 180×414, te se iz nje nasumično izrezuje isječak ciljnih dimenzija 160×368, čime se uvodi prostorna varijabilnost položaja objekta unutar fiksnih ulaznih dimenzija. Nakon toga slijedi nasumična promjena svjetline s faktorom 0.2 i kontrasta s faktorom 0.15, nakon čega se slika pretvara u tenzor s vrijednostima u [0,1] te se po kanalima standardizira statistikama ImageNet-a: $(\mu_R, \mu_G, \mu_B) = (0.485, 0.456, 0.406)$, $(\sigma_R, \sigma_G, \sigma_B) = (0.229, 0.224, 0.225)$. Naposlijetku se provodi brisanje jednog pravokutnog područja slike koje je nasumično odabrano.

6.2.2. Optimizator i funkcija gubitka

Model reidentifikacije treniran je 32000 iteracija koristeći Adam [312] optimizator s početnom stopom učenja od $3 \cdot 10^{-4}$ te koeficijentom L_2 regularizacije $5 \cdot 10^{-5}$. Stopa učenja se pritom smanjivala svakih 8000 iteracija faktorom 0.1.

U zadacima reidentifikacije osoba pokazalo se korisnim trenirati modele kombinacijom klasifikacijskog gubitka unakrsne entropije (engl. *Cross-Entropy loss*, često nazivan i ID loss) i gubitka trojki (engl. *triplet loss*) [304, 150, 313, 314]. Klasifikacijski gubitak uči mrežu ispravno razlikovati identitete u skupu za treniranje te tako potiče razvoj diskriminativnih značajki koje dobro razlikuju poznate identitet, ali sam po sebi ne jamči dobru generalizaciju na nove, neviđene identitete. S druge strane, gubitak trojki organizira metrički prostor tako da su pozitivni primjeri istog identiteta bliže, a negativni dalje, što je dobro za generalizaciju, ali njegovo samostalno korištenje u praksi otežava treniranje te sam proces čini nestabilnim. Kombinacijom ova dva gubitka postiže se optimalan učinak: klasifikacijski gubitak osigurava stabilnu optimizaciju modela, a gubitak trojki pridonosi robusnoj generalizaciji na nove identitete. Shodno tome, u ovom radu se koristi varijanta kombinacije s izglađivanjem oznaka (engl. *label smoothing*) i odabirom teških pozitivnih i negativnih primjera (engl. *hard positive/negative mining*).

Neka je I ulazna slika trenutne mini-serije B, $\mathbf{p}(I) = (p_1(I), \ldots, p_N(I))$ predviđena vjerojatnosna distribucija po klasama/identitetima za sliku I gdje je N broj identiteta u skupu za treniranje. Nadalje, neka je $y \in \{1, \ldots, N\}$ stvarni identitet plovila na slici I. Tada se ***gubitak unakrsne entropije s izglađivanjem oznaka*** [315] definira s:

$$\mathcal{L}_{CE}(\theta; I, y) = -\sum_{i=1}^{N} q_i(y) \cdot \log\left(p_i(I)\right), \tag{6.1}$$

gdje θ predstavlja trenutne parametre modela,

$$q_i(y) = \begin{cases} 1 - \varepsilon, & \text{ako } i = y, \\ \dfrac{\varepsilon}{N-1}, & \text{inače.} \end{cases} \tag{6.2}$$

je izglađena verzija oznake y, a $\varepsilon = 0.1$ faktor izglađivanja. Korištenje unakrsne entropije s izglađivanjem oznaka pokazalo se korisnim jer smanjuje prenaučenost, ublažava pretjerano samouvjerena predviđanja i potiče bolju generalizaciju na nove podatke [315, 150].

Za ulaznu sliku I, neka I_p^* označava pozitivni primjer iz trenutne mini-serije koji je najudaljeniji od I, a I_n^* negativni primjer koji je najbliži I. Formalno, neka je B_y skup svih slika dane mini-serije B koje odgovaraju identitetu y, f_θ preslikavanje koje danoj slici pridruži m-dimenzionalni vektor značajki dobiven ReID modelom s parametrima θ, a $d : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, kvadrat euklidske udaljenosti $d(f_1, f_2) = \|f_1 - f_2\|_2^2$. Tada su teški pozitivni i teški negativni primjer redom dani s:

$$I_p^* = \max_{I_b \in B_y,\, I_b \neq I} d\left(f_\theta(I),\, f_\theta(I_b)\right), \quad I_n^* = \min_{I_b \in B \setminus B_y} d\left(f_\theta(I),\, f_\theta(I_b)\right), \tag{6.3}$$

dok je ***trostruki gubitak s odabirom teških pozitivnih i teških negativnih primjera*** [304] definiran kao:

$$\mathcal{L}_{triplet}(\theta; I, y) = \max\left\{0,\ \ d\left(f_\theta(I),\, f_\theta(I_p^*)\right) - d\left(f_\theta(I),\, f_\theta(I_n^*)\right) + m\right\}. \tag{6.4}$$

Na ovaj način model se prisiljava da uči iz najzahtjevnijih primjera kako bi se ubrzalo i stabiliziralo učenje diskriminativnih značajki. Konačno, ***ukupan gubitak*** modela za dani ulaz I i pripadajuću oznaku y je onda zbroj gubitaka:

$$\mathcal{L}(\theta;\, I,y) = \mathcal{L}_{CE}(\theta;\, I,y) + \mathcal{L}_{triplet}(\theta;\, I,y). \tag{6.5}$$

6.3. Evaluacija ReID modela

Kako bi se ocijenila učinkovitost razvijenih modela za reidentifikaciju plovila, provedena je njihova evaluacija na testnom skupu SSMOT ReID podataka, koji se standardno sastoji od skupa upita i galerije. Tijekom evaluacije provjerava se sposobnost modela da za svaku ulaznu sliku iz skupa upita pronađe odgovarajuće primjere istog identiteta u galeriji, čime se ispituje njegova točnost i robusnost u složenim uvjetima pretraživanja kakvi se očekuju u stvarnoj primjeni. Uspješnost modela kvantitativno je mjerena primjenom standardnih ReID metrika, koje omogućuju objektivnu procjenu performansi u zadatku reidentifikacije. U nastavku se najprije u pododjeljku 6.3.1 detaljno opisuju korištene metrike, a zatim su u pododjeljku 6.3.2 prikazani i analizirani rezultati evaluacije implementiranih modela.

6.3.1. Korištene metrike

Za potrebe kvantitativne procjene performansi odabrane su standardne metrike koje se uobičajeno primjenjuju u području reidentifikacije. Njihova primjena omogućuje objektivnu evaluaciju učinkovitosti razvijenih modela.

Neka $\mathcal{Q}$ označava skup upita, a $\mathcal{G}$ galeriju. Za svaki upit $q \in \mathcal{Q}$ model generira rangiranu listu slika iz galerije $R_q = (r_1,\ r_2,\ \ldots,\ r_{|\mathcal{G}|})$ sortiranu prema odabranoj mjeri sličnosti tako da je r_1 primjer galerije koji je najsličniji upitu q. Nadalje, neka $\mathcal{P}_q \subseteq \mathcal{G}$ označava skup svih pozitivnih instanci galerije za dani upit q.

Kumulativna karakteristika podudaranja (engl. *Cumulative Matching Characteristic, CMC*) ili ***Rank-k*** metrika točnosti podudaranja [313] mjeri vjerojatnost da se barem jedan primjer galerije traženog identiteta $p \in \mathcal{P}_q$ pojavi među prvih k elemenata rangirane liste $\mathcal{R}_q$. Neka je

$$Acc_k(q) = \begin{cases} 1, & \text{ako } \mathcal{P}_q \cap \{r_1, r_2, \ldots, r_k\} \neq \emptyset \\ 0, & \text{inače} \end{cases}, \tag{6.6}$$

tada je

$$Rank\text{-}k = \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} Acc_k(q). \tag{6.7}$$

U literaturi se najčešće navode *Rank-1*, *Rank-5* i *Rank-10* metrike [316, 317, 318, 319], pri čemu je *Rank-1* od osobite važnosti jer označava vjerojatnost da se primjer galerije ispravnog identiteta nalazi na samom vrhu rangirane liste.

Glavno ograničenje CMC metrike očituje se u situacijama kada galerija sadrži više pozitivnih primjera istog identiteta, kao što je slučaj i u SSMOT ReID skupu podataka. Za uspješan rezultat ($Acc_k(q) = 1$) tada je dovoljno da se barem jedan pozitivan primjer galerije pojavi unutar prvih k pozicija rangirane liste, dok se rang preostalih pozitivnih primjera u potpunosti može zanemariti. Posljedično, model može postići visoke vrijednosti ovih metrika čak i ako je većina primjera ispravnog identiteta pri samom kraju rangirane liste. Upravo zbog toga se u praksi CMC gotovo uvijek nadopunjuju *mAP* metrikom, koja u obzir uzima sve pozitivne primjere i njihove pozicije unutar rangirane liste [313].

Kako bi se u obzir uzela kvaliteta cjelokupnog rangiranja, koristi se **srednja prosječna preciznost** (engl. *mean Average Precision, mAP*). Za svaki upit q se prvo računa *prosječna preciznost (AP)* dana s:

$$AP(q) = \frac{1}{|\mathcal{P}(q)|} \sum_{i=1}^{|\mathcal{G}|} P_q(i) \cdot rel_q(i), \tag{6.8}$$

pri čemu je

$$rel_q(i) = \begin{cases} 1, & \text{ako je } r_i \in P_q \\ 0, & \text{inače} \end{cases} \tag{6.9}$$

dok je

$$P_q(i) = \frac{1}{i} \sum_{j=1}^{i} rel_q(j) \tag{6.10}$$

preciznost izračunata na prvih i pozicija rang liste $\mathcal{R}_q$. Potom se srednja prosječna preciznost računa kao aritmetička sredina dobivenih AP vrijednosti:

$$mAP = \frac{1}{|Q|} \sum_{q \in Q} AP(q). \tag{6.11}$$

6.3.2. Rezultati evaluacije

Na temelju definiranih metrika provedena je evaluacija implementiranih modela na testnom skupu SSMOT ReID podataka. U ovom su pododjeljku prikazani dobiveni rezultati, zajedno s odgovarajućom analizom i interpretacijom, kako bi se procijenila učinkovitost modela u zadatku reidentifikacije plovila.

Tablica 6.1 prikazuje usporedbu osnovnih karakteristika triju korištenih ReID modela, pri čemu svi koriste ulaze dimenzija 160×368. Uspoređeni su broj značajki u izlaznom vektoru, broj parametara modela te računska složenost izražena GFLOP vrijednostima, uz dvije vremenske mjere: (1) latenciju izračuna izlaza modela, odnosno prosječno vrijeme potrebno za generiranje vektora značajki jedne ulazne slike, te (2) vrijeme pretrage, koje predstavlja prosječno trajanje usporedbe jednog upita s galerijom od 10000 uzoraka. Mjerenja su provedena na računalnom sustavu s Intel Core i7-9850H procesorom, 16 GB radne memorije i NVIDIA Quadro RTX 3000 grafičkom karticom, uz korištenje operacijskog sustava Windows 11 Pro.

Unatoč znatno manjem broju parametara i nižoj računskoj složenosti, OSNet modeli imaju veću latenciju generiranja značajki. Navedeno se može pripisati arhitekturnim i implementacijskim čimbenicima: OSNet modeli koriste specijalizirane blokove čija paralelna struktura otežava optimizaciju i nije u potpunosti prilagođena optimizacijama u postojećim GPU bibliotekama, dok se ResNet oslanja na standardne konvolucijske operacije koje su maksimalno optimizirane u okviru CUDA/cuDNN okruženja. S druge strane, prosječno vrijeme usporedbe vektora značajki upita s vektorima značajki galerije kod ResNet modela više je od tri puta dulje zbog veće dimenzionalnosti njegovih izlaznih vektora.

Tablica 6.1: Usporedba karakteristika ReID modela.

Model	**Br. značajki**	**Parametri**	**GFLOPs**	**Latencija** (ms)	**Vrijeme pretrage** (ms)
OSNet 0.75	512	1 398 614	1.07	16.751	1.226
OSNet 1.0	512	2 282 368	1.82	16.479	1.226
ResNet50	2048	23 958 812	4.88	6.610	4.024

Tablica 6.2 prikazuje usporedne rezultate evaluacije odabranih ReID modela na SSMOT ReID testnom podskupu s galerijom od 488 slika, i to u varijanti s unaprijed naučenim težinama na ImageNet [125] skupu podataka te nakon dodatnog učenja na domenski specifičnom SSMOT ReID skupu podataka.

Tablica 6.2: Usporedba performansi ReID modela s ImageNet težinama i nakon dodatnog učenja na SSMOT ReID skupu podataka.

Model	**Skup podataka**	**Sličnost**	**mAP**	**Rank-1**	**Rank-5**	**Rank-10**
OSNet 0.75	*ImageNet*	kosinusna	57.2	82.0	94.3	96.7
		euklidska	52.1	79.5	92.6	94.3
	SSMOT ReID	kosinusna	88.0	97.5	100.0	100.0
		euklidska	86.6	97.5	100.0	100.0
OSNet 1.0	*ImageNet*	kosinusna	58.7	85.2	94.3	98.4
		euklidska	50.0	82.8	91.8	95.9
	SSMOT ReID	kosinusna	88.7	95.9	100.0	100.0
		euklidska	87.6	95.9	100.0	100.0
ResNet50	*ImageNet*	kosinusna	51.5	77.9	89.3	94.3
		euklidska	50.6	77.0	89.3	95.9
	SSMOT ReID	kosinusna	80.6	94.3	99.2	100.0
		euklidska	80.4	92.6	98.4	100.0

U pravilu se pokazalo da korištenje kosinusne sličnosti pri pretrazi u galeriji daje bolje rezultate u odnosu na euklidsku udaljenost, što je osobito izraženo kod modela evaluiranih izravno s ImageNet težinama. Mogući razlog tome je što kosinusna sličnost naglašava kutnu sličnost između vektora te je, za razliku od euklidske udaljenosti, manje osjetljiva na razlike u njihovoj magnitudi koje ne moraju nužno ukazivati na različite identitete. Nadalje, dodatno treniranje na SSMOT ReID skupu značajno poboljšava performanse svih modela, što potvrđuje važnost prilagodbe na specifičnu domenu primjene.

OSNet modeli, koji su prilagođeni zadatku reidentifikacije, pokazuju bolje performanse od općeg ResNet modela, a razlika je posebno izražena u mAP vrijednostima. Primjerice, uz kosinusnu sličnost, OSNet 1.0 postiže mAP od 88.7, OSNet 0.75 od 88.0, dok ResNet50 ostvaruje 80.6, što jasno pokazuje prednost OSNet arhitekture. Navedeni rezultati ukazuju na to da OSNet modeli dosljednije rangiraju pozitivne uzorke više u listi rezultata, što ih čini pouzdanijima u scenarijima gdje je važno osigurati precizno rangiranje svih relevantnih uzoraka. U kontekstu praćenja plovila, to znači da model visoko rangira sva pojavljivanja istog plovila, što je vrlo važno kada se koraku asocijacije za danu putanju koriste značajke plovila iz više prethodnih vremenskih trenutaka. S druge strane, Rank-1 metrika pokazuje koliko često sustav odmah odabire ispravan identitet, što je ključno za izbjegavanje pogrešnih dodjela identiteta, budući da MOT algoritmi obično biraju najvjerojatnije kandidate za nastavak praćenja. Od razmatranih modela, OSNet 0.75 postiže najbolju vrijednost Rank-1 metrike od 97.5.

Korištenjem zahtjevnijeg scenarija pretrage s proširenom galerijom koja sadrži 1138 slika, vrijednosti ReID metrika očekivano su lošije, budući da veći broj kandidata otežava pronalazak pozitivnih primjera za dani upit. Tablica 6.3 prikazuje utjecaj proširenja galerije na performanse odabranih ReID modela, pri čemu svi modeli koriste kosinusnu mjeru sličnosti prilikom rangiranja. I ovdje vrijede prethodna zapažanja: OSNet arhitekture i dalje nadmašuju ReSNet50, a domenski prilagođeni modeli imaju bolje performanse od modela s težinama naučenim na ImageNet-u. U slučaju proširene galerije najbolje rezultate postiže OSNet 1.0 treniran na SSMOT ReID skupu, s mAP vrijednošću od 82.8 i Rank-1 rezultatom od 95.1. Slika 6.5 prikazuje šest najbolje rangiranih primjera (Top-6) iz proširene galerije za dva upita (označena lijevo), dobivenih različitim ReID modelima. Rezultati OSNet 1.0 modela s kosinusnom mjerom sličnosti uspoređeni su s onima temeljenima na euklidskoj udaljenosti. Nadalje, prikazani su i TOP-6 rezultati OSNet 1.0 modela s ImageNet težinama, kao i usporedbe s OSNet 0.75 te ResNet50 modelom. Zelenim okvirom označeni su pozitivni primjeri čiji identitet odgovara identitetu upita, dok su crvenim okvirom označeni negativni primjeri. Iz danih primjera jasno je vidljiv zaostatak ImageNet modela za onima treniranim na SSMOT ReID podacima: kod drugog upita ImageNet model prvi pozitivan primjer stavlja tek na šestu poziciju, dok se kod prvog upita uopće ne pojavljuje u Top-6 rezultata.

Dodatni primjeri upita i pripadajućih TOP-6 rangiranih rezultata iz proširene galerije, dobivenih različitim modelima, prikazani su u Dodatku D.

Tablica 6.3: Rezultati evaluacije ReID modela uz i bez proširenja galerije.

Model	**Skup podataka**	**Proširena galerija**	**mAP**	**Rank-1**	**Rank-5**	**Rank-10**
OSNet 0.75	*ImageNet*	✗	57.2	82.0	94.3	96.7
		✓	47.5	76.2	91.0	94.3
	SSMOT ReID	✗	88.0	97.5	100.0	100.0
		✓	81.7	94.3	100.0	100.0
OSNet 1.0	*ImageNet*	✗	58.7	85.2	94.3	98.4
		✓	49.1	77.9	90.2	94.3
	SSMOT ReID	✗	88.7	95.9	100.0	100.0
		✓	82.8	95.1	99.2	100.0
ResNet50	*ImageNet*	✗	51.5	77.9	89.3	94.3
		✓	41.8	72.1	84.4	88.5
	SSMOT ReID	✗	80.6	94.3	99.2	100.0
		✓	75.9	91.8	97.5	100.0

Slika 6.5: Šest najbolje rangiranih rezultata u galeriji za dva upita, dobivenih različitim ReID modelima i mjerama sličnosti.

7. VOc-SORT ALGORITAM PRAĆENJA S POBOLJŠANIM PERFORMANSAMA U SLUČAJU DUGOTRAJNIH OKLUZIJA

Iako detekcija i reidentifikacija predstavljaju temeljne komponente sustava za praćenje plovila, one same po sebi nisu dovoljne za osiguravanje konzistentnog identiteta objekata kroz vrijeme. Detekcija djeluje na razini pojedinačnih okvira videozapisa te daje informaciju o trenutnom položaju i klasi plovila u danom okviru. S druge strane, reidentifikacija procjenjuje sličnost plovila koja se pojavljuju u različitim okvirima videozapisa. No, bez dodatne logike povezivanja opažanja plovila iz različitih okvira videozapisa, koja je dana algoritmom praćenja, nije moguće uspostaviti stabilan kontinutet identiteta plovila kroz vrijeme. To je nužno jer bez takvog kontinuiteta sustav ne može izvući smislene informacije o putanji i ponašanju pojedinih plovila, niti prepoznati odstupanja u njihovom ponašanju, što je od presudne važnosti za nadzor, sigurnost i analizu pomorskog prometa.

Unatoč značajnom napretku postignutom u domeni automatskog praćenja, postojeći algoritmi i dalje se suočavaju s izazovima kada je objekt koji se prati određeni vremenski period zaklonjen. Dok vodeće metode praćenja obično mogu uspješno riješiti problem kratkotrajnih okluzija, dugotrajne okluzije i dalje ozbiljno narušavaju stabilnost praćenja te često rezultiraju gubitkom putanja, pogrešnim dodjelama identiteta ili fragmentacijama putanja. Problem dugotrajnih okluzija posebno je izražen u kontekstu praćenja plovila, gdje dinamika scene obično uključuje međusobna zaklanjanja objekata različitih dimenzija. Manja plovila mogu na dulje vrijeme potpuno nestati iza većih, čime se stvaraju dugotrajne okluzije koje nadilaze mogućnosti postojećih algoritama.

Kako bi se prevladala ograničenja postojećih metoda, razvijen je **VOc-SORT** (*Vessel **Oc**clusion SORT*) algoritam za praćenje plovila, utemeljen na klasičnom pristupu praćenja temeljenom na detekciji, ali proširen nizom ciljanih prilagodbi koje zajednički doprinose njegovoj većoj otpornosti na dugotrajne okluzije. Predloženi algoritam uvodi domensko znanje u postupak odlučivanja o uklanjanju izgubljenih putanja te poboljšava proces asocijacije detekcija i postojećih putanja korištenjem dvostupanjske metode povezivanja. U prvoj fazi asocijacije vizualne se značajke kombiniraju s geometrijskima kako bi se pouzdano pridružili jasni parovi putanja i detekcija, dok se u drugoj fazi za povezivanje preostalih, nejasnih slučajeva integriraju vizualne i dinamičke značajke, čime se omogućuje robusna reasocijacija u uvjetima okluzija, nestabilnih detekcija i naglih promjena kretanja. Time VOc-SORT doprinosi ublažavanju određenih ograničenja postojećih metoda i povećava stabilnost praćenja u zahtjevnim pomorskim scenama, zadržavajući pritom jednostavan i praktičan dizajn. Duboki modeli koriste se isključivo za detekciju i izdvajanje vizualnih značajki, dok ostatak algoritma ostaje minimalno računski zahtjevan čime se postiže uravnotežen odnos između računske učinkovitosti i kvalitete samog praćenja.

U ovom se poglavlju najprije predstavlja dizajn predloženog algoritma praćenja, pri čemu se ističu i detaljno opisuju ključne prilagodbe u odnosu na postojeće metode te se daje pregled osnovnih koraka samog algoritma. Nakon toga slijedi usporedba performansi predloženog algoritma s relevantnim postojećim metodama praćenja. Zatim se provodi ablacijska studija u kojoj se analizira doprinos pojedinih komponenti algoritma - korištenog detektora, modela za ekstrakciju vizualnih značajki, dodatnog kriterija za uklanjanje putanja te načina asocijacije putanja i detekcija, zajedno s pripadnim pragovima vizualne sličnosti. Poglavlje završava raspravom u kojoj se interpretiraju dobiveni rezultati i ističu prednosti, ali i ograničenja predloženog algoritma.

7.1. Dizajn VOc-SORT algoritma

U nastavku je detaljno opisan predloženi VOc-SORT algoritam praćenja, zajedno sa svim njegovim sastavnim koracima. Detaljno je pojašnjen način na koji se informacije dobivene detektorom, modelom za ekstrakciju značajki i Kalmanovim filterom međusobno integriraju kako bi se osigurao stabilan kontinuitet identiteta plovila, čak i u uvjetima dugotrajnih okluzija.

7.1.1. Ključne prilagodbe u odnosu na postojeće metode

Kako bi se smanjila osjetljivost algoritma praćenja na dugotrajne okluzije koje uzrokuju prekid putanja i gubitak identiteta plovila, VOc-SORT uvodi niz modifikacija usmjerenih na povećanje robusnosti sustava upravo u takvim situacijama. Naglasak je na omogućavanju pouzdane ponovne identifikacije plovila nakon duljih razdoblja bez opažanja i na smanjenje pogreške u koraku asocijacije. Sljedeće prilagodbe čine ključne razlike predloženog algoritma u odnosu na postojeće pristupe.

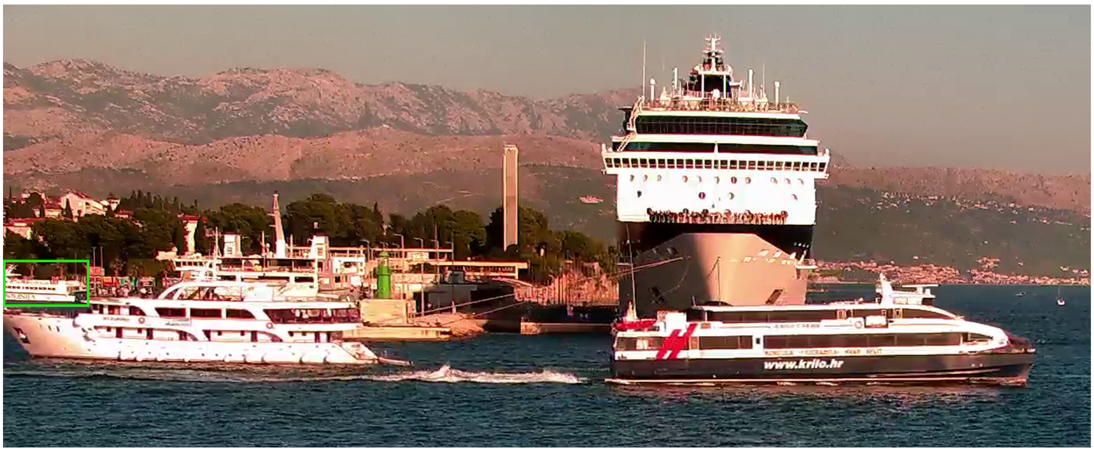
1. Produljeno vrijeme zadržavanja izgubljenih putanja

U većini standardnih algoritama praćenja putanja se zadržava vrlo kratko vrijeme nakon što joj je posljednji put bila pridružena detekcija, obično oko 30 uzastopnih okvira [67, 61, 66, 152] ili čak kraće [58, 27]. Prag zadržavanja putanja od 30 okvira može biti dovoljan za kratkotrajne okluzije, ali je nedostatan u scenarijima praćenja plovila, gdje objekti često ostaju zaklonjeni znatno dulje.

Kako bi se omogućila pouzdana rekonstrukcija putanja plovila i nakon dugotrajnih prekida vidljivosti, u VOc-SORT algoritmu znatno je produljeno vremensko razdoblje tijekom kojega se putanja zadržava i u odsutnosti detekcija. Novi prag od 625 okvira odgovara razdoblju od 25 sekundi pri standardnoj frekvenciji od 25 FPS, što je značajan porast u odnosu na otprilike 1.2 sekunde koliko dopuštaju standardne implementacije. Time algoritam postaje tolerantniji na dulje okluzije i smanjuje se vjerojatnost nepotrebnog prekida putanja.

2. Dodatni kriterij upravljanja putanjama temeljen na domenskom znanju

Uz osnovno pravilo brisanja putanja nakon što im 625 uzastopnih okvira nije pridružena detekcija, uveden je i dodatni kriterij koji koristi domensko znanje o uobičajenom kretanju plovila u promatranoj luci. Ako putanja pet uzastopnih okvira ne dobije niti jednu pridruženu detekciju, a posljednja detekcija koja joj je pridružena nalazi se u blizini lijevog ili desnog ruba kadra, putanja se odmah uklanja. Ovakva situacija najčešće upućuje na to da je plovilo napustilo vidno polje kamere, pa se bržim uklanjanjem takvih putanja sprječava bespotrebno gomilanje izgubljenih putanja i smanjuje mogućnost pogrešnih asocijacija u narednim okvirima. Prije ranijeg uklanjanja provodi se provjera statičnosti putanje. Ako je plovilo bilo praktički nepomično, putanja se ne uklanja nakon samo pet izostanaka detekcije. Time se sprječava pogrešno uklanjanje putanja koje pripadaju privezanim plovilima u dijelu luke uz lijevi rub kadra (kao na primjeru sa Slike 7.1), a koja privremeno mogu biti zaklonjena drugim objektima.



Slika 7.1: Primjer privezanog plovila u luci smještenog uz lijevi rub kadra, koje će u narednim okvirima biti zaklonjeno drugim prolaznim plovilom te zbog toga privremeno neće biti detektirano.

3. Dvostupanjska VG-VD asocijacija detekcija i putanja

U okviru predloženog VOc-SORT algoritma uveden je postupak poboljšane dvostupanjske asocijacije, nazvan **VG-VD (*Vizualno Geometrijska - Vizualno Dinamička*) asocijacija**, koji objedinjuje komplementarne izvore informacija pri povezivanju postojećih putanja i novih detekcija.

Prva faza asocijacije objedinjuje IoU udaljenost, kao geometrijski pokazatelj prostornog preklapanja detekcije i predviđanja dane putanje, te kosinusnu udaljenost vektora značajki koji opisuju njihovu vizualnu sličnost. Nasuprot tome, druga faza asocijacije kombinira kosinusnu udaljenost vizualnih značajki s Mahalanobisovom udaljenošću, koja opisuje koliko je detekcija konzistentna s očekivanom dinamikom kretanja praćenog objekta, pri čemu se u obzir uzima i nesigurnost modela kretanja. Takva kombinacija omogućuje uspješno povezivanje i u slučajevima kada je prostorno preklapanje slabo ili nepostojeće, primjerice nakon okluzija ili naglih pomaka, čime se nadopunjuju ograničenja prve faze i povećava ukupna robusnost asocijacije.

Neka je $\mathcal{T} = \{T_1, \ldots, T_n\}$ skup postojećih putanja, a $\mathcal{D}_t = \{D_1, \ldots, D_m\}$ skup detekcija u okviru t danog videozapisa. U nastavku se oznaka $d_{i,j}$ koristi kao pokrata za $d(T_i, D_j)$, gdje je d odabrana mjera udaljenosti, $T_i \in \mathcal{T}$ i $D_j \in \mathcal{D}$.

Prva faza asocijacije (VG)

Cijena pridruživanja $C_1(T_i, D_j)$ detekcije D_j putanji T_i u prvoj fazi asocijacije dana je s:

$$C_1(T_i, D_j) = \hat{d}^{cos}_{i,j} + d^{IoU}_{i,j}, \tag{7.1}$$

gdje je

$$\hat{d}^{cos}_{i,j} = \begin{cases} \frac{1}{2} d^{cos}_{i,j}, & \text{ako je } (d^{cos}_{i,j} < \theta_{vis}) \land (d^{IoU}_{i,j} < \theta_{IoU}) \\ 1, & \text{inače} \end{cases}. \tag{7.2}$$

Pritom $d^{cos}_{i,j}$ označava kosinusnu udaljenost vektora značajki detekcije D_j i eksponencijalnog pomičnog prosjeka vektora značajki detekcija pridruženih putanji T_i, a $d^{IoU}_{i,j}$ IoU udaljenost predviđenog graničnog okvira putanje T_i i graničnog okvira detekcije D_j. Pridruživanja za koja vrijedi $C_1 > 0.8$ odbacuju se jer ukazuju na izraženo vizualno ili geometrijsko odstupanje. Prag $\theta_{IoU} = 0.7$ služi kao dodatno ograničenje na preklapanja graničnih okvira, te isključuje parove s nedostatnom prostornom podudarnošću. S druge strane, $\theta_{vis} = 0.4$ predstavlja prag vizualne sličnosti kojim se odbacuju parovi sa znatno različitim vizualnim značajkama. Navedeni pragovi odabrani su temeljem inicijalnih eksperimenata i kvalitativne procjene, te ne predstavljaju nužno optimalne vrijednosti.

Druga faza asocijacije (VD)

U drugoj fazi asocijacije *neuparene* putanje i detekcije iz prve faze pokušavaju se još jednom povezati, ali sada koristeći cijenu $C_2(T_i, D_j)$ pridruživanja detekcije $D_j \in \mathcal{D}$ putanji $T_i \in \mathcal{T}$. Ova cijena kombinira vizualnu udaljenost ${d^*_{i,j}}^{cos}$ s konzistentnošću dinamike kretanja objekta opisanom Mahalanobisovom udaljenošću $d^{MhD}_{i,j}$, prema [62]. Funkcija cijene definirana je na sljedeći način:

$$C_2(T_i, D_j) = \begin{cases} \infty, & \text{ako je } d_{i,j}^{MhD} > \theta_{MhD} \\ \lambda d_{i,j}^{*\,cos} + (1-\lambda) d_{i,j}^{MhD}, & \text{inače} \end{cases}, \tag{7.3}$$

gdje je vizualna udaljenost $d_{i,j}^{*\,cos}$ dana s:

$$d_{i,j}^{*\,cos} = \begin{cases} 2, & \text{ako je } d_{i,j}^{cos} > \theta_{vis} \\ d_{i,j}^{cos}, & \text{inače} \end{cases}, \tag{7.4}$$

pri čemu $d_{i,j}^{cos}$, kao i u prvoj fazi, označava kosinusnu udaljenost vektora značajki detekcije D_j i eksponencijalnog pomičnog prosjeka vektora značajki detekcija pridruženih putanji T_i.

U implementaciji VOc-SORT algoritma koristi se $\lambda = 0.98$ kojim se naglasak stavlja na vizualnu sličnost, dok se Mahalanobisova udaljenost koristi kao dopunski uvjet konzistentnosti. Granična vrijednost θ_{MhD} jednaka je vrijednosti 95%-tnog kvantila χ^2 distribucije s četiri stupnja slobode, koji iznosi 9.4877, te se koristi za odbacivanje pridruživanja detekcija putanjama koja nisu dinamički konzistentna s njihovim prethodnim kretanjem[1]. Prag vizualne sličnosti, θ_{vis}, i u drugoj fazi asocijacije postavljen je na vrijednost 0.4.

Vrijednost $d_{i,j}^{MhD}$ iz jednadžbe (7.3) kvadrat je Mahalanobisove udaljenosti [138] koji mjeri usklađenost nove detekcije D_j i predviđanja za putanju T_i dobivenog Kalmanovim filterom. Definirana je izrazom:

$$d_{i,j}^{MhD} = (D_j - \hat{\mathbf{z}}_i)^T S_i^{-1} (D_j - \hat{\mathbf{z}}_i), \tag{7.5}$$

gdje je $\hat{\mathbf{z}}_i$ projekcija predviđenog stanja $\hat{\mathbf{x}}_i$ putanje T_i u prostor mjerenja, a S_i pripadna matrica kovarijance pogreške predviđanja u tom prostoru. $\hat{\mathbf{z}}_i$ i S_i dane su sljedećim relacijama:

$$\hat{\mathbf{z}}_i = H\hat{\mathbf{x}}_i, \quad S_i = HP_iH^T + R, \tag{7.6}$$

pri čemu sve matrice dolaze iz Kalmanovog filtera: H je matrica koja preslikava vektor stanja u prostor mjerenja, P_i je matrica kovarijance pogreške predviđenog stanja, a R matrica kovarijance mjernog šuma detektora.

Mahalanobisova udaljenost omogućuje da se svako odstupanje detekcije od predviđenog stanja putanje vrednuje u odnosu na sigurnost Kalmanovog filtera za dano predviđanje: ako je model siguran u svoje predviđanje, prihvatljiva su samo mala odstupanja, a u slučajevima veće nesigurnosti toleriraju se i veća odstupanja. Ovo je posebno važno u slučaju dugotrajnih okluzija, tijekom kojih nisu dostupna nova mjerenja za korekciju vrijednosti Kalmanovog fil-

[1] Budući da se dinamička konzistentnost procjenjuje nad prostorom mjerenja koji obuhvaća četiri komponente graničnog okvira: dvije koordinate njegovog središta (x_c, y_c), širinu w i visinu h, test odbacivanja se temelji na χ^2 distribuciji s četiri stupnja slobode [320].

tera, čime se njegova nesigurnost u dana predviđanja postupno povećava. Kako nesigurnost raste, tako se povećava i područje prihvatljivih vrijednosti Mahalanobisove udaljenosti, što algoritmu omogućuje da nakon dužeg izostanka detekcija prihvati detekciju koja je možda znatnije udaljena od početnog predviđanja. Time se povećava vjerojatnost ispravne reidentifikacije plovila kada se ono ponovno pojavi nakon duljeg perioda zaklonjenosti.

4. OSNet ReID model treniran na domeni plovila

Za ekstrakciju vizualnih značajki koristi se *OSNet 1.0* ReID model, dodatno treniran na SSMOT podskupu za reidentifikaciju kako bi se poboljšala kvaliteta dobivene vizualne reprezentacije dodatnom prilagodbom specifičnostima izgleda plovila u promatranoj luci. Time se izbjegavaju ograničenja općih modela koji su najčešće treniranih na skupovima podataka s pješacima, a zbog znatnih razlika u vizualnim obilježjima plovila i pješaka obično ne mogu pouzdano razlikovati pojedina plovila. Dodatnim treniranjem ReID modela na domeni od interesa postiže se veća diskriminativnost dobivenih značajki, što izravno doprinosi boljim performansama algoritma praćenja te stabilnijem održavanju identiteta objekata koji se prate.

5. Veći prag tolerancije na udaljenost vizualnih značajki

Zbog mogućih promjena u izgledu plovila nakon okluzije (djelomična zaklonjenost, različita orijentacija ili slično), koristi se nešto veća granična vrijednost vizualne udaljenosti od 0.4 za prihvatljiva pridruživanja novih detekcija i postojećih putanja, dok se u postojećim implementacijama popularnih algoritama uobičajeno primjenjuju niže vrijednosti od $\approx 0.2 - 0.3$ [61, 62, 63].

7.1.2. Opis koraka algoritma

Slijedi pregled svih koraka predloženog VOc-SORT algoritma, od inicijalne detekcije plovila pa sve do koraka upravljanja putanjama. Kako bi se dobio što jasniji uvid u samu strukturu i način rada algoritma, dijagramom toka na Slici 7.2 prikazani su odnosi među navedenim koracima i ključne odluke u procesu praćenja.

Detekcija

Za lokalizaciju i klasifikaciju plovila u svakom okviru videozapisa u fazi detekcije koristi se **YOLO11m** [288] detektor, opisan u Poglavlju 5, koji je treniran na SSMOT skupu podataka. Odabrani detektor generira granične okvire i pripadajuće vrijednosti pouzdanosti za sva plovila prisutna u sceni, čime se osigurava početna informacija potrebna za daljnje korake algoritma praćenja. Treniranje na SSMOT podskupu za detekciju omogućuje prilagodbu detektora specifičnostima dinamike pomorskog prometa u splitskoj luci, čime se postiže veća preciznost i robusnost detekcija u stvarnim uvjetima nadzora plovila u toj luci.

Ekstrakcija vizualnih značajki

Za potrebe reidentifikacije plovila i poboljšanja robusnosti algoritama praćenja u uvjetima dugotrajnih okluzija, nakon detekcije provodi se ekstrakcija vizualnih značajki svakog detektiranog plovila. U tu svrhu koristi se **OSNet 1.0** [228] reidentifikacijski model, koji je u Poglavlju 6 pokazao bolje performanse od općeg ResNet modela. Model je treniran na SSMOT ReID skupu podataka, čime je osigurana prilagođenost karakteristikama plovila prisutnim u stvarnim uvjetima nadzora. OSNet 1.0 generira 512-dimenzionalne diskriminativne vektore značajki koji opisuju vizualni izgled svakog plovila, čime omogućuje razlikovanje plovila vrlo sličnog izgleda i olakšava ponovno povezivanje putanja nakon razdoblja bez detekcija.

Predviđanje sljedeće pozicije objekta

Predviđanje budućeg položaja objekta važan je korak u algoritmima praćenja jer omogućuje zadržavanje kontinuiteta putanja čak i u trenutcima kada detektor privremeno ne uspijeva lokalizirati plovilo. U implementaciji predloženog VoC-SORT algoritma za ovaj korak koristi se **Kalmanov filter** (KF) [26] s vektorom stanja

$$\mathbf{x} = (x_c,\, y_c,\, w,\, h,\, \dot{x}_c,\, \dot{y}_c,\, \dot{w},\, \dot{h}), \tag{7.7}$$

gdje su $(x_c,\, y_c)$ koordinate središta graničnog okvira, w i h njegova širina i visina, a komponente $\dot{x}_c,\, \dot{y}_c,\, \dot{w},\, \dot{h}$ predstavljaju stope promjene položaja i dimenzija graničnog okvira u vremenu. Na ovaj način omogućena je istovremena procjena položaja i dinamike gibanja objekta.

Predviđanje se temelji na linearnom modelu gibanja, pri čemu Kalmanov filter koristi informacije o prethodnom stanju, kako bi procijenio očekivanu poziciju i veličinu plovila u trenutnom okviru videozapisa (`KF.predict` na vizualizaciji sa Slike 7.2). To u ovom slučaju nije ograničavajuće jer se promjene položaja plovila između uzastopnih okvira videozapisa obično odvijaju postepeno, bez naglih pomaka, pa ih linearni model može dovoljno dobro aproksimirati. Čak i kada plovilo značajno mijenja smjer, ta se promjena odvija kroz veći broj uzastopnih okvira, što omogućuje Kalmanovom filteru da je pravovremeno registrira i prilagodi svoje procjene. U slučaju da detekcija izostane, predviđanje Kalmanovog filtera koristi se za privremenu procjenu pozicije plovila, čime se sprječava trenutni prekid putanje i omogućuje održavanje njezina kontinuiteta.

Asocijacija i korištene mjere sličnosti

Jedna od glavnih razlika predloženog VOc-SORT algoritma u odnosu na postojeće metode je u načinu povezivanja detekcija s postojećim putanjama koji omogućuje robusnije praćenje i u uvjetima dugotrajnih okluzija.

Kako bi se nove detekcije iz trenutnog okvira videozapisa povezale s postojećim putanjama koristi se dvostupanjska VG-VD asocijacija predstavljena ranije u 7.1.1. U prvoj fazi asocijacije cilj je utvrditi najpouzdanija pridruživanja između detekcija i postojećih putanja, odnosno parove koji su istovremeno prostorno blizu i vizualno vrlo slični. Za to se koristi mađarski algoritam [140] s cijenom pridruživanja C_1, definiranom jednadžbom (7.1), koja spaja IoU udaljenost i kosinusnu udaljenost vizualnih značajki u jedinstvenu metriku temeljem koje se donosi odluka o optimalnom povezivanju putanja i detekcija.

Detekcije i putanje koje nisu uspješno povezane u prvoj fazi, prosljeđuju se u drugu fazu asocijacije u kojoj ih se ponovno pokušava povezati mađarskim algoritmom, ali sada kombinirajući kosinusnu udaljenost vizualnih značajki s Mahalanobisovom udaljenošću u cijeni pridruživanja C_2 iz (7.3). U situacijama kada se objekt ponovno pojavljuje nakon okluzije, IoU više ne pruža korisnu informaciju zbog izostanka preklapanja, a sama kosinusna udaljenost nije dostatna jer vizualne značajke nakon okluzija ili u prisutnosti vizualno sličnih plovila mogu postati nepouzdane i dovesti do većeg broja zamjena identiteta. Stoga druga faza asocijacije kombinira kosinusnu udaljenost s većom graničnom vrijednošću 0.4, koja omogućuje povezivanje istog objekta i u slučaju određenih odstupanja u vizualnom izgledu, zajedno s Mahalanobisovom udaljenošću koja provjerava je li nova detekcija dinamički usklađena s očekivanim položajem putanje, čime se smanjuje rizik od pogrešnih dodjela identiteta pri ponovnom pojavljivanju objekta nakon okluzije. Time se postiže ravnoteža: prva faza sprječava pogrešne asocijacije, a druga faza smanjuje fragmentaciju putanja nakon dugotrajnih okluzija.

Nakon dvije faze asocijacije u kojima se postojećim putanjama nastoje pridružiti nove detekcije, slijedi dodatna faza asocijacije u kojoj se preostale neuparene detekcije pokušavaju povezati s nepotvrđenim putanjama koje su još u probnom razdoblju. U ovoj fazi također se primjenjuje mađarski algoritam, koristeći istu funkciju cijene pridruživanja C_1 kao i u prvoj fazi asocijacije, ali s duplo manjom graničnom vrijednosti θ_{vis}.

Upravljanje putanjama

Upravljanje putanjama obuhvaća ažuriranje stanja postojećih putanja kojima je uspješno pridružena nova detekcija plovila, uklanjanje neaktivnih putanja te inicijalizaciju novih putanja za neuparene detekcije.

Ažuriranje stanja putanja

Kada je putanji $T_i \in \mathcal{T}$ uspješno pridružena detekcija $D_j \in \mathcal{D}_t$ plovila iz trenutnog okvira F_t danog videozapisa, provodi se korak korekcije Kalmanovog filtera (`KF.update` na Slici 7.2), pri čemu se procjena stanja putanje i pripadajuća matrica kovarijance pogreške procjene ažuriraju temeljem novog mjerenja D_j. Potom se ažuriraju i reidentifikacijske značajke putanje (`features.update` na Slici 7.2). Kao vizualna informacija putanje T_i koristi se eksponencijalni pomični prosjek:

$$e_i^t = \alpha e_i^{t-1} + (1-\alpha) f_j, \tag{7.8}$$

gdje je f_j vektor značajki nove detekcije D_j koja je u okviru F_t pridružena putanji T_i, a $\alpha = 0.9$ faktor izglađivanja. Eksponencijalni pomični prosjek integrira vizualne informacije sadržane u uzastopnim okvirima videozapisa smanjujući šum prisutan u detekcijama, čime se poboljšava kvaliteta asocijacije uz smanjenu vremensku potrošnju [62].

Uklanjanje putanja

Ako broj uzastopnih okvira u kojima putanji $T_i \in \mathcal{T}$ nije pridružena detekcija premaši zadani prag `max_age` $= 625$, objekt se smatra izgubljenim, te se putanja uklanja iz skupa $\mathcal{T}$. Nadalje, u slučaju da tijekom 5 uzastopnih okvira putanji nije pridružena niti jedna detekcija, a posljednja detekcija pridružena toj putanji nalazi se u neposrednoj blizini lijevog ili desnog ruba okvira te se ne radi o statičnom objektu koji bi mogao predstavljati plovilo privezano u luci, putanja se uklanja odmah, bez čekanja isteka praga `max_age`.

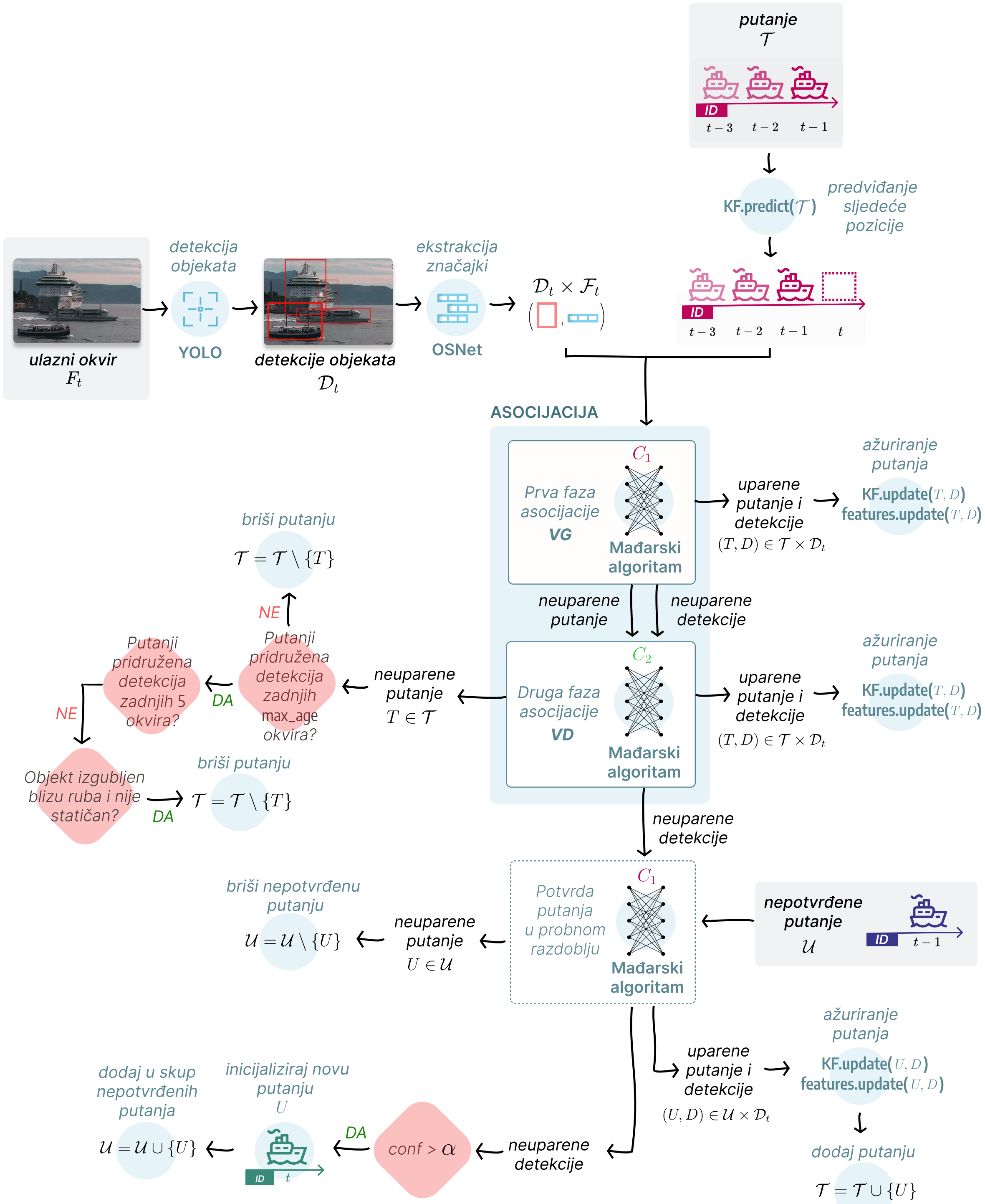
Inicijalizacija putanja

Kada se u okviru pojavi detekcija koja nije pridružena niti jednoj postojećoj putanji, ona se prvo dodaje u skup probnih, nepotvrđenih putanja $\mathcal{U}$. Te putanje služe kao privremeni kandidati te još ne ulaze u glavni skup putanja $\mathcal{T}$. Ako se istoj putanji iz $\mathcal{U}$ detekcija uspješno pridruži i u sljedećem uzastopnom okviru, putanja se smatra potvrđenom i inicijalizira se kao nova aktivna putanja u $\mathcal{T}$ ako je pouzdanost detekcije veća od $\alpha = 0.6$. Suprotno tome, ako u idućem okviru ne dođe do uspješnog pridruživanja, probna se putanja uklanja iz skupa $\mathcal{U}$. Na taj se način osigurava da se nova putanja kreira tek kada postoji potvrda pojave objekta u dva uzastopna okvira, čime se smanjuje mogućnost stvaranja putanja temeljenih na lažno pozitivnim detekcijama.

7.2. Usporedba performansi predloženog algoritma s postojećim metodama

Procjena učinkovitosti predloženog VOc-SORT algoritma provedena je usporednom analizom s reprezentativnim algoritmima temeljenim na detekciji, koji predstavljaju vodeći pristup rješavanju zadatka praćenja više objekata [147, 148]. Cilj ovog potpoglavlja je prikazati u kojoj mjeri predložene inovacije doprinose poboljšanju performansi u odnosu na postojeće metode, posebno u zahtjevnim scenarijima dugotrajnih okluzija.

Usporedba je provedena nad skupom algoritama čije su referentne implementacije dostupne u BoxMOT okviru [321], standardiziranom sustavu otvorenog koda koji okuplja suvre-



Slika 7.2: Dijagram toka predloženog VOc-SORT algoritma.

mene predstavnike algoritama praćenja više objekata. Primjena ovog okvira omogućuje evaluaciju svih metoda u jedinstvenom i reproducibilnom eksperimentalnom okruženju, čime se uklanjaju implementacijske razlike i osigurava objektivna usporedba njihovih performansi.

7.2.1. Korištene metrike

Za evaluaciju različitih metoda praćenja korišten je standardizirani evaluacijski okvir **TrackEval** [322], koji omogućuje ujednačenu i reproducibilnu procjenu performansi algoritama na temelju široko prihvaćenih metrika, uključujući CLEAR MOT [153] metrike MOTA i MOTP, HOTA metriku [154], IDF1 metriku [155], te IDSW [142] metriku. Detaljan opis navedenih metrika dan je u Poglavlju 2, u dijelu 2.4.1.

Uz prethodno navedene metrike, koristi se i FPS (engl. *Frames Per Second*) mjera, koja daje uvid u brzinu izvođenja algoritma praćenja. Pri tome je važno naglasiti da vrijeme potrebno za detekciju nije uključeno u FPS, jer su detekcije za sve videozapise unaprijed generirane odabranim detektorom, te se tijekom evaluacije samo učitavaju bez ponovnih izračuna. Na taj se način ubrzava proces evaluacije i omogućuje konzistentna usporedba performansi različitih algoritama praćenja.

Kako bi se ocijenila uspješnost modela u ponovnoj identifikaciji objekata nakon okluzije, uvodi se OKL^+ metrika. Za njezin izračun najprije je potrebno za svaku anotiranu okluziju identificirati posljednju dostupnu detekciju zaklonjenog plovila u zadanom vremenskom prozoru prije početka stvarne okluzije, a zatim i prvu ponovnu detekciju u odgovarajućem prozoru nakon okvira koji označuje njezin stvarni završetak. Uvođenje ovih prozora nužno je zbog ograničenja detekcijskih modela, koji često ne uspijevaju pouzdano detektirati djelomično zaklonjene objekte neposredno prije ulaska u okluziju i neposredno nakon izlaska iz nje. U ovoj analizi primjenjuje se vremenski prozor od 75 okvira prije i 75 okvira nakon okluzije, čime se osigurava dovoljna tolerancija na izostanke detekcija uslijed ograničenja detektora. Vrijednost OKL^+ metrike odgovara broju uspješno reidentificiranih objekata nakon okluzije, pri čemu se reidentifikacija smatra uspješnom ako je objekt detektiran i prije i nakon stvarnog razdoblja okluzije te ako je algoritam praćenja zadržao konzistentan identifikator u oba promatrana detekcijska trenutka.

7.2.2. Rezultati evaluacije

Kako bi se uklonile razlike u performansama koje proizlaze iz korištenja različitih detektora i ekstraktora vizualnih značajki, svi algoritmi koriste iste modele za detekciju i ekstrakciju značajki kao i predloženi VOc-SORT algoritam. Konkretno, svi modeli kao polaznu točku praćenja koriste iste detekcije dobivene YOLO11m detektorom treniranim na SSMOT skupu podataka za detekciju. U slučajevima kada algoritam praćenja koristi vizualne značajke, njihova ekstrakcija provodi se pomoću OSNet 1.0 modela treniranog na SSMOT ReID skupu podataka. Osim toga, za sve algoritme produljeno je vrijeme prije brisanja izgubljenih putanja na 625 okvira kako bi im se pružila mogućnost uspješnog nastavka praćenja objekata i nakon duljih okluzija.

Kvantitativna analiza

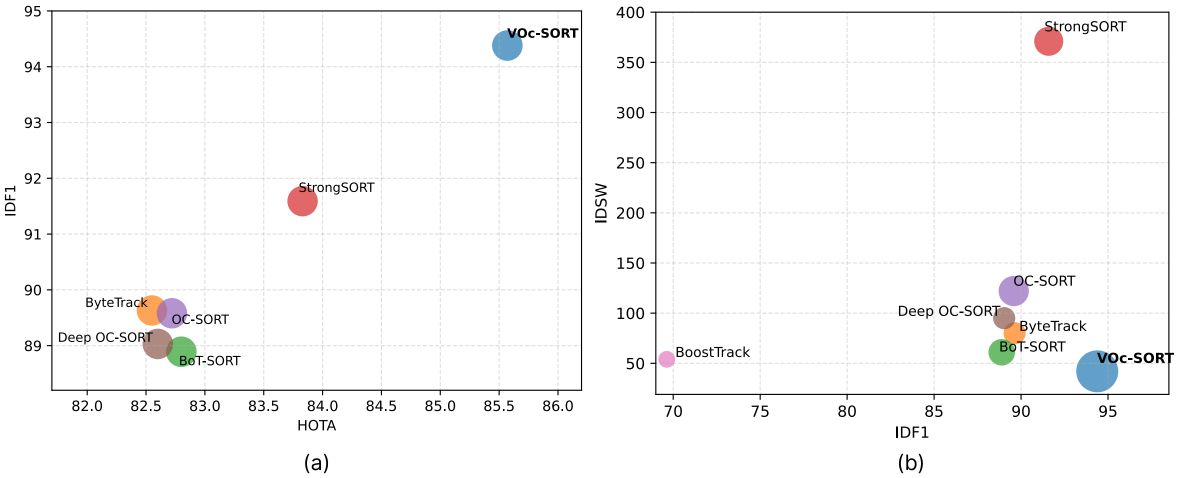
U Tablici 7.1 prikazani su rezultati kvantitativne evaluacije predloženog algoritma i odabranih postojećih metoda praćenja na SSMOT videozapisima. Simbol "✓" u stupcu *ReID* označuje da navedeni algoritam koristi reidentifikacijski model za ekstrakciju vizualnih značajki, dok simbol "✗" naznačuje da algoritam ne koristi vizualne značajke. Nadalje, strelica "↑" uz nazive pojedinih metrika ukazuje na to da su poželjne veće vrijednosti te metrike, dok strelica "↓" označuje da su povoljnije niže vrijednosti. Evaluacija je provedena na računalnom sustavu s procesorom Intel Core i7-9850H, 16 GB RAM-a i grafičkom karticom NVIDIA Quadro RTX 3000.

Tablica 7.1: Usporedna analiza performansi predloženog VOc-SORT algoritma i postojećih metoda praćenja na SSMOT skupu videozapisa.

Tracker	**ReID**	**HOTA** (↑)	**MOTA** (↑)	**MOTP** (↑)	**IDSW** (↓)	**IDF1** (↑)	**OKL$^+$** (↑)	**FPS** (↑)
ByteTrack [60]	✗	82.55	92.51	89.30	80	89.63	7	**201**
BoT-SORT [61]	✓	82.80	92.95	89.83	61	88.89	10	14
StrongSORT [62]	✓	83.83	92.21	89.78	371	91.59	12	13
OC-SORT [27]	✗	82.72	92.37	**89.89**	122	89.58	13	188
Deep OC-SORT [152]	✓	82.59	92.61	89.88	95	89.03	7	15
BoostTrack [63]	✓	65.79	56.49	89.70	**54**	69.63	4	47
VOc-SORT	✓	**85.57**	**93.03**	89.82	**42**	**94.38**	**25**	14

Predloženi VOc-SORT postiže najviše vrijednosti HOTA (85.57), MOTA (93.03) i IDF1 (94.38) metrika, čime nadmašuje ostale metode praćenja prema ključnim identifikacijskim i asocijacijskim metrikama. U pogledu točnosti lokalizacije objekata (MOTP), svi algoritmi ostvaruju slične vrijednosti, pri čemu najbolji rezultat postiže OC-SORT (89.89). VOc-SORT postiže najmanji broj zamjena identiteta (42) tijekom praćenja te ostvaruje znatno veći broj uspješnih reidentifikacija nakon okluzije (25/31) u odnosu na ostale algoritme. Po pitanju brzine izvođenja, s najvišom vrijednosti FPS ističu se ByteTrack (201) i OC-SORT (188) algoritmi, dok VOC-SORT postiže 14 FPS, što je usporedivo s ostalim metodama koje koriste vizualnu reidentifikaciju. Generalno najslabije performanse pokazuje BoostTrack, s HOTA vrijednošću od tek 65.79 i MOTA vrijednošću od 56.49.

Prikaz na Slici 7.3 (a) uspoređuje odabrane algoritme koristeći HOTA i IDF1 metrike, pri čemu veličina markera odražava vrijednost MOTA metrike. VOc-SORT algoritam se vidno izdvaja po vrijednostima HOTA i IDF1 metrika, slijedi ga StrongSORT algoritam koji se također izdvaja po vrijednostima ovih metrika, dok su preostali algoritmi grupirani u užem rasponu vrijednosti. S druge strane, na Slici 7.3 (b) prikazan je odnos između IDF1 metrike i broja zamjena identiteta (IDSW), pri čemu veličina markera označava broj uspješnih reidentifikacija nakon okluzije (OKL^+). Riječ je o skupini izdvojenih identitetnih metrika



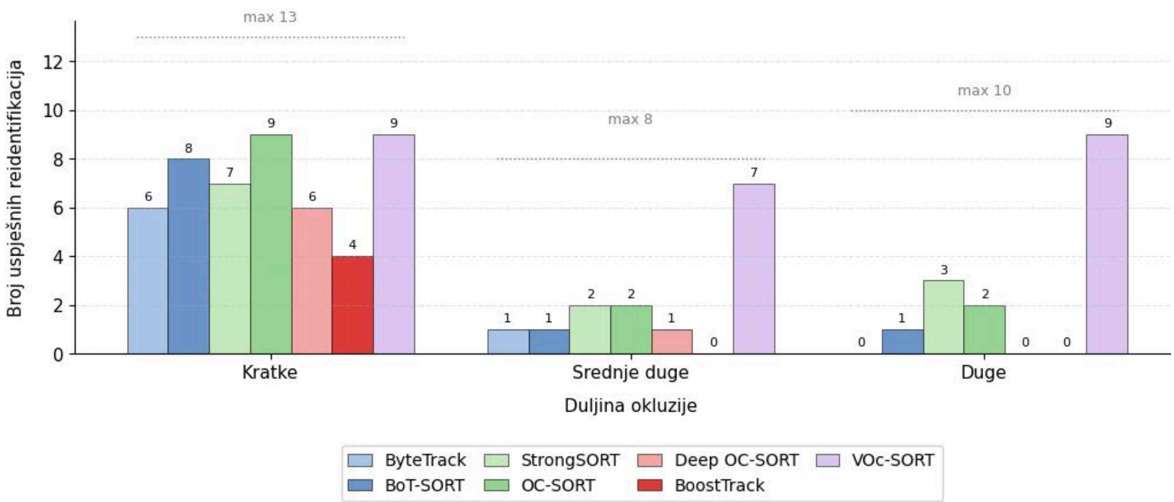
Slika 7.3: Vizualna usporedba performansi odabranih algoritama praćenja. (a) HOTA-IDF1-MOTA: x-os HOTA, y-os IDF1, veličina markera proporcionalna MOTA (zbog preglednosti izostavljen BoostTrack algoritam koji postiže znatno lošije vrijednosti odabranih metrika), (b) IDF1-IDSW-OKL$^+$: x-os IDF1, y-os IDSW, veličina markera proporcionalna OKL$^+$.

koje zajednički opisuju stabilnost i konzistentnost održavanja identiteta plovila kroz vrijeme. VOc-SORT ostvaruje najmanji broj IDSW te najveći broj uspješnih reidentifikacija, uz najvišu IDF1 vrijednost, čime se jasno izdvaja od ostalih algoritama. StrongSORT postiže visoku vrijednost IDF1, ali uz znatno veći broj zamjena identiteta, dok BoostTrack pokazuje najslabiju identitetsku stabilnost.

Kako bi se dobio uvid u sposobnost pojedinih algoritama praćenja da ispravno reidentificiraju objekt nakon okluzije ovisno o njezinom trajanju, na Slici 7.4 prikazan je broj uspješnih reidentifikacija razvrstan po trima kategorijama okluzija definiranih u Poglavlju 4, dio 4.4.3: kratke, srednje duge i duge. Prikaz obuhvaća sve razmatrane algoritme i omogućuje izravan uvid u njihove performanse za svaku od navedenih kategorija okluzija. U kategoriji kratkih okluzija broj uspješnih reidentifikacija kreće se od četiri (BoostTrack) do devet (StrongSORT i VOc-SORT), od ukupno 13 slučajeva. Kod srednje dugih okluzija većina algoritama ostvaruje jednu do dvije uspješne reidentifikacije, dok je kod dugih okluzija taj broj u rasponu od nula do tri. Izuzetak je predloženi VOc-SORT algoritam, koji u srednje dugim ostvaruje sedam uspješnih reidentifikacija, a u dugim njih devet.

Kvalitativna analiza

U okviru evaluacije prikazana su i četiri vizualna primjera ponašanja odabranih algoritama, i to predloženog VOc-SORT algoritma te sljedeća tri algoritma koja su najbolje rangirana prema vrijednosti HOTA metrike: StrongSORT, BoT-SORT i OC-SORT. Prvi primjer, prikazan na Slici 7.5, odnosi se na scenu u kojoj je putnički brod tijekom duljeg vremenskog



Slika 7.4: Usporedba broja uspješnih reidentifikacija po kategorijama okluzija predloženog VOc-SORT algoritma i postojećih metoda.

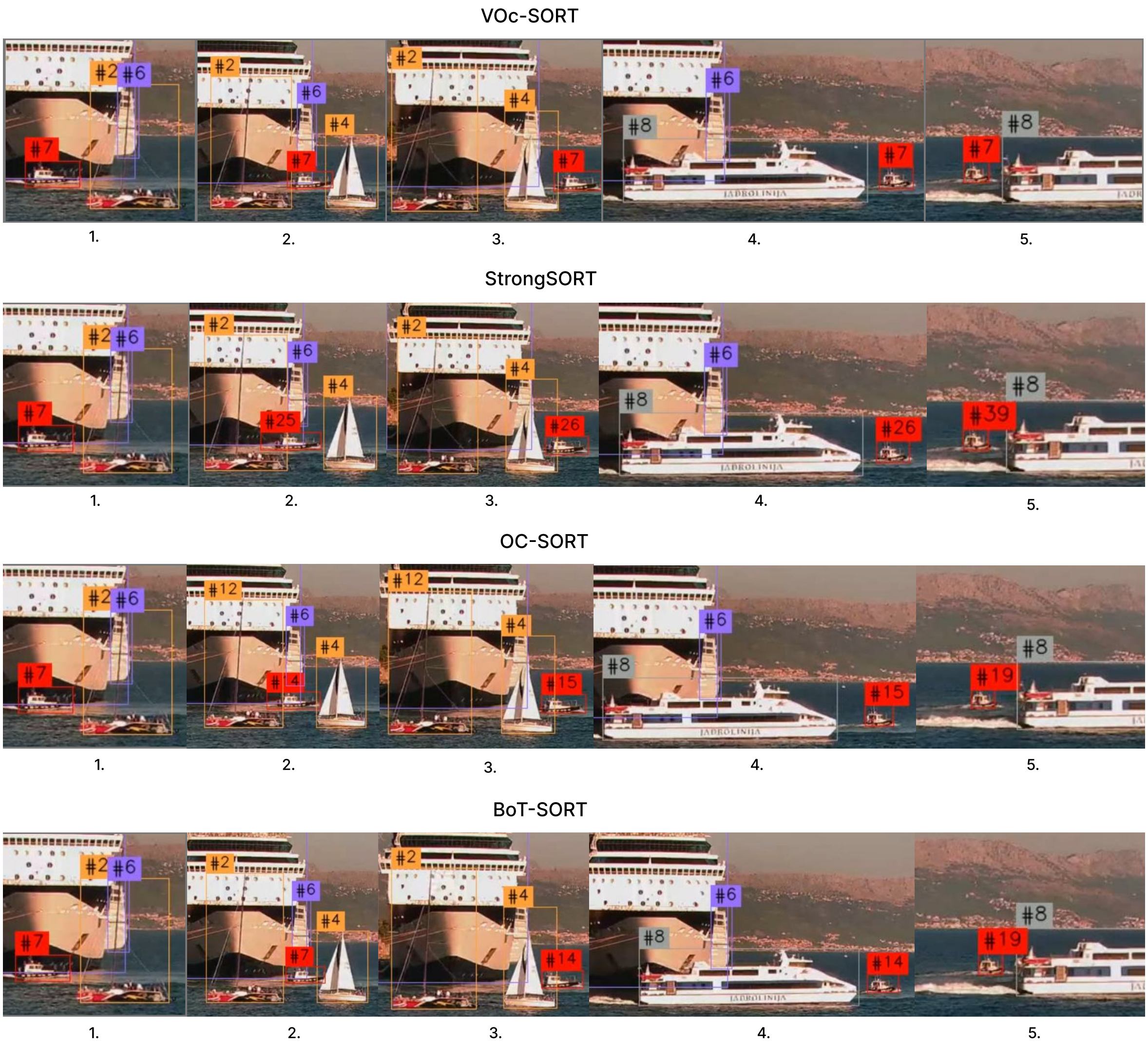
razdoblja od gotovo jedanaest sekundi u potpunosti zaklonjen trajektom. VOc-SORT i StrongSORT uspješno rekonstruiraju ispravan identitet putničkog broda nakon završetka okluzije. Međutim, StrongSORT pri izlasku iz okluzije, dok je putnički brod još djelomično zaklonjen, privremeno dodjeljuje pogrešan identitet, da bi mu tek nakon potpunog otkrivanja ponovno pridružio ispravan. Nasuprot tome, OC-SORT i BoT-SORT ne uspijevaju povratiti originalni identitet putničkog broda nakon okluzije. S druge strane, StrongSORT i OC-SORT pogrešno mijenjaju identitet trajekta nakon što je u prolazu djelomično zaklonjen katamaranom, dok VOc-SORT i BoT-SORT ispravno održavaju konzistentan identitet.

U preostalim vizualnim primjerima (Slike 7.6, 7.7, 7.8) zbog bolje preglednosti prikazani su isključivo identiteti plovila. Oznake klasa više nisu istaknute, iako se granični okviri i dalje razlikuju bojama prema pripadnosti pojedinoj klasi. Budući da su identiteti primarni pokazatelj uspješnosti praćenja, dok su klase rezultat detekcijskog modela, ovakav prikaz omogućuje lakšu usporedbu koliko dobro pojedini algoritmi održavaju ispravne identitete plovila kroz vrijeme. Na slikama je za svaki od algoritama praćenja prikazan niz različitih trenutaka iz danog videozapisa, koji su označenih brojevima 1-5 (Slika 7.6), odnosno 1-4 (Slika 7.7 i Slika 7.8).

Slika 7.6 prikazuje scenu iz SSMOT_9 videozapisa u kojoj jedno plovilo, s početno dodijeljenim identitetom 7, uzastopno ulazi u dvije okluzije: prvo u kratku okluziju u trajanju od 0.40 sekundi, a potom u srednje dugu okluziju od 6.91 sekundi. Nakon djelomične zaklonjenosti i preklapanja okvira plovila 7 s jedrilicom identiteta 2 (odnosno 12 u slučaju OC-SORT algoritma), u trenutku $t = 2$ VOc-SORT i BoT-SORT algoritmi uspijevaju održati ispravan identitet plovila, dok OC-SORT i StrongSORT plovilu 7 dodjeljuju novi, pogrešni identitet. Nakon toga plovilo inicijalnog identiteta 7 ulazi u kratku okluziju s bijelom jedrilicom identiteta 4. Pri izlasku iz okluzije u trenutku $t = 3$ jedino VOc-SORT uspijeva plovilu ponovno

Slika 7.5: Primjer ponašanja odabranih algoritama praćenja (VOc-SORT, StrongSORT, OC-SORT i BoT-SORT) na dijelu SSMOT_9 videozapisa s međusobnim preklapanjem plovila i primjerom duge okluzije putničkog broda trajektom.

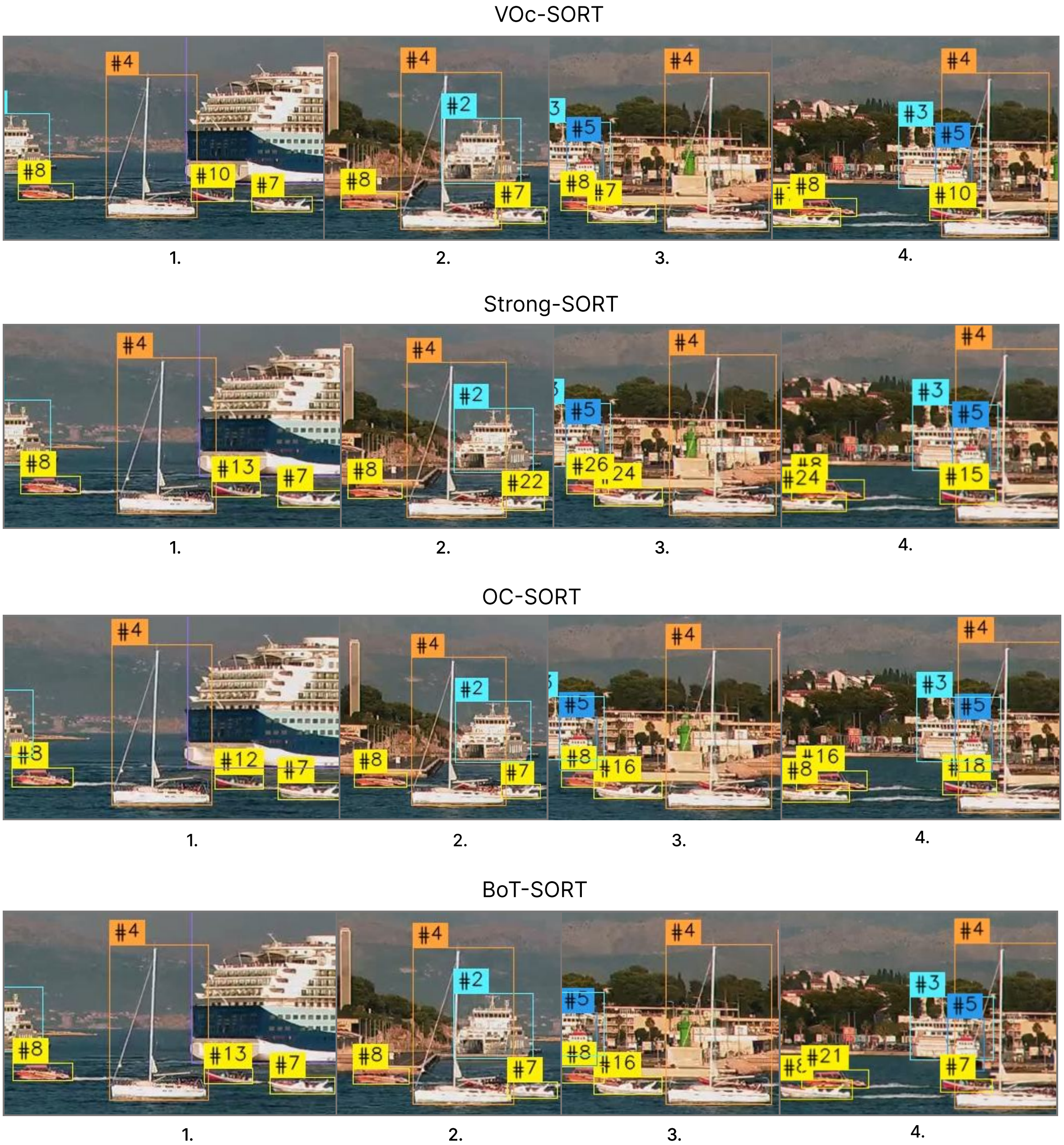
dodijeliti identitet jednak onome prije okluzije, dok preostali algoritmi u tome ne uspijevaju. Sličan obrazac se ponavlja i sa sljedećom, srednje dugom okluzijom: jedino je kod VOc-SORT algoritma identitet plovila u trenutku $t = 5$ nakon okluzije plovila katamaranom, jednak identitetu netom pred okluziju u trenutku $t = 4$.



Slika 7.6: Primjer ponašanja odabranih algoritama praćenja (VOc-SORT, StrongSORT, OC-SORT i BoT-SORT) na sceni iz SSMOT_9 videozapisa, u kojoj isto plovilo najprije ulazi u kratku okluziju u trajanju od 0.40 s (iza bijele jedrilice), a zatim u srednje dugu okluziju u trajanju od 6.91 s (iza katamarana).

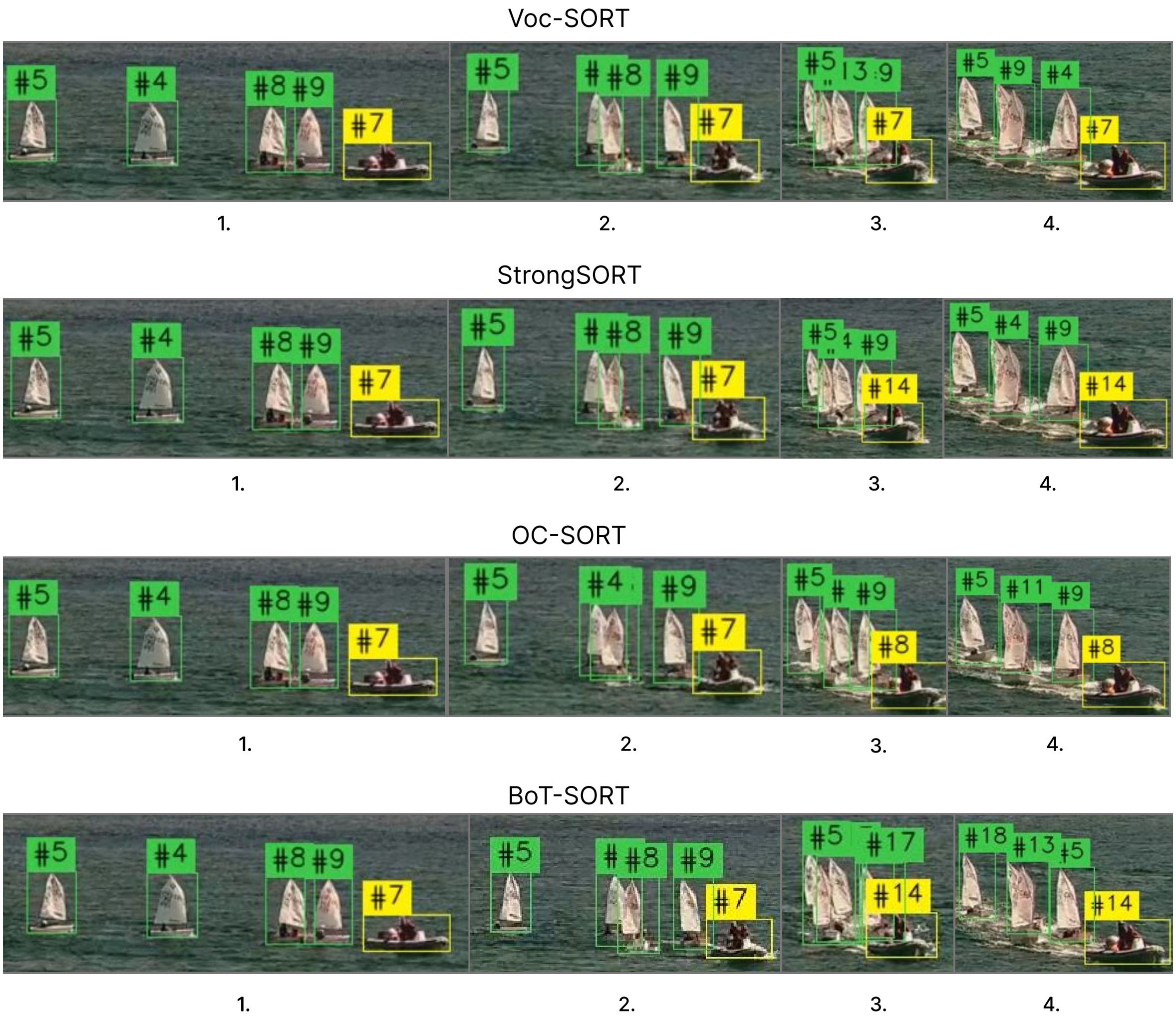
Na Slici 7.7 prikazana je scena obilježena višestrukim mimoilaženjima i jednom potpunom okluzijom. Algoritam VOc-SORT uspješno održava identitete plovila 7 i 8 u prikazanoj sceni, čak i u trenutku $t = 4$ njihova mimoilaženja. Nasuprot tome, kod algoritama BoT-SORT i OC-SORT u trenutku $t = 4$ dolazi do pogrešne dodjele identiteta: plovilu kojemu je u $t = 1$ bio pridružen identitet 7 dodijeljen je identitet plovila 8, s kojim se mimoilazilo. Nadalje, plovilo identiteta 7 u $t = 2$ (identitet 22 u slučaju OC-SORT algoritma), koje ulazi u okluziju s jedrilicom, jedino VOc-SORT uspješno ponovno identificira u trenutku $t = 3$.

Posljednji primjer, prikazan na Slici 7.8, obuhvaća scenu u kojoj je potrebno pratiti više vizualno gotovo identičnih plovila (trening jedrilica) koja se u kretanju međusobno preklapaju. U trenutku $t = 4$ detektirane su tri od četiri jedrilice; nije detektirana ona koja je gotovo potpuno zaklonjena. Za te tri vidljive jedrilice najtočnije identitete dodjeljuje algori-



Slika 7.7: Prikaz vizualno zahtjevne scene iz videozapisa SSMOT_6 obilježene preklapanjima, djelomičnim zaklonjenostima i mimoilaženjima sličnih plovila, uz jednu srednje dugu okluziju u trajanju od 3.24 s.

tam StrongSORT. Algoritam VOc-SORT zamjenjuje redoslijed identiteta 4 i 9 te u trenutku $t = 3$ pogrešno uvodi novi identitet 13. OC-SORT u $t = 4$ također pogrešno dodjeljuje novi identitet (11). Najveće odstupanje pokazuje BoT-SORT, koji u trenutcima $t = 3$ i $t = 4$ pogrešno uvodi tri nova identiteta (17, 18 i 13) te identitet 5 pridružuje neispravnom objektu. Suprotno tome, jedino VOc-SORT uspijeva održati konzistentan identitet plovila označenog žutim okvirom (identitet 7 u trenutku $t = 1$) kroz sva četiri prikazana trenutka.



Slika 7.8: Primjer scene iz videozapisa SSMOT_12 koja zahtijeva praćenje vizualno gotovo identičnih plovila s međusobnim preklapanjima.

7.3. Ablacijska studija

Ablacijska studija provodi se kako bi se ispitao utjecaj pojedinih komponenti predloženog algoritma na ukupne performanse praćenja. Kao polazna točka koristi se implementacija VOc-SORT algoritma opisana u 7.1. Sustavnim uklanjanjem, zamjenom ili modificiranjem pojedinih elemenata algoritma procjenjuje se njihov utjecaj na ukupnu učinkovitost praćenja, pri čemu se posebna pozornost posvećuje njihovoj sposobnosti uspješne reidentifikacije plovila nakon srednje dugih i dugih okluzija.

7.3.1. Utjecaj korištenog detektora

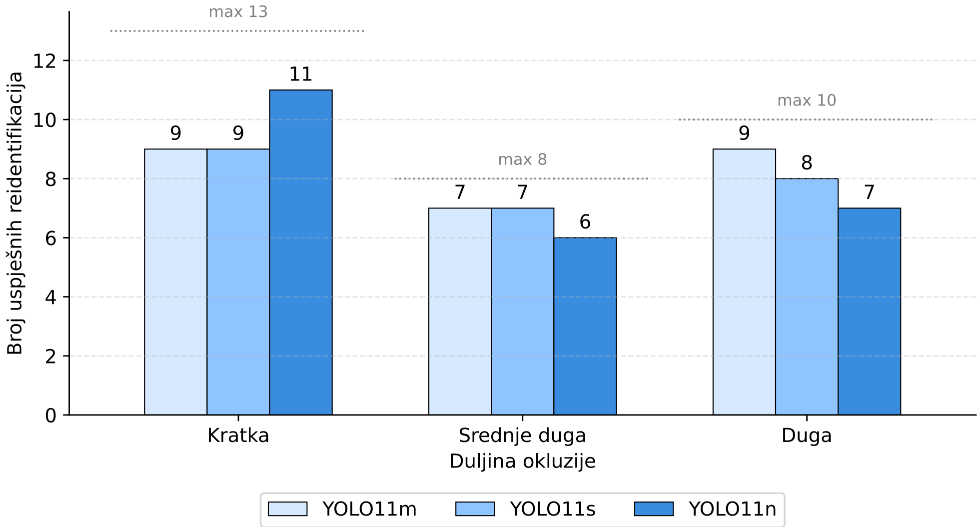
U okviru ove ablacijske studije prvo se ispituje utjecaj korištenog detektora na učinkovitost VOc-SORT algoritma. U Tablici 7.2 prikazani su rezultati evaluacije algoritma pri primjeni triju varijanti YOLO11 [288] detektora (YOLO11n, YOLO11s i YOLO11m), treniranih na

SSMOT skupu podataka za detekciju, kako je opisano u poglavlju 5, dio 5.2.3. Na testnom skupu SSMOT detekcijskog podskupa, detektor YOLO11m postigao je najvišu vrijednost detekcijske metrike $mAP_{50:95}$ od 0.829, zatim slijedi YOLO11s s vrijednošću 0.799, dok YOLO11n ostvaruje najnižu vrijednost od 0.764. U skladu s tim, YOLO11m ostvaruje i najbolje rezultate kvantitativnih metrika praćenja na SSMOT videozapisima, nakon njega slijedi YOLO11s, dok najmanji model YOLO11n postiže najslabije rezultate.

Tablica 7.2: Usporedba performansi algoritma praćenja pri korištenju različitih varijanti YOLO11 detektora.

	HOTA (↑)	**MOTA** (↑)	**MOTP** (↑)	**IDSW** (↓)	**IDF1** (↑)	**OKL^{+}** (↑)
YOLO11m	**85.57**	**93.03**	**89.82**	**42**	**94.38**	**25**
YOLO11s	83.92	91.72	89.54	45	93.18	24
YOLO11n	81.34	89.19	88.78	77	90.73	24

Na Slici 7.9 prikazana je usporedba broja uspješnih reidentifikacija za tri spomenute varijante YOLO11 detektora u odnosu na tri kategorije duljine okluzije. U kategoriji kratkih okluzija model praćenja koji koristi YOLO11n detektor ostvaruje najveći broj uspješnih reidentifikacija, njih 11, dok kod srednje dugih i dugih okluzija bilježi slabije rezultate u odnosu na preostala dva detektora. YOLO11s i YOLO11m jednako su uspješni u reidentifikaciji plovila nakon kratkih i srednje dugih okluzija, dok YOLO11m ostvaruje najbolji rezultat kada su u pitanju duge okluzije.



Slika 7.9: Prikaz broja uspješnih reidentifikacija algoritma praćenja po kategorijama okluzija pri korištenju različitih varijanti YOLO11 detektora.

7.3.2. Utjecaj integracije vizualnih ReID značajki s prostorno-dinamičkim informacijama

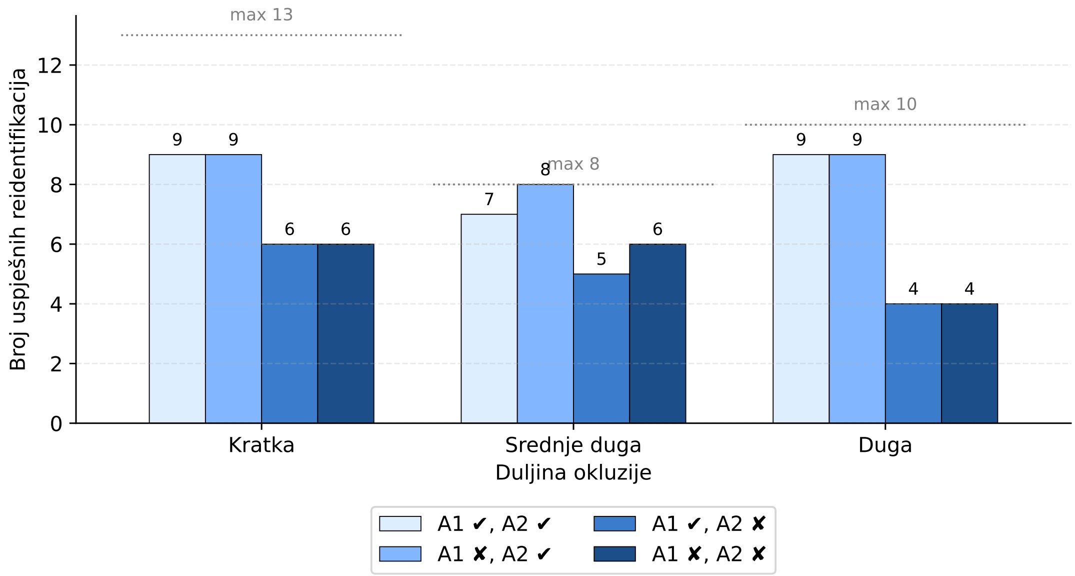
U ovom dijelu provedena je ablacijska analiza utjecaja integracije vizualnih ReID značajki u postupku asocijacije unutar predloženog algoritma praćenja. Analiza je usredotočena na ispitivanje doprinosa vizualnih značajki u prvoj (A1) i drugoj (A2) fazi asocijacije, pri čemu su razmatrane različite konfiguracije u kojima su te značajke uključene ("✓") ili izostavljene ("✗"). Usporedba je provedena uz zadržavanje istih postavki ostalih komponenti sustava, s ciljem izoliranja utjecaja vizualnih informacija na performanse praćenja, osobito u kontekstu očuvanja identiteta nakon okluzija.

Rezultati prikazani u Tablici 7.3 jasno ukazuju na utjecaj integracije vizualnih ReID značajki u pojedinim fazama asocijacije na ukupne performanse praćenja. Najbolji rezultati ostvareni su u konfiguraciji u kojoj su vizualne značajke uključene u obje faze asocijacije, pri čemu su postignute najviše vrijednosti HOTA, MOTA i IDF1 metrika, uz istodobno najmanji broj zamjena identiteta. Izostavljanje vizualnih značajki u prvoj fazi asocijacije, uz njihovo zadržavanje u drugoj fazi, rezultira tek umjerenim smanjenjem globalnih metrika praćenja, dok se pritom zadržava visok broj uspješno reidentificiranih plovila nakon okluzija. Suprotno tome, konfiguracije u kojima su vizualne značajke izostavljene u drugoj fazi asocijacije bilježe izraženije pogoršanje performansi, osobito u pogledu očuvanja identiteta i uspješnosti reidentifikacije nakon okluzija. Najslabiji rezultati ostvareni su u slučaju kada se asocijacija temelji isključivo na prostorno-dinamičkim informacijama.

Tablica 7.3: Usporedba performansi algoritma praćenja pri različitim konfiguracijama integracije vizualnih ReID značajki u prvoj (A1) i drugoj (A2) fazi asocijacije.

A1	A2	HOTA (↑)	MOTA (↑)	MOTP (↑)	IDSW (↓)	IDF1 (↑)	OKL$^+$ (↑)
✓	✓	**85.57**	**93.03**	89.82	**42**	**94.38**	25
✗	✓	84.41	93.02	89.83	48	92.54	**26**
✓	✗	83.48	92.97	89.83	57	90.98	15
✗	✗	83.21	92.97	89.83	62	90.56	16

Na Slici 7.10 prikazana je uspješnost reidentifikacije plovila nakon okluzija različitog trajanja za različite konfiguracije integracije vizualnih ReID značajki po fazama asocijacije VOc-SORT algoritma. Iz prikazanih rezultata vidljivo je da konfiguracije koje uključuju vizualne ReID značajke u drugoj fazi asocijacije ostvaruju veći broj uspješnih reidentifikacija u svim kategorijama trajanja okluzije. Posebno je izražena razlika u slučaju dugotrajnih okluzija, gdje konfiguracije s uključenim vizualnim značajkama u drugoj fazi zadržavaju znatno višu uspješnost u odnosu na konfiguraciju koje se oslanjaju samo na prostorno-dinamičke informacije, te na konfiguraciju koja vizualnu informaciju koristi samo u prvoj fazi asocijacije.

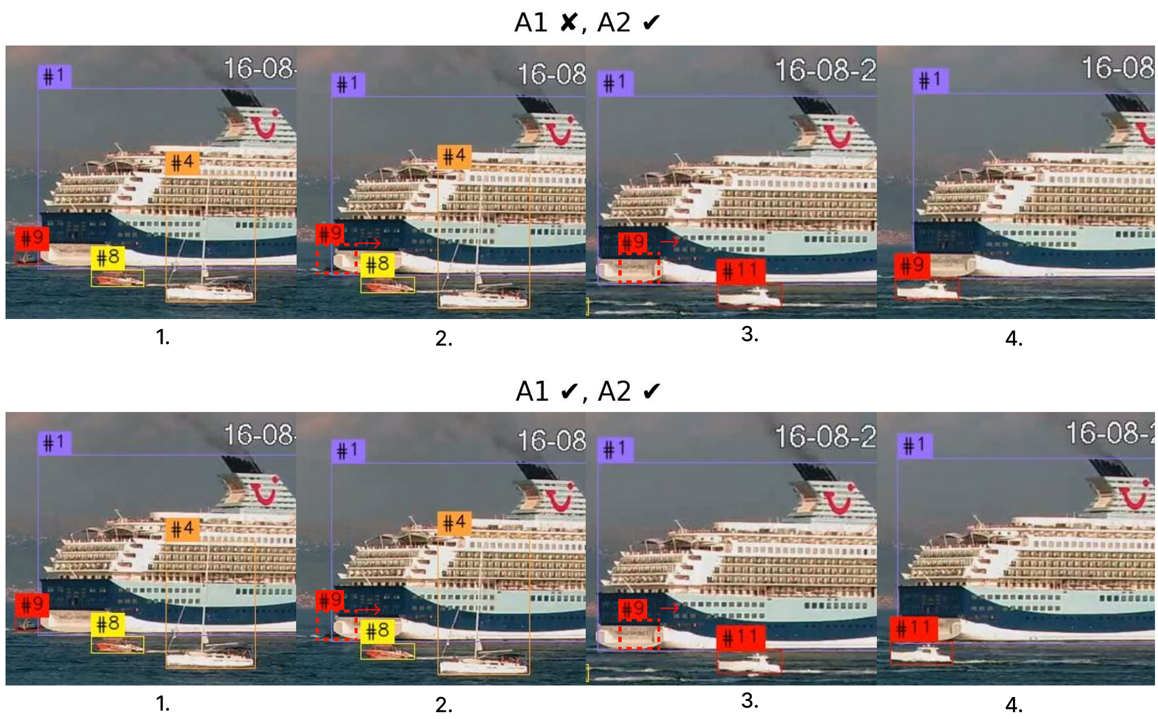


Slika 7.10: Prikaz broja uspješnih reidentifikacija algoritma praćenja po kategorijama okluzija pri različitim konfiguracijama integracije vizualnih ReID značajki u prvoj i drugoj fazi asocijacije.

Primjer prikazan na Slici 7.11 ilustrira kako izostavljanje vizualnih informacija u prvoj fazi asocijacije može utjecati na stabilnost praćenja. Slika prikazuje četiri trenutka praćenja, $t = 1, \ldots, 4$, iz videozapisa SSMOT_6. U gornjem retku prikazana je konfiguracija u kojoj se u prvoj fazi asocijacije koriste isključivo geometrijske informacije temeljene na preklapanju graničnih okvira, dok druga faza asocijacije kombinira vizualne značajke i dinamičke informacije kvantificirane Mahalanobisovom udaljenošću. Donji redak prikazuje konfiguraciju u kojoj se vizualne značajke koriste u obje faze. U trenutku $t = 2$ plovilo s identitetom 9 privremeno nestaje iza kruzera. Isprekidani granični okvir označava izgubljeni objekt čija putanja još nije uklonjena iz sustava praćenja, dok strelica prikazuje procijenjeni smjer kretanja objekta dobiven predviđanjem Kalmanovog filtera. U sljedećem trenutku, $t = 3$, drugo plovilo s identitetom 11 približava se području u kojem je plovilo identiteta 9 prethodno izgubljeno. U konfiguraciji koja ne koristi vizualne informacije u prvoj fazi asocijacije dolazi do pogrešne dodjele identiteta u trenutku $t = 4$, pri čemu se identitet 9 pridružuje plovilu koje je prethodno imalo identitet 11. Suprotno tome, konfiguracija koja koristi vizualne značajke u prvoj fazi asocijacije uspješno zadržava ispravan identitet praćenog plovila.

7.3.3. Utjecaj korištenog modela za ekstrakciju vizualnih značajki

Kako bi se ispitao utjecaj odabranog modela za ekstrakciju vizualnih značajki na performanse praćenja, s posebnim naglaskom na održavanje stabilnih identiteta objekata, provedena je ablacijska analiza u kojoj se uspoređuje pet različitih konfiguracija ReID modela unutar VOc-SORT algoritma. U analizi su razmotrene dvije arhitekture za ekstrakciju značajki: ResNet50 [180] i OSNet 1.0 [228].



Slika 7.11: Primjer koji ilustrira utjecaj izostavljanja vizualnih informacija u prvoj fazi asocijacije na performanse praćenja.

ResNet50 predstavlja opći model duboke neuronske mreže, izvorno razvijen za zadatke klasifikacije slika, koji se u praksi često koristi i kao osnova za ekstrakciju vizualnih značajki u ReID zadacima. U tom kontekstu uspoređene su dvije varijante ResNet50 modela: model treniran na općem klasifikacijskom skupu podataka ImageNet [125] te model dodatno treniran na domenski specifičnom SSMOT ReID skupu podataka. S druge strane, OSNet 1.0 je arhitektura posebno dizajnirana za zadatak reidentifikacije objekata. Za ovaj model analizirane su tri varijante treniranja: model treniran na SSMOT ReID skupu podataka, model treniran na ImageNet skupu podataka te model treniran na MSMT17 [323] skupu podataka za reidentifikaciju pješaka. Problem reidentifikacije pješaka najzastupljeniji je u ovom području istraživanja, zbog čega su skupovi podataka iz te domene najbrojniji i najčešće korišteni. Posljedično, modeli trenirani na podacima za reidentifikaciju pješaka, kao i na ImageNet skupu podataka, često se primjenjuju kao univerzalni ekstraktori značajki u zadacima praćenja više objekata, osobito kada domenski specifični ReID podaci nisu dostupni.

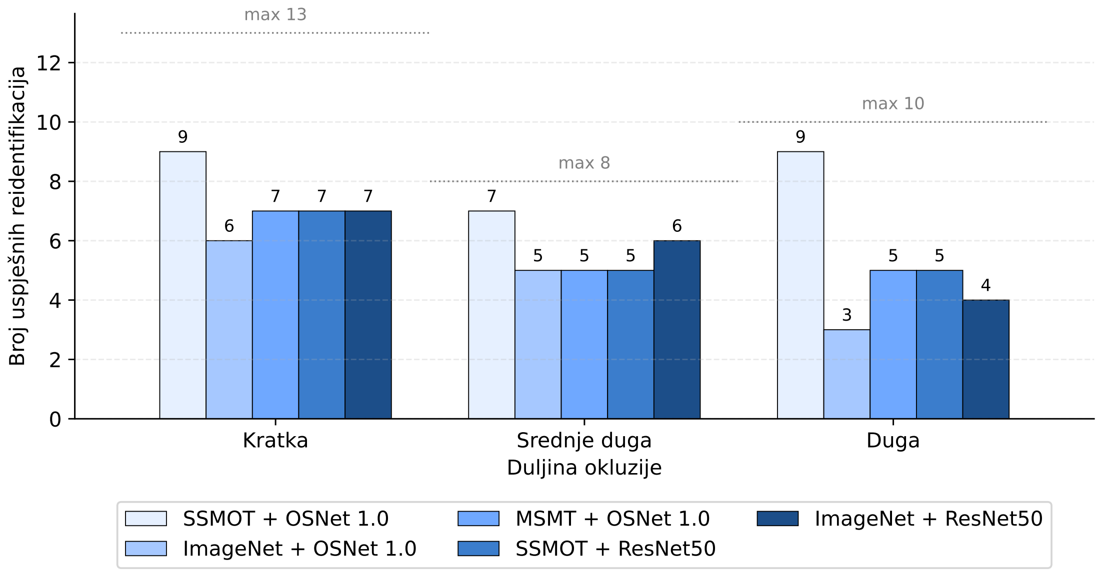
U Tablici 7.4 prikazani su rezultati evaluacije VOc-SORT algoritma pri korištenju različitih modela za ekstrakciju vizualnih značajki te različitih skupova podataka na kojima su ti modeli trenirani. Najbolje performanse pokazuje OSNet 1.0 model treniran na SSMOT ReID skupu podataka s najvećim vrijednostima HOTA (85.57), MOTA (93.03) i IDF1 (94.38) metrika, najmanjim brojem promjena identiteta (42) te znatno više ispravnih reidentifikacija nakon okluzija (25) od ostalih modela. Varijante OSNet 1.0 modela trenirane na ImageNet i MSMT17 skupovima podataka ostvaruju niže vrijednosti HOTA i IDF1 metrika te veći broj promjena identiteta u odnosu na model treniran na SSMOT ReID skupu podataka, neznatno niže vrijednosti MOTA metrike, te usporedive vrijednosti MOTP metrike. Kod ResNet50 arhitekture, razlike između modela treniranih na različitim podacima manje su izražene. Model treniran na domenski specifičnom SSMOT skupu podataka ostvaruje nešto bolje rezultate u odnosu na ImageNet varijantu u metrikama povezanim s konzistentnošću identiteta (IDSW i IDF1). Međutim, obje ResNet50 konfiguracije ostvaruju slabije ukupne rezultate u usporedbi s OSNet 1.0 modelom treniranim na SSMOT ReID skupu podataka. Vrijednosti MOTA i MOTP metrika relativno su stabilne među svim razmatranim konfiguracijama, dok se najveće razlike u performansama očituju u metrikama povezanim s održavanjem identiteta.

Tablica 7.4: Usporedba performansi algoritma praćenja pri korištenju različitih ReID modela i skupova podataka na kojima su oni trenirani.

Podaci	**Model**	**HOTA** (↑)	**MOTA** (↑)	**MOTP** (↑)	**IDSW** (↓)	**IDF1** (↑)	**OKL**$^{+}$ (↑)
SSMOT	OSNet 1.0	**85.57**	**93.03**	89.82	**42**	**94.38**	**25**
ImageNet	OSNet 1.0	83.11	92.97	89.82	63	90.43	14
MSMT17	OSNet 1.0	83.65	93.00	**89.83**	51	91.21	17
SSMOT	ResNet50	83.57	92.98	**89.83**	58	91.36	17
ImageNet	ResNet50	83.59	92.96	**89.83**	60	91.25	17

Slika 7.12 prikazuje broj uspješnih reidentifikacija objekata nakon okluzija kada se različit ReID modeli koriste za ekstrakciju vizualnih značajki u VOc-SORT algoritmu, grupirane prema duljini okluzije. U svim kategorijama okluzija OSNet model treniran na skupu podataka SSMOT ostvaruje najbolje rezultate. Najizraženije razlike u performansama mogu se uočiti kod dugih okluzija, gdje su performanse ostalih ReID modela znatno slabije.

Vizualni primjeri na Slici 7.13 dodatno prikazuju razlike u primjeni različitih ReID modela na dva testna primjera iz SSMOT skupa videozapisa. U oba primjera OSNet model treniran na SSMOT ReID skupu podataka uspijeva održati ispravno identitete, dok ostali modeli imaju poteškoće u očuvanju konzistentnih identiteta. Na desnom primjeru iz videozapisa SSMOT_6 prikazana su tri plovila klase *speed craft* sličnog oblika (identiteti 6, 10 i 7 u trenutku $t = 1$). VOc-SORT, pri korištenju svih ReID modela osim domenski prilagođenog OsNet 1.0 modela, u trenutku $t = 2$ ne uspijeva niti jednom od tih plovila dodijeliti isti identitet. Pogreške istih modela vidljive su i na lijevom primjeru iz SSMOT_8 videozapisa. Model OsNet 1.0 treniran na MSMT17 skupu podataka dodjeljuje identitet 8, koji je u trenutku $t = 1$ pripadao trajektu, novom plovilu (jahti) koje je ušlo u scenu. Kod ResNet50 modela identitet trajekta pogrešno se dodjeljuje katamaranu s kojim se mimoilazi, dok se jahti koja je tek ušla u scenu ispravno dodjeljuje novi identitet.



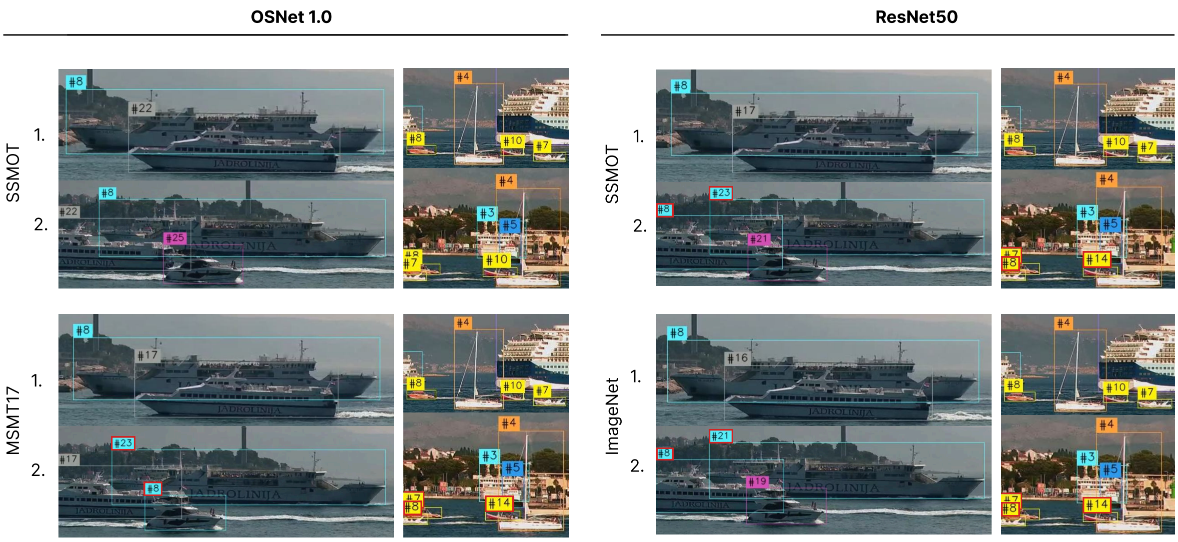
Slika 7.12: Prikaz broja uspješnih reidentifikacija algoritma praćenja po kategorijama okluzija pri korištenju različitih modela za ekstrakciju vizualnih značajki.

7.3.4. Utjecaj korištenja dodatnog kriterija za ranije uklanjanje izgubljenih putanja

Varijanta VOc-SORT algoritma koja uključuje dodatni uvjet za ranije uklanjanje putanja plovila izgubljenih u blizini ruba kadra pokazuje usporedive performanse u odnosu na verziju algoritma koja taj uvjet ne koristi. Iz Tablice 7.5 vidljivo je da varijanta s dodatnim uvjetom ostvaruje nešto bolje vrijednosti metrika HOTA (85.57 u odnosu na 85.35) i IDF1 (94.38 u odnosu na 94.04). S druge strane, varijanta bez dodatnog uvjeta bilježi jednu promjenu identiteta manje, odnosno uspješnu dodatnu reidentifikaciju nakon jedne kratke okluzije. Za ostale kategorije okluzija broj uspješnih reidentifikacija jednak je u obje varijante, što se može vidjeti na Slici 7.14. Iz navedenog slijedi da dodatan uvjet za ranije uklanjanje putanja učinkovito ograničava broj izgubljenih putanja uz neznatan utjecaj na uspješnost reidentifikacije. Stoga ga je preporučljivo zadržati, osobito u scenama s velikim brojem plovila i čestim izlascima iz kadra.

Tablica 7.5: Razlika u performansama algoritma praćenja kada se koristi dodatan uvjet za brisanje putanja i kada se on ne koristi.

Dodatni uvjet	**HOTA** (↑)	**MOTA** (↑)	**MOTP** (↑)	**IDSW** (↓)	**IDF1** (↑)	**OKL$^+$** (↑)
✓	**85.57**	93.03	89.82	42	**94.38**	25
✗	85.35	93.03	89.82	**41**	94.04	**26**



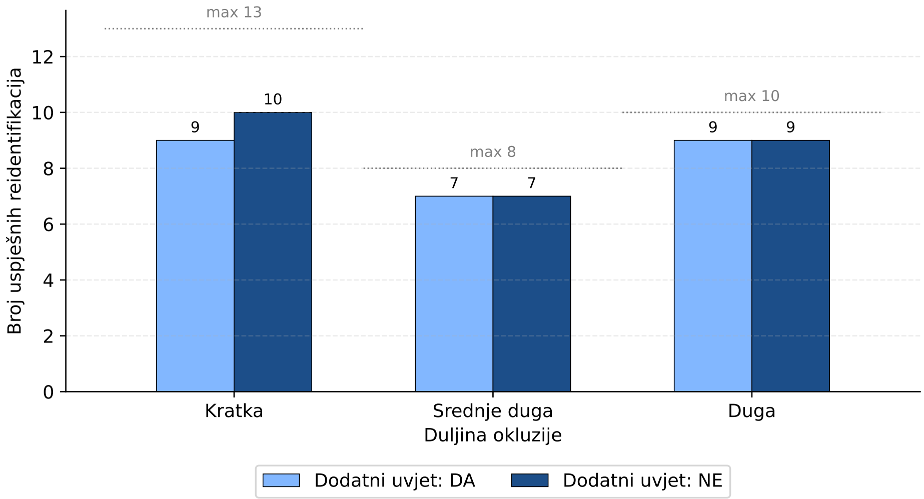
Slika 7.13: Usporedni prikaz uspješnosti različitih ReID modela u održavanju identiteta plovila na dva testna primjera. Crvenim okvirom istaknute su pogreške reidentifikacije.

7.3.5. Utjecaj korištene metode asocijacije i odabira granične vrijednosti za vizualnu sličnost

Predloženi VOc-SORT algoritam temelji se na dvostupanjskom postupku asocijacije. U prvoj fazi asocijacije za povezivanje putanja i detekcija koristi se kombinacija IoU udaljenosti i kosinusne udaljenosti vizualnih vektora značajki, dok se u drugoj fazi asocijacije kosinusna udaljenost kombinira s Mahalanobisovom udaljenošću. U obje faze asocijacije koristi se nešto viša granična vrijednost za vizualnu udaljenost (0.4) u odnosu na uobičajene vrijednosti. U okviru ove ablacijske studije analizira se utjecaj ovakvog pristupa asocijaciji na ukupne performanse praćenja. Posebno se razmatra opravdanost primjene dviju uzastopnih faza asocijacije, utjecaj njihova redoslijeda, doprinos svake pojedinačne faze te osjetljivost algoritma na odabrani prag vizualne udaljenosti.

U Tablici 7.6 prikazani su rezultati ablacijske analize primjene različitih metoda asocijacije putanja i novih detekcija, te utjecaj odabrane granične vrijednosti vizualne sličnosti. U prvom stupcu naveden je primijenjeni postupak asocijacije, pri čemu "A1" označava jednostupanjsku asocijaciju temeljenu na kombinaciji IoU i kosinusne udaljenosti, dok "A2" označava asocijaciju koja kombinira kosinusnu i Mahalanobisovu udaljenost. Oznake "A1, A2" i "A2, A1" ukazuju na dvostupanjski postupak asocijacije, pri čemu redoslijed oznaka odgovara redoslijedu primjene pojedinih faza. Drugi stupac prikazuje korištene granične vrijednosti vizualne sličnosti za svaku fazu asocijacije. U slučaju jednostupanjske asocijacije navedena je jedna vrijednost praga, dok su kod dvostupanjskih metoda prikazane dvije vrijednosti koje se redom odnose na prvu i drugu fazu asocijacije.

Predloženi dvostupanjski pristup asocijaciji detekcija i putanja (A1, A2 s graničnim vrijednostima 0.4 i 0.4) postiže ukupno najbolje rezultate praćenja s najvišim vrijednostima



Slika 7.14: Prikaz broja uspješnih reidentifikacija algoritma praćenja po kategorijama okluzija kada se koristi dodatni uvjet za brisanje putanja i kada se on ne koristi.

Tablica 7.6: Usporedba performansi algoritma praćenja za različite metode asocijacije putanja i detekcija te utjecaj odabrane granične vrijednosti vizualne sličnosti na konačne rezultate praćenja.

Asocijacija	**Granične vrijednosti**	**HOTA** (↑)	**MOTA** (↑)	**MOTP** (↑)	**IDSW** (↓)	**IDF1** (↑)	**OKL$^+$** (↑)
A1	0.2	83.97	93.01	89.83	67	91.08	6
A1	0.4	84.11	93.01	89.83	66	91.36	6
A2	0.2	84.45	92.89	**89.85**	119	92.20	20
A2	0.4	84.64	92.90	89.84	110	92.54	25
A1, A2	0.2, 0.2	84.97	93.04	89.83	47	93.22	20
A1, A2	0.2, 0.4	85.43	93.03	89.82	43	94.10	25
A1, A2	0.4, 0.2	85.11	**93.05**	89.83	46	93.50	20
A1, A2	0.4, 0.4	**85.57**	93.03	89.82	**42**	**94.38**	25
A2, A1	0.2, 0.2	84.72	93.01	89.83	63	92.71	21
A2, A1	0.2, 0.4	84.86	93.02	89.83	62	92.99	21
A2, A1	0.4, 0.2	84.96	92.99	89.83	63	93.20	**26**
A2, A1	0.4, 0.4	85.10	92.99	89.83	62	93.48	**26**

HOTA (85.57) i IDF1 (94.38) metrika, uz najmanji broj zamjena identiteta (42). Jednostupanjske asocijacije (A1 i A2) u pravilu ostvaruju niže vrijednosti promatranih metrika u odnosu na konfiguracije koje koriste dvije faze asocijacije, pri čemu je iznimka metrika MOTP, kod koje samostalna primjena asocijacije A2 s graničnom vrijednošću 0.4 ostvaruje najvišu vrijednost.

Usporedba jednostupanjskih pristupa pokazuje da samostalna primjena asocijacije A2 rezultira boljim ukupnim performansama praćenja u odnosu na samostalnu primjenu asocijacije A1, ali uz znatno veći broj zamjene identiteta praćenih plovila. Nasuprot tome, samostalna primjena asocijacije A1 pokazuje ograničenu učinkovitost u održavanju konzistentnih identiteta plovila nakon okluzija, što se očituje u vrlo malom broju uspješnih reidentifikacija - svega njih šest. Kombiniranjem ovih dviju asocijacija u dvostupanjski postupak navedeni nedostaci se ublažavaju. Primjena asocijacije A2 u drugoj fazi, nakon A1, znatno povećava broj plovila koji je uspješno identificiran nakon okluzije (s 6 se povećava na 20 pri nižem pragu vizualne sličnosti, a na 25 pri višem). S druge strane, primjena asocijacije A1 u drugoj fazi, nakon asocijacije A2, rezultira značajnim smanjenjem broja zamjena identiteta tijekom praćenja, pri čemu se broj zamjena identiteta smanjuje sa 119 i 110 na 63 i 62, ovisno o odabranoj graničnoj vrijednosti. Istodobno se postiže i uspješna reidentifikacija jednog dodatnog slučaja okluzije.

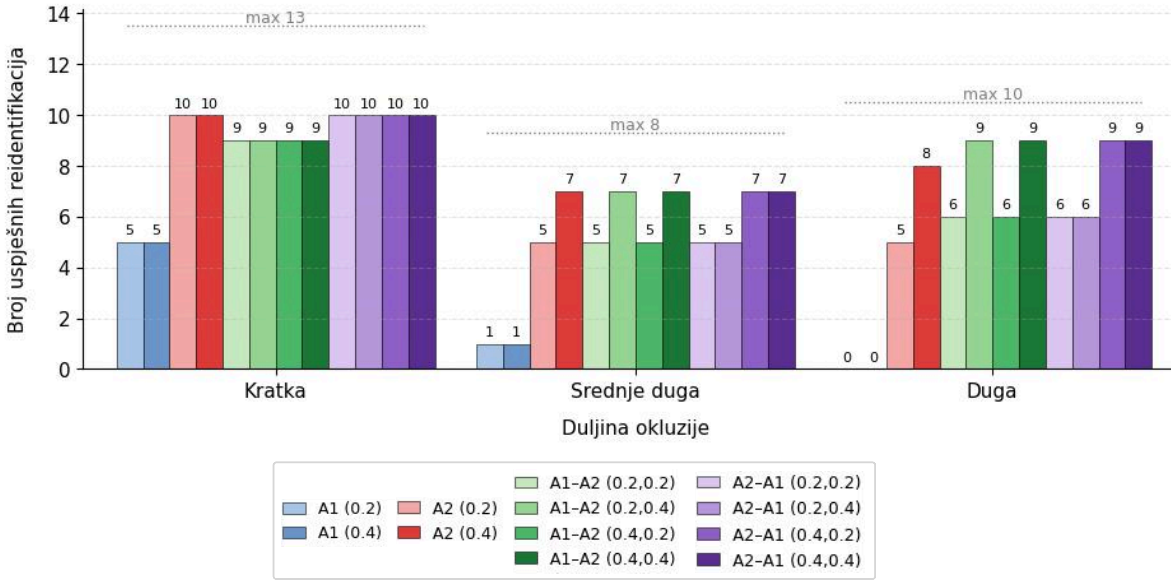
Analiza redoslijeda primjene asocijacija u dvostupanjskom postupku pokazuje da se u oba razmatrana slučaja najbolji rezultati postižu kada se u obje faze koristi granična vrijednost vizualne sličnosti od 0.4. Iako konfiguracija u kojoj se asocijacija A2 primjenjuje prije A1 omogućuje ispravno razrješavanje jedne okluzije više, konfiguracija u kojoj se A1 primjenjuje u prvoj, a A2 u drugoj fazi, u pravilu ostvaruje bolje ukupne performanse praćenja. To se očituje u višim vrijednostima HOTA, MOTA i IDF1 metrika, kao i u manjem broju zamjena identiteta.

Na Slici 7.15 prikazan je broj uspješno reidentificiranih plovila u ovisnosti o duljini okluzije, za različite konfiguracije metode asocijacije. Najlošije rezultate pokazuje samostalna primjena asocijacije A1 s tek jednom uspješnom reidentifikacijom nakon okluzije srednje duljine te niti jednom uspješnom reidentifikacijom u slučaju dugih okluzija. Dvostupanjski pristupi asocijacije koji kombiniraju metode A1 i A2 pokazuju znatno bolje rezultate u reidentifikaciji plovila nakon okluzija, neovisno o redoslijedu njihove primjene. Nadalje, kod konfiguracija koje uključuju asocijaciju A2 povećanjem granične vrijednosti na vizualnu sličnost s 0.2 na 0.4 kod okluzija srednje duljine uspješno se reidentificiraju dva dodatna slučaja, dok su kod dugih okluzija zabilježena tri dodatna slučaja uspješne reidentifikacije.

Budući da je druga faza asocijacije VOc-SORT algoritma inspirirana načinom asocijacije StrongSORT algoritma uz povećanu graničnu vrijednost vizualne sličnosti, proveden je dodatni eksperiment kako bi se ispitalo je li za poboljšanje performansi praćenja dovoljno prilagoditi prag vizualne sličnosti unutar izvornog StrongSORT-a, ili je pak nužno i dodatno modificirati sam postupak asocijacije. Rezultati tog eksperimenta prikazani su u Tablici 7.7, gdje oznaka StrongSORT*[2] označava varijantu StrongSORT algoritma s povećanom graničnom vrijednošću vizualne sličnosti, s 0.2 na 0.4.

Povećanje granične vrijednosti vizualne sličnosti u StrongSORT algoritmu dovodi do porasta broja uspješnih reidentifikacija nakon okluzije, s 13 na 19 slučajeva. Na Slici 7.16 porast je vidljiv u svim kategorijama okluzija (kratke: 7 $\rightarrow$ 9, srednje duge: 2 $\rightarrow$ 4, duge:

[2]Razlika između StrongSORT* algoritma i samostalne asocijacije A2 s pragom vizualne sličnosti 0.4 je u varijanti Kalmanovog filtera korištenoj za predviđanje sljedećeg stanja putanje i u načinu potvrđivanja putanja u probnom razdoblju



Slika 7.15: Prikaz broja uspješnih reidentifikacija algoritma praćenja po kategorijama okluzija ovisno o korištenoj metodi asocijacije postojećih putanja i novih detekcija.

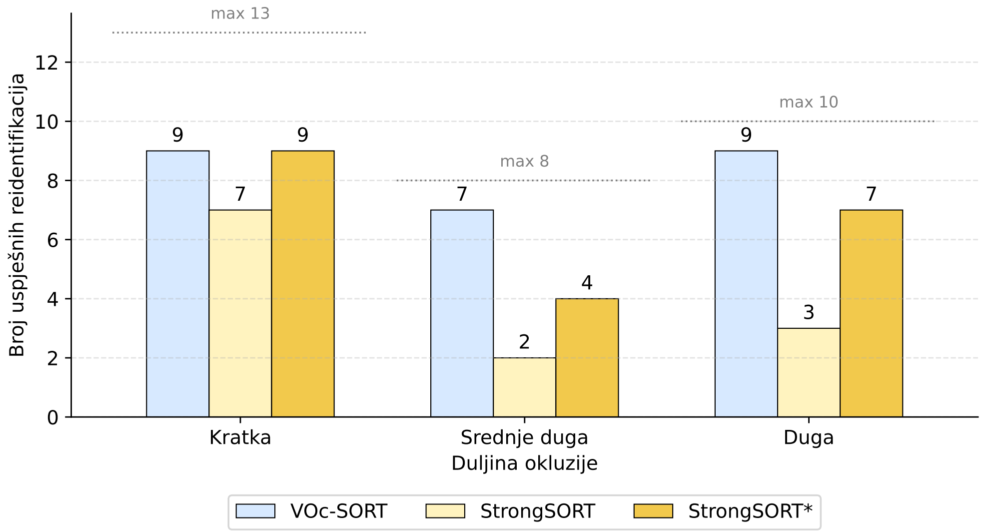
$3 \rightarrow 7$), pri čemu je najizraženiji upravo kod dugih okluzija. Međutim, ovo poboljšanje ne prati značajan porast ostalih kvantitativnih metrika HOTA, MOTA, MOTP i IDF1, a broj zamjena identiteta i dalje ostaje vrlo visok.

Tablica 7.7: Utjecaj povećanja praga vizualne sličnosti u StrongSORT algoritmu.

	θ_{vis}	**HOTA** (↑)	**MOTA** (↑)	**MOTP** (↑)	**IDSW** (↓)	**IDF1** (↑)	**OKL**$^{+}$ (↑)
StrongSORT	0.2	83.83	92.21	89.78	371	91.59	13
StrongSORT*	0.4	83.70	92.25	89.77	374	91.33	19
VOc-SORT	0.4, 0.4	**85.57**	**93.03**	**89.82**	**42**	**94.38**	**25**

7.4. Rasprava

U ovom poglavlju predstavljen je VOc-SORT algoritam praćenja, razvijen s ciljem poboljšanja robusnosti praćenja u scenarijima dugotrajnih okluzija u pomorskim okruženjima. Provedena evaluacija i opsežna ablacijska studija omogućile su preciznu analizu doprinosa pojedinih komponenti algoritma njegovim ukupnim performansama, pri čemu je poseban naglasak stavljen na održavanje stabilnih identiteta praćenih objekata, osobito nakon razdoblja zaklonjenosti. Dobiveni rezultati potvrđuju da pouzdano praćenje plovila u složenim pomorskim scenama zahtjeva integriran pristup koji objedinjuje kvalitetnu detekciju, domenski prilagođenu ekstrakciju vizualnih značajki te pažljivo odabranu cijenu pridruživanja i strategiju asocijacije putanja i detekcija.



Slika 7.16: Prikaz broja uspješnih reidentifikacija algoritma praćenja po kategorijama okluzija za VOc-SORT, StrongSORT i StrongSORT algoritme.*

7.4.1. Utjecaj detekcije na performanse praćenja

Rezultati ablacijske studije pokazuju da kvaliteta detekcija izravno utječe na ukupnu učinkovitost algoritma praćenja, ali ima ograničen doprinos u rješavanju problema reidentifikacije plovila nakon okluzije. Iako bolji detektor povećava vjerojatnost ponovne detekcije plovila nakon okluzije, on ne rješava problem njegova ispravnog povezivanja s prethodnim identitetom, što potvrđuju rezultati iz Tablice 7.2: u odnosu na najslabiji detektor YOLO11n, najsnažniji detektor YOLO11m nakon okluzije uspješno reidentificira svega jedno plovilo više. Međutim, značajno poboljšanje u vidu stabilnosti i točnosti praćenja jasno je izraženo povećanjem općih metrika praćenja (HOTA: 81.34–85.57, MOTA: 89.19–93.03, IDF1: 90.73–94.38) i smanjenju broja promjena identiteta (IDSW: 77–42). Iz navedenog se može zaključiti da iako detekcija predstavlja nužan uvjet za uspješnu reidentifikaciju objekta nakon okluzije, ona sama po sebi nije dovoljna. Stoga je ključne izazove reidentifikacije, osobito u slučajevima duljih okluzija, potrebno rješavati na razini drugih komponenti algoritma, poput asocijacije i ekstrakcije vizualnih značajki, a ne isključivo kroz unaprjeđenje detekcijskog modela. U skladu s navedenim, može se zaključiti da je hipoteza H_2 potvrđena.

7.4.2. Uloga integracije vizualnih ReID značajki i informacija o kretanju u asocijaciji putanja i detekcija

Rezultati ablacijske studije 7.3.2 ukazuju na važnost zajedničke integracije vizualnih značajki i informacija o kretanju u postupku asocijacije putanja i detekcija. Uključivanje vizualnih značajki u drugoj fazi asocijacije pokazalo se ključnim za uspješnu reidentifikaciju plovila nakon okluzije, što je u skladu s njezinom funkcionalnom ulogom u povezivanju izgubljenih putanja s novim detekcijama nakon duljih prekida vidljivosti. U toj fazi, prostorno-dinamičke informacije dobivene iz modela kretanja često nisu dovoljne za pouzdanu asocijaciju, osobito u scenarijima s većim brojem plovila koja imaju slične obrasce kretanja kada vizualne značajke predstavljaju nužan dodatni izvor diskriminativne informacije.

S druge strane, izostavljanje vizualnih ReID značajki samo u prvoj fazi asocijacije ne dovodi do izrazitog pogoršanja performansi predloženog algoritma u vidu reidentifikacije plovila nakon okluzije. U promatranoj konfiguraciji zabilježena je čak i neznatno veća uspješnost reidentifikacije u kategoriji srednje dugih okluzija, dok je broj uspješno reidentificiranih plovila u ostalim kategorijama jednak konfiguraciji u kojoj se vizualne značajke koriste u obje faze asocijacije. Takvo ponašanje može se objasniti činjenicom da se prva faza asocijacije u velikoj mjeri oslanja na geometrijsku blizinu detekcija i dobivenih predviđanja putanja, koja je u uvjetima malih vremenskih razmaka često dovoljna za ispravnu asocijaciju. Međutim, kako ilustrira kvalitativni primjer prikazan na Slici 7.11, u situacijama koje uključuju privremeni gubitak objekta i istodobnu prisutnost drugih plovila u istom prostornom području, izostavljanje vizualnih informacija u prvoj fazi asocijacije može povećati vjerojatnost pogrešne dodjele identiteta te narušiti stabilnost praćenja. Takve se pogreške očituju u degradaciji HOTA, MOTA, IDSW i IDF1 metrika, što je vidljivo iz rezultata prikazanih u Tablici 7.3. Kako se povećava broj izgubljenih plovila (čije se putanje u sustavu zadržavaju određeni vremenski period), raste i rizik od pogrešnih asocijacija temeljenih isključivo na preklapanju graničnih okvira, budući da se prostorne pozicije izgubljenih i aktivno praćenih objekata mogu preklapati. Uključivanjem vizualnih ReID značajki u prvu fazu asocijacije sustavu se pruža dodatna diskriminativna informacija koja omogućuje razlikovanje trenutno detektiranog plovila od vizualno različitih izgubljenih plovila, čime se smanjuje vjerojatnost pogrešne dodjele identiteta.

Iz Tablice 7.3 može se vidjeti da konfiguracija u kojoj su vizualne ReID značajke uključene u obje faze asocijacije dosljedno ostvaruje najbolje ukupne rezultate prema standardnim evaluacijskim metrikama, uz najmanji broj zamjena identiteta. Izostavljanjem vizualnih značajki iz jedne od faza asocijacije dolazi do pogoršanja vrijednosti metrika, pri čemu je to pogoršanje izraženije kada se vizualne značajke izostave iz druge faze asocijacije, koja je namijenjena reidentifikaciji objekata nakon duljih okluzija. Najslabije performanse zabilježene su u konfiguraciji koja ne koristi vizualne značajke ni u jednoj fazi asocijacije. U cjelini, dobiveni rezultati pružaju empirijsku potporu hipotezi H_4, ukazujući na važnost integracije vizualnih ReID značajki i informacija o kretanju za stabilnije praćenje i pouzdanu asocijaciju detekcija i putanja u uvjetima dugotrajnih okluzija.

7.4.3. Uloga ReID modela i domenske prilagodbe

Analiza utjecaja korištenog modela za ekstrakciju vizualnih značajki ukazuje na to da stabilna reidentifikacija plovila ovisi o odabiru prikladne arhitekture ReID modela, kao i o njegovoj domenskoj prilagodbi. Iako su oba promatrana modela trenirana na domenski specifičnom SSMOT ReID skupu podataka, OSNet, kao arhitektura izvorno dizajnirana za zadatak reidentifikacije, postiže ukupno bolje rezultate praćenja od općeg ResNet50 modela. To se očituje u višim vrijednostima HOTA (85.57-83.57) i IDF1 (94.38-91.36) metrika, kao i u većem broju uspješnih reidentifikacija nakon okluzija (25-17), pri čemu je razlika osobito izražena u slučaju okluzija duljeg trajanja (9-4).

Dodatna analiza učinka domenske prilagodbe ukazuje na njezin izražen pozitivan utjecaj kod OSNet arhitekture, dok se kod ResNet50 arhitekture ne uočavaju značajna poboljšanja u odnosu na model treniran na općem skupu podataka. Dobiveni rezultati upućuju na to da učinak domenske prilagodbe ReID modela nije univerzalan, već je uvjetovan arhitekturom modela, pri čemu su pozitivni učinci izraženiji kod arhitektura specijaliziranih za zadatak reidentifikacije. U tom se smislu hipoteza H_3 o pozitivnoj ulozi domenske prilagodbe može smatrati djelomično potvrđenom.

S druge strane, ReID modeli trenirani na općim skupovima podataka mogu predstavljati prihvatljivo rješenje u situacijama kada domenski prilagođeni modeli nisu dostupni. Međutim, dobiveni rezultati ukazuju na njihova ograničenja u složenijim uvjetima praćenja, osobito u prisutnosti dugotrajnih okluzija, mimoilaženja plovila te u scenarijima s vizualno sličnim objektima.

7.4.4. Strategija asocijacije i robusnost na dugotrajne okluzije

Kao jedan od važnijih doprinosa ove disertacije može se istaknuti predloženi dvostupanjski pristup asocijaciji, s pažljivo odabranim mjerama sličnosti u pojedinim fazama asocijacije, koji pokazuje potencijal za poboljšanje performansi algoritma u uvjetima dugotrajnih okluzija, uz očuvanje stabilnosti praćenja.

Prva asocijacija (A1) zasniva se na geometrijsko-vizualnim kriterijima koji prioritiziraju visoku vizualnu sličnost i preklapanje između detektiranih i predviđenih graničnih okvira, čime se osigurava stabilan kontinuitet putanja u situacijama kada su nove detekcije dostupne i pouzdane. Međutim, kada zbog okluzije detekcije izostaju tijekom određenog vremenskog perioda, pogreška u predviđanju Kalmanovanovog filtera postupno se akumulira, zbog čega se predviđeni granični okvir u pravilu više ne preklapa s novom detekcijom u trenutku ponovnog pojavljivanja plovila, što ograničava sposobnost ove faza asocijacije da samostalno vrati ispravan identitet plovila nakon okluzije. To potvrđuju rezultati ablacijske studije, prema kojima samostalna primjena asocijacije A1 postiže svega 6 uspješnih reidentifikacija nakon okluzije, dok se u slučaju dugotrajnih okluzija ne bilježi niti jedna uspješna reidentifikacija.

S druge strane, kombinacija Mahalanobisove i kosinusne udaljenosti (A2) pokazuje se znatno učinkovitijom u fazi reidentifikacije, budući da u obzir uzima i dinamiku kretanja plovila te nesigurnost predviđanja Kalmanovog filtera. Time se omogućuje tolerancija na veća odstupanja između predviđenog i stvarnog položaja objekta, što je osobito važno nakon duljih okluzija, kada geometrijsko preklapanje graničnih okvira često izostaje. Međutim, kada se takav fleksibilniji kriterij koristi samostalno dolazi do znatnog porasta zamjena identiteta, što potvrđuje vrijednost IDSW metrike asocijacije A2 u Tablici 7.6.

Predloženi dvostupanjski pristup koristi prostorno strožu asocijaciju u prvoj fazi kako bi se osigurala stabilnost identiteta kod pouzdanih pridruživanja, dok se prostorno fleksibilnija asocijacija s vizualnim značajkama i Mahalanobisovom udaljenošću koristi tek u drugoj fazi kako bi se omogućila reidentifikacija plovila nakon duljih prekida. Time se istodobno smanjuje broj zamjena identiteta za vrijeme praćenja, zadržava sposobnost reidentifikacije plovila nakon duljih okluzija i postiže stabilnost praćenja. To potvrđuju rezultati ablacijske studije prikazani u Tablici 7.6. Pri tome je važno, osobito u drugoj fazi asocijacije koja je usmjerena na rješavanje problema reidentifikacije objekta, koristiti nešto višu graničnu vrijednost za udaljenost vektora vizualnih značajki, kako se potencijalno ispravna pridruživanja ne bi odbacila u slučajevima većih vizualnih odstupanja koja mogu nastati uslijed duljih okluzija. Samim povećanjem granične vrijednosti u drugoj asocijaciji A2 s 0.2 na 0.4 broj uspješno reidentificiranih plovila nakon okluzije raste s 20 na 25.

Dodatnu potporu tezi o važnosti povećanja granične vrijednosti vizualne sličnosti u drugoj fazi asocijacije za pouzdanu reidentifikaciju plovila nakon dugotrajnih okluzija pružaju i rezultati ablacijske studije 7.3.5 (Tablica 7.7, Slika 7.16) u kojoj su uspoređene performanse izvornog StrongSORT algoritma, koji je bio inspiracija za drugu fazu asocijacije VOc-SORT-a, s originalnim pragom vizualne sličnosti od 0.2, te njegove modificirane varijante s povišenom graničnom vrijednosti od 0.4. Rezultati pokazuju da povećanje granične vrijednosti vizualne sličnosti doprinosi većoj uspješnosti reidentifikacije plovila nakon okluzija, osobito u slučaju dugotrajnih okluzija. Međutim, kao i u konfiguracijama "A2" koje se oslanjaju isključivo na drugu fazu asocijacije, modificirana varijanta StrongSORT algoritma i dalje bilježi povećan broj zamjena identiteta te generalno ostvaruje slabije performanse praćenja. Time se dodatno ističu prednosti dvostupanjske strategije asocijacije primijenjene u VOc-SORT algoritmu, koja omogućuje učinkovitije očuvanje identiteta uz ravnotežu pouzdanosti reidentifikacije i stabilnosti praćenja.

Prethodno razmatranje i dobiveni rezultati potvrđuju da unaprjeđenje postupka asocijacije detekcija i putanja može doprinijeti boljem očuvanju identiteta plovila tijekom dugotrajnih okluzija, u skladu s hipotezom H_5.

Produljeno vrijeme čekanja na uklanjanje putanja uvedeno je kako bi se omogućila reidentifikacija plovila i nakon duljih okluzija. Takav pristup, međutim, može dovesti do nakupljanja izgubljenih putanja plovila koja su vjerojatno trajno napustila kadar. Uvođenjem dodatnog kriterija za ranije uklanjanje putanja izgubljenih u blizini ruba kadra taj se problem reducira. Rezultati ablacijske studije pokazuju da navedeni kriterij ne utječe nepovoljno na uspješnost reidentifikacije, štoviše rezultira nešto višim vrijednostima HOTA i IDF1 metrika, te doprinosi stabilnijem radu algoritma i sprječava akumulaciju nerelevantnih putanja tijekom produljenog vremena čekanja.

7.4.5. Položaj predloženog algoritma u odnosu na postojeće metode

Jedna od ključnih pretpostavki ovog istraživanja, formulirana hipotezom H_1, jest da postojeće metode praćenja plovila pokazuju smanjenu sposobnost očuvanja identiteta u uvjetima dugotrajnih okluzija. Kvantitativni rezultati evaluacije na SSMOT skupu videozapisa, prikazani u Tablici 7.1 i dodatno ilustrirani analizom broja uspješnih reidentifikacija prema duljini okluzije sa Slike 7.4, potvrđuju navedenu hipotezu. Sve analizirane postojeće metode bilježe jako nizak broj uspješnih reidentifikacija u slučajevima srednje dugih i dugih okluzija, što ukazuje na ograničenu sposobnost očuvanja identiteta u takvim uvjetima.

Usporedba s postojećim metodama praćenja pokazuje da se predloženi VOc-SORT algoritam pozicionira kao rješenje koje uravnotežuje stabilnost praćenja i sposobnost reidentifikacije plovila u složenim pomorskim scenarijima. Posebno se ističe njegova učinkovitost u očuvanju identiteta nakon okluzija, s ukupno 25 uspješnih reidentifikacija, u usporedbi s 13 koliko ostvaruje sljedeća najbolja metoda, pri čemu je razlika još izraženija kod srednje dugih i dugih okluzija.

Usporedba VOc-SORT algortima s postojećim pristupima koji koriste vizualne značajke, poput BoT-SORT-a, StrongSORT-a i Deep OC-SORT-a, ukazuje da razlike u performansama ne proizlaze isključivo iz korištenja ReID modela, već ponajprije iz načina na koji su vizualne informacije integrirane u proces asocijacije. Nadalje, usporedba performansi varijante VOc-SORT algoritma u kojoj su vizualne informacije u potpunosti izostavljene (A1 ✗, A2 ✗ u Tablici 7.3) s postojećim pristupima koji također koriste samo prostorno-dinamičke informacije (Tablica 7.1, ReID ✗) pokazuje da VOc-SORT i u toj konfiguraciji ostvaruje više vrijednosti HOTA, MOTA i IDF1 metrika, uz manji broj zamjena identiteta te veći broj ispravno reidentificiranih plovila nakon okluzija. Štoviše, u promatranom eksperimentalnom okruženju, performanse varijante VOc-SORT algoritma koja ne koristi ReID značajke usporedive su, a u nekim slučajevima i bolje u odnosu na neke postojeće pristupe koji ih koriste.

Predloženi dvostupanjski postupak asocijacije razlikuje se od postojećih pristupa u algoritmima BoT-SORT i ByteTrack. U tim metodama povezivanje u prvoj fazi asocijacije temelji se isključivo na IoU udaljenosti (ByteTrack) ili na njenoj kombinaciji s vizualnim značajkama BoT-SORT), dok oba algoritma u drugoj fazi koriste samo IoU. OC-SORT algoritam također koristi dvije faze asocijacije, no u slučaju OC-SORT-a obje faze asocijacije koriste isključivo prostorne informacije i informaciju o kretanju: kombinaciju IoU i konzistontnosti smjera kretanja u prvoj fazi, te samo IoU u drugoj fazi. Nasuprot tome, u predloženom pristupu vizualne značajke koriste se u obje faze asocijacije, te se u drugoj fazi kriterij preklapanja graničnih okvira zamjenjuje se Mahalanobisovom udaljenošću koja je ključna za reidentifikaciju objekata nakon duljih zaklonjenosti kada eksplicitno preklapanje graničnih okvira predviđene putanje i detekcije izostaje. Navedeno je potkrijepljeno kvantitativnim rezultatima evaluacije iz Tablice 7.1, kao i analizom prikazanom na Slici 7.4, iz kojih je vidljivo da VOc-SORT algoritam pokazuje bolje rezultate u reidentifikaciji plovila nakon okluzija, osobito u slučajevima srednje dugih i dugih razdoblja zaklonjenosti.

U odnosu na postojeće pristupe, VOc-SORT se pokazuje kao uravnoteženo rješenje za praćenje više objekata u pomorskom okruženju. Složeniji duboki modeli neuronskih mreža koriste se isključivo ondje gdje su doista nužni, odnosno u fazama detekcije i ekstrakcije vizualnih značajki. Umjesto oslanjanja na dodatne duboke arhitekture, poboljšane performanse predloženog algoritma postignute su sustavnom kombinacijom postojećih jednostavnih koncepata kroz pažljivo osmišljen postupak asocijacije i domenski prilagođene vizualne značajke. Na ovaj način poboljšane su performanse praćenja u vidu stabilnosti i reidentifikacije, a zadržana je računalna učinkovitost koja je usporediva s ostalim metodama praćenja koje koriste ReID modele.

7.4.6. Ograničenja predloženog pristupa

Unatoč postignutim poboljšanjima, predloženi pristup ima određena ograničenja. Performanse VOc-SORT algoritma i dalje su u velikoj mjeri uvjetovane kvalitetom ulaznih detekcija, što predstavlja temeljno ograničenje pristupa temeljenih na detekciji. Poboljšana strategija asocijacije pomaže u reidentifikaciji plovila nakon okluzije, ali ne može ispravno ponovno identificirati objekt koji uopće nije detektiran. Nadalje, mogućnost reidentifikacije ograničena je maksimalnim vremenom čekanja putanja od 625 okvira, zbog čega okluzije ekstremnog trajanja i dalje predstavljaju otvoren izazov. Pojedini pragovi, uključujući i graničnu vrijednost vizualne sličnosti, određeni su empirijski, a njihov učinak van odabranih vrijednosti nije analiziran. U okviru rada također je zadržan standardni Kalmanov filter za predviđanje stanja putanja, dok potencijalni učinci alternativnih modela predviđanja nisu razmatrani. Konačno, evaluacija je provedena isključivo u pomorskoj domeni, na konkretnom primjeru praćenja plovila u splitskoj luci, generalizacija predloženog pristupa na druga pomorska okruženja i domene nije ispitana.

8. ZAKLJUČAK

Automatska detekcija i praćenje plovila predstavljaju ključne preduvjete za razvoj autonomnih sustava nadzora pomorskog prometa kojima bi se ublažila ograničenja postojećih rješenja temeljenih na radarskim sustavima i AIS-u te smanjila potreba za ljudskom intervencijom. U tom se kontekstu metode računalnog vida temeljene na dubokom učenju nameću kao relevantan istraživački smjer. Međutim, pouzdano praćenje plovila u stvarnim operativnim uvjetima ostaje izazovno zbog čestih zaklonjenosti i privremenih gubitaka detekcija, što otežava očuvanje identiteta i kontinuiteta putanja, osobito tijekom duljih prekida vidljivosti. Dodatno, u području praćenja plovila općenito je izražen nedostatak javno dostupnih i općeprihvaćenih referentnih skupova podataka, što ograničava objektivnu evaluaciju postojećih metoda i usporava sustavni razvoj novih algoritamskih rješenja.

Cilj ove doktorske disertacije je istražiti i eksperimentalno validirati sustav za automatsku detekciju i praćenje plovila u RGB videozapisima koji kombinira metode dubokog učenja i tradicionalne algoritamske pristupe, poput Kalmanovog filtera i mađarskog algoritma, s ciljem povećanja stabilnosti praćenja i očuvanja identiteta plovila u uvjetima dugotrajnih okluzija. U okviru tog cilja, disertacija adresira i izazov nedostatka referentnih skupova podataka za praćenje plovila razvojem novog SSMOT skupa podataka, koji omogućuje objektivnu usporedbu različitih pristupa za praćenje plovila te analizu njihovih performansi u uvjetima okluzija različitog trajanja.

Provedenim istraživanjem, u skladu s postavljenim istraživačkim ciljem, ostvareni su očekivani znanstveni doprinosi. Prvi doprinos odnosi se na izradu referentnog skupa označenih slika i videozapisa za detekciju, reidentifikaciju i praćenje plovila, koji uključuje jasno definirane primjere okluzija različitog trajanja. Takav skup podataka omogućuje objektivnu evaluaciju postojećih metoda, kao i sustavnu analizu njihovih performansi tijekom okluzija, te predstavlja vrijednu osnovu za daljnja istraživanja u području praćenja plovila. Drugi doprinos ostvaren je razvojem algoritma za praćenje plovila u RGB videozapisima s poboljšanim performansama u uvjetima dugotrajnih okluzija, čime se postiže stabilnije očuvanje identiteta i smanjenje fragmentacije putanja u odnosu na postojeće pristupe te se dodatno unapređuje robusnost praćenja u složenim pomorskim scenarijima.

Rezultati provedenih eksperimenata ukazuju na ograničenja postojećih algoritama praćenja u scenarijima koji uključuju dulje prekide vidljivosti plovila. Iako poboljšanje kvalitete detekcija pozitivno utječe na opće performanse praćenja, ono samo po sebi nije dostatno za pouzdanu reidentifikaciju plovila nakon dugotrajnih okluzija. U tom je kontekstu u sklopu disertacije predložen algoritam praćenja plovila VOc-SORT, koji integrira domenski prilagođene vizualne značajke i prostorno-dinamičke informacije unutar dvostupanjskog postupka asocijacije, čime se dodatno unapređuje sposobnost očuvanje identiteta plovila. Dobiveni rezultati potvrđuju da takav pristup omogućuje stabilnije praćenje plovila, uz smanjenje zamjena identiteta i povećanje uspješnosti reidentifikacije nakon duljih razdoblja zaklonjenosti. Na temelju dobivenih rezultata može se smatrati da su postavljene pomoćne istraživačke hipoteze empirijski utemeljene, što ujedno podupire i glavnu istraživačku hipotezu disertacije.

Unatoč postignutim poboljšanjima i potvrđenoj učinkovitosti predloženog pristupa, provedeno istraživanje ima određena ograničenja koja je potrebno uzeti u obzir. Jedno od ključnih ograničenja odnosi se na prostornu i kontekstualnu ograničenost skupa podataka SSMOT, koji je prikupljen na jednoj lokaciji i s jednom fiksnom kamerom, zbog čega generalizacija dobivenih rezultata na različita pomorska okruženja, konfiguracije kamera i klimatske uvjete zahtijeva dodatnu eksperimentalnu provjeru. Nadalje, predloženi sustav temelji se isključivo na vizualnim informacijama RGB kamere, pa se u uvjetima izrazito loše vidljivosti, poput guste magle, noćnih uvjeta bez rasvjete ili jakih atmosferskih smetnji, može očekivati degradacija performansi. Iako se prilikom implementacije VOc-SORT algoritma pazilo na njegovu računsku složenost, znatno povećanje broja istovremeno praćenih plovila ili primjena na sustave s ograničenim hardverskim resursima može predstavljati izazov. Također, zbog definiranog ograničenja na maksimalnu duljinu čekanja pojavljivanja plovila nakon okluzije, ovako implementirani algoritam nije otporan na okluzije ekstremnog trajanja, koje i dalje predstavljaju otvoren izazov.

Polazeći od navedenog, buduća istraživanja mogu biti usmjerena u nekoliko smjerova. Prije svega, preporučuje se proširiti referentni SSMOT skup podataka na veći broj lokacija i raznovrsnija pomorska okruženja, uključujući otvoreno more i više luka, kako bi se omogućila robusnija evaluacija i poboljšala generalizacija predloženog pristupa. Nadalje, integracija dodatnih senzora, poput termalnih kamera, radara ili AIS podataka, otvara mogućnost razvoja višesenzorskih sustava povećane otpornosti na loše uvjete vidljivosti. Buduća istraživanja mogu se dodatno usmjeriti na ispitivanje naprednijih pristupa predviđanju kretanja plovila, razvoj adaptivnih strategija asocijacije prilagođenih promjenjivim uvjetima praćenja te na analizu potencijala primjene transformerskih modela, uz poseban naglasak na optimizaciju njihove računalne učinkovitosti i mogućnost primjene u stvarnom vremenu.

LITERATURA

[1] B. Iancu, V. Soloviev, L. Zelioli i J. Lilius, Aboships—an inshore and offshore maritime vessel detection dataset with precise annotations, *Remote Sensing*, 13, 5, 988, 2021.

[2] B. Zhang, J. Liu, R. W. Liu i Y. Huang, Deep-learning-empowered visual ship detection and tracking: Literature review and future direction, *Engineering Applications of Artificial Intelligence*, 141, 109754, 2025.

[3] M. H. Zwemer, R. G. Wijnhoven i P. H. de With, Ship detection in harbour surveillance based on large-scale data and cnns., *VISIGRAPP (5: VISAPP)*, 153–160, 2018.

[4] P. Kaur, A. Aziz, D. Jain, H. Patel, J. Hirokawa, L. Townsend, C. Reimers i F. Hua, Sea situational awareness (seasaw) dataset, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2579–2587, 2022.

[5] F. L. Chevalier, Radar, without tears, 2017.

[6] G. Longo, E. Russo, A. Armando i A. Merlo, Attacking (and defending) the maritime radar system, *IEEE Transactions on Information Forensics and Security*, 18, 3575–3589, 2023.

[7] E. Badurina, Automatski identifikacijski sustav (ais), *Pomorski zbornik*, 40, 1, 79–94, 2002.

[8] M. Tomović i Z. Lušić, e-navigacija, *Kapetanov Glasnik*, , 37, 20–29, 2019.

[9] D. Weihmayr, F. Sezgin, L. Tolksdorf, C. Birkner i R. N. Jazar, Predicting the influence of adverse weather on pedestrian detection with automotive radar and lidar sensors, 2024.

[10] C. Aydogdu, G. K. Carvajal, O. Eriksson, H. Hellsten, H. Herbertsson, M. F. Keskin, E. Nilsson, M. Rydström, K. Vanäs i H. Wymeersch, Radar interference mitigation for automated driving, 2019.

[11] D. Xu i H. Zhang, Study of low-altitude slow and small target detection on radar, *2017 5th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2017)*, 529–532, Atlantis Press, 2017.

[12] A. Androjna, M. Perkovič, I. Pavic i J. Mišković, Ais data vulnerability indicated by a spoofing case-study, *Applied Sciences*, 11, 11, 5015, 2021.

[13] D. Yang, M. I. Solihin, Y. Zhao, B. Yao, C. Chen, B. Cai i A. Machmudah, A review of intelligent ship marine object detection based on rgb camera, *IET Image Processing*, 18, 2, 281–297, 2024.

[14] Y. LeCun, Y. Bengio i G. Hinton, Deep learning, *nature*, 521, 7553, 436–444, 2015.

[15] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard i L. D. Jackel, Backpropagation applied to handwritten zip code recognition, *Neural computation*, 1, 4, 541–551, 1989.

[16] D. H. Hubel i T. N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, *The Journal of physiology*, 195, 1, 215–243, 1968.

[17] K. Fukushima, Neocognitron, *Scholarpedia*, 2, 1, 1717, 2007, revision #91558.

[18] R. Zhang, S. Li, G. Ji, X. Zhao, J. Li i M. Pan, Survey on deep learning-based marine object detection, *Journal of Advanced Transportation*, 2021, 1, 5808206, 2021.

[19] Y. Shi, H. Wang i S. Li, Ship detection algorithm based on structural reparameterize dilated large-kernel convolution and spatial selective kernel mechanism, *Journal of Real-Time Image Processing*, 22, 2, 70, 2025.

[20] R. D. S. Moreira, N. F. Ebecken, A. S. Alves, F. Livernet i A. Campillo-Navetti, A survey on video detection and tracking of maritime vessels, *International Journal of Recent Research and Applied Studies*, 20, 1, 2014.

[21] B. Kiefer, M. Kristan, J. Perš, L. Žust, F. Poiesi, F. Andrade, A. Bernardino, M. Dawkins, J. Raitoharju, Y. Quan et al., 1st workshop on maritime computer vision (macvi) 2023: Challenge results, *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 265–302, 2023.

[22] X. Zhou, S. Chan, C. Qiu, X. Jiang i T. Tang, Multi-target tracking based on a combined attention mechanism and occlusion sensing in a behavior-analysis system, *Sensors*, 23, 6, 2956, 2023.

[23] P. Dendorfer, V. Yugay, A. Osep i L. Leal-Taixé, Quo vadis: Is trajectory forecasting the key towards long-term multi-object tracking?, *Advances in Neural Information Processing Systems*, 35, 15657–15671, 2022.

[24] B. Dominguez-Dager, F. Escalona, F. Gomez-Donoso i M. Cazorla, Chirla: Comprehensive high-resolution identification and re-identification for large-scale analysis, *arXiv preprint arXiv:2502.06681*, 2025.

[25] S. Zhang, K. Zheng i S. Huaiyuan, Analysis of the occlusion interference problem in target tracking, *Mathematical Problems in Engineering*, 2022, 1, 4605111, 2022.

[26] R. E. Kalman et al., Contributions to the theory of optimal control, *Bol. soc. mat. mexicana*, 5, 2, 102–119, 1960.

[27] J. Cao, J. Pang, X. Weng, R. Khirodkar i K. Kitani, Observation-centric sort: Rethinking sort for robust multi-object tracking, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9686–9696, 2023.

[28] D. M. Jiménez-Bravo, Álvaro Lozano Murciego, A. Sales Mendes, H. Sánchez San Blás i J. Bajo, Multi-object tracking in traffic environments: A systematic literature review, *Neurocomputing*, 494, 43–55, 2022.

[29] J. Ferryman i A. Shahrokni, Pets2009: Dataset and challenge, *2009 Twelfth IEEE international workshop on performance evaluation of tracking and surveillance*, 1–6, IEEE, 2009.

[30] L. Leal-Taixé, A. Milan, I. Reid, S. Roth i K. Schindler, Motchallenge 2015: Towards a benchmark for multi-target tracking, *arXiv preprint arXiv:1504.01942*, 2015.

[31] A. Milan, L. Leal-Taixé, I. Reid, S. Roth i K. Schindler, Mot16: A benchmark for multi-object tracking, *arXiv preprint arXiv:1603.00831*, 2016.

[32] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler i L. Leal-Taixe, Cvpr19 tracking and detection challenge: How crowded can it get?, 2019.

[33] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler i L. Leal-Taixé, Mot20: A benchmark for multi object tracking in crowded scenes, *arXiv preprint arXiv:2003.09003*, 2020.

[34] A. Geiger, P. Lenz i R. Urtasun, Are we ready for autonomous driving? the kitti vision benchmark suite, *2012 IEEE conference on computer vision and pattern recognition*, 3354–3361, IEEE, 2012.

[35] A. Geiger, P. Lenz, C. Stiller i R. Urtasun, Vision meets robotics: The kitti dataset, *The International Journal of Robotics Research*, 32, 11, 1231–1237, 2013.

[36] L. Wen, D. Du, Z. Cai, Z. Lei, M.-C. Chang, H. Qi, J. Lim, M.-H. Yang i S. Lyu, Ua-detrac: A new benchmark and protocol for multi-object detection and tracking, *Computer Vision and Image Understanding*, 193, 102907, 2020.

[37] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan i T. Darrell, Bdd100k: A diverse driving dataset for heterogeneous multitask learning, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2636–2645, 2020.

[38] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine et al., Scalability in perception for autonomous driving: Waymo open dataset, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2446–2454, 2020.

[39] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan i O. Beijbom, nuscenes: A multimodal dataset for autonomous driving, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11621–11631, 2020.

[40] Y. Liao, J. Xie i A. Geiger, Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45, 3, 3292–3310, 2022.

[41] Z. Soleimanitaleb i M. A. Keyvanrad, Single Object Tracking: A Survey of Methods, Datasets, and Evaluation Metrics, 1–15, jan 2022.

[42] D. Meimetis, I. Daramouskas, I. Perikos i I. Hatzilygeroudis, *Real-time multiple object tracking using deep learning methods*, 35, Springer London, 2023.

[43] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu i T. K. Kim, Multiple object tracking: A literature review, *Artificial Intelligence*, 293, 103448, 2021.

[44] G. Ciaparrone, F. Luque Sánchez, S. Tabik, L. Troiano, R. Tagliaferri i F. Herrera, Deep learning in video multi-object tracking: A survey, *Neurocomputing*, 381, 61–88, 2020.

[45] H. Wang, S. Wang, J. Lv, C. Hu i Z. Li, Non-local attention association scheme for online multi-object tracking, *Image and Vision Computing*, 102, 103983, 2020.

[46] X. Gao, Z. Wang, X. Wang, S. Zhang, S. Zhuang i H. Wang, Dettrack: An algorithm for multiple object tracking by improving occlusion object detection, *Electronics*, 13, 1, 91, 2023.

[47] Z. Sun, J. Chen, L. Chao, W. Ruan i M. Mukherjee, A survey of multiple pedestrian tracking based on tracking-by-detection framework, *IEEE Transactions on Circuits and Systems for Video Technology*, 31, 5, 1819–1833, 2021.

[48] S. Guo, S. Wang, Z. Yang, L. Wang, H. Zhang, P. Guo, Y. Gao i J. Guo, A review of deep learning-based visual multi-object tracking algorithms for autonomous driving, *Applied Sciences*, 12, 21, 10741, 2022.

[49] S. Guo, S. Wang, Z. Yang, L. Wang, H. Zhang, P. Guo, Y. Gao i J. Guo, A Review of Deep Learning-Based Visual Multi-Object Tracking Algorithms for Autonomous Driving, *Applied Sciences*, 12, 21, 10741, oct 2022.

[50] W. Zhang, W. Deng, Z. Cui, J. Liu i L. Jiao, Object knowledge distillation for joint detection and tracking in satellite videos, *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[51] X. Gong, Z. Le, Y. Wu i H. Wang, Real-time multiobject tracking based on multiway concurrency, *Sensors*, 21, 3, 2021.

[52] Z. Cao, J. Li, D. Zhang, M. Zhou i A. Abusorrah, A multi-object tracking algorithm with center-based feature extraction and occlusion handling, *IEEE Transactions on Intelligent Transportation Systems*, 24, 4, 4464–4473, 2022.

[53] Q. Wang, C. Lu, L. Gao i G. He, Transformer-based multiple-object tracking via anchor-based-query and template matching, *Sensors (Basel, Switzerland)*, 24, 2023.

[54] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser i I. Polosukhin, Attention is all you need, *Advances in neural information processing systems*, 30, 2017.

[55] X. Zhu, Y. Jia, S. Jian, L. Gu i Z. R. Pu, Vitt: Vision transformer tracker, *Sensors (Basel, Switzerland)*, 21, 2021.

[56] T. Liu, P. Li, Y.-Z. Gu, P. Liu i H. Wang, Adaptive offloading of transformer inference for weak edge devices with masked autoencoders, *ACM Transactions on Sensor Networks*, 2024.

[57] Y. Xu, Y. Ban, G. Delorme, C. Gan, D. Rus i X. Alameda-Pineda, Transcenter: Transformers with dense representations for multiple-object tracking, *IEEE transactions on pattern analysis and machine intelligence*, 45, 6, 7820–7835, 2022.

[58] A. Bewley, Z. Ge, L. Ott, F. Ramos i B. Upcroft, Simple online and realtime tracking, *2016 IEEE international conference on image processing (ICIP)*, 3464–3468, IEEE, 2016.

[59] N. Wojke, A. Bewley i D. Paulus, Simple online and realtime tracking with a deep association metric, *2017 IEEE international conference on image processing (ICIP)*, 3645–3649, IEEE, 2017.

[60] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu i X. Wang, Byte-track: Multi-object tracking by associating every detection box, *European Conference on Computer Vision*, 1–21, Springer, 2022.

[61] N. Aharon, R. Orfaig i B.-Z. Bobrovsky, Bot-sort: Robust associations multi-pedestrian tracking, *arXiv preprint arXiv:2206.14651*, 2022.

[62] Y. Du, Z. Zhao, Y. Song, Y. Zhao, F. Su, T. Gong i H. Meng, Strongsort: Make deepsort great again, *IEEE Transactions on Multimedia*, 2023.

[63] V. D. Stanojevic i B. T. Todorovic, Boosttrack: Boosting the similarity measure and detection confidence for improved multiple object tracking, *Machine Vision and Applications*, 35, 3, 53, 2024.

[64] P. Bergmann, T. Meinhardt i L. Leal-Taixe, Tracking without bells and whistles, *Proceedings of the IEEE/CVF international conference on computer vision*, 941–951, 2019.

[65] X. Zhou, V. Koltun i P. Krähenbühl, Tracking objects as points, *European conference on computer vision*, 474–490, Springer, 2020.

[66] Y. Zhang, C. Wang, X. Wang, W. Zeng i W. Liu, Fairmot: On the fairness of detection and re-identification in multiple object tracking, *International Journal of Computer Vision*, 129, 3069–3087, 2021.

[67] Z. Wang, L. Zheng, Y. Liu, Y. Li i S. Wang, Towards Real-Time Multi-Object Tracking, *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, 107–122, Springer, 2020.

[68] P. Sun, J. Cao, Y. Jiang, R. Zhang, E. Xie, Z. Yuan, C. Wang i P. Luo, Transtrack: Multiple object tracking with transformer, *arXiv preprint arXiv:2012.15460*, 2020.

[69] T. Meinhardt, A. Kirillov, L. Leal-Taixe i C. Feichtenhofer, Trackformer: Multi-object tracking with transformers, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8844–8854, 2022.

[70] P. Chu, J. Wang, Q. You, H. Ling i Z. Liu, Transmot: Spatial-temporal graph transformer for multiple object tracking, *Proceedings of the IEEE/CVF Winter Conference on applications of computer vision*, 4870–4880, 2023.

[71] P. Voigtlaender, M. Krause, A. Osep, J. Luiten, B. B. G. Sekar, A. Geiger i B. Leibe, Mots: Multi-object tracking and segmentation, *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, 7942–7951, 2019.

[72] J. Peng, C. Wang, F. Wan, Y. Wu, Y. Wang, Y. Tai, C. Wang, J. Li, F. Huang i Y. Fu, Chained-tracker: Chaining paired attentive regression results for end-to-end joint multiple-object detection and tracking, *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*, 145–161, Springer, 2020.

[73] Y. Xiao, Z. Tian, J. Yu, Y. Zhang, S. Liu, S. Du i X. Lan, A review of object detection based on deep learning, *Multimedia Tools and Applications*, 79, 23729–23791, 2020.

[74] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng i R. Qu, A survey of deep learning-based object detection, *IEEE Access*, 7, 128837–128868, 2019.

[75] F. Yu, W. Li, Q. Li, Y. Liu, X. Shi i J. Yan, Poi: Multiple object tracking with high performance detection and appearance feature, *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II 14*, 36–42, Springer, 2016.

[76] N. Ran, L. Kong, Y. Wang i Q. Liu, A robust multi-athlete tracking algorithm by exploiting discriminant features and long-term dependencies, *MultiMedia Modeling: 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8–11, 2019, Proceedings, Part I 25*, 411–423, Springer, 2019.

[77] Y. Lu, C. Lu i C.-K. Tang, Online video object detection using association lstm, *Proceedings of the IEEE international conference on computer vision*, 2344–2352, 2017.

[78] D. Zhao, H. Fu, L. Xiao, T. Wu i B. Dai, Multi-object tracking with correlation filter for autonomous vehicle, *Sensors*, 18, 7, 2004, 2018.

[79] I. Ahmed, S. Din, G. Jeon, F. Piccialli i G. Fortino, Towards collaborative robotics in top view surveillance: A framework for multiple object tracking by detection using deep learning, *IEEE/CAA Journal of Automatica Sinica*, 8, 7, 1253–1270, 2021.

[80] J. Wu, J. Cao, L. Song, Y. Wang, M. Yang i J. Yuan, Track to detect and segment: An online multi-object tracker, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12352–12361, 2021.

[81] T. Ogawa, T. Shibata i T. Hosoi, Frog-mot: Fast and robust generic multiple-object tracking by iou and motion-state associations, *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 6563–6572, 2024.

[82] A. Moksyakov, Y. Wu, S. A. Gadsden, J. Yawney i M. AlShabi, Object detection and tracking with yolo and the sliding innovation filter, *Sensors (Basel, Switzerland)*, 24, 2024.

[83] M. A. Islam, W. Xing, J. Zhou, Y. Gao i K. K. Paliwal, Hy-tracker: A novel framework for enhancing efficiency and accuracy of object tracking in hyperspectral videos, *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[84] R. Girshick, J. Donahue, T. Darrell i J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 580–587, 2014.

[85] R. Girshick, Fast r-cnn, *Proceedings of the IEEE international conference on computer vision*, 1440–1448, 2015.

[86] S. Ren, K. He, R. Girshick i J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, *Advances in neural information processing systems*, 28, 2015.

[87] K. He, G. Gkioxari, P. Dollár i R. Girshick, Mask r-cnn, *Proceedings of the IEEE international conference on computer vision*, 2961–2969, 2017.

[88] J. Redmon, S. Divvala, R. Girshick i A. Farhadi, You only look once: Unified, real-time object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788, 2016.

[89] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu i A. C. Berg, Ssd: Single shot multibox detector, *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, 21–37, Springer, 2016.

[90] T.-Y. Lin, P. Goyal, R. Girshick, K. He i P. Dollár, Focal loss for dense object detection, *Proceedings of the IEEE international conference on computer vision*, 2980–2988, 2017.

[91] X. Zhou, D. Wang i P. Krähenbühl, Objects as points, *arXiv preprint arXiv:1904.07850*, 2019.

[92] M. Tan, R. Pang i Q. V. Le, Efficientdet: Scalable and efficient object detection, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10781–10790, 2020.

[93] J. Xiang, G. Zhang i J. Hou, Online multi-object tracking based on feature representation and bayesian filtering within a deep learning architecture, *IEEE Access*, 7, 27923–27935, 2019.

[94] N. Mahmoudi, S. M. Ahadi i M. Rahmati, Multi-target tracking using cnn-based features: Cnnmtt, *Multimedia Tools and Applications*, 78, 6, 7077–7096, 2019.

[95] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier i L. Van Gool, Robust tracking-by-detection using a detector confidence particle filter, *2009 IEEE 12th International Conference on Computer Vision*, 1515–1522, 2009.

[96] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier i L. Van Gool, Online multiperson tracking-by-detection from a single, uncalibrated camera, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33, 9, 1820–1833, 2011.

[97] T. Kokul, A. Ramanan i U. Pinidiyaarachchi, Online multi-person tracking-by-detection method using acf and particle filter, *2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS)*, 529–536, 2015.

[98] A. Milan, S. H. Rezatofighi, A. Dick, I. Reid i K. Schindler, Online multi-target tracking using recurrent neural networks, *Proceedings of the AAAI conference on Artificial Intelligence*, 31, 2017.

[99] M. Babaee, Z. Li i G. Rigoll, Occlusion handling in tracking multiple people using rnn, *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2715–2719, 2018.

[100] C. Kim, F. Li i J. M. Rehg, Multi-object tracking with neural gating using bilinear lstm, *Proceedings of the European conference on computer vision (ECCV)*, 200–215, 2018.

[101] J. Yang, H. Ge, S. Su i G. Liu, Transformer-based two-source motion model for multi-object tracking, *Applied Intelligence*, 1–13, 2022.

[102] M. Alapić i I. Velčić, Izvod jednadžbi diskretnog kalmanovog filtera, *Osječki matematički list*, 18, 2, 105–122, 2018.

[103] G. Welch, G. Bishop et al., An introduction to the kalman filter, 1995.

[104] Y. Du, J. Wan, Y. Zhao, B. Zhang, Z. Tong i J. Dong, Giaotracker: A comprehensive framework for mcmot with global information and optimizing strategies in visdrone 2021, *Proceedings of the IEEE/CVF International conference on computer vision*, 2809–2819, 2021.

[105] G. L. Smith, S. F. Schmidt i L. A. McGee, *Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle*, 135, National Aeronautics and Space Administration, 1962.

[106] S. J. Julier i J. K. Uhlmann, New extension of the kalman filter to nonlinear systems, *Signal processing, sensor fusion, and target recognition VI*, 3068, 182–193, Spie, 1997.

[107] E. N. Chatzi i A. W. Smyth, The unscented kalman filter and particle filter methods for nonlinear structural system identification with non-collocated heterogeneous sensing, *Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and Monitoring and of the European Association for the Control of Structures*, 16, 1, 99–123, 2009.

[108] F. Gustafsson i G. Hendeby, Some relations between extended and unscented kalman filters, *IEEE Transactions on Signal Processing*, 60, 2, 545–555, 2011.

[109] K. György, A. Kelemen i L. Dávid, Unscented kalman filters and particle filter methods for nonlinear state estimation, *Procedia Technology*, 12, 65–74, 2014.

[110] A. Bera, D. Wolinski, J. Pettré i D. Manocha, Realtime pedestrian tracking and prediction in dense crowds, *Group and Crowd Behavior for Computer Vision*, 391–415, Elsevier, 2017.

[111] A. Basia, Z. Simeu-Abazi, E. Gascard i P. Zwolinski, Review on state of health estimation methodologies for lithium-ion batteries in the context of circular economy, *CIRP Journal of Manufacturing Science and Technology*, 32, 517–528, 2021.

[112] C. Kim, F. Li, A. Ciptadi i J. M. Rehg, Multiple hypothesis tracking revisited, *2015 IEEE International Conference on Computer Vision (ICCV)*, 4696–4704, 2015.

[113] S. Chaib, H. Liu, Y. Gu i H. Yao, Deep feature fusion for vhr remote sensing scene classification, *IEEE Transactions on Geoscience and Remote Sensing*, 55, 8, 4775–4784, 2017.

[114] X. Cao, S. Guo, J. Lin, W. Zhang i M. Liao, Online tracking of ants based on deep association metrics: method, dataset and evaluation, *Pattern Recognition*, 103, 107233, 2020.

[115] W. Gong, H. Chen, Z. Zhang, M. Zhang i H. Gao, A data-driven-based fault diagnosis approach for electrical power dc-dc inverter by using modified convolutional neural network with global average pooling and 2-d feature image, *Ieee Access*, 8, 73677–73697, 2020.

[116] I. Marin, S. Mladenović, S. Gotovac i G. Zaharija, Deep-feature-based approach to marine debris classification, *Applied Sciences*, 11, 12, 5644, 2021.

[117] K. P. F.R.S., Liii. on lines and planes of closest fit to systems of points in space, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2, 11, 559–572, 1901.

[118] M. Ullah i F. A. Cheikh, Deep feature based end-to-end transportation network for multi-target tracking, *2018 25th IEEE international conference on image processing (ICIP)*, 3738–3742, IEEE, 2018.

[119] H. Sheng, Y. Zhang, J. Chen, Z. Xiong i J. Zhang, Heterogeneous association graph fusion for target association in multiple object tracking, *IEEE Transactions on Circuits and Systems for Video Technology*, 29, 11, 3269–3280, 2019.

[120] L. Wen, D. Du, S. Li, X. Bian i S. Lyu, Learning non-uniform hypergraph for multi-object tracking, *Proceedings of the AAAI conference on artificial intelligence*, 33, 8981–8988, 2019.

[121] L. Chen, X. Peng i M. Ren, Recurrent metric networks and batch multiple hypothesis for multi-object tracking, *IEEE Access*, 7, 3093–3105, 2018.

[122] B. Shuai, A. G. Berneshawi, D. Modolo i J. Tighe, Multi-object tracking with siamese track-rcnn, *arXiv preprint arXiv:2004.07786*, 2020.

[123] S. Tang, M. Andriluka, B. Andres i B. Schiele, Multiple people tracking by lifted multicut and person re-identification, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[124] W. Zhang, H. Zhou, S. Sun, Z. Wang, J. Shi i C. C. Loy, Robust multi-modality multi-object tracking, *Proceedings of the IEEE/CVF international conference on computer vision*, 2365–2374, 2019.

[125] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li i L. Fei-Fei, Imagenet: A large-scale hierarchical image database, *2009 IEEE conference on computer vision and pattern recognition*, 248–255, Ieee, 2009.

[126] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger i R. Shah, Signature verification using a" siamese" time delay neural network, *Advances in neural information processing systems*, 6, 1993.

[127] R. Hadsell, S. Chopra i Y. LeCun, Dimensionality reduction by learning an invariant mapping, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2, 1735–1742, 2006.

[128] F. Schroff, D. Kalenichenko i J. Philbin, Facenet: A unified embedding for face recognition and clustering, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 815–823, 2015.

[129] M. Kim, S. Alletto i L. Rigazio, Similarity mapping with enhanced siamese network for multi-object tracking, *arXiv preprint arXiv:1609.09156*, 2016.

[130] B. Wang, L. Wang, B. Shuai, Z. Zuo, T. Liu, K. L. Chan i G. Wang, Joint learning of convolutional neural networks and temporally constrained metrics for tracklet association, *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 386–393, 2016.

[131] L. Leal-Taixé, C. Canton-Ferrer i K. Schindler, Learning by tracking: Siamese cnn for robust target association, *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 33–40, 2016.

[132] Z. Zhou, J. Xing, M. Zhang i W. Hu, Online multi-target tracking with tensor-based high-order graph matching, *2018 24th International Conference on Pattern Recognition (ICPR)*, 1809–1814, IEEE, 2018.

[133] L. Chen, H. Ai, Z. Zhuang i C. Shang, Real-time multiple people tracking with deeply learned candidate selection and person re-identification, *2018 IEEE international conference on multimedia and expo (ICME)*, 1–6, IEEE, 2018.

[134] S. Zhang, Y. Gong, J.-B. Huang, J. Lim, J. Wang, N. Ahuja i M.-H. Yang, Tracking persons-of-interest via adaptive discriminative features, B. Leibe, J. Matas, N. Sebe i M. Welling, editori, *Computer Vision – ECCV 2016*, 415–433, Springer International Publishing, Cham, 2016.

[135] S.-H. Bae i K.-J. Yoon, Confidence-based data association and discriminative deep appearance learning for robust online multi-object tracking, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40, 3, 595–610, 2018.

[136] E. Bochinski, V. Eiselein i T. Sikora, High-speed tracking-by-detection without using image information, *2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS)*, 1–6, IEEE, 2017.

[137] D. Stadler, A detailed study of the association task in tracking-by-detection-based multi-person tracking, *Proceedings of the 2022 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory*, 62, 59–85, KIT Scientific Publishing Karlsruhe, Germany, 2023.

[138] M. P. Chandra et al., On the generalised distance in statistics, *Proceedings of the National Institute of Sciences of India*, 2, 49–55, 1936.

[139] D. Stadler i J. Beyerer, An improved association pipeline for multi-person tracking, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3170–3179, 2023.

[140] H. W. Kuhn, The hungarian method for the assignment problem, *Naval research logistics quarterly*, 2, 1-2, 83–97, 1955.

[141] J. Munkres, Algorithms for the assignment and transportation problems, *Journal of the society for industrial and applied mathematics*, 5, 1, 32–38, 1957.

[142] B. Wu i R. Nevatia, Tracking of multiple, partially occluded humans based on static body part detection, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 1, 951–958, 2006.

[143] C. Kim, L. Fuxin, M. Alotaibi i J. M. Rehg, Discriminative appearance modeling with multi-track pooling for real-time multi-object tracking, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9553–9562, 2021.

[144] H. Azizpour, A. Sharif Razavian, J. Sullivan, A. Maki i S. Carlsson, From generic to specific deep representations for visual recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 36–45, 2015.

[145] H. Liu, H. Zhang i C. Mertz, Deepda: Lstm-based deep data association network for multi-targets tracking in clutter, *2019 22th International Conference on Information Fusion (FUSION)*, 1–8, IEEE, 2019.

[146] K. Yoon, D. Y. Kim, Y.-C. Yoon i M. Jeon, Data association for multi-object tracking via deep neural networks, *Sensors*, 19, 3, 559, 2019.

[147] M. M. Morsali, Z. Sharifi, F. Fallah, S. Hashembeiki, H. Mohammadzade i S. B. Shouraki, Sfsort: Scene features-based simple online real-time tracker, *arXiv preprint arXiv:2404.07553*, 2024.

[148] M. Yang, G. Han, B. Yan, W. Zhang, J. Qi, H. Lu i D. Wang, Hybrid-sort: Weak cues matter for online multi-object tracking, *Proceedings of the AAAI conference on artificial intelligence*, 38, 6504–6512, 2024.

[149] G. D. Evangelidis i E. Z. Psarakis, Parametric image alignment using enhanced correlation coefficient maximization, *IEEE transactions on pattern analysis and machine intelligence*, 30, 10, 1858–1865, 2008.

[150] H. Luo, W. Jiang, Y. Gu, F. Liu, X. Liao, S. Lai i J. Gu, A strong baseline and batch normalization neck for deep person re-identification, *IEEE Transactions on Multimedia*, 22, 10, 2597–2609, 2019.

[151] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, Y. Sun, T. He, J. Mueller, R. Manmatha et al., Resnest: Split-attention networks, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2736–2746, 2022.

[152] G. Maggiolino, A. Ahmad, J. Cao i K. Kitani, Deep oc-sort: Multi-pedestrian tracking by adaptive re-identification, *2023 IEEE International conference on image processing (ICIP)*, 3025–3029, IEEE, 2023.

[153] K. Bernardin i R. Stiefelhagen, Evaluating multiple object tracking performance: the clear mot metrics, *EURASIP Journal on Image and Video Processing*, 2008, 1–10, 2008.

[154] J. Luiten, A. Ošep, P. Dendorfer, P. Torr, A. Geiger, L. Leal-Taixé i B. Leibe, Hota: A higher order metric for evaluating multi-object tracking, *International Journal of Computer Vision*, 129, 2, 548–578, listopad 2020.

[155] E. Ristani, F. Solera, R. Zou, R. Cucchiara i C. Tomasi, Performance measures and a data set for multi-target, multi-camera tracking, *European conference on computer vision*, 17–35, Springer, 2016.

[156] T. Wang, K. Chen, W. Lin, J. See, Z. Zhang, Q. Xu i X. Jia, Spatio-temporal point process for multiple object tracking, *IEEE Transactions on Neural Networks and Learning Systems*, 34, 4, 1777–1788, travanj 2023.

[157] N.-D. Nguyen, T. Do, T. D. Ngo i D.-D. Le, An evaluation of deep learning methods for small object detection, *Journal of electrical and computer engineering*, 2020, 1, 3189691, 2020.

[158] Q. Feng, X. Xu i Z. Wang, Deep learning-based small object detection: A survey, *Mathematical Biosciences and Engineering*, 20, 4, 6551–6590, 2023.

[159] R. Sapkota, R. Qureshi, M. F. Calero, C. Badjugar, U. Nepal, A. Poulose, P. Zeno, U. B. P. Vaddevolu, S. Khan, M. Shoman et al., Yolov11 to its genesis: a decadal and comprehensive review of the you only look once (yolo) series, *arXiv preprint arXiv:2406.19407*, 2025.

[160] L. Zhang, G. Ding, G. Li, Y. Jiang, Z. Li i D. Li, An anti-occlusion optimization algorithm for multiple pedestrian tracking, *PLoS one*, 19, 1, e0291538, 2024.

[161] P. Sun, J. Cao, Y. Jiang, Z. Yuan, S. Bai, K. Kitani i P. Luo, Dancetrack: Multi-object tracking in uniform appearance and diverse motion, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20993–21002, 2022.

[162] W. Zhan, W. Yu, Y. Wang, T. Hu i B. Zhu, E-gnn: An enhanced method for multi-object tracking with collective motion patterns, *IEEE Robotics and Automation Letters*, 9, 4, 3403–3410, 2024.

[163] A. Krizhevsky, G. Hinton et al., Learning multiple layers of features from tiny images, 2009.

[164] G. Griffin, A. Holub i P. Perona, Caltech-256 object category dataset, 2007.

[165] M. Everingham, L. Van Gool, C. K. Williams, J. Winn i A. Zisserman, The pascal visual object classes (voc) challenge, *International journal of computer vision*, 88, 303–338, 2010.

[166] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár i C. L. Zitnick, Microsoft coco: Common objects in context, *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 740–755, Springer, 2014.

[167] M. M. Zhang, J. Choi, K. Daniilidis, M. T. Wolf i C. Kanan, Vais: A dataset for recognizing maritime imagery in the visible and infrared spectrums, *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 10–16, 2015.

[168] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally i C. Quek, Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey, *IEEE Transactions on Intelligent Transportation Systems*, 18, 8, 1993–2016, 2017.

[169] E. Gundogdu, B. Solmaz, V. Yücesoy i A. Koc, Marvel: A large-scale image dataset for maritime vessels, *Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part V 13*, 165–180, Springer, 2017.

[170] Z. Shao, W. Wu, Z. Wang, W. Du i C. Li, Seaships: A large-scale precisely annotated dataset for ship detection, *IEEE transactions on multimedia*, 20, 10, 2593–2604, 2018.

[171] Kaggle, Game of deep learning ship dataset, 2019, https://www.kaggle.com/datasets/arpitjain007/game-of-deep-learning-ship-datasets, [Zadnji pristup: 12. ožujka 2025.].

[172] Y. Zheng i S. Zhang, Mcships: A large-scale ship dataset for detection and fine-grained categorization in the wild, *2020 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6, IEEE, 2020.

[173] Z. Shao, J. Wang, L. Deng, X. Huang, T. Lu, F. Luo, R. Zhang, X. Lv, C. Dang, Q. Ding et al., Glsd: The global large-scale ship database and baseline evaluations, *arXiv preprint arXiv:2106.02773*, 2021.

[174] Y. Shan, S. Liu, Y. Zhang, M. Jing i H. Xu, Lmd-tship: vision based large-scale maritime ship tracking benchmark for autonomous navigation applications, *IEEE Access*, 9, 74370–74384, 2021.

[175] M. Ribeiro, B. Damas i A. Bernardino, Real-time ship segmentation in maritime surveillance videos using automatically annotated synthetic datasets, *Sensors*, 22, 21, 8090, 2022.

[176] M. Petković, I. Vujović, Z. Lušić i J. Šoda, Image dataset for neural network performance estimation with application to maritime ports, *Journal of marine science and engineering*, 11, 3, 578, 2023.

[177] Y. Guo, R. W. Liu, J. Qu, Y. Lu, F. Zhu i Y. Lv, Asynchronous trajectory matching-based multimodal maritime data fusion for vessel traffic surveillance in inland waterways, *IEEE Transactions on Intelligent Transportation Systems*, 24, 11, 12779–12792, 2023.

[178] N. Wang, Y. Wang, Y. Wei, B. Han i Y. Feng, Marine vessel detection dataset and benchmark for unmanned surface vehicles, *Applied Ocean Research*, 142, 103835, 2024.

[179] H. Fu, Y. Li, Y. Wang i L. Han, Maritime target detection method based on deep learning, *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, 878–883, 2018.

[180] K. He, X. Zhang, S. Ren i J. Sun, Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778, 2016.

[181] S. Ioffe i C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, *International conference on machine learning*, 448–456, pmlr, 2015.

[182] J. Zou, W. Yuan i M. Yu, Maritime target detection of intelligent ship based on faster r-cnn, *2019 Chinese Automation Congress (CAC)*, 4113–4117, 2019.

[183] L. Qi, B. Li, L. Chen, W. Wang, L. Dong, X. Jia, J. Huang, C. Ge, G. Xue i D. Wang, Ship target detection algorithm based on improved faster r-cnn, *Electronics*, 8, 9, 959, 2019.

[184] S.-J. Lee, M.-I. Roh, H.-W. Lee, J.-S. Ha i I.-G. Woo, Image-based ship detection and classification for unmanned surface vehicle using real-time object detection neural networks, *ISOPE International Ocean and Polar Engineering Conference*, ISOPE–I, ISOPE, 2018.

[185] Y. Li, J. Guo, X. Guo, K. Liu, W. Zhao, Y. Luo i Z. Wang, A novel target detection method of the unmanned surface vehicle under all-weather conditions with an improved yolov3, *Sensors*, 20, 17, 4885, 2020.

[186] G. Huang, Z. Liu, L. Van Der Maaten i K. Q. Weinberger, Densely connected convolutional networks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708, 2017.

[187] J. Zhang, J. Jin, Y. Ma i P. Ren, Lightweight object detection algorithm based on yolov5 for unmanned surface vehicles, *Frontiers in marine science*, 9, 1058401, 2023.

[188] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu i C. Xu, Ghostnet: More features from cheap operations, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1580–1589, 2020.

[189] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye i D. Ren, Distance-iou loss: Faster and better learning for bounding box regression, *Proceedings of the AAAI conference on artificial intelligence*, 34, 12993–13000, 2020.

[190] M. Cafaro, I. Epicoco, M. Pulimeno i E. Sansebastiano, Toward enhanced support for ship sailing, *IEEE Access*, 11, 87047–87061, 2023.

[191] W. Wu, X. Li, Z. Hu i X. Liu, Ship detection and recognition based on improved yolov7, *Comput. Mater. Contin*, 76, 1, 489–498, 2023.

[192] Z. Jiang, L. Su i Y. Sun, Yolov7-ship: A lightweight algorithm for ship object detection in complex marine environments, *Journal of Marine Science and Engineering*, 12, 1, 190, 2024.

[193] X. Jiang, J. Cai i B. Wang, Yoloseaship: a lightweight model for real-time ship detection, *European Journal of Remote Sensing*, 57, 1, 2307613, 2024.

[194] Y. Jin, P. Wang, S. Liu, K. Xue, Q. Li i H. Wang, Yolo-gcv: a lightweight algorithm for ship object detection in complex inland waterway environments, *Journal of Real-Time Image Processing*, 22, 1, 32, 2025.

[195] X. Zhao i Y. Song, Improved ship detection with yolov8 enhanced with mobilevit and gsconv, *Electronics*, 12, 22, 4666, 2023.

[196] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov i L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510–4520, 2018.

[197] H. Li, J. Li, H. Wei, Z. Liu, Z. Zhan i Q. Ren, Slim-neck by gsconv: A better design paradigm of detector architectures for autonomous vehicles, *arXiv preprint arXiv:2206.02424*, 10, 2022.

[198] K. Cheng, M. Song, Y. Xu, L. Zhuo, X. Yan, S. Zheng, Z. Yang, J. Huang, H. Song i S. Gao, Yolo-fe: A lightweight ship detection algorithm based on improved yolov8, *2023 5th International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI)*, 1082–1087, IEEE, 2023.

[199] J. Chen, S.-h. Kao, H. He, W. Zhuo, S. Wen, C.-H. Lee i S.-H. G. Chan, Run, don't walk: chasing higher flops for faster neural networks, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12021–12031, 2023.

[200] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan i Z. Huang, Efficient multi-scale attention module with cross-spatial learning, *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5, IEEE, 2023.

[201] S. Wang, Y. Li i S. Qiao, Alf-yolo: Enhanced yolov8 based on multiscale attention feature fusion for ship detection, *Ocean Engineering*, 308, 118233, 2024.

[202] G. Yang, J. Lei, Z. Zhu, S. Cheng, Z. Feng i R. Liang, Afpn: Asymptotic feature pyramid network for object detection, *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2184–2189, IEEE, 2023.

[203] D. Yang, M. I. Solihin, I. Ardiyanto, Y. Zhao, W. Li, B. Cai i C. Chen, A streamlined approach for intelligent ship object detection using el-yolo algorithm, *Scientific Reports*, 14, 1, 15254, 2024.

[204] Y. Yuan, X. Yu, X. Rong i X. Wang, Aff-lightnet: A lightweight ship detection architecture based on attentional feature fusion, *Journal of Marine Science and Engineering*, 13, 1, 44, 2025.

[205] Z. Gevorgyan, Siou loss: More powerful learning for bounding box regression, *arXiv preprint arXiv:2205.12740*, 2022.

[206] M. Liu i C. Zhu, Residual yolox-based ship object detection method, *2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, 427–431, IEEE, 2022.

[207] B. Iancu, J. Winsten, V. Soloviev i J. Lilius, A benchmark for maritime object detection with centernet on an improved dataset, aboships-plus, *Journal of Marine Science and Engineering*, 11, 9, 1638, 2023.

[208] A. Li, X. Zhu, S. He i J. Xia, Water surface object detection using panoramic vision based on improved single-shot multibox detector, *EURASIP Journal on Advances in Signal Processing*, 2021, 1–15, 2021.

[209] K. Simonyan i A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*, 2014.

[210] S. Moosbauer, D. Konig, J. Jakel i M. Teutsch, A benchmark for deep learning based object detection in maritime environments, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 0–0, 2019.

[211] N. Wang, Y. Wang, Y. Feng i Y. Wei, Aodemar: Attention-aware occlusion detection of vessels for maritime autonomous surface ships, *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[212] C. Zhao, R. W. Liu, J. Qu i R. Gao, Deep learning-based object detection in maritime unmanned aerial vehicle imagery: Review and experimental comparisons, *Engineering Applications of Artificial Intelligence*, 128, 107513, 2024.

[213] A. Correira, F. Ferreira i N. Mišković, Comparing different yolo versions for boat detection and classification in real datasets, *OCEANS 2024-Singapore*, 1–4, IEEE, 2024.

[214] D. Heller, M. Rizk, R. Douguet, A. Baghdadi i J.-P. Diguet, Marine objects detection using deep learning on embedded edge devices, *2022 IEEE International Workshop on Rapid System Prototyping (RSP)*, 1–7, 2022.

[215] M. Kristan, V. S. Kenk, S. Kovačič i J. Perš, Fast image-based obstacle detection from unmanned surface vehicles, *IEEE transactions on cybernetics*, 46, 3, 641–654, 2015.

[216] M. Kristan, J. Perš, V. Sulič i S. Kovačič, A graphical model for rapid obstacle image-map estimation from unmanned surface vehicles, *Computer Vision–ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part II 12*, 391–406, Springer, 2015.

[217] Y. Jie, L. Leonidas, F. Mumtaz i M. Ali, Ship detection and tracking in inland waterways using improved yolov3 and deep sort, *Symmetry*, 13, 2, 308, 2021.

[218] Z. Zhou, J. Zhao, X. Chen i Y. Chen, A ship tracking and speed extraction framework in hazy weather based on deep learning, *Journal of Marine Science and Engineering*, 11, 7, 1353, 2023.

[219] Y. Li, H. Yuan, Y. Wang i B. Zhang, Maritime vessel detection and tracking under uav vision, *2022 International Conference on Artificial Intelligence and Computer Information Technology (AICIT)*, 1–4, IEEE, 2022.

[220] J. Liu i C. Li, Maritime video ship detection and tracking based on improved yolox and deepsort, *Journal of Electronic Imaging*, 32, 1, 013042–013042, 2023.

[221] B. Qi, P. Zhang i W. Huang, Research on improved yolo and deepsort ship detection and tracking algorithms, *2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI)*, 816–819, IEEE, 2024.

[222] X. Chen, M. Wang, J. Ling, H. Wu, B. Wu i C. Li, Ship imaging trajectory extraction via an aggregated you only look once (yolo) model, *Engineering Applications of Artificial Intelligence*, 130, 107742, 2024.

[223] G. Zhang, J. Liu, Y. Zhao, W. Luo, K. Mei, P. Wang, Y. Song i X. Li, A reliable unmanned aerial vehicle multi-ship tracking method, *PloS one*, 20, 1, e0316933, 2025.

[224] C. Wu, Y. Song i Q. Ge, Maritime ship tracking algorithm based on improved yolov8 and strongsort, *2024 IEEE International Conference on Unmanned Systems (ICUS)*, 1043–1048, IEEE, 2024.

[225] Y.-S. Han i J.-Y. Jung, A stable multi-object tracking method for unstable and irregular maritime environments., *Journal of Marine Science & Engineering*, 12, 12, 2024.

[226] F. Yang, S. Odashima, S. Masui i S. Jiang, Hard to track objects with irregular motions and similar appearances? make it easier by buffering the matching space, *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 4799–4808, 2023.

[227] K. Hao, Z. Deng, B. Wang, Z. Jin, Z. Li i X. Zhao, Lightweight multiobject ship tracking algorithm based on trajectory association and improved yolov7tiny, *Expert Systems with Applications*, 259, 125129, 2025.

[228] K. Zhou, Y. Yang, A. Cavallaro i T. Xiang, Omni-scale feature learning for person re-identification, *Proceedings of the IEEE/CVF international conference on computer vision*, 3702–3712, 2019.

[229] B. Xing, W. Wang, J. Qian, C. Pan i Q. Le, A lightweight model for real-time monitoring of ships, *Electronics*, 12, 18, 3804, 2023.

[230] Y. Chen, Z. Chen, Z. Zhang i S. Bian, Adaptrack: An adaptive fairmot tracking method applicable to marine ship targets, *AI Communications*, 36, 2, 127–145, 2023.

[231] S. Zou, J. Liu i W. Dai, An anti-occlusion ship tracking method based on yolox and improved matching algorithm, *2025 2nd International Conference on Electronic Engineering and Information Systems (EEISS)*, 1–5, IEEE, 2025.

[232] S. Zou, J. Liu, X. Zhang, Z. Wu, J. Liu i B. Han, Joint feature representation optimization and anti-occlusion for robust multi-vessel tracking in inland waterways, *Complex & Intelligent Systems*, 11, 7, 294, 2025.

[233] J. Wu, C. Cao, Y. Zhou, X. Zeng, Z. Feng, Q. Wu i Z. Huang, Multiple ship tracking in remote sensing images using deep learning, *Remote Sensing*, 13, 18, 3601, 2021.

[234] H. Park, S.-H. Ham, T. Kim i D. An, Object recognition and tracking in moving videos for maritime autonomous surface ships, *Journal of Marine Science and Engineering*, 10, 7, 841, 2022.

[235] G. Wang, Z. Wen, T. Wu, Y. Qian i Q. Pan, Knowledge-assisted intelligent maritime multi-ship tracking, *2023 China Automation Congress (CAC)*, 2971–2976, IEEE, 2023.

[236] J. Cen, J.-H. Chen, X. Liu, J.-X. Li, H.-S. Li, W.-S. Huang i J.-X. Kang, A lightweight multi-target ship tracking model based on yolov7, *Physica Scripta*, 99, 3, 035010, 2024.

[237] Y. Shan, X. Zhou, S. Liu, Y. Zhang i K. Huang, Siamfpn: A deep learning method for accurate and real-time maritime ship tracking, *IEEE Transactions on Circuits and Systems for Video Technology*, 31, 1, 315–325, 2020.

[238] W. Luo, Y. Xia i T. He, Video-based identification and prediction techniques for stable vessel trajectories in bridge areas, *Sensors*, 24, 2, 372, 2024.

[239] X. Yang, H. Zhu, H. Zhao i D. Yang, Coastal ship tracking with memory-guided perceptual network, *Remote Sensing*, 15, 12, 3150, 2023.

[240] N. Karaev, I. Rocco, B. Graham, N. Neverova, A. Vedaldi i C. Rupprecht, Cotracker: It is better to track together, *arXiv preprint arXiv:2307.07635*, 2023.

[241] D. Stadler i J. Beyerer, Improving multiple pedestrian tracking by track management and occlusion handling, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10958–10967, 2021.

[242] S. Han, H. Wang, E. Yu i Z. Hu, Ort: Occlusion-robust for multi-object tracking, *Fundamental Research*, 2023.

[243] Z. Sun, J. Chen, M. Mukherjee, W. Ruan, C. Liang, Y. Yu i D. Zhang, Long-short term prediction for occluded multiple object tracking, *2021 IEEE Global Communications Conference (GLOBECOM)*, 1–6, IEEE, 2021.

[244] X. Tan, Z. Li, Q. Liang, W. Sun, Y. Wang i D. Zhang, Sequence-tracker: Multiple object tracking with sequence features in severe occlusion scene, *Journal of Visual Communication and Image Representation*, 79, 103250, 2021.

[245] X. Gao i T. Jiang, Osmo: Online specific models for occlusion in multiple object tracking under surveillance scene, *Proceedings of the 26th ACM international conference on Multimedia*, 201–210, 2018.

[246] R. Zhao, X. Zhang i J. Zhang, Psmot: Online occlusion-aware multi-object tracking exploiting position sensitivity, *Sensors*, 24, 4, 1199, 2024.

[247] K. He, X. Zhang, S. Ren i J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, *IEEE transactions on pattern analysis and machine intelligence*, 37, 9, 1904–1916, 2015.

[248] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin i B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022, 2021.

[249] X. Chen, X. Xu, Y. Yang, H. Wu, J. Tang i J. Zhao, Augmented ship tracking under occlusion conditions from maritime surveillance videos, *IEEE Access*, 8, 42884–42897, 2020.

[250] S. Chen, M. Gao, P. Shi, X. Zeng i A. Zhang, Target ship recognition and tracking with data fusion based on bi-yolo and oc-sort algorithms for enhancing ship navigation assistance, *Journal of Marine Science and Engineering*, 13, 2, 366, 2025.

[251] M. Zhang i Z. Zhang, Research on vehicle target detection method based on improved yolov8, *Applied Sciences*, 15, 10, 5546, 2025.

[252] H. Perreault, G.-A. Bilodeau, N. Saunier i M. Héritier, Ffavod: Feature fusion architecture for video object detection, *Pattern Recognition Letters*, 151, 294–301, 2021.

[253] Z. Zhang, C. Lan, W. Zeng i Z. Chen, Multi-granularity reference-aided attentive feature aggregation for video-based person re-identification, *Proceedings of the IEEE/-CVF conference on computer vision and pattern recognition*, 10407–10416, 2020.

[254] L. Zhu, T. Chen, D. Ji, J. Ye i J. Liu, Not every patch is needed: Towards a more efficient and effective backbone for video-based person re-identification, *IEEE Transactions on Image Processing*, 2025.

[255] D. Komorčec i D. Matika, Small crafts role in maritime traffic and detection by technology integration, *Pomorstvo*, 30, 1, 3–11, 2016.

[256] M. Petković, I. Vujović, N. Kaštelan i J. Šoda, Every vessel counts: Neural network based maritime traffic counting system, *Sensors*, 23, 15, 6777, 2023.

[257] D. Technology, Dahua tpc-pt8620a-tb, 2023, https://material.dahuasecurity.com/product/20180320/DH-TPC-PT8620A-T_Datasheet_20180320.pdf, [Zadnji pristup: 31. kolovoza 2025.].

[258] Lučka uprava Split, Bazen – gradska luka, https://portsplit.hr/luka-split/lucka-podrucja/bazen-gradska-luka/, [Zadnji pristup: 4. rujna 2025.].

[259] CVAT.ai, Cvat: Computer vision annotation tool, 2024, open-source software for computer vision data annotation, https://www.cvat.ai [Zadnji pristup: 6. rujna 2025.].

[260] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang i Q. Tian, Scalable person re-identification: A benchmark, *Proceedings of the IEEE international conference on computer vision*, 1116–1124, 2015.

[261] Z. Zheng, L. Zheng i Y. Yang, Unlabeled samples generated by gan improve the person re-identification baseline in vitro, *Proceedings of the IEEE international conference on computer vision*, 3754–3762, 2017.

[262] L. Wei, S. Zhang, W. Gao i Q. Tian, Person transfer gan to bridge domain gap for person re-identification, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 79–88, 2018.

[263] W. Li, R. Zhao i X. Wang, Human reidentification with transferred metric learning, *Asian conference on computer vision*, 31–44, Springer, 2012.

[264] W. Li, R. Zhao, T. Xiao i X. Wang, Deepreid: Deep filter pairing neural network for person re-identification, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 152–159, 2014.

[265] C. Guo, Y. Lu i T. Ikenaga, Multiple likelihood models based particle filter for long-term full occlusion, *The Journal of the Institute of Image Electronics Engineers of Japan*, 39, 5, 580–589, 2010.

[266] N. Dalal i B. Triggs, Histograms of oriented gradients for human detection, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 886–893 vol. 1, 2005.

[267] D. G. Lowe, Object recognition from local scale-invariant features, *Proceedings of the seventh IEEE international conference on computer vision*, 2, 1150–1157, Ieee, 1999.

[268] H. Bay, T. Tuytelaars i L. Van Gool, Surf: Speeded up robust features, *Lecture notes in computer science*, 3951, 404–417, 2006.

[269] P. Viola i M. Jones, Rapid object detection using a boosted cascade of simple features, *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 1, I–I, 2001.

[270] W. Chen, J. Luo, F. Zhang i Z. Tian, A review of object detection: Datasets, performance evaluation, architecture, applications and current trends, *Multimedia Tools and Applications*, 83, 24, 65603–65661, 2024.

[271] X. Wu, D. Sahoo i S. C. Hoi, Recent advances in deep learning for object detection, *Neurocomputing*, 396, 39–64, 2020.

[272] J. Redmon i A. Farhadi, Yolo9000: better, faster, stronger, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7263–7271, 2017.

[273] J. Redmon i A. Farhadi, Yolov3: An incremental improvement, *arXiv preprint arXiv:1804.02767*, 2018.

[274] A. Bochkovskiy, C.-Y. Wang i H.-Y. M. Liao, Yolov4: Optimal speed and accuracy of object detection, *arXiv preprint arXiv:2004.10934*, 2020.

[275] G. Jocher, Ultralytics yolov5, 2020, https://github.com/ultralytics/yolov5, [Zadnji pristup: 10. rujna 2025.].

[276] G. Jocher, A. Chaurasia i J. Qiu, Ultralytics yolov8, 2023, https://github.com/ultralytics/ultralytics, [Zadnji pristup: 10. rujna 2025.].

[277] C.-Y. Wang, I.-H. Yeh i H.-Y. M. Liao, Yolov9: Learning what you want to learn using programmable gradient information, *arXiv preprint arXiv:2402.13616*, 2024.

[278] X. Long, K. Deng, G. Wang, Y. Zhang, Q. Dang, Y. Gao, H. Shen, J. Ren, S. Han, E. Ding et al., Pp-yolo: An effective and efficient implementation of object detector, *arXiv preprint arXiv:2007.12099*, 2020.

[279] C.-Y. Wang, I.-H. Yeh i H.-Y. M. Liao, You only learn one representation: Unified network for multiple tasks, *arXiv preprint arXiv:2105.04206*, 2021.

[280] Z. Ge, S. Liu, F. Wang, Z. Li i J. Sun, Yolox: Exceeding yolo series in 2021, *arXiv preprint arXiv:2107.08430*, 2021.

[281] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke i A. Rabinovich, Going deeper with convolutions, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9, 2015.

[282] S. Lloyd, Least squares quantization in pcm, *IEEE transactions on information theory*, 28, 2, 129–137, 1982.

[283] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan i S. Belongie, Feature pyramid networks for object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125, 2017.

[284] Y. Tian, Q. Ye i D. Doermann, Yolov12: Attention-centric real-time object detectors, *arXiv preprint arXiv:2502.12524*, 2025.

[285] S. Aharon, Louis-Dupont, Ofri Masad, K. Yurkova, Lotem Fridman, Lkdci, E. Khvedchenya, R. Rubin, N. Bagrov, B. Tymchenko, T. Keren, A. Zhilko i Eran-Deci, Supergradients, 2021, https://github.com/Deci-AI/super-gradients, [Zadnji pristup: 10. rujna 2025.].

[286] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai i S. Chintala, Pytorch: An imperative style, high-performance deep learning library, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox i R. Garnett, editori, *Advances in Neural Information Processing Systems 32*, 8024–8035, Curran Associates, Inc., 2019.

[287] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar i B. Lee, A survey of modern deep learning based object detection models, *Digital Signal Processing*, 126, 103514, 2022.

[288] G. Jocher i J. Qiu, Ultralytics yolo11, 2024.

[289] N. Jegham, C. Y. Koh, M. Abdelatti i A. Hendawi, Yolo evolution: A comprehensive benchmark and architectural review of yolov12, yolo11, and their previous versions, *arXiv preprint arXiv:2411.00201*, 2024.

[290] Y. Qu, C. Wang, Y. Xiao, H. Ju i J. Wu, Dynamically optimized object detection algorithms for aviation safety, *Electronics*, 14, 17, 3536, 2025.

[291] R. Khanam i M. Hussain, Yolov11: An overview of the key architectural enhancements, *arXiv preprint arXiv:2410.17725*, 2024.

[292] Ultralytics, Yolov11 documentation, https://docs.ultralytics.com/models/yolo11/, 2024, [Zadnji pristup: 10. rujna 2025.].

[293] R. Sapkota, Z. Meng, M. Churuvija, X. Du, Z. Ma i M. Karkee, Comprehensive performance evaluation of yolov12, yolo11, yolov10, yolov9 and yolov8 on detecting and counting fruitlet in complex orchard environments, *arXiv preprint arXiv:2407.12040*, 2024.

[294] N. Simic i A. Gavrovska, Comparative analysis of yolov11 and yolov12 for ai-powered aerial people detection, *2025 12th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN)*, 1–4, IEEE, 2025.

[295] Sveučilišni računski centar Srce, Napredno računanje – usluga napredno računanje, https://www.srce.unizg.hr/napredno-racunanje, [Zadnji pristup: 11. rujna 2025.].

[296] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han i G. Ding, Yolov10: Real-time end-to-end object detection, 2024.

[297] I. Marin, A. Kuzmanic Skelin i T. Grujic, Empirical evaluation of the effect of optimization and regularization techniques on the generalization performance of deep convolutional neural network, *Applied Sciences*, 10, 21, 7817, 2020.

[298] C. Shorten i T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, *Journal of big data*, 6, 1, 1–48, 2019.

[299] E. D. Cubuk, B. Zoph, J. Shlens i Q. V. Le, Randaugment: Practical automated data augmentation with a reduced search space, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 702–703, 2020.

[300] M. Raghu, C. Zhang, J. Kleinberg i S. Bengio, Transfusion: Understanding transfer learning for medical imaging, *Advances in neural information processing systems*, 32, 2019.

[301] J.-J. Wang, Y.-F. Liu, X. Nie i Y. Mo, Deep convolutional neural networks for semantic segmentation of cracks, *Structural Control and Health Monitoring*, 29, 1, e2850, 2022.

[302] M. F. Hashmi, M. Devipriyanka, B. Raghavaram i A. A. Husain, Advanced semantic lung segmentation with a hybrid segnet-resnet50 network, *Journal of Electrical Systems and Information Technology*, 12, 1, 64, 2025.

[303] S. Alam, N. K. Tomar, A. Thakur, D. Jha i A. Rauniyar, Automatic polyp segmentation using u-net-resnet50, *arXiv preprint arXiv:2012.15247*, 2020.

[304] A. Hermans, L. Beyer i B. Leibe, In defense of the triplet loss for person re-identification, *arXiv preprint arXiv:1703.07737*, 2017.

[305] Y. Sun, L. Zheng, Y. Yang, Q. Tian i S. Wang, Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline), *Proceedings of the European conference on computer vision (ECCV)*, 480–496, 2018.

[306] X. Liu, W. Liu, H. Ma i H. Fu, Large-scale vehicle re-identification in urban surveillance videos, *2016 IEEE international conference on multimedia and expo (ICME)*, 1–6, IEEE, 2016.

[307] M. Lin, Q. Chen i S. Yan, Network in network, *arXiv preprint arXiv:1312.4400*, 2013.

[308] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto i H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, *arXiv preprint arXiv:1704.04861*, 2017.

[309] K. Zhou i T. Xiang, Torchreid: A library for deep learning person re-identification in pytorch, *arXiv preprint arXiv:1910.10093*, 2019.

[310] Z. Yang, D. Wu, C. Wu, Z. Lin, J. Gu i W. Wang, A pedestrian is worth one prompt: Towards language guidance person re-identification, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17343–17353, 2024.

[311] W. He, Y. Deng, S. Tang, Q. Chen, Q. Xie, Y. Wang, L. Bai, F. Zhu, R. Zhao, W. Ouyang et al., Instruct-reid: A multi-purpose person re-identification task with instructions, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17521–17531, 2024.

[312] D. P. Kingma, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*, 2014.

[313] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao i S. C. Hoi, Deep learning for person re-identification: A survey and outlook, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44, 6, 2872–2893, 2022.

[314] X. Zhang, Y. Wang i L. Zhao, Global-local attention with triplet loss and label smoothed cross-entropy for person re-identification, *Multimedia Tools and Applications*, 82, 36743–36761, 2023.

[315] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens i Z. Wojna, Rethinking the inception architecture for computer vision, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826, 2016.

[316] X. Bai, M. Yang, T. Huang, Z. Dou, R. Yu i Y. Xu, Deep-person: Learning discriminative deep features for person re-identification, *Pattern Recognition*, 98, 107036, 2020.

[317] G. Wang, Y. Yuan, X. Chen, J. Li i X. Zhou, Learning discriminative features with multiple granularities for person re-identification, *Proceedings of the 26th ACM international conference on Multimedia*, 274–282, 2018.

[318] L. Zhang, Y. Xu, L. Zhao i F. Qin, Resolution independent person re-identification network, *IET Computer Vision*, 19, 1, e12140, 2025.

[319] C. Lyu, T. Xu, K. Wang i J. Chen, Person re-identification based on human semantic parsing and message passing., *Journal of Supercomputing*, 79, 5, 2023.

[320] Y. Bar-Shalom, X. R. Li i T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*, John Wiley & Sons, 2001.

[321] M. Broström i contributors, Boxmot: Pluggable sota multi-object tracking modules for segmentation, object detection and pose estimation models, `https://github.com/mikel-brostrom/boxmot`, open-source software [Zadnji pristup: 10. prosinca 2025.].

[322] A. H. Jonathon Luiten, Trackeval, `https://github.com/JonathonLuiten/TrackEval`, 2020. [Zadnji pristup: 10. prosinca 2025.].

[323] L. Wei, S. Zhang, W. Gao i Q. Tian, Person transfer gan to bridge domain gap for person re-identification, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 79–88, 2018.

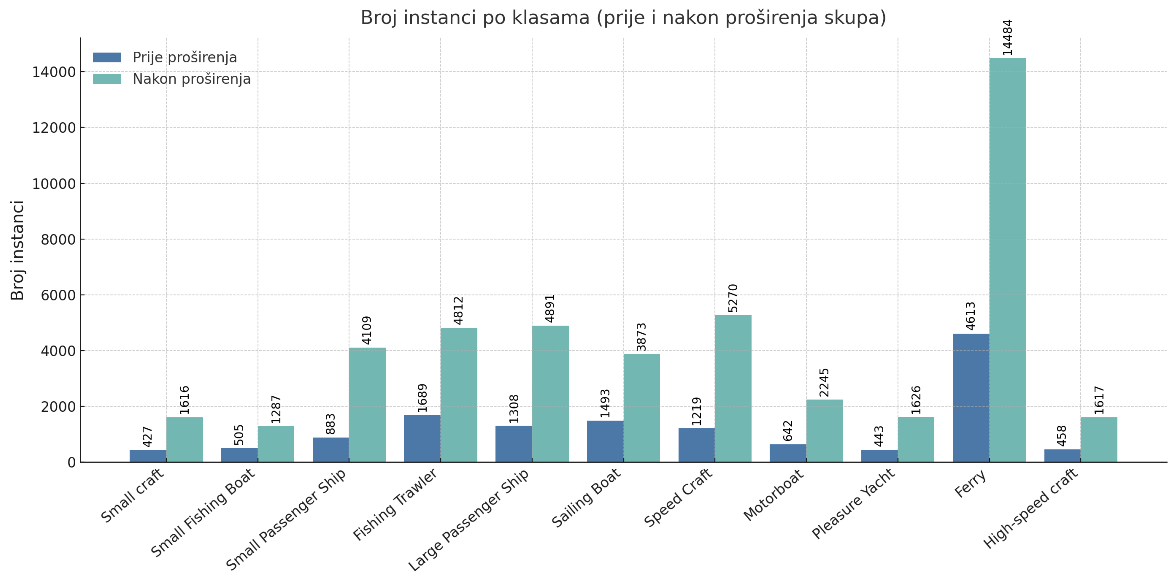
A. Struktura SSMOT detekcijskog podskupa prije i nakon proširenja

U ovom dodatku detaljno je prikazana struktura SSMOT detekcijskog podskupa prije i nakon proširenja.

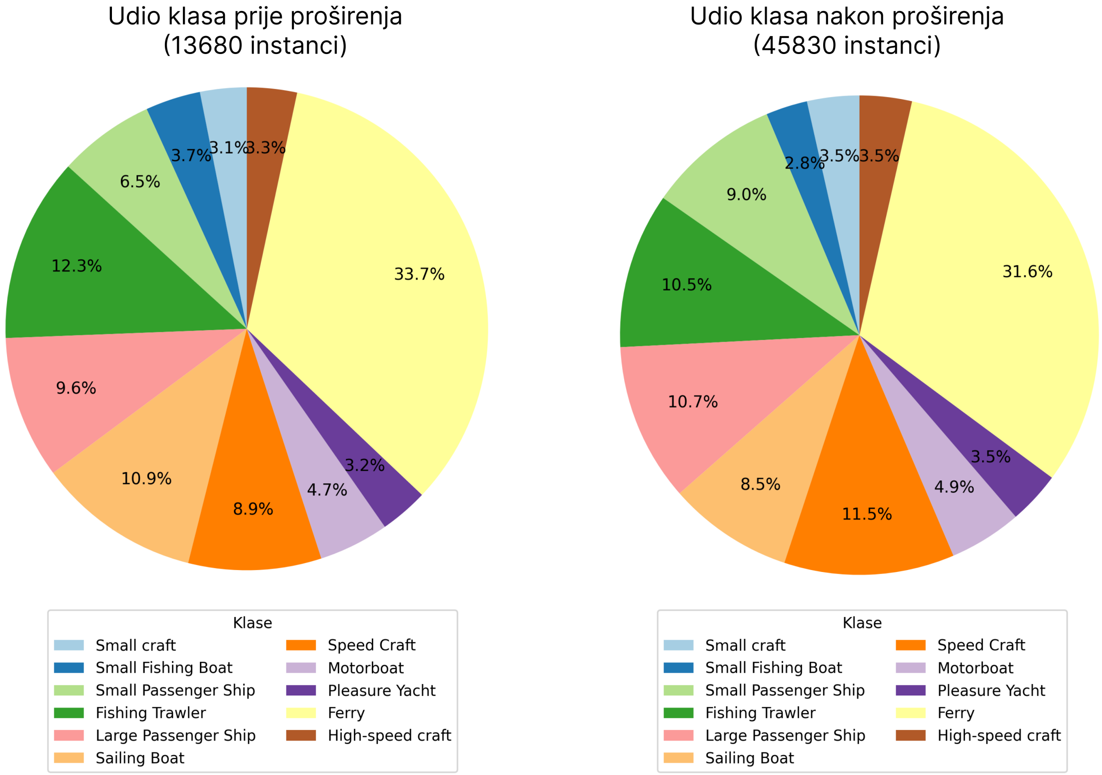
Slika A.1 daje pregled broja instanci objekata po klasama te jasno pokazuje razliku u obimu podataka prije i nakon proširenja. Dodatno, Slika B.1 daje usporedni prikaz udjela instanci pojedine klase u detekcijskom skupu podataka prije i nakon proširenja. S druge strane, Tablica A.1 pruža detaljan prikaz broja slika na kojima se pojavljuju plovila pojedinih klasa te broja instanci po klasama, razdvojeno prema skupovima za treniranje i validaciju, kao i sveukupno. Na taj način omogućena je usporedba distribucije objekata te uvid u značajno povećanje veličine detekcijskog podskupa nakon provedenog postupka proširenja.

Napomena. *U retku "Ukupno" Tablice A.1 u stupcima naziva "Slike“ naveden je ukupan broj slika u pojedinom skupu podataka. Ova vrijednost ne mora odgovarati zbroju po svim klasama, budući da se na istoj slici može nalaziti više različitih objekata.*

Analizom nije obuhvaćen nezavisni testni skup koji se vodi odvojeno, budući da je fiksan i nije bio predmet proširenja.



Slika A.1: Broj instanci objekata pojedine klase prije i nakon proširenja podskupa za detekciju.



Slika A.2: Udio instanci objekata pojedine klase prije i nakon proširenja podskupa za detekciju.

Tablica A.1: Pregled broja slika i instanci po klasama u SSMOT detekcijskom podskupu prije i nakon proširenja.

		Train			**Val**			**Sveukupno**		
	Klasa	Slike	Inst.	Inst. (%)	Slike	Inst.	Inst. (%)	Slike	Inst.	Inst. (%)
Prije proširenja	*Small craft*	169	343	2.97	39	84	3.92	208	427	3.12
	Small Fishing Boat	410	428	3.71	71	77	3.60	481	505	3.69
	Small Passenger Ship	635	754	6.53	110	129	6.03	745	883	6.45
	Fishing Trawler	1109	1414	12.25	209	275	12.84	1318	1689	12.35
	Large Passenger Ship	886	1113	9.65	154	195	9.11	1040	1308	9.56
	Sailing Boat	669	1270	11.01	124	223	10.42	793	1493	10.91
	Speed Craft	750	1027	8.90	140	192	8.97	890	1219	8.91
	Motorboat	506	532	4.61	102	110	5.14	608	642	4.69
	Pleasure Yacht	345	382	3.31	57	61	2.85	402	443	3.24
	Ferry	2321	3901	33.81	416	712	33.26	2737	4613	33.72
	High-speed craft	357	375	3.25	77	83	3.88	434	458	3.35
	Ukupno	**2661**	**11539**	**100.00**	**470**	**2141**	**100.00**	**3131**	**13680**	**100.00**
Nakon proširenja	*Small craft*	517	1417	3.63	90	199	2.92	607	1616	3.53
	Small Fishing Boat	1023	1101	2.82	168	186	2.73	1191	1287	2.81
	Small Passenger Ship	2755	3496	8.96	483	613	9.00	3238	4109	8.97
	Fishing Trawler	3751	4098	10.50	652	714	10.48	4403	4812	10.50
	Large Passenger Ship	3360	4150	10.64	604	741	10.88	3964	4891	10.67
	Sailing Boat	2139	3261	8.36	390	612	8.98	2529	3873	8.45
	Speed Craft	3203	4501	11.54	540	769	11.29	3743	5270	11.50
	Motorboat	1780	1902	4.87	320	343	5.03	2100	2245	4.90
	Pleasure Yacht	1282	1383	3.54	225	243	3.57	1507	1626	3.55
	Ferry	7088	12351	31.66	1242	2133	31.31	8330	14484	31.60
	High-speed craft	1274	1357	3.48	233	260	3.82	1507	1617	3.53
	Ukupno	**7632**	**39017**	**100.00**	**1347**	**6813**	**100.00**	**8979**	**45830**	**100.00**

B. Detalji o SSMOT videozapisima za praćenje

Ovaj dodatak nadopuna je Potpoglavlju 4.4, u kojem je predstavljen podskup SSMOT skupa podataka namijenjen evaluaciji algoritama praćenja, koji sadrži ukupno 18 videozapisa. Na Slici B.1 prikazani su primjeri kadrova iz tih videozapisa, raspoređeni od *SSMOT_1* u gornjem lijevom kutu do *SSMOT_18* u donjem desnom.

Slika B.1: Primjer kadrova iz SSMOT videozapisa.

U Tablici B.1 prikazan je točan broj putanja i odgovarajućih graničnih okvira za svaku klasu plovila u pojedinim videozapisima SSMOT skup podataka. Ovakav pregled omogućuje bolje razumijevanje strukture samih videozapisa, odnosno raspodjele objekata po klasama i videozapisima, čime se može procijeniti raznolikost i složenost pojedinih scena.

Tablica B.2 daje detaljan pregled svih zabilježenih okluzija u videozapisima SSMOT skupa podataka. Za svaku okluziju naveden je identifikator zaklonjenog objekta, pripadajuća klasa plovila, početni i završni okvir te ukupno trajanje izraženo u okvirima i sekundama. Boje korištene u tablici označavaju kategoriju okluzije prema trajanju: kratke (zelene), srednje duge (žute) i duge (crvene) okluzije. Ove su informacije posebno su važne za procjenu otpornosti algoritama praćenja na okluzije različitog trajanja. Usporedba njihovih performansi na sekvencama s kratkim, srednje dugim i dugim okluzijama omogućuje analizu ograničenja i prednosti u realnim uvjetima praćenja. Tablica pritom služi kao referenca za kvantitativnu i kvalitativnu evaluaciju, jer precizno definira situacije potpunog prekida vidljivosti objekata.

Tablica B.1: Pregled broja putanja i graničnih okvira za različite tipove plovila u SSMOT videozapisima.

		Small Craft	***Small Fishing Boat***	***Small Passenger Ship***	***Fishing Trawler***	***Large Passenger Ship***	***Sailing Boat***	***Speed Craft***	***Motorboat***	***Pleasure Yacht***	***Ferry***	***High-speed craft***	
SSMOT_1	*putanje*						1	1			2	1	**5**
	gr. okviri						287	287			574	287	1435
SSMOT_2	*putanje*			1		1		4		1	2		9
	gr. okviri			545		604		1516		542	1208		4415
SSMOT_3	*putanje*				1	2		1	1	1	1		7
	gr. okviri				982	2074		882	1037	1037	1037		7049
SSMOT_4	*putanje*				1	1	1	2			2	1	8
	gr. okviri				370	370	370	630			740	370	2850
SSMOT_5	*putanje*			1	1	1		2	1		1		7
	gr. okviri			368	433	433		801	433		433		2901
SSMOT_6	*putanje*				1	1	1	4	2		2		11
	gr. okviri				856	856	856	2903	807		1712		7990
SSMOT_7	*putanje*		1		1			1			1		4
	gr. okviri		451		451			451			451		1804
SSMOT_8	*putanje*			1	1		3			3	2	1	11
	gr. okviri			345	1697		1691			2547	3394	820	10494

		Small Craft	Small Fishing Boat	Small Passenger Ship	Fishing Trawler	Large Passenger Ship	Sailing Boat	Speed Craft	Motorboat	Pleasure Yacht	Ferry	High-speed craft	
SSMOT_9	*putanje*			1	1		1	1			2	2	8
	gr. okviri			922	872		472	922			1607	1562	6357
SSMOT_10	*putanje*				1		2	4	3		2		12
	gr. okviri				836		1672	849	1542		1672		6571
SSMOT_11	*putanje*	7			1			3			1		12
	gr. okviri	1914			277			717			277		3185
SSMOT_12	*putanje*	6						1			2	1	10
	gr. okviri	3783						733			1856	558	6930
SSMOT_13	*putanje*			1		1			2		2	1	7
	gr. okviri			646		646			776		1277	302	3647
SSMOT_14	*putanje*				1	2	1	1			3	1	9
	gr. okviri				598	1196	598	288			1794	598	5072
SSMOT_15	*putanje*			1	1	2			1		1	1	7
	gr. okviri			585	585	1170			585	585	585		4095
SSMOT_16	*putanje*				1	2	3		1		2	1	10
	gr. okviri				940	1880	2331		940		1741	910	8742
SSMOT_17	*putanje*			2				1			1		4
	gr. okviri			1882				941			941		3764

		Small Craft	***Small Fishing Boat***	***Small Passenger Ship***	***Fishing Trawler***	***Large Passenger Ship***	***Sailing Boat***	***Speed Craft***	***Motorboat***	***Pleasure Yacht***	***Ferry***	***High-speed craft***	
SSMOT_18	*putanje*		1	1	1	1		2			2		8
	gr. okviri		332	538	300	538		921			1076		3705
Ukupno	***putanje***	**13**	**2**	**9**	**13**	**14**	**13**	**28**	**11**	**5**	**31**	**10**	**149**
	gr. okviri	**5697**	**783**	**5831**	**9197**	**9767**	**8277**	**12841**	**6120**	**4126**	**22375**	**5992**	**91006**

Tablica B.2: Detaljan pregled okluzija u SSMOT skupu videozapisa. Zelenom su označene kratke ($\leq 2\,s$), žutom srednje duge ($> 2\,s$ i $\leq 8\,s$), a crvenom duge ($> 8\,s$) okluzije plovila.

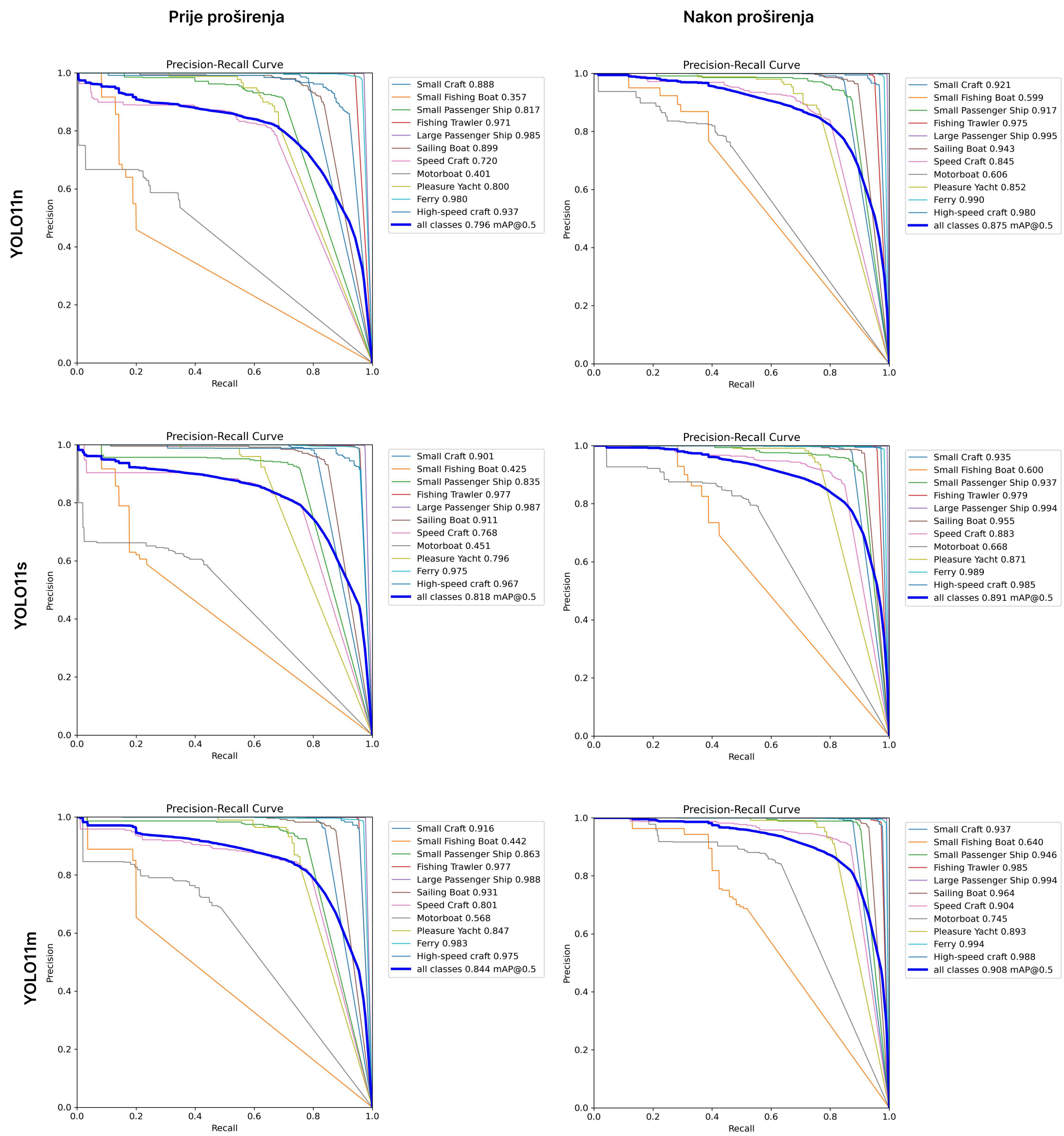
Video	**ID**	**Kategorija**	**Početni okvir**	**Završni okvir**	**Trajanje** (okviri)	**Trajanje** (s)
SSMOT_1	5	*Speed Craft*	103	199	97	3.87
SSMOT_3	4	*Pleasure Yacht*	421	740	320	12.79
	5	*Motorboat*	251	810	560	22.38
SSMOT_5	6	*Motorboat*	64	115	52	2.08
SSMOT_6	6	*Speed Craft*	535	615	81	3.24
SSMOT_7	4	*Small Fishing Boat*	177	210	34	1.36
SSMOT_8	2	*Fishing Trawler*	1395	1408	14	0.56
SSMOT_9	1	*High-speed Craft*	652	686	35	1.40
	3	*Small Passenger Ship*	181	455	275	10.99
	4	*Speed Craft*	213	507	295	11.79
	4	*Speed Craft*	911	920	10	0.40
	6	*Fishing Trawler*	574	610	37	1.48
	6	*Fishing Trawler*	647	660	14	0.56
SSMOT_12	4	*Small Craft*	553	583	31	1.24
	8	*Small Craft*	74	84	11	0.44
SSMOT_13	1	*Ferry*	369	535	167	16.67
	3	*Small Passenger Ship*	416	457	42	4.19
	4	*Motorboat*	75	147	73	7.29
	4	*Motorboat*	308	364	57	5.69
	4	*Motorboat*	478	618	141	14.08
SSMOT_14	7	*Large Passenger Ship*	36	307	272	27.15
	8	*Speed Craft*	80	181	102	10.18
	9	*Sailing Boat*	207	497	291	29.05
SSMOT_15	6	*Motorboat*	131	415	285	11.38
SSMOT_16	2	*Fishing Trawler*	137	161	25	1.00
	2	*Fishing Trawler*	766	775	10	0.40
	6	*Motorboat*	221	230	10	0.40
	6	*Motorboat*	545	717	173	6.91
SSMOT_17	2	*Small Passenger Ship*	441	478	38	1.27
	4	*Speed Craft*	351	555	205	6.83
SSMOT_18	8	*Speed Craft*	209	218	10	0.40

C. Detaljni rezultati evaluacije YOLO11n/m/s detektora prije i nakon proširenja detekcijskog podskupa

U ovom dodatku prikazani su detaljni rezultati evaluacije YOLO11n/m/s detektora na skupu za testiranje. Tablica C.1 sadrži pregled performansi modela po pojedinačnim klasama i ukupno, kako prije tako i nakon treniranja na proširenom SSMOT skupu podataka za detekciju. Uz to, data je i usporedna analiza u vidu preciznost–odziv krivulja na Slici C.1, čime se dodatno ilustrira utjecaj proširenja skupa podataka i korištene arhitekture na konačne performanse modela.

Tablica C.1: Detaljan pregled rezultata evaluacije YOLO11n/s/m detektora (po klasama i ukupno) prije i nakon treniranja na proširenom SSMOT skupu podataka za detekciju.

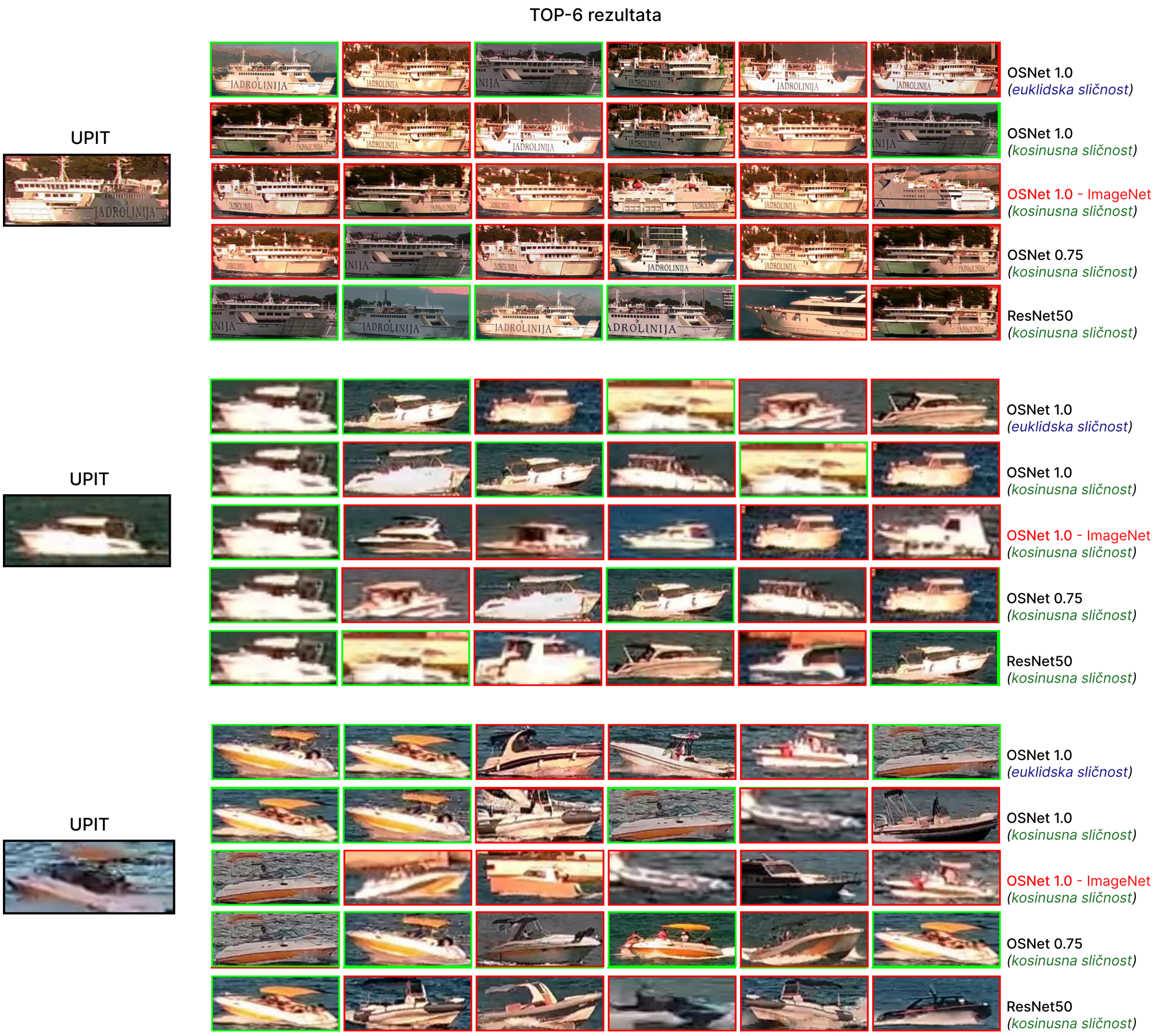
		YOLO11n				**YOLO11s**				**YOLO11m**			
	Klasa	P	R	mAP_{50}	$mAP_{50:95}$	P	R	mAP_{50}	$mAP_{50:95}$	P	R	mAP_{50}	$mAP_{50:95}$
Prije proširenja	*Small craft*	0.968	0.784	0.888	0.670	0.969	0.810	0.901	0.747	0.958	0.840	0.916	0.781
	Small Fishing Boat	0.459	0.200	0.357	0.297	0.588	0.235	0.425	0.372	0.654	0.200	0.442	0.416
	Small Passenger Ship	0.909	0.701	0.817	0.706	0.919	0.754	0.835	0.753	0.918	0.777	0.863	0.783
	Fishing Trawler	0.991	0.942	0.971	0.721	0.984	0.958	0.977	0.722	0.993	0.954	0.977	0.777
	Large Passenger Ship	0.990	0.972	0.985	0.935	0.997	0.975	0.987	0.950	0.997	0.997	0.988	0.960
	Sailing Boat	0.886	0.838	0.899	0.815	0.931	0.850	0.911	0.851	0.951	0.877	0.931	0.883
	Speed Craft	0.777	0.677	0.720	0.559	0.795	0.758	0.768	0.621	0.802	0.792	0.801	0.680
	Motorboat	0.537	0.348	0.401	0.306	0.569	0.452	0.451	0.369	0.689	0.486	0.568	0.493
	Pleasure Yacht	0.824	0.682	0.800	0.677	0.914	0.636	0.796	0.706	0.832	0.755	0.847	0.746
	Ferry	0.975	0.966	0.980	0.859	0.985	0.955	0.975	0.866	0.986	0.970	0.983	0.901
	High-speed craft	0.856	0.922	0.937	0.717	0.906	0.961	0.967	0.824	0.961	0.956	0.975	0.854
	Zajedno	0.834	0.730	**0.796**	**0.660**	0.869	0.759	**0.818**	**0.707**	0.885	0.780	**0.844**	**0.752**
Nakon proširenja	*Small craft*	0.974	0.848	0.921	0.747	0.979	0.874	0.935	0.768	0.983	0.877	0.937	0.801
	Small Fishing Boat	0.767	0.388	0.599	0.485	0.692	0.424	0.600	0.530	0.688	0.518	0.640	0.573
	Small Passenger Ship	0.910	0.874	0.917	0.798	0.907	0.910	0.937	0.830	0.961	0.904	0.946	0.860
	Fishing Trawler	0.982	0.951	0.975	0.816	0.983	0.959	0.979	0.845	0.986	0.973	0.985	0.877
	Large Passenger Ship	0.995	0.993	0.995	0.975	0.998	0.993	0.994	0.974	0.998	0.992	0.992	0.977
	Sailing Boat	0.958	0.895	0.943	0.866	0.974	0.916	0.955	0.903	0.969	0.931	0.964	0.917
	Speed Craft	0.835	0.801	0.845	0.680	0.861	0.850	0.883	0.740	0.901	0.869	0.904	0.786
	Motorboat	0.748	0.467	0.606	0.500	0.780	0.557	0.668	0.580	0.842	0.633	0.745	0.647
	Pleasure Yacht	0.878	0.762	0.852	0.743	0.928	0.768	0.871	0.798	0.904	0.815	0.893	0.829
	Ferry	0.985	0.983	0.990	0.928	0.984	0.983	0.989	0.932	0.983	0.990	0.994	0.949
	High-speed craft	0.956	0.967	0.980	0.866	0.978	0.972	0.985	0.888	0.951	0.978	0.988	0.907
	Zajedno	0.908	0.812	**0.875**	**0.764**	0.915	0.837	**0.891**	**0.799**	0.924	0.862	**0.908**	**0.829**



Slika C.1: Uporedan prikaz preciznost–odziv krivulja YOLO11n/m/s detektora treniranih na početnom i proširenom SSMOT detekcijskom skupu podataka.

D. Vizualizacije dodatnih TOP-6 rezultata ReID modela

U ovom dodatku prikazani su dodatni primjeri upita i pripadajućih TOP-6 rangiranih rezultata iz proširene galerije, dobivenih primjenom različitih ReID modela. Primjeri služe kao dopuna glavnim rezultatima u radu te omogućuju detaljniji uvid u ponašanje modela u zahtjevnijim scenarijima pretrage.



Slika D.1: Primjeri upita i pripadajućih TOP-6 rezultata iz proširene galerije, dobivenih različitim ReID modelima (1).

Slika D.2: Primjer upita i pripadajućih TOP-6 rezultata iz proširene galerije, dobivenih različitim ReID modelima (2).

Slika D.3: Primjer upita i pripadajućih TOP-6 rezultata iz proširene galerije, dobivenih različitim ReID modelima (3).

Životopis

Ivana Marin

Ivana Marin rođena je 31. srpnja 1994. godine u Splitu, Hrvatska. Nakon završetka opće gimnazije u Srednjoj školi Ivana Meštrovića u Drnišu upisuje prijediplomski studij matematike na Prirodoslovno-matematičkom fakultetu Sveučilišta u Splitu, koji završava 2017. godine. Iste godine na istom fakultetu upisuje diplomski studij matematike, računarski smjer, te dvije godine kasnije stječe akademski naziv magistre matematike. Diplomski studij završava s izvrsnim uspjehom te joj je za postignute rezultate dodijeljena dekanova nagrada. Nakon završetka studija 2019. godine zapošljava se kao asistentica na Odjelu za informatiku Prirodoslovno-matematičkog fakulteta u Splitu, gdje je zaposlena i danas. Tijekom dosadašnjeg rada sudjelovala je u izvođenju nastave i pripremi nastavnih materijala iz više kolegija, uključujući Uvod u umjetnu inteligenciju, Arhitekture neuronskih mreža, Strojno učenje te Uvod u podatkovnu znanost.

Poslijediplomski studij Elektrotehnike i informacijske tehnologije u polju računarstva upisuje 2019. godine na Fakultetu elektrotehnike, strojarstva i brodogradnje u Splitu. Autorica je i koautorica ukupno osam znanstvenih radova s međunarodnom recenzijom, od kojih su tri objavljena u zbornicima skupova, a pet u znanstvenim časopisima. Sudjelovala je u recenzentskom postupku znanstvenih radova za međunarodne časopise i konferencije, uključujući časopise Artificial Intelligence Review i Neural Processing Letters te međunarodnu konferenciju International Conference on Software, Telecommunications and Computer Networks. Njezini znanstveni interesi obuhvaćaju područja računalnog vida, umjetne inteligencije i dubokog učenja.

Supruga je i majka dvoje djece.

Curriculum Vitae

Ivana Marin

Ivana Marin was born on July 31, 1994, in Split, Croatia. After completing a general secondary education at Ivana Meštrović Secondary School in Drniš, she enrolled in the undergraduate program in Mathematics at the Faculty of Science, University of Split, which she completed in 2017. In the same year, she enrolled in the graduate program in Mathematics, computer science track, at the same faculty, and two years later received the academic degree of Master of Mathematics. She completed her graduate studies with excellent academic performance and was awarded the Dean's Award. After completing her studies, she was employed in 2019 as a Teaching and Research Assistant at the Department of Informatics, Faculty of Science, University of Split, where she is currently employed. During her academic engagement, she participated in teaching activities and in the preparation of teaching materials for several courses, including Introduction to Artificial Intelligence, Neural Network Architectures, Machine Learning, and Introduction to Data Science.

In 2019, she enrolled in the doctoral program in Electrical Engineering and Information Technologies in the field of Computer Science at the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split. She is the author and co-author of a total of eight internationally peer-reviewed scientific papers, three of which were published in conference proceedings and five in scientific journals. She has contributed as a reviewer for international journals and conferences, including Artificial Intelligence Review and Neural Processing Letters, as well as the International Conference on Software, Telecommunications and Computer Networks. Her research interests include computer vision, artificial intelligence, and deep learning.

She is married and the mother of two children.