SVEUČILIŠTE U SPLITU
FAKULTET ELEKTROTEHNIKE, STROJARSTVA I BRODOGRADNJE

DOKTORSKI STUDIJ
ELEKTROTEHNIKA I INFORMACIJSKA TEHNOLOGIJA

KVALIFIKACIJSKI DOKTORSKI ISPIT

Metode detekcije, praćenja i reidentifikacije plovila u složenim pomorskim scenama

Paula Muslim

Split, listopada 2026.

SADRŽAJ

1 UVOD 1
2 PREGLED ALGORITAMA ZA PRAĆENJE VIŠE OBJEKATA 2
2.1 Zajednička struktura algoritama praćenja 2
2.2 Kritički prikaz odabranih MOT algoritama 3
2.2.1 SORT - Simple Online and Realtime Tracking 3
2.2.2 DeepSORT - Simple Online and Realtime Tracking with a Deep Association Metric 4
2.2.3 FairMOT – On the Fairness of Detection and Re-Identification in Multiple Object Tracking 6
2.2.4 ByteTrack - Multi-Object Tracking by Associating Every Detection Box .. 7
2.2.5 StrongSORT i StrongSORT++ – Unaprjeđenje praćenja kombiniranjem vizualnih značajki, modela gibanja i naknadne obrade putanja 9
2.2.6 BoT-SORT - Robust Associations Multi-Pedestrian Tracking 12
2.2.7 OC-SORT - Rethinking SORT for Robust Multi-Object Tracking 14
2.2.8 Deep OC-SORT - Multi-Pedestrian Tracking by Adaptive Re-Identification 16
2.2.9 VOc-SORT – Vessel Occlusion SORT 18
2.2.10 OcclusionTrack - Multi-Object Tracking in Dense Scenes 21
2.2.11 OA-SORT – Occlusion-Aware SORT: Observing Occlusion for Robust Multi-Object Tracking 23
2.3 Usporedba prema komponentama i okluzijama 27
2.4 Evaluacijske metrike za praćenje više objekata i njihova ograničenja 28
2.4.1 MOTA – Multiple Object Tracking Accuracy 28
2.4.2 IDF1 – Identification F1 Score 29
2.4.3 HOTA – Higher Order Tracking Accuracy 30
2.4.4 Usporedba metrika i ograničenja u pomorskim scenama 31
2.5 Zaključak 32
3 NVIDIA DEEPSTREAM SDK: ARHITEKTURA, KOMPONENTE I KONFIGURACIJA SUSTAVA ZA INTELIGENTNU VIDEOANALITIKU 33
3.1 Uvod i položaj DeepStreama u videoanalitici 33
3.2 Programska arhitektura i GStreamer pipeline 34
3.2.1 Prihvat izvora i hardversko dekodiranje 35
3.2.2 Grupiranje kadrova elementom nvstreammux 35
3.3 Izvođenje modela dubokog učenja 36
3.3.1 Primarna inferencija i nvinfer 36

3.3.2 Sekundarna inferencija i nvinferserver 36
3.3.3 Pretprocesiranje i naknadna obrada 36
3.4 Praćenje objekata: Gst-nvtracker i biblioteka NvMultiObjectTracker 37
3.4.1 IoU tracker i NvSORT 37
3.4.2 NvDeepSORT 37
3.4.3 NvDCF 38
3.4.4 Upravljanje životnim ciklusom putanja 38
3.5 Reidentifikacija, metapodaci i analitički sloj 39
3.6 Konfiguriranje sustava i ključni parametri 40
3.6.1 Primjer konfiguracije glavnih elemenata 41
3.6.2 Primjer skupina postavki NvDCF-a 42
3.7 Performanse, latencija i višestruki videotokovi 42
3.8 Proširivost i razvoj vlastitih komponenti 43
3.9 Prednosti, ograničenja i prikladnost za detekciju i praćenje objekata 43
3.10 Zaključak 44
4 SKUPOVI PODATAKA ZA DETEKCIJU, REIDENTIFIKACIJU I PRAĆENJE PLOVILA 45
4.1 Uloga skupova podataka u istraživanju računalnog vida za pomorski nadzor 45
4.2 Opći skupovi podataka i njihova ograničenja 45
4.3 Pomorski skupovi podataka prikazani u dosadašnjim istraživanjima 46
4.3.1 Klasifikacijske baze 46
4.3.2 Skupovi za detekciju plovila 46
4.3.3 Video baze za praćenje 46
4.4 Split Ship MOT (SSMOT): struktura, nastanak i znanstvena vrijednost 48
4.4.1 Izvor snimaka i specifičnosti splitske luke 48
4.4.2 Kategorizacija na temelju vizualnih obilježja 49
4.4.3 Detekcijski podskup: odabir slika i raspodjela klasa 49
4.4.4 Reidentifikacijski podskup: fizički identitet iznad kategorije 51
4.4.5 Video podskup za praćenje više plovila 52
4.4.6 Okluzije kao najvažniji dio SSMOT-a za analizu identiteta 53
4.5 Noviji relevantni skupovi 54
4.5.1 ShipsMOT (2025.) 54
4.5.2 ViV-ReID (2025./2026.) 54
4.5.3 HOSS ReID (2025.) i ostale multimodalne baze 55
4.6 Kritička usporedba primjenjivosti baza za buduću evaluaciju 55

4.7 Zaključak 56
5 EKSPERIMENTALNA ANALIZA VIZUALNE SLIČNOSTI KLASA PLOVILA I EVALUACIJA YOLO DETEKTORA 57
5.1 Polazište i istraživačko pitanje 57
5.2 Eksperimentalni skup i postupak anotiranja – relevantne razlike 57
5.3 Višeklasna detekcija: YOLO11m na 640 px 58
5.4 Višeklasna detekcija pri povećanoj rezoluciji 59
5.5 Redukcija klasa: eksperiment s jedinstvenom klasom „plovilo” 60
5.6 Usporedni prikaz provedenih pokusa 62
5.7 Metodološka ograničenja i implikacije za praćenje plovila 63
5.8 Zaključak poglavlja i poveznica s budućim istraživanjima 63
6 ZAKLJUČAK 64
LITERATURA 65

1 UVOD

Razvoj računalnog vida i metoda dubokog učenja omogućio je značajan napredak u automatiziranoj analizi videosignala, uključujući detekciju, klasifikaciju i praćenje objekata u složenim okruženjima. Jedno od posebno zahtjevnih područja primjene jest pomorski nadzor, gdje je potrebno pouzdano prepoznavati i pratiti plovila različitih veličina, oblika i namjena, često u uvjetima promjenjivog osvjetljenja, refleksija na morskoj površini, velikih udaljenosti od kamere te međusobnih djelomičnih i potpunih zaklanjanja. Za razliku od detekcije, koja određuje prisutnost i položaj objekta u pojedinačnom kadru, praćenje više objekata (*Multi-Object Tracking*, MOT) zahtijeva očuvanje njihova identiteta tijekom cijelog videoslijeda, uključujući razdoblja privremenog gubitka vidljivosti.

Suvremeni algoritmi praćenja temelje se na različitim kombinacijama detekcije objekata, predikcije gibanja, geometrijske asocijacije i vizualne reidentifikacije (*Re-Identification*, ReID). Njihova učinkovitost značajno ovisi o kvaliteti ulaznih detekcija, pouzdanosti procjene gibanja i mogućnosti razlikovanja vizualno sličnih objekata. Poseban izazov predstavljaju okluzije tijekom kojih se mijenjaju položaj i dimenzije detekcijskih pravokutnika, dok potpuna zaklonjenost uzrokuje prekid opažanja i povećanje nesigurnosti predikcije. U pomorskim scenama navedeni problemi dodatno su izraženi zbog sličnosti pojedinih plovila, sporog i promjenjivog gibanja te mogućnosti dugotrajnog zaklanjanja manjeg plovila većim objektima ili lučkom infrastrukturom.

Cilj ovoga kvalifikacijskog rada jest sustavno prikazati i kritički analizirati postojeće metode detekcije, praćenja i reidentifikacije plovila, s posebnim naglaskom na njihove mogućnosti i ograničenja u složenim pomorskim scenama. Razmatraju se temeljni i suvremeni MOT algoritmi, njihove arhitekturne značajke, postupci povezivanja detekcija i mehanizmi očuvanja identiteta tijekom okluzija. Posebna pozornost posvećena je evaluacijskim metrikama MOTA, IDF1 i HOTA te njihovoj prikladnosti za procjenu kvalitete praćenja u uvjetima dugotrajnih prekida vidljivosti. Uz algoritamske pristupe analizira se i NVIDIA DeepStream SDK kao programski okvir za razvoj i izvođenje sustava inteligentne videoanalitike u stvarnom vremenu.

Važan dio rada obuhvaća pregled dostupnih skupova podataka za detekciju, reidentifikaciju i praćenje plovila, s posebnim osvrtom na skup Split Ship MOT (SSMOT), razvijen za analizu zahtjevnih scenarija praćenja u području splitske luke. U okviru rada prikazani su i rezultati prethodno provedenih eksperimenata s modelom YOLO11m, u kojima je analiziran utjecaj vizualne sličnosti kategorija plovila, razlučivosti ulaznih slika i objedinjavanja klasa na kvalitetu detekcije. Ti rezultati omogućuju dodatno sagledavanje odnosa između anotacije, detekcije i pouzdanosti daljnjeg praćenja objekata.

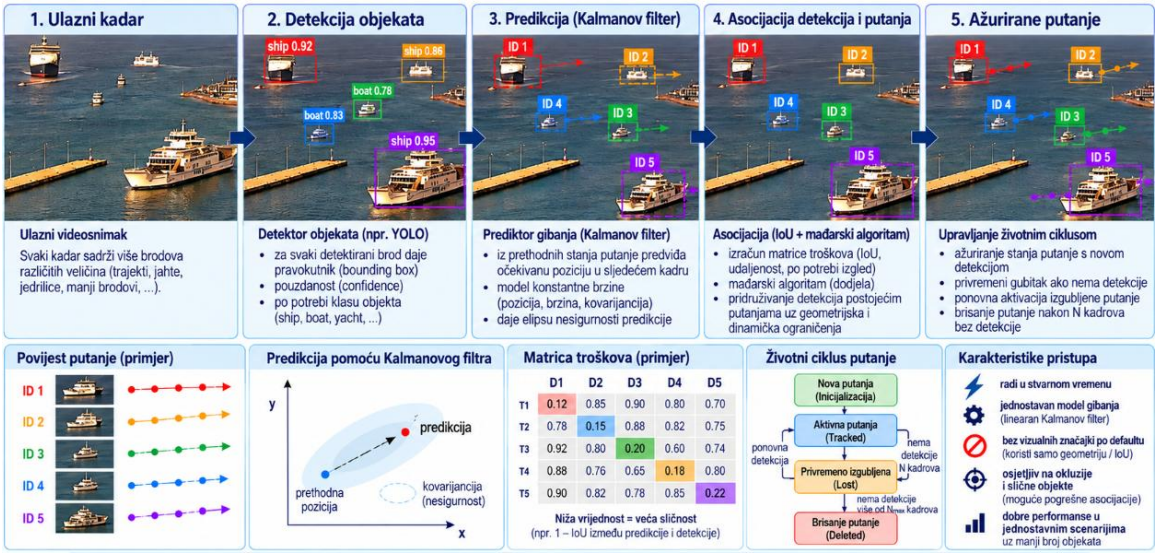
Rad je organiziran u šest poglavlja. Nakon uvoda, drugo poglavlje donosi pregled i usporedbu algoritama za praćenje više objekata te evaluacijskih metrika. Treće poglavlje opisuje arhitekturu i komponente sustava NVIDIA DeepStream SDK. Četvrto poglavlje razmatra relevantne skupove podataka, dok peto prikazuje eksperimentalnu analizu detekcije plovila modelima YOLO. U završnom poglavlju sažimaju se najvažnija saznanja i ograničenja razmatranih pristupa. Time rad pruža pregled teorijskih, algoritamskih i eksperimentalnih osnova relevantnih za znanstveno istraživanje detekcije, praćenja i reidentifikacije plovila.

2 PREGLED ALGORITAMA ZA PRAĆENJE VIŠE OBJEKATA

Praćenje više objekata (engl. Multi-Object Tracking, MOT) predstavlja važnu komponentu suvremenih sustava računalnog vida jer omogućuje kontinuirano određivanje položaja i očuvanje identiteta objekata tijekom videoslijeda. Iako su razvijeni brojni algoritmi za praćenje, njihova pouzdanost i dalje predstavlja izazov u složenim scenama, osobito pri međusobnom preklapanju objekata, privremenom gubitku vidljivosti i pojavi objekata sličnih vizualnih obilježja. Navedeni problemi posebno su izraženi u pomorskom okruženju, gdje promjene perspektive, udaljenosti, osvjetljenja i dinamike gibanja dodatno otežavaju održavanje kontinuiteta putanja. U ovom poglavlju daje se pregled najznačajnijih algoritama za praćenje više objekata, od klasičnih pristupa temeljenih na geometrijskoj asocijaciji i predikciji gibanja do suvremenih metoda koje uključuju duboke vizualne značajke, reidentifikaciju i posebne mehanizme za obradu okluzija. Njihovom usporedbom nastoje se identificirati prednosti, ograničenja i otvoreni problemi koji su posebno važni za praćenje plovila. Radi razumijevanja razlika među pojedinim pristupima, najprije se razmatra zajednička struktura algoritama za praćenje, nakon čega slijedi detaljniji pregled odabranih metoda i njihova kritička usporedba.

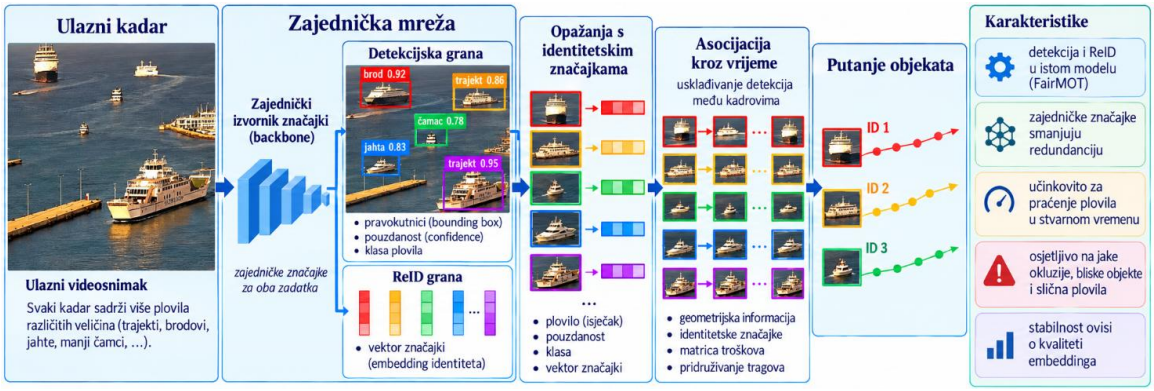
2.1 Zajednička struktura algoritama praćenja

U paradigmi "*tracking-by-detection*" za svaki kadar detektor daje skup opažanja: pravokutnike, pouzdanost i po potrebi klasu. Prediktor gibanja iz povijesti putanje računa očekivanu poziciju. Modul asocijacije pridružuje nova opažanja postojećim putanjama uz ograničenja geometrijske kompatibilnosti, gibanja i/ili izgleda. Upravljanje životnim ciklusom odlučuje kada se putanja uspostavlja, privremeno gubi, ponovno aktivira ili briše.



Slika 2.1. Model praćenja na temelju detekcije - "tracking-by-detection" (generirano AI).

Alternativa su zajednički modeli detekcije i pridruživanja, kao FairMOT. Za razliku od klasičnog pristupa "tracking-by-detection", modeli poput FairMOT-a istodobno uče detekciju objekata i identitetske značajke, čime se pridruživanje među kadrovima izvodi izravnije i učinkovitije.



Slika 2.2. Zajednički model detekcije i pridruživanja – primjer FairMOT (generirano AI).

U pomorskim scenama uspjeh ovisi o smjeru gibanja plovila, perspektivnom smanjivanju objekata, refleksijama te dugotrajnom preklapanju. Djelomična okluzija istodobno mijenja geometriju detekcijskog pravokutnika i sadržaj slike iz kojeg se izdvaja vizualni vektor. Stoga nije dovoljno svesti evaluaciju na jednu agregatnu metriku.

2.2 Kritički prikaz odabranih MOT algoritama

2.2.1 SORT - Simple Online and Realtime Tracking

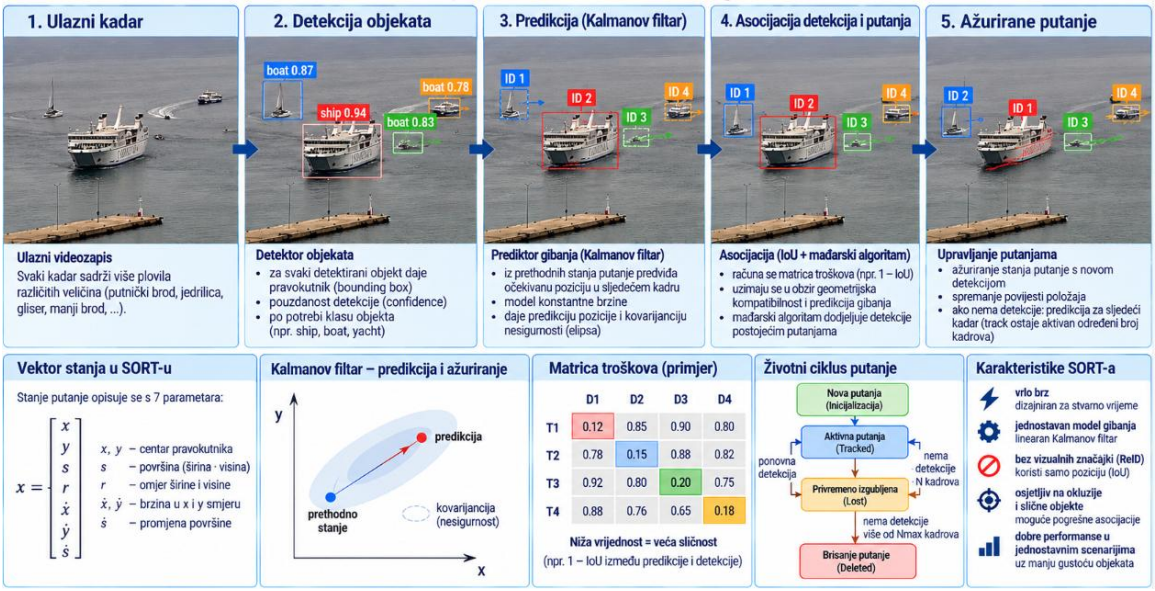
Algoritam SORT (Simple Online and Realtime Tracking) predstavili su Bewley i suradnici 2016. godine s ciljem razvoja jednostavnog i računalno učinkovitog pristupa praćenju više objekata u stvarnom vremenu [1]. Temelji se na paradigmi tracking-by-detection, pri čemu se detekcija objekata provodi zasebnim algoritmom, dok je SORT odgovoran za povezivanje detekcija između uzastopnih kadrova i održavanje njihovih putanja. Osnovna je pretpostavka algoritma da se gibanje objekata tijekom kratkih vremenskih intervala može dovoljno dobro opisati jednostavnim linearnim modelom, bez potrebe za analizom njihovih vizualnih značajki.

SORT koristi Kalmanov filtar za procjenu položaja i gibanja svakog praćenog objekta. Stanje pojedinog objekta opisuje se vektorom koji sadrži koordinate središta detekcijskog pravokutnika, njegovu površinu i omjer širine i visine te pripadajuće brzine promjene položaja i površine. Pretpostavlja se približno konstantna brzina gibanja između uzastopnih kadrova, što omogućuje predviđanje položaja objekta čak i kada nova detekcija privremeno nije dostupna.

Za određivanje međusobne podudarnosti SORT koristi mjeru preklapanja pravokutnika Intersection over Union (IoU). Ona predstavlja omjer površine presjeka i površine unije predviđenog i detektiranog pravokutnika. Veća vrijednost IoU-a upućuje na veću prostornu podudarnost, a time i na veću vjerojatnost da detekcija pripada odgovarajućoj putanji. Na temelju izračunatih vrijednosti formira se matrica troškova pridruživanja, pri čemu se najpovoljnije globalno uparivanje određuje primjenom mađarskog algoritma (Hungarian algorithm).

Nakon završetka pridruživanja ažuriraju se stanja uspješno povezanih putanja. Kalmanov filtar pritom kombinira prethodnu predikciju s novim mjerenjem, uzimajući u obzir nesigurnost procijenjenog stanja i opažanja. Ako se određena detekcija ne može pridružiti nijednoj postojećoj putanji, inicijalizira se nova putanja koja mora zadovoljiti zadane uvjete potvrđivanja prije nego što se smatra pouzdano praćenim objektom. S druge strane, putanje kojima nije pridružena nova detekcija privremeno se održavaju

na temelju predikcije gibanja. Ako izostanak detekcija traje dulje od zadanog broja kadrova, odgovarajuća putanja uklanja se iz aktivnog skupa.



Slika 2.3. SORT - Simple Online and Realtime Tracking (generirano AI).

Glavna prednost algoritma SORT proizlazi iz njegove jednostavne arhitekture i male računalne složenosti. Kombinacija Kalmanova filtra, IoU mjere i mađarskog algoritma omogućuje učinkovito praćenje većeg broja objekata bez izvođenja dodatnih neuronskih mreža za izdvajanje vizualnih značajki. Zbog toga je SORT prikladan za sustave koji zahtijevaju obradu videosignala u stvarnom vremenu, a često se koristi i kao referentni algoritam pri evaluaciji složenijih metoda praćenja.

Poseban problem predstavljaju djelomične i potpune okluzije. Tijekom djelomične okluzije detekcijski pravokutnik može promijeniti dimenzije i položaj jer detektor prepoznaje samo vidljivi dio objekta. Takvo opažanje može uzrokovati pogrešno ažuriranje Kalmanova filtra, čime se narušava procjena položaja i brzine. Kod potpune okluzije algoritam se oslanja isključivo na predikciju, a njezina nesigurnost raste s trajanjem izostanka mjerenja. Nakon uklanjanja putanje SORT nema ugrađen mehanizam reidentifikacije kojim bi ponovno pojavljenom objektu mogao pouzdano pridružiti prethodni identitet.

Navedena ograničenja posebno dolaze do izražaja u pomorskom okruženju. Primjerice, dva plovila sličnih dimenzija koja se međusobno približavaju mogu imati vrlo slične detekcijske pravokutnike. Ako jedno plovilo postupno zaklanja drugo, promjena vidljivog dijela trupa može uzrokovati pomak središta detekcije i pogrešnu procjenu brzine, iako stvarno gibanje plovila ostaje nepromijenjeno. Nakon njihovog razdvajanja algoritam može zamijeniti identitete ili uspostaviti nove putanje.

2.2.2 DeepSORT - Simple Online and Realtime Tracking with a Deep Association Metric

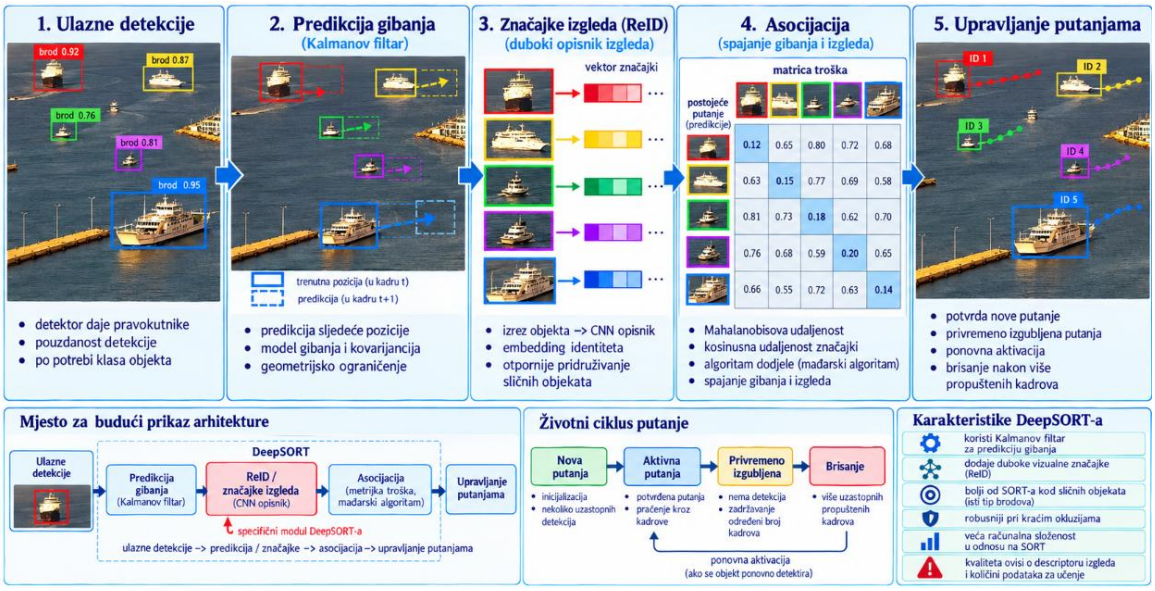
Algoritam DeepSORT (Simple Online and Realtime Tracking with a Deep Association Metric) predstavili su Wojke, Bewley i Paulus 2017. godine kao proširenje algoritma SORT s ciljem poboljšanja pouzdanosti pridruživanja detekcija i smanjenja broja zamjena identiteta tijekom praćenja više objekata [2]. Osnovna novost algoritma temelji se na uključivanju dubokih vizualnih značajki (deep appearance features) u postupak asocijacije, čime se prevladava jedno od glavnih ograničenja izvornog SORT-

a, koji objekte povezuje gotovo isključivo na temelju geometrijske podudarnosti i predikcije gibanja. DeepSORT zadržava Kalmanov filtar za procjenu dinamičkog stanja objekta, ali uvodi dodatni model dubokog učenja koji iz slikovnog prikaza objekta izdvaja karakterističan vektor značajki. Kombiniranjem prostornih, dinamičkih i vizualnih informacija omogućuje se pouzdanije razlikovanje objekata, osobito u situacijama kada se njihove putanje približavaju ili međusobno preklapaju.

Prostorna kompatibilnost određuje se Mahalanobisovom udaljenošću između predviđenog stanja postojeće putanje i nove detekcije. Za razliku od obične euklidske udaljenosti, Mahalanobisova udaljenost uzima u obzir kovarijanciju predikcije, odnosno nesigurnost procijenjenog položaja i dimenzija objekta. Time se omogućuje razlikovanje detekcija koje su, s obzirom na trenutačnu nesigurnost modela gibanja, prostorno prihvatljive od onih koje odstupaju od očekivanog ponašanja. Detekcije koje ne zadovoljavaju zadani statistički prag isključuju se iz daljnjeg postupka pridruživanja.

Vizualna sličnost određuje se kosinusnom udaljenošću između vektora nove detekcije i prethodno pohranjenih vektora značajki pridruženih pojedinoj putanji. Za svaki praćeni objekt održava se galerija prethodnih vizualnih reprezentacija, čime se omogućuje usporedba s više ranijih opažanja, a ne isključivo s posljednjom detekcijom. U izvornoj metodi vizualni trošak pridruživanja određuje se prema najmanjoj kosinusnoj udaljenosti između nove detekcije i pohranjenih značajki odgovarajuće putanje.

Posebna značajka DeepSORT-a jest kaskadni postupak pridruživanja (matching cascade). Umjesto istodobnog pridruživanja svih postojećih putanja, algoritam daje prednost putanjama koje su nedavno uspješno povezane s detekcijama. Najprije se obrađuju putanje potvrđene u prethodnom kadru, a zatim one koje dulje vrijeme nisu imale odgovarajuće opažanje. Time se smanjuje mogućnost da putanja s velikom nesigurnošću predikcije preuzme detekciju koja pripada pouzdanije praćenom objektu. Nakon kaskadnog pridruživanja preostale se detekcije i odgovarajuće putanje mogu povezivati dodatnim postupkom temeljenim na IoU mjeri.



Slika 2.4. DeepSORT - Simple Online and Realtime Tracking with a Deep Association Metric (generirano AI).

Najvažnija prednost algoritma DeepSORT u odnosu na SORT jest uključivanje vizualnih značajki u postupak pridruživanja objekata. Time se smanjuje ovisnost o

prostornoj blizini i geometrijskom preklapanju detekcijskih pravokutnika, što je osobito korisno u scenama s većim brojem objekata čije se putanje međusobno približavaju ili presijecaju. Dodatna vizualna informacija može spriječiti pogrešno pridruživanje detekcija koje su prostorno vrlo bliske, ali pripadaju objektima različitog izgleda.

Međutim, uvođenje dubokih vizualnih značajki povećava računalnu složenost sustava. Za svaku novu detekciju potrebno je izdvojiti odgovarajući slikovni izrez i provesti dodatnu inferenciju neuronske mreže. Pri većem broju objekata ili istodobnoj obradi više videotokova takav postupak može značajno povećati zahtjeve za GPU resursima i utjecati na ukupnu latenciju obrade.

Poseban problem nastaje tijekom postupnog ulaska u okluziju. Kada drugo plovilo počne zaklanjati praćeni objekt, njegov detekcijski pravokutnik može obuhvatiti samo vidljivi dio trupa ili uključivati dijelove objekta koji uzrokuje zaklanjanje. Vizualni vektor izdvojen iz takvog područja više ne predstavlja pouzdan opis izvornog plovila. Ako se takav vektor pohrani u galeriju značajki, povećava se mogućnost pogrešnog pridruživanja pri kasnijoj reidentifikaciji.

2.2.3 FairMOT – On the Fairness of Detection and Re-Identification in Multiple Object Tracking

Algoritam FairMOT (*Fair Multi-Object Tracking*) predstavili su Zhang i suradnici 2021. godine s ciljem razvoja učinkovitog sustava za praćenje više objekata koji objedinjuje detekciju i reidentifikaciju unutar jedinstvene neuronske mreže [3]. Za razliku od algoritama SORT i DeepSORT, kod kojih se detekcija objekata provodi zasebno od postupka praćenja, FairMOT primjenjuje pristup zajedničkog učenja (*joint detection and embedding*, JDE), pri čemu se istodobno određuju položaji objekata i izdvajaju njihove identitetske značajke.

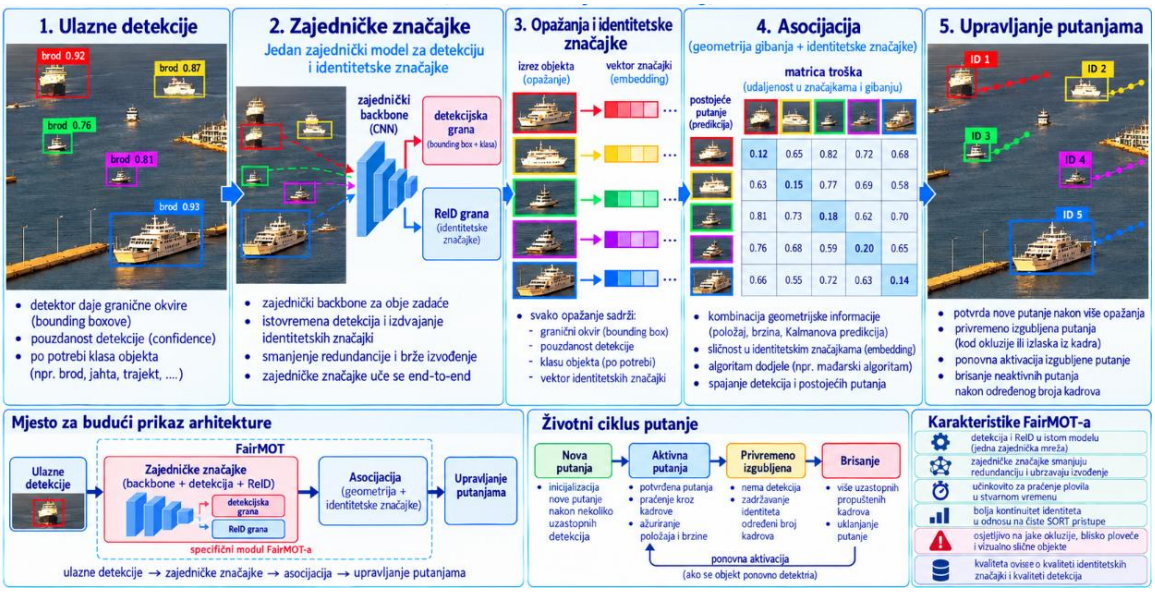
Arhitektura algoritma FairMOT temelji se na neuronskoj mreži sa zajedničkim ekstraktorom značajki i dvjema paralelnim granama koje obavljaju detekciju objekata i izdvajanje njihovih identitetskih reprezentacija. Kao osnova koristi se pristup detekciji bez unaprijed definiranih sidrenih pravokutnika (*anchor-free detection*), zasnovan na arhitekturi CenterNet. Umjesto generiranja velikog broja potencijalnih detekcijskih pravokutnika, mreža izravno predviđa središta objekata te njihove prostorne dimenzije.

Važna prednost ovakvog pristupa jest izbjegavanje unaprijed definiranih sidrenih pravokutnika (*anchors*). U algoritmima koji koriste sidrene pravokutnike potrebno je odrediti veći broj mogućih oblika i dimenzija objekata, nakon čega se odabiru najprikladnije detekcije. FairMOT pojednostavljuje taj postupak jer lokaciju objekta određuje prema njegovu središtu. Time se smanjuje problem prostornog nepodudaranja između regije korištene za detekciju i regije iz koje se izdvajaju identitetske značajke.

Za razliku od DeepSORT-a, kod kojega se za svaku detekciju zasebno izdvaja slikovni izrez i provodi inferencija ReID mreže, FairMOT identitetske značajke dobiva tijekom istog prolaska kroz zajedničku neuronsku mrežu. Takva organizacija smanjuje potrebu za ponovljenom obradom slikovnih podataka i omogućuje učinkovitije korištenje računalnih resursa.

Upravljanje životnim ciklusom putanja temelji se na potvrđivanju novih objekata, održavanju aktivnih putanja i privremenom zadržavanju objekata kojima nije pridružena nova detekcija. Ako objekt kratkotrajno nestane iz vidnog polja, njegovo se stanje nastavlja predviđati Kalmanovim filtrom. Nakon ponovne pojave pokušava se povezati s prethodnom putanjom na temelju vizualnih značajki i prostorne

kompatibilnosti. Ako povezivanje nije uspješno ili je razdoblje nevidljivosti dulje od dopuštenog, može doći do uspostavljanja novog identiteta.



Slika 2.5. FairMOT - On the Fairness of Detection and Re-Identification in Multiple Object Tracking (generirano AI).

Najvažnija prednost algoritma FairMOT proizlazi iz zajedničke obrade detekcije i reidentifikacije objekata. Za razliku od pristupa koji koriste odvojene neuronske mreže, FairMOT omogućuje dijeljenje računalno zahtjevnog ekstraktora značajki između dviju zadaća. Time se smanjuje potreba za dodatnom inferencijom nad svakim detektiranim objektom, što je osobito značajno u sustavima koji moraju obrađivati videosignal u stvarnom vremenu.

Navedeni problem posebno dolazi do izražaja kod malih objekata. Kada plovilo zauzima vrlo malo područje slike, broj dostupnih piksela može biti nedovoljan za pouzdano određivanje njegova središta i izdvajanje diskriminativnih vizualnih značajki. U pomorskom okruženju to se često događa pri praćenju udaljenih brodica i jedrilica, čije se dimenzije zbog perspektive znatno smanjuju. U takvim uvjetima različiti objekti mogu imati vrlo slične embedding vektore, što povećava mogućnost pogrešnog pridruživanja.

Problemi nastaju i tijekom djelomičnih okluzija, kada je središte vidljivog dijela objekta pomaknuto u odnosu na središte njegova stvarnog položaja. Budući da FairMOT identitetske značajke izdvaja iz područja povezanog s detektiranim središtem, takav pomak može uzrokovati izdvajanje značajki koje ne odgovaraju prethodnim opažanjima istog objekta. Istodobno, promjena detekcijskog pravokutnika može narušiti Kalmanovu predikciju, čime se smanjuje prostorna kompatibilnost između postojeće putanje i nove detekcije.

2.2.4 ByteTrack - Multi-Object Tracking by Associating Every Detection Box

Algoritam ByteTrack (Multi-Object Tracking by Associating Every Detection Box) predstavili su Zhang i suradnici 2022. godine s ciljem poboljšanja pouzdanosti praćenja više objekata kroz učinkovitije korištenje rezultata detekcije [4]. Za razliku od prethodno opisanih algoritama, čiji se razvoj prvenstveno temeljio na poboljšanju predikcije gibanja ili uključivanju vizualnih značajki, ByteTrack uvodi drugačiji pristup

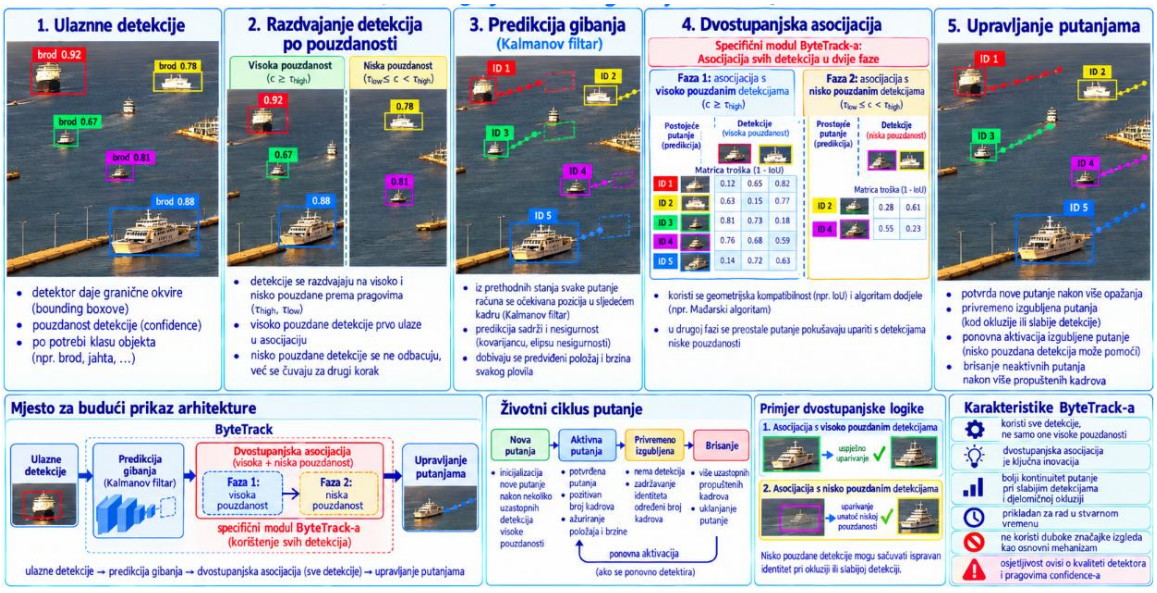
asocijaciji detekcija. Njegova osnovna pretpostavka jest da detekcije s niskom razinom pouzdanosti ne moraju nužno predstavljati pogrešne rezultate detektora, nego mogu odgovarati stvarnim objektima koji su djelomično zaklonjeni, udaljeni ili prikazani u nepovoljnim uvjetima. Odbacivanje takvih detekcija može uzrokovati nepotreban prekid putanja i gubitak identiteta objekata.

Za razliku od pristupa koji zadržavaju samo detekcije čija pouzdanost prelazi jedan unaprijed zadani prag, ByteTrack koristi dva praga pouzdanosti. Na temelju njih detekcije se razdvajaju u dvije skupine: detekcije visoke pouzdanosti (high-score detections) i detekcije niže pouzdanosti (low-score detections). Detekcije čija pouzdanost ne doseže niži prag odbacuju se jer se smatra da je vjerojatnost njihove pogrešnosti previsoka. Takvim razdvajanjem omogućuje se dodatno iskorištavanje potencijalno korisnih opažanja koja bi u standardnim postupcima bila uklonjena prije faze praćenja.

Nakon predikcije slijedi prvi stupanj pridruživanja, u kojem se postojeće putanje povezuju s detekcijama visoke pouzdanosti. Prostorna podudarnost određuje se na temelju mjere Intersection over Union (IoU), koja opisuje stupanj preklapanja predviđenog i detektiranog pravokutnika. Na temelju izračunatih vrijednosti formira se matrica troškova, a optimalno pridruživanje određuje se primjenom mađarskog algoritma.

Upravo se na takve putanje odnosi drugi stupanj pridruživanja, koji predstavlja temeljnu inovaciju ByteTracka. Putanje koje nisu uspješno povezane u prvom stupnju uspoređuju se s detekcijama niže pouzdanosti. Pritom se ponovno koristi prostorna kompatibilnost izvedena iz IoU mjere, uz zaseban prag prihvatljivosti pridruživanja. Ako se pronađe odgovarajuća detekcija, putanja se ažurira unatoč tome što je detektor njezinu pouzdanost procijenio relativno niskom.

Važno je istaknuti da se detekcije niske pouzdanosti u izvornom ByteTracku ne koriste na isti način kao detekcije visoke pouzdanosti. Njihova je prvenstvena uloga očuvanje već uspostavljenih putanja, a ne pokretanje novih. Time se smanjuje mogućnost inicijalizacije velikog broja lažnih objekata, što bi predstavljalo značajan rizik kada bi se sve detekcije tretirale jednako.



Slika 2.6. ByteTrack - Multi-Object Tracking by Associating Every Detection Box (generirano AI).

Najvažnija prednost algoritma ByteTrack proizlazi iz učinkovitijeg korištenja rezultata detekcije. Za razliku od metoda koje unaprijed odbacuju sve detekcije ispod relativno visokog praga pouzdanosti, ByteTrack zadržava mogućnost njihova korištenja ako su prostorno kompatibilne s već uspostavljenim putanjama. Time se smanjuje broj nepotrebnih prekida praćenja u situacijama kada detektor privremeno gubi sigurnost u prisutnost objekta.

Međutim, temeljno ograničenje ByteTracka proizlazi iz činjenice da se pridruživanje u osnovnoj implementaciji oslanja prvenstveno na prostornu kompatibilnost i predikciju gibanja. Algoritam ne uključuje zaseban model vizualne reidentifikacije kojim bi se uspoređivale karakteristične značajke pojedinačnih objekata. Posljedično, u scenama s velikim brojem međusobno bliskih objekata može doći do pogrešnog pridruživanja, osobito kada se njihove putanje presijecaju ili kada imaju slične dimenzije.

U opisanim okolnostima ByteTrack se suočava s dvostrukim problemom. S jedne strane, dvostupanjsko pridruživanje omogućuje prihvaćanje detekcija niže pouzdanosti, što povećava vjerojatnost očuvanja putanje. S druge strane, prihvaćanje prostorno pomaknutih ili deformiranih detekcijskih pravokutnika može dovesti do postupnog narušavanja dinamičkog modela. Zbog toga kontinuirano praćenje ne mora nužno značiti i točno praćenje položaja ili očuvanje identiteta.

Važno ograničenje predstavlja i ovisnost o konfiguraciji pragova detekcijske pouzdanosti. Ako je gornji prag postavljen previsoko, velik broj stvarnih objekata može biti prebačen u skup detekcija niže pouzdanosti, čime se smanjuje mogućnost pouzdanog inicijaliziranja novih putanja. S druge strane, prenizak donji prag može povećati broj potencijalno pogrešnih detekcija koje se razmatraju tijekom drugog stupnja pridruživanja. Odabir odgovarajućih pragova stoga ovisi o karakteristikama detektora i promatranog okruženja.

2.2.5 StrongSORT i StrongSORT++ – Unaprjeđenje praćenja kombiniranjem vizualnih značajki, modela gibanja i naknadne obrade putanja

Algoritme StrongSORT i StrongSORT++ predstavili su Du i suradnici u radu StrongSORT: Make DeepSORT Great Again, objavljenom u časopisu IEEE Transactions on Multimedia 2023. godine [5]. Polazište njihova razvoja bila je analiza ograničenja algoritma DeepSORT te mogućnosti poboljšanja njegove učinkovitosti primjenom kvalitetnijih detekcijskih modela, robusnijih vizualnih značajki i naprednijih postupaka pridruživanja objekata. Za razliku od pristupa koji uvode potpuno novu arhitekturu praćenja, StrongSORT zadržava osnovnu strukturu DeepSORT-a, ali unaprjeđuje pojedine komponente odgovorne za procjenu gibanja, reidentifikaciju i upravljanje putanjama.

Uz osnovni StrongSORT, autori predlažu i proširenu inačicu StrongSORT++, koja uvodi dva dodatna mehanizma: AFLink (Appearance-Free Link) za povezivanje prethodno razdvojenih dijelova putanja i GSI (Gaussian-Smoothed Interpolation) za procjenu nedostajućih položaja objekata. Time se nastoje riješiti dva različita problema praćenja: prekid povezanosti putanja zbog pogrešne asocijacije te nedostatak opažanja tijekom pojedinih vremenskih intervala. Dok StrongSORT prvenstveno poboljšava postupak online praćenja, StrongSORT++ dodatno obrađuje dobivene putanje kako bi se poboljšala njihova vremenska povezanost i prostorna potpunost.

Jedno od važnih unaprjeđenja StrongSORT-a predstavlja kompenzacija gibanja kamere (camera motion compensation, CMC). U izvornom radu koristi se postupak ECC

(Enhanced Correlation Coefficient) za procjenu geometrijske transformacije između uzastopnih kadrova. Naime, promjena položaja objekta u slici može biti posljedica njegova stvarnog gibanja, ali i pomicanja kamere. Ako se te dvije komponente ne razlikuju, model gibanja može pogrešno protumačiti promjenu koordinata objekta kao njegovu vlastitu brzinu ili promjenu smjera.

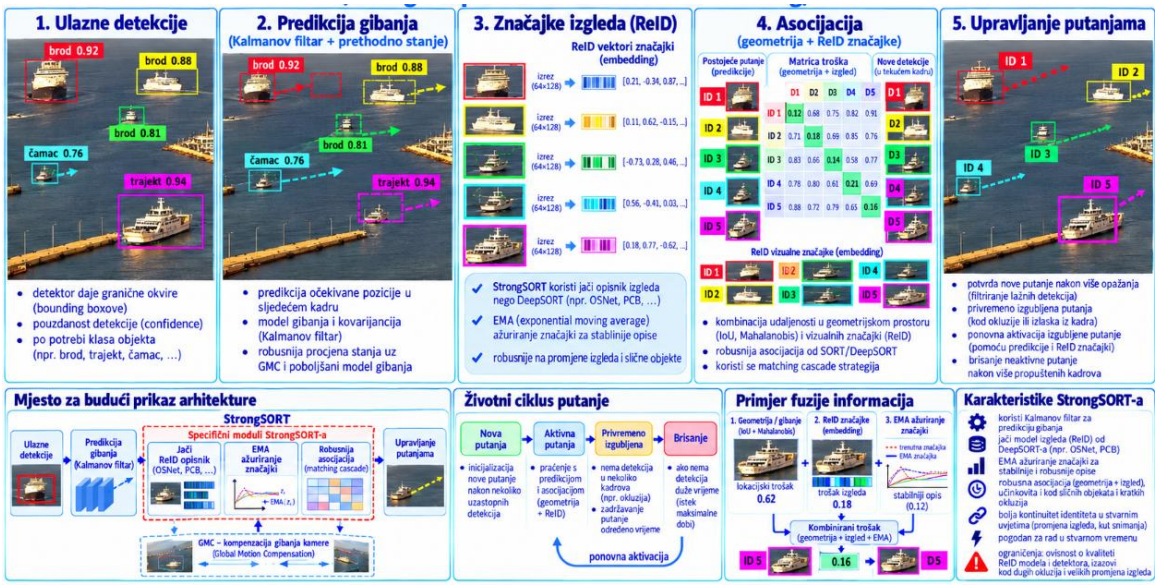
Sljedeće unaprjeđenje odnosi se na model predikcije gibanja. Umjesto standardnog Kalmanova filtra, StrongSORT koristi njegovu prilagođenu inačicu NSA (Noise Scale Adaptive) Kalman, koja pri određivanju nesigurnosti mjerenja uzima u obzir pouzdanost detekcije. Osnovna je ideja da detekcije različite pouzdanosti ne bi trebale jednako utjecati na ažuriranje procijenjenog stanja objekta.

Posebno važan dio StrongSORT-a predstavlja poboljšano upravljanje vizualnim značajkama objekata. DeepSORT za svaku putanju održava galeriju prethodno izdvojenih vektora, pri čemu se nova detekcija uspoređuje s pohranjenim reprezentacijama. StrongSORT umjesto takvog pristupa koristi eksponencijalni pomični prosjek (Exponential Moving Average, EMA), kojim se prethodna vizualna reprezentacija postupno ažurira novim opažanjima.

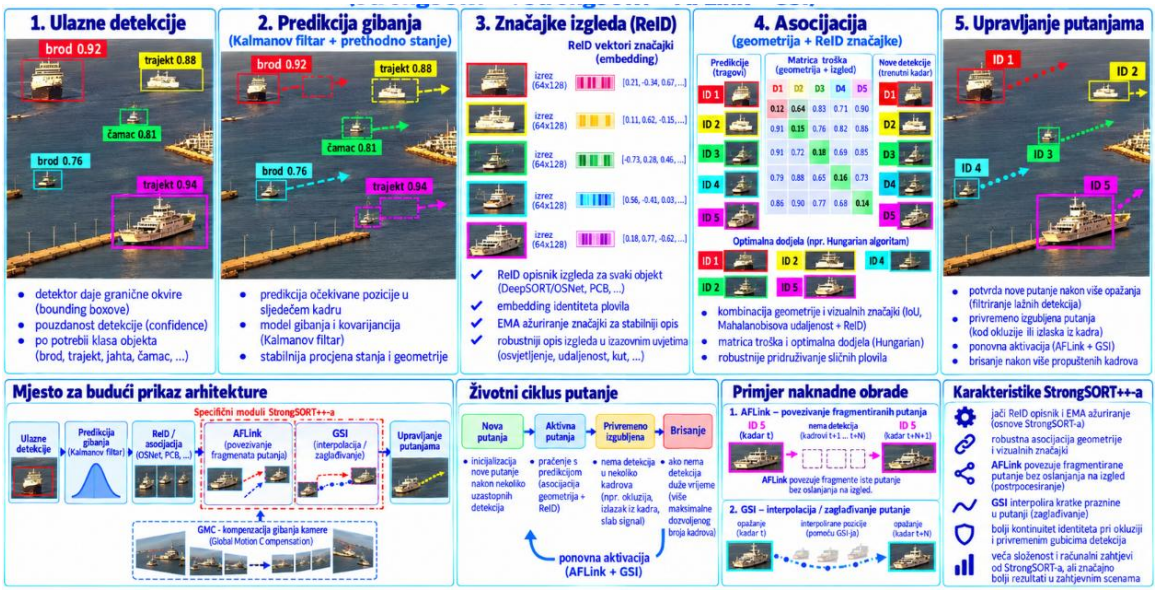
Postupak asocijacije također je izmijenjen u odnosu na izvorni DeepSORT. Dok se kod DeepSORT-a informacije o gibanju prvenstveno koriste za ograničavanje mogućih pridruživanja, StrongSORT uključuje vizualnu i dinamičku udaljenost u zajedničku funkciju troška. Time se nastoji postići uravnoteženija procjena kompatibilnosti detekcija i postojećih putanja. Vizualna sličnost pritom ostaje dominantna komponenta, dok model gibanja pruža dodatne informacije potrebne za razrješavanje nejasnih pridruživanja.

Prvi postupak, AFLink (Appearance-Free Link), namijenjen je povezivanju razdvojenih dijelova putanja, odnosno trackleta. Tijekom praćenja isti objekt može dobiti više različitih identifikatora ako njegovo praćenje bude prekinuto zbog okluzije, pogreške detekcije ili neuspješne asocijacije. AFLink pokušava utvrditi pripadaju li dva vremenski razdvojena dijela putanje istom objektu na temelju njihovih prostornih i vremenskih karakteristika.

Drugi dodatni postupak, GSI (Gaussian-Smoothed Interpolation), namijenjen je procjeni položaja objekta tijekom vremenskih intervala u kojima nisu dostupne odgovarajuće detekcije. Standardna linearna interpolacija pretpostavlja jednostavnu promjenu položaja između dvaju poznatih opažanja, što može biti nedovoljno za objekte čije se gibanje mijenja tijekom promatranog intervala.



Slika 2.7. StrongSORT (generirano AI).



Slika 2.8. StrongSORT++ (generirano AI).

Glavna prednost StrongSORT-a proizlazi iz sustavnog unaprjeđenja postojećih komponenti algoritma DeepSORT. Umjesto uvođenja potpuno nove arhitekture, autori pokazuju da se značajna poboljšanja mogu postići kvalitetnijom detekcijom, diskriminativnijim vizualnim značajkama, prilagodbom Kalmanova filtra i učinkovitijim postupkom asocijacije. Takav pristup potvrđuje da uspješnost praćenja više objekata ovisi o međusobnoj usklađenosti pojedinih komponenti sustava.

Posebnu pozornost zahtijeva EMA mehanizam ažuriranja vizualnih značajki. Iako omogućuje smanjenje utjecaja slučajnih promjena izgleda, njegova učinkovitost pretpostavlja da su nova opažanja pravilno pridružena odgovarajućoj putanji i da sadrže dovoljno pouzdanu vizualnu reprezentaciju objekta. Tijekom postupne okluzije detekcijski pravokutnik može obuhvaćati samo dio plovila ili uključivati elemente susjednog objekta.

Slično ograničenje prisutno je i kod NSA Kalmanova filtra. Prilagodba kovarijancije mjerenja prema pouzdanosti detekcije omogućuje smanjenje utjecaja opažanja za koja detektor izražava manju sigurnost. Međutim, visoka pouzdanost detekcije ne jamči nužno da pravokutnik odgovara stvarnom položaju i dimenzijama objekta. Tijekom djelomične okluzije detektor može s relativno visokom pouzdanošću prepoznati vidljivi dio plovila, iako je njegov pravokutnik geometrijski pomaknut u odnosu na cjeloviti objekt.

Međutim, povezivanje putanja isključivo na temelju prostorno-vremenskih odnosa može biti nepouzdano u scenama s više objekata sličnog gibanja. Primjerice, ako dvije jedrilice plove usporednim putanjama i obje privremeno nestanu iza većeg plovila ili lučke infrastrukture, AFLink može imati poteškoća pri određivanju odgovarajućeg nastavka pojedine putanje. Budući da model ne koristi vizualne značajke, ne raspolaže dodatnim informacijama kojima bi razlikovao prostorno i dinamički slične objekte.

Važna razlika između StrongSORT-a i StrongSORT++-a odnosi se na mogućnost primjene u stvarnom vremenu. Osnovni StrongSORT omogućuje online obradu kadrova, dok AFLink i GSI u predloženoj izvedbi predstavljaju postupke naknadne obrade koji koriste širi vremenski kontekst i mogu zahtijevati informacije iz budućih kadrova. Njihova neposredna primjena u sustavima koji zahtijevaju minimalnu latenciju stoga nije moguća bez odgovarajuće prilagodbe arhitekture i prihvaćanja dodatnog vremenskog kašnjenja.

2.2.6 BoT-SORT - Robust Associations Multi-Pedestrian Tracking

Algoritam BoT-SORT (Robust Associations Multi-Pedestrian Tracking) predstavili su Aharon, Orfaig i Bobrovsky 2022. godine s ciljem poboljšanja pouzdanosti praćenja više objekata kombiniranjem informacija o gibanju, prostornoj podudarnosti, vizualnim značajkama i pomicanju kamere [6]. Algoritam se temelji na paradigmi tracking-by-detection te predstavlja nadogradnju pristupa razvijenih u algoritmima SORT, DeepSORT i ByteTrack. Njegova osnovna ideja jest povećati pouzdanost povezivanja detekcija i postojećih putanja poboljšanim modelom gibanja, kompenzacijom pomaka kamere i učinkovitijom asocijacijom objekata. U inačici BoT-SORT-ReID dodatno se koriste duboke vizualne značajke za razlikovanje pojedinačnih identiteta.

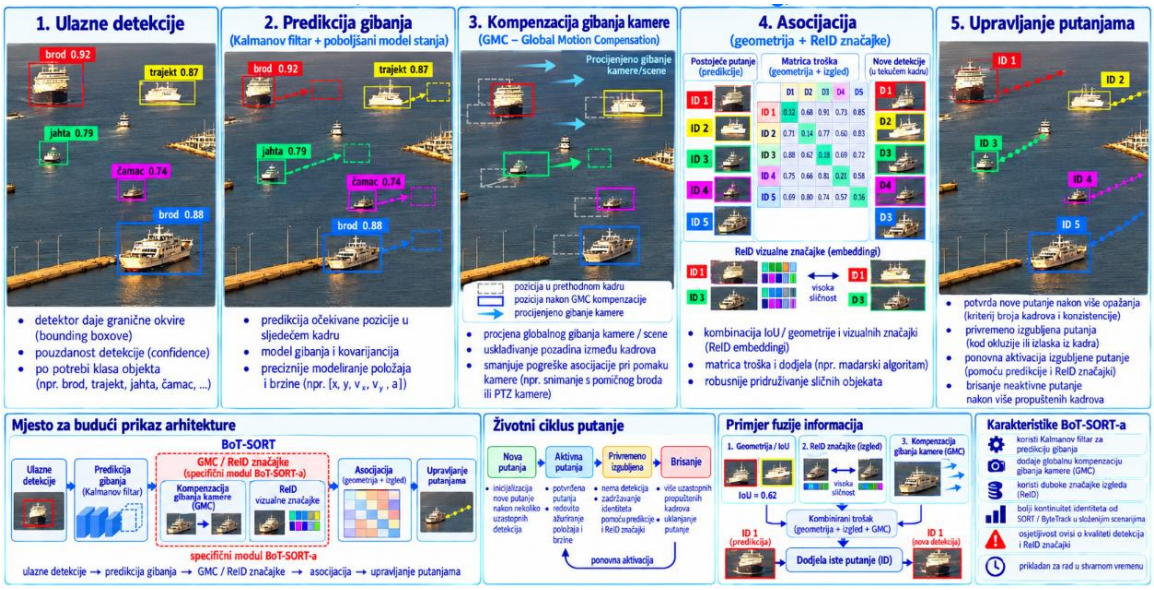
Jedno od temeljnih poboljšanja algoritma odnosi se na model stanja Kalmanova filtra. U algoritmima poput DeepSORT-a stanje objekta opisuje se koordinatama središta detekcijskog pravokutnika, omjerom njegove širine i visine, visinom te pripadajućim brzinama promjene. BoT-SORT umjesto omjera stranica koristi zasebne vrijednosti širine i visine, čime omogućuje njihovo neovisno modeliranje. Takav prikaz bolje odgovara situacijama u kojima se dimenzije objekta u slici mijenjaju zbog perspektive, promjene orijentacije ili drugih geometrijskih utjecaja.

Procjena gibanja kamere temelji se na prostornim značajkama slike i njihovim promjenama između uzastopnih kadrova. Dobivena transformacija koristi se za korekciju stanja putanja prije postupka pridruživanja. Time se nastoji odvojiti prividno gibanje uzrokovano kamerom od stvarnog gibanja objekata, što poboljšava pouzdanost predikcije i prostorne asocijacije.

U drugom stupnju pokušavaju se povezati preostale putanje s detekcijama niže pouzdanosti. Takav pristup omogućuje korištenje opažanja koja bi standardnim postupkom filtriranja mogla biti odbačena. Posebno je koristan kada je objekt djelomično zaklonjen ili kada zbog udaljenosti, osvjetljenja ili drugih čimbenika detektor ne može ostvariti visoku pouzdanost. Uspješnim pridruživanjem takvih

detekcija moguće je očuvati kontinuitet postojeće putanje i smanjiti broj nepotrebnih prekida praćenja.

Upravljanje putanjama temelji se na inicijalizaciji novih objekata, održavanju potvrđenih putanja i privremenom zadržavanju izgubljenih objekata. Detekcije visoke pouzdanosti koje nisu povezane s postojećim putanjama mogu inicijalizirati nove putanje, dok se objekti bez odgovarajuće detekcije privremeno održavaju na temelju predikcije gibanja. Ako se objekt ponovno pojavi unutar dopuštenog intervala i zadovolji uvjete pridruživanja, moguće je nastaviti prethodnu putanju. Nakon prekoračenja vremenskog praga izgubljena se putanja uklanja iz aktivnog skupa.



Slika 2.9. BoT-SORT: Robust Associations Multi-Pedestrian Tracking (Generirano AI).

Međutim, učinkovitost kompenzacije ovisi o pouzdanosti procjene globalnog gibanja slike. U pomorskom okruženju veliki dio vidnog polja može zauzimati morska površina, čiji se vizualni sadržaj neprestano mijenja zbog valova, refleksija i promjena osvjetljenja. Takve strukture ne predstavljaju stabilne prostorne značajke, zbog čega procjena globalne transformacije može biti nepouzdana. U slučaju pogrešne kompenzacije mogu se narušiti položaji svih praćenih objekata i povećati mogućnost pogrešne asocijacije.

Inačica BoT-SORT-ReID dodatno povećava pouzdanost praćenja korištenjem vizualnih značajki. Takav pristup može biti koristan pri praćenju više plovila čije se putanje približavaju ili presijecaju, jer omogućuje njihovo razlikovanje na temelju izgleda, a ne isključivo položaja. Međutim, učinkovitost reidentifikacije ovisi o diskriminativnosti izdvojenih značajki i njihovoj stabilnosti tijekom promjena perspektive i vidljivosti.

U slučaju dugotrajne potpune okluzije ograničenja BoT-SORT-a postaju još izraženija. Tijekom razdoblja bez detekcija Kalmanov filtar nastavlja predviđati položaj objekta, ali nesigurnost takve procjene raste s trajanjem izostanka opažanja. Ako plovilo tijekom okluzije uspori, zaustavi se ili promijeni smjer gibanja, predviđeni položaj može značajno odstupati od stvarnog. Nakon ponovnog pojavljivanja prostorna podudarnost može biti nedovoljna za ponovno povezivanje prethodne putanje.

Sa stajališta praćenja plovila, posebno je važno istaknuti da BoT-SORT ne uključuje eksplicitni model stanja okluzije kojim bi se razlikovalo normalno praćenje od

postupnog ulaska u okluziju, potpune zaklonjenosti i ponovnog pojavljivanja. Zbog toga algoritam ne raspolaže posebnim mehanizmom kojim bi na temelju procijenjene vidljivosti objekta mogao privremeno ograničiti ažuriranje modela gibanja i pohranjene identitetske reprezentacije.

2.2.7 OC-SORT - Rethinking SORT for Robust Multi-Object Tracking

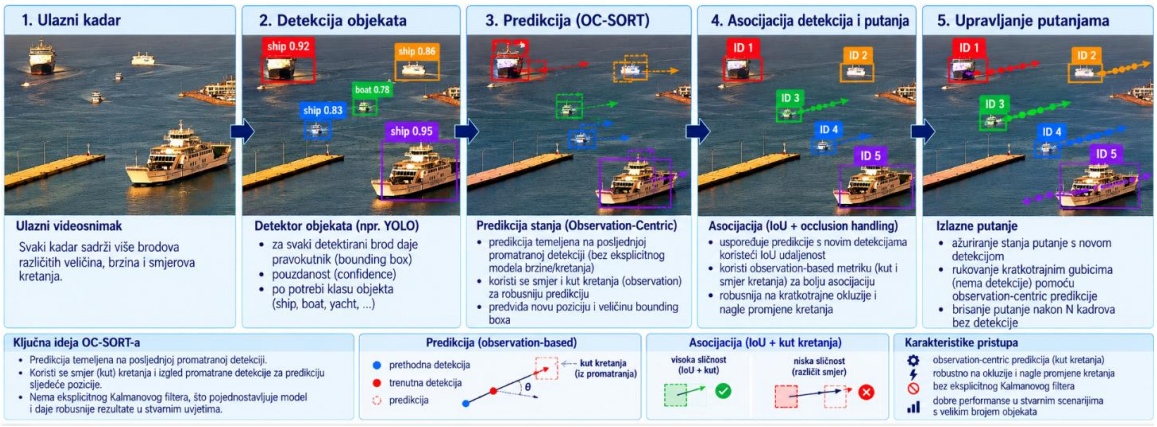
Algoritam OC-SORT (Observation-Centric Simple Online and Realtime Tracking) predstavili su Cao i suradnici 2023. godine u radu Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking [7]. Algoritam je razvijen s ciljem poboljšanja robusnosti praćenja više objekata u uvjetima djelomičnih i potpunih okluzija, promjenjivog gibanja i nepouzdanih predikcija položaja. Za razliku od pristupa koji povećavaju složenost praćenja uvođenjem dodatnih neuronskih mreža za izdvajanje vizualnih značajki, OC-SORT zadržava jednostavnu arhitekturu izvornog SORT-a, ali mijenja način korištenja opažanja i korekcije modela gibanja.

Prvi karakteristični mehanizam predstavlja OCM (Observation-Centric Momentum), čija je svrha povećati pouzdanost asocijacije uključivanjem informacije o smjeru gibanja objekta. Kod standardnog Kalmanova filtra smjer gibanja procjenjuje se iz komponenata brzine koje su dio dinamičkog stanja. Takva procjena može biti osjetljiva na male pogreške položaja, osobito kada je pomak objekta između dvaju uzastopnih kadrova relativno malen. U tim uvjetima čak i mala promjena detekcijskog pravokutnika može uzrokovati značajnu promjenu procijenjenog smjera gibanja.

Informacija o smjeru uključuje se u funkciju troška pridruživanja zajedno s mjerom prostornog preklapanja, najčešće IoU. Na temelju tako oblikovane matrice troškova provodi se globalno pridruživanje postojećih putanja i novih detekcija. OCM time poboljšava postupak asocijacije bez potrebe za zasebnim modelom vizualne reidentifikacije.

Kada se objekt nakon određenog broja kadrova ponovno uspješno poveže s prethodnom putanjom, OC-SORT koristi posljednje opažanje prije gubitka i prvo novo opažanje nakon ponovnog pronalaska kako bi konstruirao virtualnu putanju tijekom razdoblja nevidljivosti. Virtualna putanja predstavlja niz procijenjenih međupoložaja između dviju stvarnih detekcija, dobivenih uz pretpostavku približno linearnog gibanja tijekom promatranog intervala.

ORU se aktivira nakon uspješnog ponovnog pridruživanja izgubljene putanje, a ne tijekom same potpune okluzije. To predstavlja važnu razliku jer mehanizam ne omogućuje neposredno određivanje stvarnog položaja nevidljivog objekta, nego koristi naknadno dostupne informacije za ispravljanje akumulirane pogreške. Njegov doprinos stoga se prvenstveno očituje u stabilizaciji nastavka praćenja nakon ponovnog pronalaska objekta.



Slika 2.10. Observation-Centric SORT - Rethinking SORT for Robust Multi-Object Tracking (Generirano AI).

Takav dodatni postupak povećava mogućnost oporavka putanja nakon kraćih prekida, osobito kada se objekt tijekom okluzije nije značajno udaljio od posljednjeg opaženog položaja. Nakon uspješnog ponovnog povezivanja može se primijeniti ORU kako bi se korigiralo dinamičko stanje i omogućio stabilniji nastavak praćenja.

Najvažnija prednost algoritma OC-SORT proizlazi iz prepoznavanja ograničenja standardnog Kalmanova filtra tijekom prekida opažanja. Umjesto da poboljšanje praćenja temelji prvenstveno na složenijoj vizualnoj reprezentaciji, OC-SORT nastoji učinkovitije iskoristiti informacije o stvarno opaženim položajima objekata. Time pokazuje da se kvaliteta praćenja može značajno poboljšati već odgovarajućim promjenama načina procjene gibanja i pridruživanja detekcija.

Poseban izazov predstavljaju postupne okluzije tijekom kojih se mijenja geometrija detekcijskog pravokutnika. Kada drugo plovilo počne zaklanjati praćeni objekt, detektor može prepoznati samo njegov vidljivi dio. Središte takvog pravokutnika može se postupno pomicati, iako se stvarni položaj i brzina plovila nisu značajno promijenili. Budući da OC-SORT koristi stvarna opažanja za procjenu smjera gibanja, nepouzdana geometrija detekcija može neposredno utjecati na OCM.

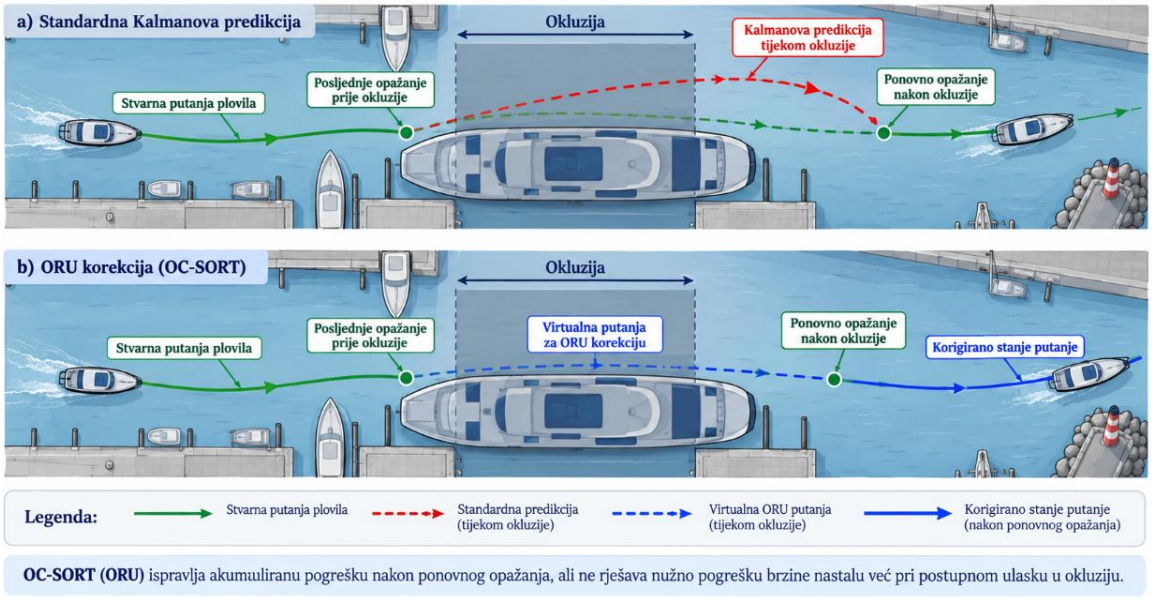
Dodatno ograničenje odnosi se na ORU mehanizam. Virtualna putanja tijekom okluzije rekonstruira se na temelju posljednjeg opažanja prije prekida i prvog uspješno pridruženog opažanja nakon ponovnog pojavljivanja. Takva rekonstrukcija pretpostavlja približno linearno gibanje između tih dviju točaka. Međutim, stvarno gibanje plovila tijekom okluzije može uključivati usporavanje, zaustavljanje, ubrzavanje ili promjenu smjera.

Važno je također istaknuti da ORU zahtijeva uspješno ponovno povezivanje objekta s prethodnom putanjom. Ako algoritam nakon okluzije ne uspije prepoznati novu detekciju kao nastavak prethodno praćenog objekta, ORU se ne može primijeniti na odgovarajući identitet. Drugim riječima, ORU predstavlja mehanizam korekcije stanja nakon uspješnog pridruživanja, ali ne rješava samostalno problem reidentifikacije objekata.

Sa stajališta primjene u pomorskim scenama, OC-SORT predstavlja posebno zanimljivo polazište jer se izravno bavi jednim od ključnih problema praćenja tijekom okluzije: akumulacijom pogrešaka u modelu gibanja. Njegov ORU mehanizam pokazuje da se dinamičko stanje objekta može poboljšati korištenjem opažanja prije i

nakon prekida, dok OCM potvrđuje važnost robusne procjene smjera iz vremenski razdvojenih mjerenja.

U analizi algoritma OC-SORT posebno se izdvaja razlika između standardne predikcije Kalmanovim filtrom i naknadne korekcije stanja primjenom mehanizma ORU. Dok standardni pristup tijekom okluzije nastavlja propagirati stanje na temelju prethodno procijenjene brzine, OC-SORT nakon ponovnog opažanja objekta koristi virtualno rekonstruiranu putanju radi korekcije akumulirane pogreške. Takav pristup može povećati stabilnost nastavka praćenja nakon kraćeg ili srednje dugog prekida opažanja, ali ne rješava u potpunosti problem nepouzdanih detekcija tijekom postupnog ulaska u okluziju. Upravo je ta razlika od posebnog značaja za praćenje plovila, gdje promjena geometrije detekcijskog pravokutnika može narušiti procjenu gibanja i prije potpune zaklonjenosti objekta. Zbog toga sljedeća slika prikazuje konceptualnu razliku između standardne predikcije i ORU korekcije kao jednu od ključnih značajki algoritma OC-SORT.



Slika 2.11. Usporedni prikaz standardne Kalmanove predikcije i ORU korekcije u algoritmu OC-SORT tijekom okluzije plovila: prikazana je stvarna putanja objekta, posljednje opažanje prije okluzije, pogrešna predikcija tijekom zaklonjenosti, novo opažanje nakon ponovnog pojavljivanja te virtualna putanja koja se koristi za korekciju stanja.

2.2.8 Deep OC-SORT - Multi-Pedestrian Tracking by Adaptive Re-Identification

Algoritam Deep OC-SORT (Multi-Pedestrian Tracking by Adaptive Re-Identification) predstavili su Maggiolino i suradnici 2023. godine kao proširenje algoritma OC-SORT s ciljem poboljšanja očuvanja identiteta objekata kombiniranjem informacija o gibanju i dubokih vizualnih značajki [8]. Rad je predstavljen na konferenciji IEEE International Conference on Image Processing (ICIP 2023). Za razliku od izvornog OC-SORT-a, koji se prvenstveno oslanja na prostorne i dinamičke informacije, Deep OC-SORT uvodi adaptivni postupak reidentifikacije kojim se vizualne značajke uključuju u asocijaciju ovisno o njihovoj procijenjenoj pouzdanosti i sposobnosti razlikovanja pojedinačnih objekata.

Deep OC-SORT nastoji riješiti navedeni problem zadržavanjem robusnog modela gibanja algoritma OC-SORT i uvođenjem triju dodatnih komponenti: kompenzacije gibanja kamere (Camera Motion Compensation, CMC), dinamičkog ažuriranja vizualnih značajki (Dynamic Appearance, DA) i adaptivnog određivanja težine vizualne sličnosti (Adaptive Weighting, AW). Takva arhitektura omogućuje da doprinos vizualnih značajki postupku asocijacije ovisi o njihovoj informativnosti, umjesto da bude određen isključivo unaprijed zadanim konstantnim težinskim faktorom.

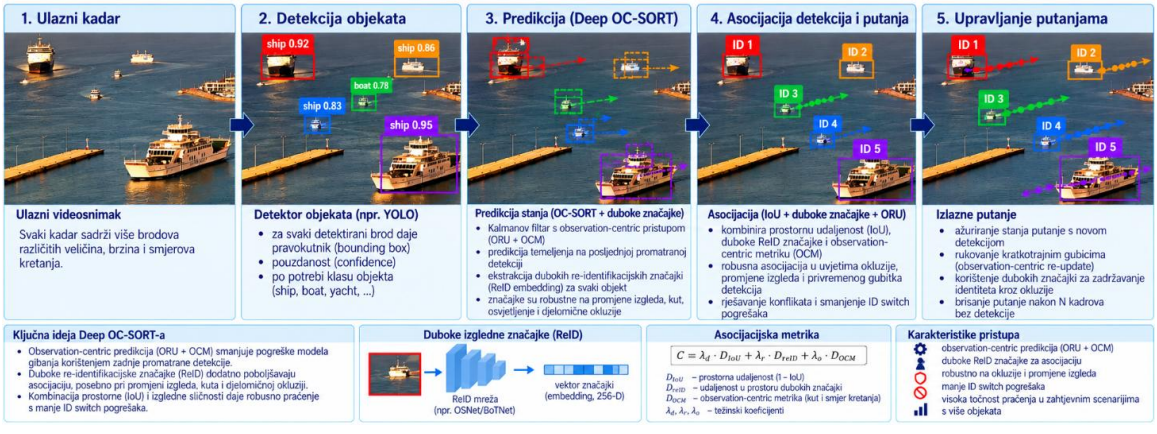
Prva dodatna komponenta algoritma jest kompenzacija gibanja kamere (CMC). Tijekom obrade uzastopnih kadrova promjena položaja objekta može biti posljedica njegova stvarnog gibanja ili pomicanja kamere. Budući da se OC-SORT oslanja na opažene položaje objekata i njihove međusobne vremenske odnose, globalni pomak slike može narušiti procjenu smjera i prostornu podudarnost detekcija.

Druga i posebno važna komponenta jest dinamičko ažuriranje vizualnih značajki (DA). U algoritmima poput StrongSORT-a i BoT-SORT-a za održavanje vizualne reprezentacije putanje često se koristi eksponencijalni pomični prosjek (Exponential Moving Average, EMA). Nova značajka postupno se uključuje u prethodno pohranjeni vektor prema unaprijed određenom težinskom faktoru.

Deep OC-SORT proširuje taj pristup tako da težina novog opažanja ovisi o pouzdanosti pripadajuće detekcije. Osnovna je pretpostavka da detekcije niže pouzdanosti češće odgovaraju situacijama smanjene vidljivosti, zamućenja ili djelomične zaklonjenosti te da izdvojene vizualne značajke u takvim okolnostima mogu biti manje pouzdane.

Treća komponenta algoritma jest adaptivno određivanje težine vizualne sličnosti (AW). U standardnim pristupima koji kombiniraju gibanje i izgled objekta matrica troškova pridruživanja često se formira linearnom kombinacijom prostorne i vizualne udaljenosti, pri čemu je odnos njihovih doprinosa unaprijed zadan. Takav pristup pretpostavlja da vizualne značajke imaju približno jednaku važnost u svim situacijama, što nije nužno opravdano.

Deep OC-SORT zato analizira diskriminativnost vizualnih podudaranja usporedbom najboljih i sljedećih najboljih kandidata u matrici vizualne sličnosti. Za svaku putanju procjenjuje se koliko je najbolja detekcija vizualno podudarnija od ostalih dostupnih detekcija. Analogno tome, za svaku detekciju razmatra se razlika između najbolje i druge najbolje odgovarajuće putanje.



Slika 2.12. Deep OC-SORT - Multi-Pedestrian Tracking by Adaptive Re-Identification (Generirano AI).

Najvažnija prednost algoritma Deep OC-SORT proizlazi iz kombiniranja robusnog modela gibanja s adaptivnim korištenjem vizualnih značajki. Za razliku od izvornog OC-SORT-a, koji ne raspolaže dodatnim vizualnim kriterijem za razlikovanje objekata, Deep OC-SORT omogućuje usporedbu identitetskih reprezentacija tijekom pridruživanja detekcija. Time se povećava mogućnost očuvanja identiteta kada se putanje više objekata međusobno približavaju ili presijecaju.

Ipak, osnovno ograničenje takvog pristupa proizlazi iz pretpostavke da je detekcijska pouzdanost dovoljno dobar pokazatelj kvalitete vizualnih značajki. Ta pretpostavka nije uvijek zadovoljena. Detektor može s visokom pouzdanošću prepoznati djelomično vidljiv objekt, iako pripadajući pravokutnik ne obuhvaća njegovo cjelovito tijelo ili uključuje dijelove susjednog objekta. U takvim situacijama izdvojeni embedding može biti nepouzdan unatoč visokoj detekcijskoj pouzdanosti.

Druga važna prednost Deep OC-SORT-a proizlazi iz adaptivnog određivanja težine vizualne sličnosti. Za razliku od metoda koje unaprijed određuju konstantan doprinos vizualnih i dinamičkih informacija, AW omogućuje veći utjecaj izgleda u situacijama kada vizualne značajke jasno razlikuju pojedine objekte. Time se smanjuje mogućnost da nepouzdana ili nedovoljno diskriminativna vizualna informacija dominira postupkom pridruživanja.

Kod plovila koja iza velike prepreke ostaju zaklonjena tijekom duljeg vremenskog intervala, navedeno ograničenje može biti posebno izraženo. Tijekom razdoblja nevidljivosti Kalmanova predikcija postaje sve manje pouzdana, a nakon ponovnog pojavljivanja plovilo se može nalaziti izvan očekivanog prostornog područja. Ako istodobno postoji više vizualno sličnih kandidata, ni adaptivna reidentifikacija ne mora omogućiti jednoznačno povezivanje s prethodnim identitetom.

U pomorskom okruženju ovaj nedostatak može imati značajne posljedice. Tijekom postupnog ulaska plovila u okluziju nepouzdane detekcije mogu narušiti procjenu njegova položaja i brzine prije nego što objekt potpuno nestane. Ako se takvo stanje ne prepozna, Kalmanov filtar nastavlja ažurirati dinamičko stanje prema opažanjima koja ne odgovaraju stvarnoj geometriji plovila. Nakon početka potpune okluzije daljnja predikcija temelji se na prethodno narušenom stanju, što povećava mogućnost odstupanja od stvarnog gibanja.

2.2.9 VOc-SORT – Vessel Occlusion SORT

Algoritam VOc-SORT (Vessel Occlusion SORT) razvili su Marin i Gotovac s ciljem poboljšanja pouzdanosti praćenja plovila tijekom dugotrajnih okluzija u složenim pomorskim scenama [9]. Za razliku od prethodno opisanih algoritama, koji su prvenstveno razvijeni i evaluirani na skupovima podataka za praćenje osoba i cestovnih vozila, VOc-SORT izravno je prilagođen specifičnostima praćenja plovila, uključujući njihove različite dimenzije, promjene vizualnog izgleda, međusobna preklapanja i dugotrajne prekide vidljivosti. Algoritam se temelji na paradigmi tracking-by-detection, ali uvodi dodatne mehanizme za pridruživanje detekcija i održavanje identiteta tijekom razdoblja u kojima objekt nije vidljiv.

Arhitektura algoritma VOc-SORT temelji se na povezivanju detekcije plovila, izdvajanja vizualnih značajki, predikcije gibanja i dvostupanjskog postupka asocijacije. U izvornom radu za detekciju je korišten model YOLO11m, treniran na skupu podataka SSMOT (Split Ship Multi-Object Tracking), koji sadrži anotirane primjere različitih kategorija plovila u stvarnim lučkim scenama. Detektor u svakom kadru određuje

položaje plovila u obliku detekcijskih pravokutnika i pripadajuće vrijednosti pouzdanosti, koje predstavljaju osnovni ulaz u postupak praćenja.

Vizualne značajke koriste se za usporedbu novih detekcija s prethodno uspostavljenim putanjama. Za svaku putanju održava se agregirana identitetska reprezentacija koja se ažurira eksponencijalnim pomičnim prosjekom (Exponential Moving Average, EMA). U izvornoj implementaciji koristi se faktor izglađivanja od 0,9, čime se prethodno pohranjene značajke kombiniraju s novim opažanjima uz nastojanje smanjenja utjecaja kratkotrajnog šuma i manjih promjena izgleda. Takav pristup omogućuje održavanje relativno stabilnog vizualnog opisa tijekom uzastopnog praćenja.

Središnju inovaciju VOc-SORT-a predstavlja dvostupanjski postupak pridruživanja detekcija postojećim putanjama. Prvi stupanj naziva se vizualno-geometrijska asocijacija (Visual-Geometric association, VG), a drugi vizualno-dinamička asocijacija (Visual-Dynamic association, VD). Njihova je osnovna svrha različito tretirati situacije u kojima postoji dovoljna prostorna podudarnost i situacije u kojima je geometrijska podudarnost narušena, primjerice nakon dugotrajne okluzije.

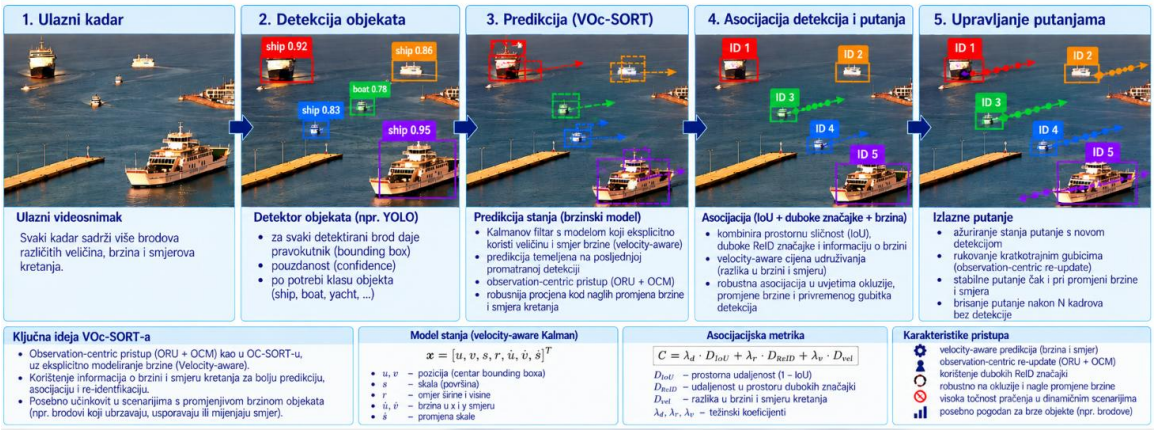
Za svaki par putanje i detekcije oblikuje se funkcija troška koja uključuje obje komponente. Vizualna udaljenost pritom se dodatno ponderira kada su zadovoljeni pragovi vizualne i prostorne kompatibilnosti. U izvornoj implementaciji primjenjuju se prag kosinusne udaljenosti od 0,4, prag IoU udaljenosti od 0,7 i najveći dopušteni ukupni trošak pridruživanja od 0,8. Na temelju izračunate matrice troškova mađarski algoritam određuje optimalno pridruživanje detekcija postojećim putanjama.

Zbog toga VOc-SORT uvodi drugi stupanj pridruživanja, odnosno vizualno-dinamičku asocijaciju. U tom se stupnju ponovno razmatraju putanje i detekcije koje nisu uspješno povezane tijekom prvog stupnja, ali se umjesto IoU udaljenosti koristi Mahalanobisova udaljenost između nove detekcije i predviđenog stanja Kalmanova filtra.

U drugom stupnju funkcija troška kombinira kosinusnu udaljenost vizualnih značajki s Mahalanobisovom udaljenošću. U izvornoj implementaciji vizualnoj se komponenti dodjeljuje dominantna težina od 0,98, dok dinamička komponenta prvenstveno služi kao dodatno ograničenje kompatibilnosti. Za odbacivanje dinamički neprihvatljivih kandidata koristi se prag Mahalanobisove udaljenosti od 9,4877, izveden iz hi-kvadrat distribucije s četiri stupnja slobode i razinom pouzdanosti od 95 %.

VOc-SORT zato povećava dopušteno trajanje putanje bez pridružene detekcije na 625 kadrova. Pri brzini od 25 kadrova u sekundi to odgovara razdoblju od 25 sekundi. Na taj način prethodno uspostavljeni identitet ostaje dostupan postupku reidentifikacije tijekom znatno duljeg vremenskog intervala.

Ako plovilo tijekom pet uzastopnih kadrova nema pridruženu detekciju, a njegovo posljednje opažanje nalazi se blizu lijevog ili desnog ruba slike, algoritam može zaključiti da je plovilo napustilo promatrano područje te ukloniti njegovu putanju prije isteka općeg vremenskog praga. Pritom se iznimka primjenjuje na nepomična plovila, primjerice ona privezana uz obalu, kako se njihovi identiteti ne bi nepotrebno uklanjali zbog privremenog zaklanjanja.



Slika 2.13. VOc-SORT - Model dubokog učenja za praćenje plovila s fokusom na rješavanje problema dugotrajnih okluzija (Generirano AI).

Najvažnija prednost algoritma VOc-SORT proizlazi iz njegove izravne prilagodbe problemu dugotrajnih okluzija plovila. Za razliku od standardnih algoritama koji održavaju izgubljene putanje samo tijekom relativno kratkog razdoblja, VOc-SORT omogućuje njihovo produljeno zadržavanje i ponovno povezivanje nakon višesekundnog prekida opažanja. Time se povećava mogućnost očuvanja identiteta objekata koji privremeno nestaju iza većih plovila ili lučke infrastrukture.

Takva organizacija posebno je značajna tijekom dugotrajnih okluzija jer standardna IoU mjera postaje nepouzdana kada se plovilo nakon ponovnog pojavljivanja nalazi daleko od predviđenog pravokutnika. Korištenjem Mahalanobisove udaljenosti algoritam uzima u obzir nesigurnost predikcije i omogućuje razmatranje kandidata s većim prostornim odstupanjem, uz istodobnu provjeru njihove vizualne sličnosti.

Učinkovitost predloženog pristupa potvrđena je eksperimentalnom evaluacijom na SSMOT skupu podataka. Evaluacija je obuhvatila 18 videosekvenci s ukupno 149 jedinstvenih putanja plovila i 31 anotiranim događajem potpune okluzije. Okluzije su podijeljene na kratke, u trajanju do dvije sekunde, srednje, između dvije i osam sekundi, te dugotrajne, dulje od osam sekundi.

U usporedbi s algoritmima ByteTrack, BoT-SORT, StrongSORT, OC-SORT, Deep OC-SORT i BoostTrack, VOc-SORT ostvario je najviše vrijednosti metrika HOTA, MOTA i IDF1. Konkretno, ostvarene su vrijednosti HOTA od 85,57 %, MOTA od 93,03 % i IDF1 od 94,38 %, uz ukupno 42 zamjene identiteta. Posebno je značajan rezultat reidentifikacije nakon okluzije, pri čemu je algoritam uspješno očuvao identitet u 25 od ukupno 31 anotiranog događaja.

Ako se takva opažanja koriste za ažuriranje Kalmanova filtra, procijenjena brzina i budući položaj mogu postati nepouzdani. Budući da VOc-SORT koristi standardni Kalmanov filtar, pogreške nastale tijekom posljednjih kadrova prije potpune okluzije mogu utjecati na daljnju predikciju tijekom razdoblja nevidljivosti.

Drugo važno ograničenje odnosi se na ažuriranje vizualne reprezentacije. VOc-SORT koristi EMA postupak kojim se novi embedding vektori kombiniraju s prethodno pohranjenim značajkama. Takav pristup omogućuje stabilizaciju reprezentacije, ali pretpostavlja da nova opažanja sadrže dovoljno pouzdane informacije o identitetu objekta.

Produljeno održavanje izgubljenih putanja također povećava broj mogućih kandidata za asocijaciju. U scenama s velikim brojem vizualno sličnih plovila može se dogoditi da se više prethodno praćenih objekata istodobno nalazi u stanju nevidljivosti. Ako se

nova detekcija može povezati s više takvih putanja, kombinacija vizualne i dinamičke udaljenosti ne mora uvijek omogućiti jednoznačno određivanje identiteta.

Sa stajališta istraživanja praćenja plovila tijekom dugotrajnih okluzija, VOc-SORT predstavlja posebno važan doprinos jer potvrđuje da se značajno poboljšanje očuvanja identiteta može postići kombiniranjem domenski prilagođene reidentifikacije, geometrijskih i dinamičkih kriterija te odgovarajućeg upravljanja životnim ciklusom putanja.

Istodobno, analiza algoritma pokazuje da pouzdano ponovno povezivanje plovila nakon okluzije ne podrazumijeva nužno pouzdanu procjenu njegova gibanja i vizualnog identiteta tijekom postupnog ulaska u okluziju. VOc-SORT prvenstveno poboljšava asocijaciju i oporavak putanja nakon prekida vidljivosti, dok se pitanje pravodobnog prepoznavanja geometrijski i vizualno nepouzdanih opažanja može dodatno istražiti.

2.2.10 OcclusionTrack - Multi-Object Tracking in Dense Scenes

Algoritam OcclusionTrack (OCCTrack) predstavili su Chen, Meng i Chen 2025. godine u radu OcclusionTrack: Multi-Object Tracking in Dense Scenes, objavljenom u časopisu Applied Sciences [10]. Algoritam je razvijen s ciljem povećanja pouzdanosti praćenja više objekata u gusto naseljenim scenama, u kojima su međusobna preklapanja česta, a djelomične i potpune okluzije uzrokuju prekide putanja i zamjene identiteta. Za razliku od pristupa koji poboljšanje praćenja temelje prvenstveno na složenijim modelima duboke reidentifikacije, OcclusionTrack nastoji povećati robusnost povezivanjem nekoliko komplementarnih mehanizama koji djeluju na procjenu gibanja, kompenzaciju pomaka kamere, pridruživanje detekcija i ponovno aktiviranje izgubljenih putanja.

OcclusionTrack nastoji ublažiti navedene probleme integracijom četiriju mehanizama: Kalmanova filtra s prilagodbom prema pouzdanosti detekcije (Confidence-Based Kalman Filter, CBKF), kompenzacije gibanja kamere (Camera Motion Compensation, CMC), kaskadnog pridruživanja temeljenog na relativnoj dubini (Depth–Cascade-Matching, DCM) i postupka ponovnog aktiviranja izgubljenih putanja (Re-activate). Algoritam pritom zadržava paradigmu tracking-by-detection i koristi geometrijsku podudarnost kao osnovni kriterij pridruživanja, bez potrebe za zasebnim modelom vizualne reidentifikacije.

Prvi karakteristični mehanizam predstavlja CBKF (Confidence-Based Kalman Filter). Njegova je osnovna svrha smanjiti negativan utjecaj nepouzdanih detekcija na procjenu dinamičkog stanja objekta. U standardnom Kalmanovu filtru kovarijacijska matrica šuma mjerenja određuje koliko se može vjerovati novom opažanju u odnosu na prethodnu predikciju. Ako se koristi konstantna kovarijancija mjerenja, sve detekcije tretiraju se na približno jednak način, neovisno o razlikama u njihovoj pouzdanosti.

Takav pristup posebno je važan tijekom djelomičnih okluzija, kada detektor može prepoznati objekt s manjom sigurnošću ili odrediti njegov pravokutnik s povećanom pogreškom. Umjesto da se takvo opažanje u potpunosti odbaci ili koristi s nepromijenjenom težinom, CBKF omogućuje njegovo uključivanje uz odgovarajuće smanjenje utjecaja na procijenjeno stanje. Time se nastoji spriječiti nagla promjena procijenjene brzine i položaja uzrokovana pojedinačnim nepouzdanim mjerenjem.

Treći mehanizam predstavlja DCM (Depth–Cascade-Matching), koji je namijenjen smanjenju neodređenosti pridruživanja u scenama s velikim brojem međusobno preklopljenih objekata. Kada se više objekata nalazi na približno jednakim položajima

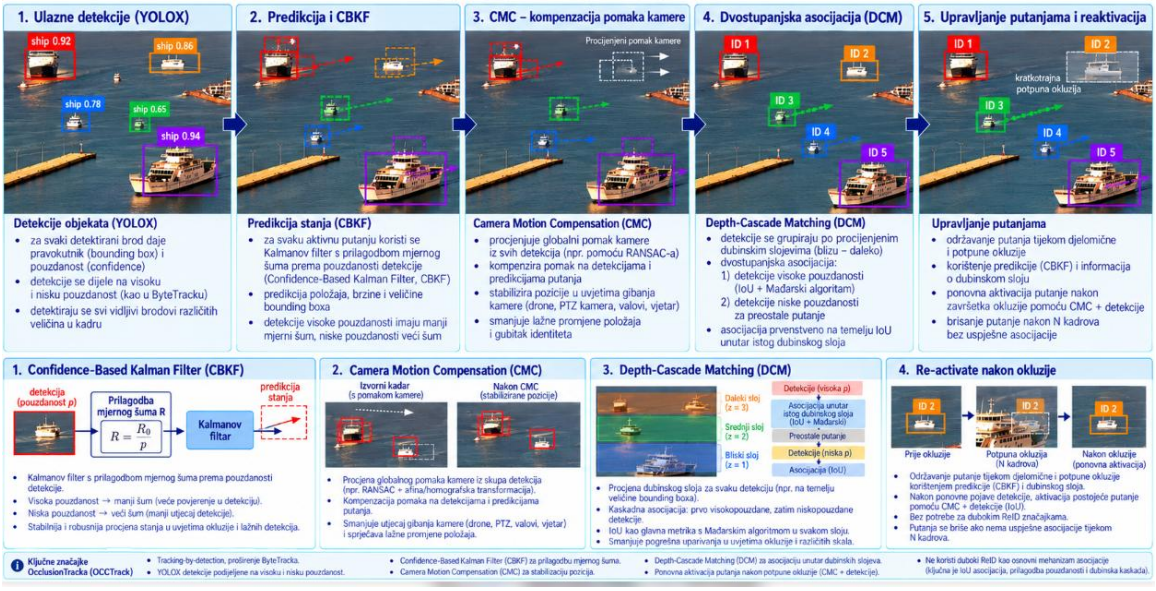
u slici, njihovi se detekcijski pravokutnici mogu značajno preklapati, zbog čega sama IoU mjera ne mora biti dovoljna za pouzdano razlikovanje odgovarajućih putanja.

Na temelju izračunate pseudodubine detekcije i postojeće putanje razvrstavaju se u nekoliko dubinskih razina. Umjesto da se sve putanje i detekcije istodobno pridružuju, DCM provodi kaskadni postupak u kojem se najprije uspoređuju objekti pripadajućih dubinskih razina. Pridruživanje se provodi korištenjem IoU udaljenosti i mađarskog algoritma, dok se nepovezane putanje i detekcije mogu razmatrati u sljedećim koracima kaskade.

U cjelokupnom postupku OcclusionTracka DCM se koristi unutar dvostupanjske asocijacije preuzete iz ByteTracka. U prvom stupnju provodi se pridruživanje postojećih putanja detekcijama visoke pouzdanosti, pri čemu se kandidati organiziraju prema dubinskim razinama. Nakon završetka prvog stupnja preostale se putanje pokušavaju povezati s detekcijama niže pouzdanosti, ponovno uz korištenje DCM postupka.

Četvrta komponenta predstavlja mehanizam ponovnog aktiviranja izgubljenih putanja (CMC-detection-based Re-activate). Njegova je svrha ispraviti akumulirane pogreške modela gibanja nakon potpune okluzije. Tijekom razdoblja u kojem objekt nije vidljiv Kalmanov filtar nastavlja predviđati njegov položaj bez korekcije novim opažanjima. Ako prekid traje dulje ili se objekt tijekom njega giba drugačije od pretpostavljenog modela, procijenjeno stanje može značajno odstupati od stvarnog.

Nakon generiranja virtualnih mjerenja algoritam vraća dinamičko stanje putanje na prethodno pohranjeno stanje i ponovno izvodi korake Kalmanova filtra kroz razdoblje izgubljene vidljivosti. Pri tome se koristi CBKF, tako da rekonstruirane detekcije utječu na procjenu stanja u skladu s interpoliranim vrijednostima pouzdanosti. Dobiveno stanje predstavlja korigiranu procjenu položaja i gibanja objekta u trenutku njegova ponovnog pojavljivanja.



Slika 2.14. OcclusionTrack - Multi-Object Tracking in Dense Scenes (Generirano AI).

Najvažnija prednost algoritma OcclusionTrack proizlazi iz sustavnog pristupa obradi različitih posljedica okluzije. Za razliku od metoda koje se oslanjaju na jedan dominantan mehanizam, poput poboljšane predikcije gibanja ili vizualne

reidentifikacije, OcclusionTrack kombinira više postupaka koji djeluju na različite faze praćenja. CBKF nastoji smanjiti utjecaj nepouzdanih detekcija, CMC poboljšava prostornu usklađenost između kadrova, DCM ograničava neodređenost pridruživanja u preklopljenim scenama, a Re-activate omogućuje korekciju akumuliranih pogrešaka nakon ponovnog pojavljivanja objekta.

Primjerice, pri postupnom prolasku manje brodice iza većeg plovila detektor može nastaviti pouzdano prepoznavati njezin vidljivi pramac, dok se središte detekcijskog pravokutnika postupno pomiče prema preostalom vidljivom dijelu. Ako detekcijska pouzdanost ostane visoka, CBKF može takvo mjerenje tretirati kao pouzdano, iako ono ne predstavlja odgovarajuću geometrijsku informaciju za procjenu stvarnog gibanja brodice.

Ako se pogrešna transformacija primijeni na sve postojeće putanje, može doći do sustavnog pomicanja njihovih predviđenih položaja. Posljedično se narušava IoU podudarnost i povećava mogućnost pogrešnog pridruživanja. Dodatni problem nastaje pri izraženim promjenama perspektive ili složenom trodimenzionalnom gibanju kamere, koje nije moguće u potpunosti opisati jednostavnom globalnom dvodimenzionalnom transformacijom.

Kod plovila donji rub detekcijskog pravokutnika ne mora odgovarati stabilnoj kontaktnoj točki s morskom površinom. Njegov položaj može se mijenjati zbog valova, nagiba plovila, različite visine nadvodnog dijela trupa i perspektive kamere. Tijekom djelomične okluzije donji dio plovila može potpuno nestati, zbog čega izračunata pseudodubina može odstupati od stvarnog prostornog odnosa između plovila.

Takvo ograničenje posebno je izraženo pri praćenju plovila u luci, gdje se manevriranje često odvija pri malim brzinama, uz povremeno zaustavljanje ili promjenu smjera. U takvim situacijama prostorna interpolacija može proizvesti putanju koja povezuje početni i završni položaj, ali ne opisuje stvarno gibanje tijekom razdoblja nevidljivosti.

Temeljno ograničenje OcclusionTracka proizlazi iz činjenice da se pri pridruživanju koristi IoU udaljenost bez zasebnog modela vizualnih značajki. Iako DCM smanjuje neodređenost pridruživanja prema relativnoj dubini, u scenama s više objekata sličnih položaja, dimenzija i smjerova gibanja geometrijske informacije mogu biti nedovoljne za očuvanje identiteta.

Sa stajališta istraživanja praćenja plovila tijekom dugotrajnih okluzija, OcclusionTrack je posebno zanimljiv jer prepoznaje potrebu za različitim postupcima obrade djelomičnih i potpunih okluzija. CBKF smanjuje utjecaj opažanja niže pouzdanosti tijekom djelomične zaklonjenosti, dok Re-activate korigira akumulirane pogreške nakon ponovnog pojavljivanja objekta. Istodobno, DCM pokazuje mogućnost korištenja kontekstnih prostornih odnosa za poboljšanje asocijacije.

2.2.11 OA-SORT – Occlusion-Aware SORT: Observing Occlusion for Robust Multi-Object Tracking

Algoritam OA-SORT (Occlusion-Aware SORT) predstavili su Li i suradnici 2026. godine u radu Occlusion-Aware SORT: Observing Occlusion for Robust Multi-Object Tracking, objavljenom na konferenciji IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2026) [11]. Algoritam je razvijen s ciljem povećanja pouzdanosti praćenja više objekata u uvjetima međusobnog preklapanja i djelomičnih okluzija, pri čemu se posebna pozornost posvećuje utjecaju zaklonjenosti na geometriju detekcijskih pravokutnika, pridruživanje opažanja i procjenu gibanja. Za razliku od algoritama koji okluzije prvenstveno obrađuju produljenim zadržavanjem putanja ili

naknadnom korekcijom stanja, OA-SORT uvodi izravnu procjenu stupnja zaklonjenosti objekta i koristi je tijekom samog postupka praćenja.

OA-SORT nastoji ublažiti navedene probleme integracijom triju komplementarnih mehanizama: modula za procjenu zaklonjenosti (Occlusion-Aware Module, OAM), korekcije postupka pridruživanja prema stupnju okluzije (Occlusion-Aware Offset, OAO) i prilagodljivog ažuriranja modela gibanja (Bias-Aware Momentum, BAM). Posebnost predloženog pristupa jest činjenica da navedeni mehanizmi ne zahtijevaju dodatno treniranje neuronske mreže, nego se temelje na geometriji detekcijskih pravokutnika, procijenjenim prostornim odnosima i postojećim informacijama sustava praćenja.

Za razliku od standardnih algoritama, kod kojih se postupak pridruživanja temelji prvenstveno na prostornoj podudarnosti detekcijskih pravokutnika, OA-SORT prethodno procjenjuje njihovu međusobnu zaklonjenost. Na temelju te procjene prilagođava se kompatibilnost pojedinih kandidata za pridruživanje, ali i način na koji nova opažanja utječu na ažuriranje dinamičkog stanja.

Kako bi se smanjio taj problem, OA-SORT uvodi Gaussovu prostornu mapu (Gaussian Map, GM). Ona pojedinim dijelovima detekcijskog pravokutnika dodjeljuje različite težine, pri čemu područja bliža središtu objekta imaju veći doprinos procjeni zaklonjenosti od rubnih područja. Na taj se način nastoji smanjiti utjecaj pozadine i perifernih dijelova pravokutnika koji ne moraju pripadati stvarnom objektu.

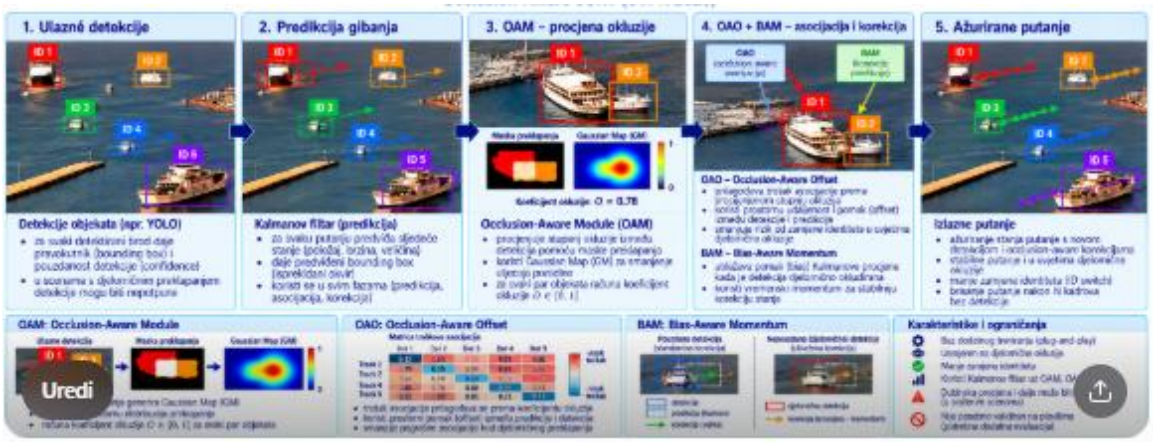
OAO zato uključuje procijenjeni stupanj zaklonjenosti u funkciju prostorne kompatibilnosti. Osnovna je ideja da se pri ocjenjivanju kandidata za pridruživanje ne koristi samo međusobno preklapanje pravokutnika, nego i informacija o tome koliko je pojedina putanja izložena mogućoj okluziji. Na taj se način nastoji ublažiti neodređenost koja nastaje kada je geometrija detekcije promijenjena zbog zaklanjanja.

Treća komponenta algoritma jest BAM (Bias-Aware Momentum), čija je svrha smanjiti utjecaj geometrijski nepouzdanih detekcija na ažuriranje Kalmanova filtra. Za razliku od OAO mehanizma, koji djeluje na postupak pridruživanja, BAM izravno utječe na oblikovanje mjerenja koje se koristi za korekciju procijenjenog dinamičkog stanja.

Nasuprot tome, kada je objekt snažno zaklonjen ili kada nova detekcija značajno odstupa od očekivanog položaja, težina novog opažanja smanjuje se. Tada se mjerenje koje se prosljeđuje Kalmanovu filtru oblikuje kao kombinacija stvarne detekcije i prethodne predikcije, pri čemu veći doprinos dobiva predviđeno stanje.

Time se nastoji postići pouzdanije ažuriranje dinamičkog stanja u situacijama kada detekcijska pouzdanost sama po sebi nije dovoljan pokazatelj točnosti položaja. Posebno se smanjuje mogućnost da promjena vidljivog dijela objekta bude pogrešno interpretirana kao nagla promjena njegove brzine ili smjera gibanja.

Eksperimentalna evaluacija provedena je na skupovima podataka DanceTrack, SportsMOT i MOT17, koji obuhvaćaju različite obrasce gibanja, gustoću objekata i učestalost međusobnih okluzija. Na testnom skupu DanceTrack OA-SORT ostvario je vrijednosti HOTA od 63,1 % i IDF1 od 64,2 %. Dodatno je pokazano da integracija predloženih mehanizama u četiri druga algoritma praćenja u prosjeku povećava HOTA za 2,08, a IDF1 za 3,05 postotnih bodova. Rezultati upućuju na važnost eksplicitnog razmatranja zaklonjenosti tijekom asocijacije i procjene gibanja.



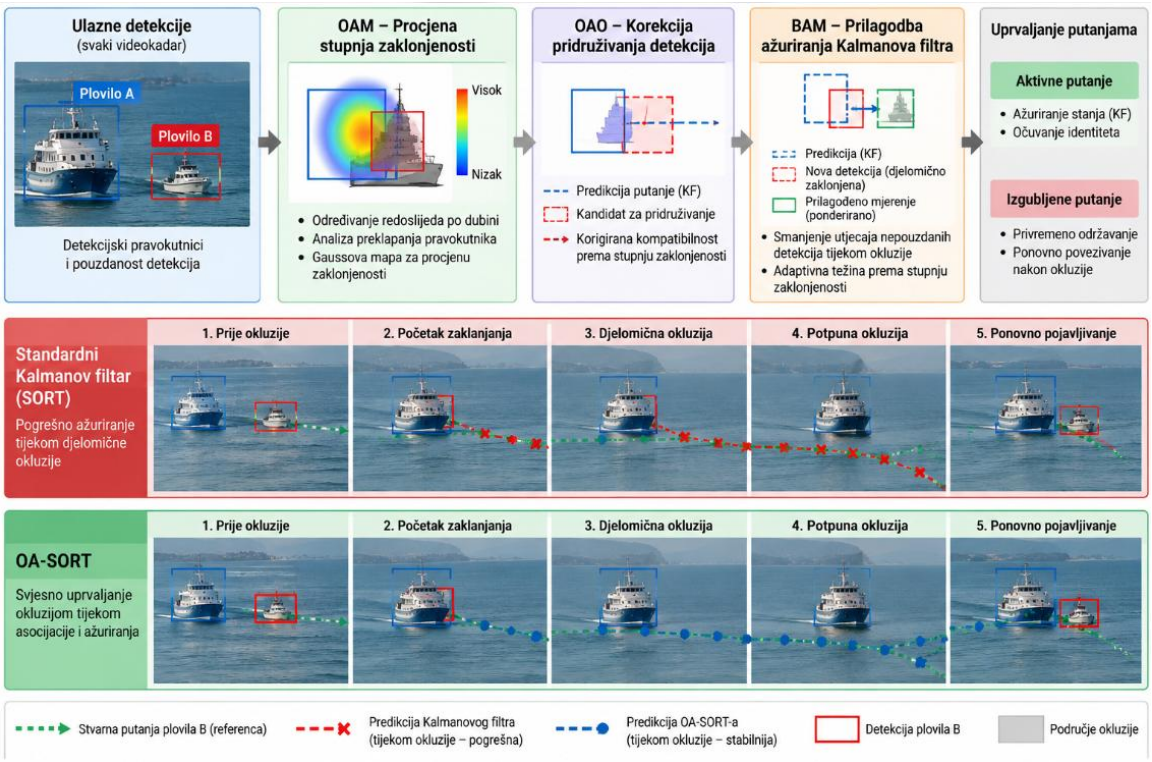
Slika 2.15. Occlusion-Aware SORT: Observing Occlusion for Robust Multi-Object Tracking (Generirano AI).

Najvažnija prednost algoritma OA-SORT proizlazi iz eksplicitnog korištenja informacije o stupnju okluzije tijekom praćenja. Za razliku od prethodno razmatranih algoritama koji se prvenstveno oslanjaju na prostornu kompatibilnost, vizualnu reidentifikaciju ili naknadnu korekciju gibanja, OA-SORT nastoji prepoznati utjecaj zaklonjenosti u trenutku kada se ona pojavljuje. Takav pristup omogućuje prilagodbu asocijacije i ažuriranja modela gibanja prije nego što pogreške detekcija uzrokuju značajnije odstupanje procijenjenog stanja.

U pomorskom okruženju upravo takva situacija predstavlja jedan od najzahtjevnijih problema. Primjerice, mala brodica koja polako ulazi iza velikog putničkog plovila može tijekom više uzastopnih kadrova imati sve manju vidljivu površinu. Njezini detekcijski pravokutnici mogu se postupno sužavati i pomicati, iako se stvarna putanja brodice ne mijenja. Standardni Kalmanov filtar takve promjene može interpretirati kao promjenu položaja i brzine, što povećava pogrešku predikcije tijekom kasnije potpune zaklonjenosti.

Ovo je ograničenje posebno izraženo u pomorskom okruženju. U scenama s plovilima različitih dimenzija položaj donjeg ruba pravokutnika ne mora pouzdano odražavati stvarnu udaljenost od kamere. Veliko plovilo na većoj udaljenosti može imati sličan vertikalni položaj detekcijskog pravokutnika kao manja brodica koja se nalazi bliže kameri. Istodobno, nadgrađe, jarboli i drugi elementi plovila mogu značajno utjecati na ukupnu geometriju pravokutnika.

Posebno je važno razlikovati stvarnu vidljivu površinu objekta od površine preklapanja detekcijskih pravokutnika. Čak i kada su dva pravokutnika značajno preklopljena, stvarna okluzija može biti relativno mala ako se preklapanje odvija kroz područja koja ne pripadaju objektima. Nasuprot tome, plovilo može biti gotovo potpuno zaklonjeno, dok detektor prepoznaje samo mali vidljivi dio i generira pravokutnik koji ne opisuje njegov prethodni oblik.

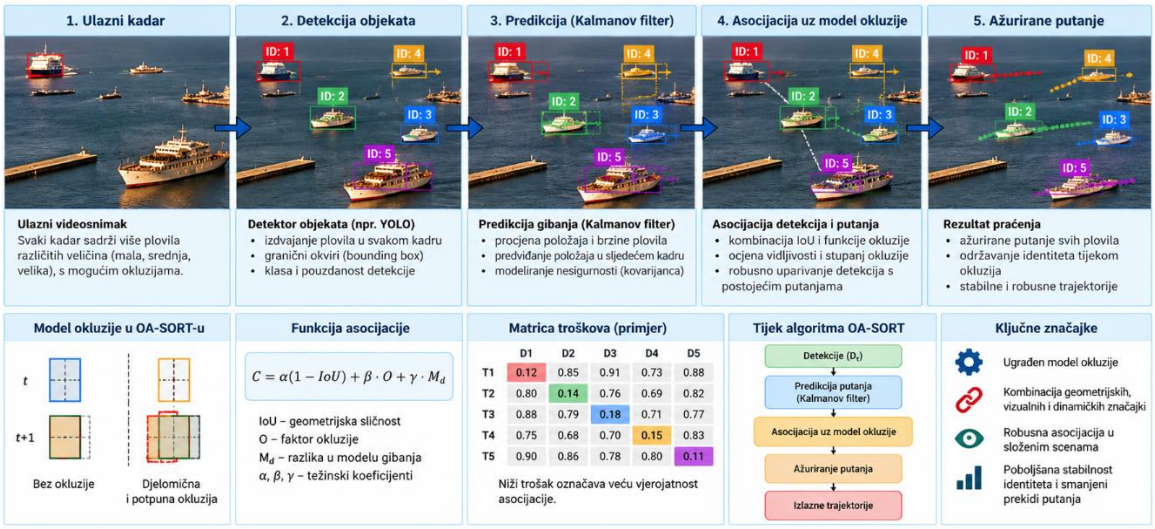


Slika 2.16. Arhitektura algoritma OA-SORT i usporedba predikcije gibanja tijekom okluzije primjenom standardnog Kalmanova filtra i pristupa svjesnog okluzije (Generirano AI).

Važno ograničenje algoritma odnosi se i na dugotrajne potpune okluzije. Kada objekt potpuno nestane iz vidnog polja, nove detekcije više nisu dostupne, a procjena gibanja nastavlja se na temelju prethodno pohranjenog stanja. OAM, OAO i BAM prvenstveno djeluju tijekom razdoblja u kojima su raspoloživi prostorni odnosi i opažanja objekata. Oni ne predstavljaju zaseban dugoročni mehanizam reidentifikacije koji bi mogao pouzdano obnoviti identitet objekta nakon proizvoljno dugog prekida vidljivosti.

Sa stajališta istraživanja pouzdanog praćenja plovila tijekom dugotrajnih okluzija, OA-SORT predstavlja jedan od posebno važnih suvremenih pristupa jer eksplicitno povezuje procjenu zaklonjenosti s asocijacijom detekcija i ažuriranjem Kalmanova filtra. Njegov doprinos potvrđuje da okluziju nije dovoljno promatrati samo kao razdoblje tijekom kojeg detekcija nije dostupna, nego kao proces koji može postupno narušavati geometriju opažanja i dinamičko stanje objekta.

U gornjem dijelu slike prikazane su glavne komponente algoritma OA-SORT: procjena stupnja zaklonjenosti pomoću modula OAM (Occlusion-Aware Module), prilagodba pridruživanja detekcija pomoću OAO (Occlusion-Aware Offset) i korekcija ažuriranja Kalmanova filtra primjenom BAM (Bias-Aware Momentum) mehanizma. Donji dio slike ilustrira konceptualnu razliku između standardnog pristupa SORT i algoritma OA-SORT tijekom postupnog zaklanjanja manjeg plovila većim. Standardni Kalmanov filtar može pogrešno interpretirati promjene detekcijskog pravokutnika kao promjene gibanja, što dovodi do odstupanja predviđene putanje. OA-SORT koristi procijenjeni stupanj zaklonjenosti kako bi prilagodio postupak pridruživanja i smanjio utjecaj geometrijski nepouzdanih opažanja na procjenu gibanja, čime nastoji očuvati stabilniju putanju tijekom okluzije. Prikaz je konceptualan i ne predstavlja eksperimentalno izmjerene putanje.



Slika 2.17. Autorski shematski prikaz prema načelima algoritma OA-SORT [11].

2.3 Usporedba prema komponentama i okluzijama

Tablica 2.1. Usporedba MOT algoritama prema komponentama i rješavanju problema okluzije

Metoda	Vizualni ReID	Ključni mehanizam	Rizik pri dugoj okluziji
SORT	Ne	KF + IoU	Gubitak identiteta
DeepSORT	Da	Appearance cascade	Kontaminirani embedding
FairMOT	Da	Zajednička detekcija+ReID	Mali objekti / centar
ByteTrack	Ne (izvorno)	Nisko pouzdane detekcije	Pogrešna asocijacija
StrongSORT	Da	Jači ReID, EMA, CMC	Drift izgleda
BoT-SORT	Opcionalno / ReID varijanta	CMC + dvostupanjska asocijacija	Loša CMC na moru
OC-SORT	Ne (izvorno)	Observation-centric motion	Pristran bbox
Deep OC-SORT	Da	Adaptive ReID + OC	Slični objekti
VOc-SORT	Da	Dvostupanjska asocijacija	Ovisno o kvaliteti značajki
OcclusionTrack	Ovisno o izvedbi	CBKF + DCM + reactivation	Domenski prijenos
OA-SORT	Neobvezan / plug-in	OAM + OAO + BAM	Kalibracija na plovilima

2.4 Evaluacijske metrike za praćenje više objekata i njihova ograničenja

Objektivna usporedba algoritama za praćenje više objekata zahtijeva primjenu evaluacijskih metrika kojima se mogu kvantificirati točnost detekcije, kontinuitet putanja i očuvanje identiteta objekata. Za razliku od detekcije objekata u pojedinačnoj slici, gdje se uspješnost najčešće izražava metrikama preciznosti, odziva i srednje prosječne preciznosti (mAP), praćenje više objekata uključuje i vremensku dimenziju. Stoga je, osim ispravnog određivanja položaja objekta, potrebno procijeniti uspješnost povezivanja njegovih uzastopnih opažanja te sposobnost održavanja izvornog identiteta tijekom cijelog videoslijeda. U literaturi se za tu svrhu najčešće koriste metrike MOTA (*Multiple Object Tracking Accuracy*), IDF1 (*Identification F1 Score*) i HOTA (*Higher Order Tracking Accuracy*). Svaka od njih vrednuje različite aspekte praćenja, zbog čega njihova zajednička primjena omogućuje potpuniju procjenu učinkovitosti algoritma.

2.4.1 MOTA – Multiple Object Tracking Accuracy

MOTA je jedna od najraširenijih evaluacijskih metrika za praćenje više objekata, uvedena u okviru CLEAR MOT metodologije [12]. Njezina je osnovna svrha objediniti tri vrste pogrešaka koje nastaju tijekom praćenja: propuštene objekte, lažno detektirane objekte i zamjene identiteta. Metrika se definira izrazom:

$$MOTA = 1 - \frac{\sum_{t=1}^{T}(FN_t + FP_t + IDSW_t)}{\sum_{t=1}^{T} GT_t} \tag{2.1}$$

gdje su:

- **(FN_t) (False Negatives)** – broj stvarnih objekata koje sustav nije uspješno detektirao ili pridružio u kadru (t);
- **(FP_t) (False Positives)** – broj pogrešno detektiranih objekata koji nemaju odgovarajući referentni objekt;
- **(IDSW_t) (Identity Switches)** – broj zamjena identiteta tijekom pridruživanja objekata postojećim putanjama;
- **(GT_t) (Ground Truth)** – broj stvarnih, referentno anotiranih objekata u kadru (t);
- **(T)** – ukupan broj evaluiranih kadrova.

Vrijednost MOTA jednaka je 1, odnosno 100 %, kada sustav ne ostvaruje nijednu od navedenih pogrešaka. S povećanjem broja pogrešaka vrijednost se smanjuje, a može postati i negativna ako njihov ukupan broj premašuje broj referentnih pojavljivanja objekata.

Prednost metrike MOTA proizlazi iz jednostavnosti interpretacije i mogućnosti usporedbe različitih algoritama na istom evaluacijskom skupu. Međutim, njezina je značajna karakteristika da se pogreške FN, FP i IDSW zbrajaju s jednakim doprinosom. Time se ne uzima u obzir da pojedine vrste pogrešaka mogu imati bitno različite posljedice za primjenu sustava. Primjerice, zamjena identiteta dvaju plovila tijekom dugotrajne okluzije može biti znatno ozbiljnija od pojedinačne lažne detekcije, premda obje pogreške u izrazu za MOTA imaju jednaku težinu.

Dodatno ograničenje proizlazi iz činjenice da je broj zamjena identiteta često mnogo manji od ukupnog broja detekcijskih pogrešaka i referentnih pojavljivanja objekata.

Posljedično, algoritam može ostvariti vrlo visoku vrijednost MOTA čak i kada ne uspijeva pouzdano održavati identitete kroz dulje vremensko razdoblje.

Primjerice, ako videoslijed sadrži ukupno 10 000 referentnih pojavljivanja objekata, uz 100 propuštenih detekcija, 50 lažnih detekcija i 20 zamjena identiteta, dobiva se:

$$MOTA = 1 - \frac{100+50+20}{10000} = 0{,}983 \tag{2.2}$$

odnosno MOTA = 98,3 %.

Iako navedena vrijednost upućuje na visoku ukupnu uspješnost, iz nje nije moguće zaključiti jesu li identiteti pojedinih plovila bili pravilno očuvani nakon okluzije. Sustav može zamijeniti identitete dvaju brodova, nastaviti ih pogrešno pratiti tijekom velikog broja kadrova, a da se pritom broj evidentiranih zamjena identiteta ne povećava razmjerno trajanju pogrešnog pridruživanja.

Zbog toga MOTA predstavlja važnu mjeru ukupne točnosti praćenja, ali nije dovoljna za samostalnu procjenu kontinuiteta identiteta.

2.4.2 IDF1 – Identification F1 Score

Za razliku od metrike MOTA, IDF1 usmjerena je na ispravnost očuvanja identiteta objekata tijekom videoslijeda. Uvedena je kao dio metodologije evaluacije praćenja temeljene na identitetima, pri čemu se procjenjuje vremenska usklađenost predviđenih putanja s referentnim putanjama [14].

IDF1 predstavlja harmonijsku sredinu identifikacijske preciznosti (IDP) i identifikacijskog odziva (IDR):

$$IDF1 = \frac{2(IDP \cdot IDR)}{IDP + IDR} \tag{2.3}$$

Identifikacijska preciznost i odziv određuju se izrazima:

$$IDP = \frac{ID_{TP}}{ID_{TP} + ID_{FP}} \tag{2.4}$$

$$IDR = \frac{ID_{TP}}{ID_{TP} + ID_{FN}} \tag{2.5}$$

gdje su:

- **IDTP (Identity True Positives)** – broj ispravno identificiranih pojavljivanja objekata u okviru optimalnog povezivanja referentnih i predviđenih putanja;
- **IDFP (Identity False Positives)** – broj predviđenih pojavljivanja koja nisu ispravno povezana s referentnim identitetima;
- **IDFN (Identity False Negatives)** – broj referentnih pojavljivanja objekata koja nisu ispravno identificirana.

Iz navedenih izraza može se dobiti izravni oblik metrike:

$$IDF1 = \frac{2 \cdot IDTP}{2 \cdot IDTP + IDFP + IDFN} \tag{2.6}$$

Temeljna prednost metrike IDF1 jest vrednovanje identitetske dosljednosti na razini vremenskih putanja, a ne samo pojedinačnih događaja zamjene identiteta. Ako algoritam nakon okluzije dodijeli plovilu pogrešan identitet i nastavi ga koristiti kroz

veći broj kadrova, IDF1 će takvu pogrešku odraziti kroz smanjenje broja ispravno identificiranih pojavljivanja.

Ova je značajka osobito važna u pomorskim scenama, gdje su plovila često vizualno slična, a dugotrajne okluzije mogu dovesti do trajne zamjene identiteta. U takvim slučajevima IDF1 pruža informaciju koja nije dovoljno izražena u vrijednosti MOTA.

Ipak, IDF1 također ima određena ograničenja. Budući da se temelji na globalnom povezivanju referentnih i predviđenih identiteta, može nedovoljno razlikovati pojedine vrste lokalnih pogrešaka pridruživanja. Nadalje, ne razdvaja izravno kvalitetu detekcije od kvalitete asocijacije, što otežava utvrđivanje uzroka smanjenja ukupne vrijednosti.

2.4.3 HOTA – Higher Order Tracking Accuracy

HOTA je novija evaluacijska metrika razvijena s ciljem uravnoteženog vrednovanja detekcije i asocijacije objekata [13]. Za razliku od metrike MOTA, koja zbraja različite vrste pogrešaka, te IDF1, koja prvenstveno vrednuje identitetsku usklađenost, HOTA nastoji istodobno procijeniti koliko su objekti uspješno detektirani i koliko su njihove detekcije pravilno povezane tijekom vremena.

Osnovna interpretacija metrike HOTA temelji se na geometrijskoj sredini dviju komponenti:

$$HOTA_\alpha = \sqrt{DetA_\alpha \cdot AssA_\alpha} \tag{2.7}$$

gdje su:

- **DetA (Detection Accuracy)** – mjera uspješnosti detekcije objekata;
- **AssA (Association Accuracy)** – mjera uspješnosti vremenskog povezivanja detekcija s odgovarajućim identitetima;
- **(α)** – prag prostornog podudaranja između predviđenih i referentnih objekata.

DetA se određuje na temelju broja točno pozitivnih, lažno pozitivnih i lažno negativnih detekcija:

$$DetA_\alpha = \frac{TP_\alpha}{TP_\alpha + FN_\alpha + FP_\alpha} \tag{2.8}$$

Komponenta AssA izračunava se kao prosječna asocijacijska točnost pravilno uparenih detekcija. Za pojedino uparivanje razmatra se koliko su često referentna i predviđena putanja ispravno povezane, a koliko je njihov kontinuitet narušen pogrešnim povezivanjima ili propuštenim pridruživanjima. Time se kvaliteta asocijacije ocjenjuje tijekom trajanja putanje, a ne samo u trenutku promjene identifikatora.

Konačna vrijednost HOTA dobiva se usrednjavanjem rezultata kroz više pragova podudaranja:

$$HOTA = \frac{1}{|A|} \sum_{\alpha \in A} HOTA_\alpha \tag{2.9}$$

gdje |A| je broj pragova u skupu.

Važna je prednost ove metrike mogućnost zasebnog prikazivanja komponenti DetA i AssA. Na taj se način može utvrditi je li slabiji rezultat praćenja posljedica nepouzdane detekcije ili pogrešnog povezivanja identiteta.

Primjerice, algoritam može pouzdano detektirati gotovo sva plovila, ali ih nakon međusobnog preklapanja pogrešno povezivati. U tom slučaju DetA može ostati visoka, dok AssA upućuje na problem održavanja identiteta. Obrnuto, algoritam može ostvariti stabilnu asocijaciju objekata koje uspije detektirati, ali propuštati velik broj malih ili udaljenih plovila.

Takvo razdvajanje posebno je korisno pri evaluaciji sustava koji kombiniraju neuronske mreže za detekciju, algoritme predikcije gibanja i modele reidentifikacije.

2.4.4 Usporedba metrika i ograničenja u pomorskim scenama

Analiza metrika MOTA, IDF1 i HOTA pokazuje da nijedna od njih samostalno ne obuhvaća sve aspekte pouzdanosti praćenja više objekata. Njihove razlike proizlaze ponajprije iz načina vrednovanja detekcijskih pogrešaka, kontinuiteta identiteta i uspješnosti povezivanja opažanja.

Tablica 2.2. Prikaz metrika, njihovo vrednovanje i ograničenja.

Metrika	**Primarno vrednovanje**	**Glavno ograničenje**
MOTA	Detekcijske pogreške i broj zamjena identiteta	Nedovoljno izražava trajanje pogrešnog identiteta
IDF1	Vremenska dosljednost identiteta	Ne razdvaja jasno detekciju i asocijaciju
HOTA	Uravnotežena kvaliteta detekcije i asocijacije	Zahtijeva detaljniju interpretaciju sastavnih pokazatelja
AssA	Kvaliteta vremenskog pridruživanja	Ne predstavlja samostalnu mjeru ukupne detekcijske uspješnosti
IDSW	Broj događaja zamjene identiteta	Ne pokazuje trajanje ni posljedice pogrešnog pridruživanja

U sustavima za praćenje plovila posebnu pozornost potrebno je posvetiti dugotrajnim okluzijama i vizualnoj sličnosti objekata. Tijekom postupnog ulaska u okluziju detekcijski pravokutnik može mijenjati položaj i dimenzije, iako stvarna putanja plovila ostaje nepromijenjena. Takva promjena može narušiti procjenu brzine i smjera gibanja te postupno povećavati pogrešku predikcije. Istodobno se vizualni vektor značajki može mijenjati zbog zaklanjanja dijela plovila ili povećanog utjecaja pozadine.

U slučaju potpune okluzije detekcija privremeno izostaje, a algoritam se oslanja na prethodno procijenjeno stanje i raspoložive vizualne značajke. Nakon ponovne pojave plovila ključno je utvrditi može li sustav obnoviti prethodni identitet ili će uspostaviti novu putanju, odnosno pogrešno pridružiti identitet drugom plovilu.

Standardne MOT metrike omogućuju kvantificiranje posljedica tih događaja, ali ne opisuju izravno sve njihove uzroke. Zbog toga je pri evaluaciji algoritama u pomorskom okruženju opravdano rezultate MOTA, IDF1 i HOTA nadopuniti brojem zamjena identiteta, analizom AssA komponente te zasebnim pokazateljima uspješnosti oporavka identiteta nakon okluzija različitog trajanja.

Posebno je korisno razlikovati postupnu djelomičnu okluziju, potpunu okluziju i ponovno pojavljivanje plovila, jer svako od tih stanja uzrokuje različite vrste pogrešaka u detekciji, predikciji i asocijaciji.

Takav pristup omogućuje kritičku usporedbu algoritama ne samo prema njihovoj ukupnoj točnosti, nego i prema sposobnosti očuvanja identiteta u zahtjevnim operativnim uvjetima. Ujedno pokazuje važnost odgovarajućih referentnih skupova podataka koji, osim anotacija položaja i identiteta objekata, sadrže raznolike scenarije preklapanja, prekida vidljivosti i ponovnog pojavljivanja plovila. Zbog toga se u jednom od sljedećih poglavlja razmatraju dostupni skupovi podataka za detekciju, reidentifikaciju i praćenje plovila te njihova prikladnost za objektivnu evaluaciju navedenih problema.

2.5 Zaključak

Provedeni pregled algoritama za praćenje više objekata pokazuje da, unatoč značajnom napretku u području detekcije, predikcije gibanja i reidentifikacije, pouzdano očuvanje identiteta tijekom dugotrajnih okluzija i dalje predstavlja otvoren istraživački problem. Posebno su zahtjevni scenariji s vizualno sličnim plovilima, postupnim ulaskom u okluziju i promjenama detekcijskih pravokutnika koje mogu uzrokovati pogreške u procjeni gibanja i izdvojenim vizualnim značajkama. Usporedba razmatranih algoritama pokazuje da pojedini pristupi različito obrađuju navedene probleme, pri čemu se sve više naglašava potreba za kombiniranjem geometrijskih, dinamičkih i vizualnih informacija. Istodobno, analiza evaluacijskih metrika MOTA, IDF1 i HOTA pokazuje da ukupna točnost praćenja ne mora nužno odražavati sposobnost očuvanja identiteta, zbog čega je potrebna višekriterijska evaluacija, osobito u uvjetima dugotrajnih okluzija. Navedene spoznaje upućuju na važnost integracije različitih algoritamskih komponenti u cjelovite sustave koji omogućuju kontinuiranu obradu videosignala i prilagodbu specifičnostima promatranog okruženja. Stoga se u sljedećem poglavlju razmatra NVIDIA DeepStream SDK kao programski okvir koji omogućuje povezivanje detekcije, praćenja i reidentifikacije objekata te njihovo učinkovito izvođenje u sustavima inteligentne videoanalitike.

3 NVIDIA DEEPSTREAM SDK: ARHITEKTURA, KOMPONENTE I KONFIGURACIJA SUSTAVA ZA INTELIGENTNU VIDEOANALITIKU

Pregled algoritama za praćenje više objekata u prethodnom poglavlju pokazao je da pouzdanost praćenja ne ovisi samo o odabranom algoritmu, nego i o kvaliteti detekcije, predikciji gibanja, reidentifikaciji te načinu povezivanja i upravljanja putanjama. Međutim, njihova primjena u stvarnim sustavima videonadzora zahtijeva odgovarajuću programsku infrastrukturu koja omogućuje učinkovitu obradu videosignala, povezivanje različitih algoritamskih komponenti i izvođenje u stvarnom vremenu. Poseban izazov pritom predstavlja istodobna obrada više videotokova uz ograničene računalne resurse, očuvanje vremenske usklađenosti i dostupnost rezultata drugim aplikacijama. U ovom poglavlju razmatra se NVIDIA DeepStream SDK kao razvojni okvir za izgradnju GPU-ubrzanih sustava inteligentne videoanalitike. Analiziraju se njegova programska arhitektura temeljena na okviru GStreamer, komponente za prihvat i obradu videosignala, primarnu i sekundarnu inferenciju, praćenje objekata i upravljanje metapodacima, kao i mogućnosti konfiguracije i optimizacije performansi. Posebna pozornost posvećuje se dostupnim algoritmima praćenja unutar komponente NvMultiObjectTracker te mogućnostima njihove integracije s modelima dubokog učenja. Kako bi se razumjela uloga pojedinih komponenti i njihova međusobna povezanost, najprije se razmatra položaj DeepStream SDK-a u cjelokupnom procesu inteligentne videoanalitike, nakon čega slijedi detaljniji prikaz njegove arhitekture, funkcionalnosti i konfiguracijskih mogućnosti.

3.1 Uvod i položaj DeepStreama u videoanalitici

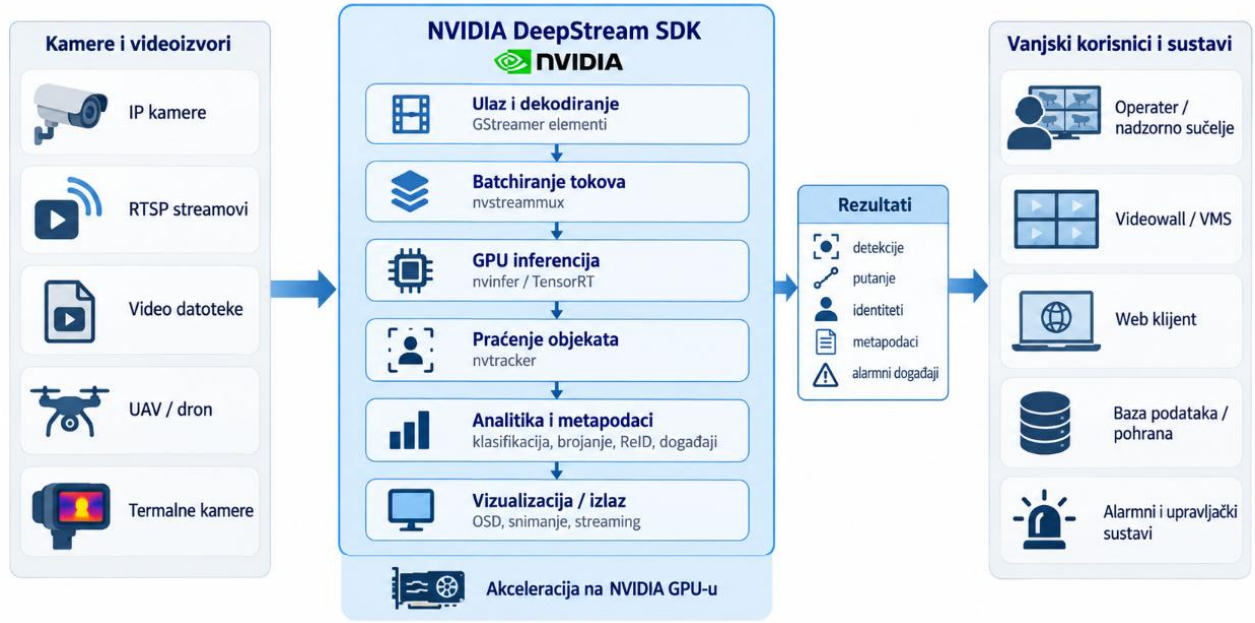
Suvremena analiza videosignala u stvarnom vremenu obuhvaća znatno više od izvođenja neuronske mreže na pojedinačnoj slici. Videosignal najprije je potrebno dohvatiti iz kamere, mrežnog protokola ili datoteke, dekodirati ga, pripremiti za obradu, organizirati ulazne kadrove te rezultate detekcije povezati kroz vrijeme. Nakon zaključivanja slijedi transformacija koordinata, vizualizacija ili distribucija rezultata prema drugim aplikacijama. Svaki od tih koraka troši računalne resurse, a njihov raspored i međusobna sinkronizacija određuju ukupnu latenciju, propusnost i pouzdanost sustava.

NVIDIA DeepStream SDK razvojni je okvir za izgradnju GPU-ubrzanih aplikacija za videoanalitiku. Temelji se na multimedijskom okviru GStreamer i povezuje elemente za prihvat videosignala, hardversko dekodiranje, grupiranje kadrova, izvođenje neuronskih mreža, praćenje objekata, analitiku te objavljivanje rezultata. DeepStream stoga nije jedan algoritam za detekciju ili praćenje, nego infrastruktura u koju se mogu uključiti različiti modeli strojnog učenja i različiti algoritmi praćenja [15–17].

Njegova je osobita vrijednost u tome što omogućuje kontinuirani protok kadrova i njihovih metapodataka između specijaliziranih modula, uz smanjivanje nepotrebnih prijenosa između memorije CPU-a i GPU-a. Ovakav pristup važan je kada se istodobno obrađuje više kamera, kada ulazi imaju različite brzine kadrova ili kada sustav mora održavati predvidivu latenciju uz ograničene resurse. DeepStream podržava NVIDIA diskretne grafičke procesore i odgovarajuće Jetson platforme, pri čemu dostupnost pojedinih značajki ovisi o izdanju SDK-a i hardverskoj platformi [15].

DeepStream 9.1 predstavlja aktualno izdanje u trenutku pripreme ovog pregleda. Međutim, struktura aplikacije, konfiguracijski ključevi i podrška za dodatke mijenjaju

se tijekom razvoja, pa svaku konkretnu implementaciju treba vezati uz odgovarajuću verziju dokumentacije, operacijskog sustava, upravljačkog programa, CUDA-e i TensorRT-a. Opisi u nastavku obuhvaćaju opće načelo rada; primjerene vrijednosti konfiguracijskih parametara nisu univerzalne.



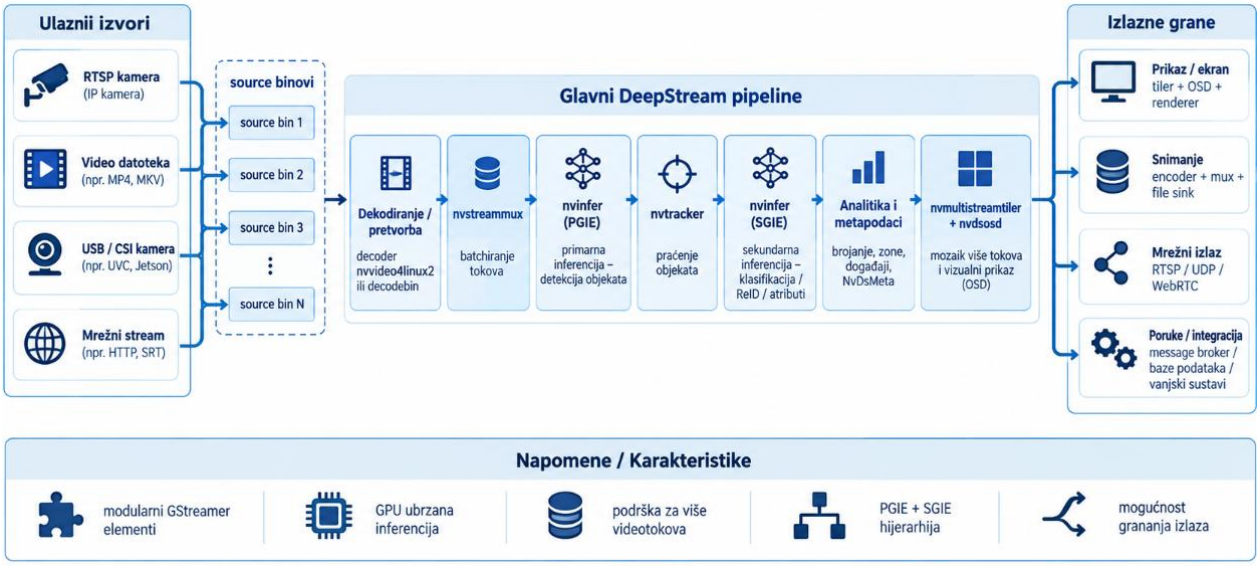
Slika 3.1Položaj DeepStream SDK-a u lancu od kamera i videoizvora do GPU inferencije, trackinga, analitike i vanjskih korisnika (Generirano AI).

3.2 Programska arhitektura i GStreamer pipeline

GStreamer omogućuje modularnu izgradnju multimedijskih aplikacija povezivanjem elemenata (engl. elements) u usmjeren graf obrade koji se naziva pipeline. Pojedini element prihvaća podatke preko ulaznih priključaka (sink pads), obrađuje ih i prosljeđuje preko izlaznih priključaka (source pads). Elementi razmjenjuju GstBuffer objekte, događaje i informacije o vremenskom označavanju. U DeepStreamu se ovakav model proširuje NVIDIA komponentama koje obrađuju kadrove u hardverski pogodnim memorijskim površinama i pridružuju strukturirane metapodatke [16, 23].

Tipična aplikacija sadrži izvor videa i depayloader/parser za mrežne ili komprimirane tokove, hardverski dekoder, element nvstreammux za organizaciju ulaza u serije, primarni inferencijski model, praćenje objekata i eventualne sekundarne modele. Na izlazu se podaci mogu slati prikazivaču, videoenkoderu, datotečnom zapisivaču ili mrežnom posredniku. Programski tok nije ograničen na jedan linearni put: pomoću tee elemenata moguće je video usmjeriti u više grana, primjerice jednu za analitiku, a drugu za prikaz ili snimanje. Takva razgranata arhitektura zahtijeva pravilno postavljanje queue elemenata radi odvajanja rasporeda izvršavanja i upravljanja povratnim pritiskom.

Uobičajeno je razlikovati podatkovnu ravninu, u kojoj putuju video površine, od metapodatkovne ravnine koja prenosi detekcije, klase, identifikatore i druge rezultate. Iako putuju kroz isti pipeline, video i metapodaci mogu se različito obrađivati ili izdvajati na izlazu. To olakšava izradu aplikacija koje izvode više analitičkih zadataka nad istim kadrom bez ponovnog dekodiranja.



Slika 3.2. Referentni GStreamer/DeepStream pipeline sa source binovima, nvstreammux, nvinfer, nvtracker, sekundarnom inferencijom i izlaznim granama (Generirano AI).

3.2.1 Prihvat izvora i hardversko dekodiranje

Izvori mogu biti lokalne video datoteke, USB ili mrežne kamere, RTSP/RTP tokovi te različiti drugi formati koje podržava odgovarajuća GStreamer instalacija. U produkcijskom sustavu potrebno je voditi računa o nestabilnim mrežnim vezama, promjeni rezolucije, varijacijama frekvencije kadrova i vremenskim oznakama. RTSP izvor obično uključuje demultipleksiranje, parsiranje i hardversko dekodiranje, dok izravna kamera može isporučivati nekomprimirane slike. Hardverski dekoder rasterećuje CPU, no ukupna učinkovitost ovisi i o formatu, rezoluciji te broju simultanih dekodiranja.

GStreamer elementi izvora i dekodera moraju proizvesti format koji prihvaća sljedeći dio obrade. NVIDIA memorijske površine (NvBufSurface) omogućuju učinkovito rukovanje slikovnim podacima. Nepotrebne konverzije formata i kopiranja podataka među procesorima mogu poništiti prednosti ubrzane inferencije, zbog čega se raspored konverzija mora planirati na razini cjelokupnog pipelinea.

3.2.2 Grupiranje kadrova elementom nvstreammux

Element nvstreammux objedinjuje kadrove iz jednog ili više ulaza u serije (engl. batches) te pridružuje informacije o izvornom toku i vremenskim oznakama. Grupiranje omogućuje učinkovitiju obradu na GPU-u, ali uvodi kompromis između propusnosti i kašnjenja: veće serije mogu povećati iskorištenost inferencijskog modela, dok čekanje na spore izvore ili na popunjavanje serije može povećati latenciju. Parametri poput batch-size i batched-push-timeout, ovisno o inačici muxera, određuju takvo ponašanje [18].

Kod višestrukih izvora potrebno je razlikovati nominalni broj kamera od broja raspoloživih kadrova u jednom trenutku. Asinkroni ulazi i nepodudarne frekvencije kadrova otežavaju strogu sinkronizaciju. Novije implementacije nvstreammux elementa podržavaju dodatne postavke i načine grupiranja; zbog razlika između

tradicionalnog i novog muxera nije preporučljivo nekritički prenositi konfiguracijske primjere iz različitih izdanja dokumentacije.

3.3 Izvođenje modela dubokog učenja

3.3.1 Primarna inferencija i nvinfer

Gst-nvinfer povezuje TensorRT inferencijske modele s DeepStreamovim tokom kadrova i metapodataka. Primarni model (Primary GIE, PGIE) najčešće obrađuje cijeli kadar ili unaprijed pripremljene regije te izvodi detekciju plovila, vozila, osoba ili drugih objekata. Rezultat uključuje pravokutnik koji omeđuje objekt, klasu i ocjenu pouzdanosti. Ovisno o modelu, moguće su i segmentacija ili drugačiji oblici izlaza. Za detektore koji nisu izravno podržani standardnim parserima često je potrebno osigurati prilagođeni parser izlaznih tenzora, postupak filtriranja i eventualno NMS [19].

Model se može učitati preko TensorRT engine datoteke ili se engine može izgraditi iz odgovarajuće reprezentacije modela, ovisno o podršci i konfiguraciji. TensorRT omogućuje optimizaciju izvođenja, između ostalog uporabom FP16 ili INT8 preciznosti kada to podržavaju model i hardver. INT8 pritom zahtijeva odgovarajući postupak kalibracije ili kvantizacije; manja preciznost nije automatski bez utjecaja na detekciju malih ili udaljenih objekata.

Posebno su značajni interval izvođenja modela, razlučivost ulazne slike, batch-size i prag pouzdanosti. Ako se detektor izvodi na svakom n-tom kadru, smanjuje se opterećenje inferencije, ali rastu zahtjevi za interpolacijom i održavanjem putanja između opažanja. U postavkama detekcije mjerilo evaluacije mora ostati jasno: prag detekcije utječe na broj lažnih pozitivnih i propuštenih objekata te posredno na kvalitetu trackinga.

3.3.2 Sekundarna inferencija i nvinferserver

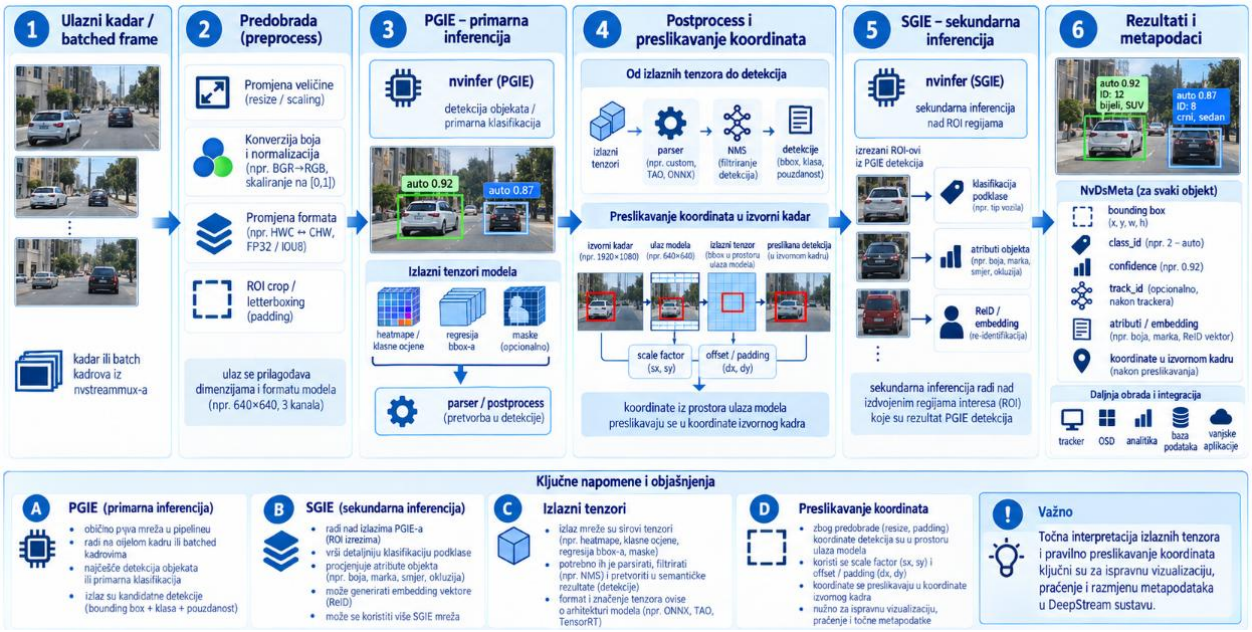
Sekundarni inferencijski element (Secondary GIE, SGIE) primjenjuje model na izdvojene objekte koje je prethodno otkrio PGIE ili koji se prate. Njime se mogu procjenjivati dodatni atributi, poput potklase, orijentacije, boje ili reprezentacije izgleda. Zbog toga se često provodi izrezivanje regije objekta, skaliranje slike i priprema tenzora, uz čuvanje veze između dobivenog rezultata i identifikatora izvornog objekta. Odvojeni sekundarni modeli korisni su kada jedan detektor ne treba izravno predviđati sva svojstva objekta.

Gst-nvinferserver integrira inferenciju s NVIDIA Triton Inference Serverom. Takva organizacija može biti prikladna za upravljanje složenijim skupovima modela i različitim backendovima, ali donosi dodatni konfiguracijski sloj. Odabir nvinfer ili nvinferserver zato se ne svodi samo na algoritamsku točnost nego i na zahtjeve za održavanjem modela, latencijom i raspoloživim hardverom [15, 19].

3.3.3 Pretprocesiranje i naknadna obrada

Pretprocesiranje obuhvaća promjenu veličine slike, normalizaciju, konverziju boja, odabir područja interesa te eventualno zadržavanje omjera stranica. Plugin nvdspreprocess može pripremati ulazne regije i tenzore prije inferencije, dok se naknadnom obradom interpretiraju izlazi neuronske mreže. Važno je da se koordinate predviđenih pravokutnika dosljedno preslikavaju u izvorni koordinatni sustav videa. Kod letterbox obrade, gdje se dodaju rubovi radi očuvanja omjera, zanemarivanje

pomaka i faktora skaliranja dovodi do sustavnih pogrešaka lokalizacije i posljedično trackinga.



Slika 3.3. Uloga PGIE, SGIE, predobrade, izlaznih tenzora i preslikavanja koordinata u DeepStreamu (Generirano AI).

3.4 Praćenje objekata: Gst-nvtracker i biblioteka NvMultiObjectTracker

Detekcija u pojedinačnom kadru sama po sebi ne osigurava vremenski kontinuitet. Plugin Gst-nvtracker prihvaća rezultate detektora i prethodno stanje putanja, a zatim radi s niskorazinskom bibliotekom praćenja koja pridružuje opažanja trajnim identifikatorima. Arhitektura razlikuje GStreamer integracijski sloj od stvarnog algoritma, pa je na istom sučelju moguće koristiti različite implementacije. NVIDIA referentna biblioteka NvMultiObjectTracker uključuje IOU, NvSORT, NvDeepSORT i NvDCF, a izbor i mogućnosti ovise o verziji sustava [20].

3.4.1 IoU tracker i NvSORT

IoU tracker povezuje objekte između kadrova pretežno prema preklapanju njihovih pravokutnika. Takav pristup ima malu računalnu složenost, ali mu uspješnost brzo opada kada se objekti ubrzano pomiču, kada nedostaju detekcije ili kada su brodovi međusobno blizu. NvSORT proširuje geometrijski pristup predikcijom stanja, tipično primjenom Kalmanova filtra, čime omogućuje povezivanje detekcija i pri umjerenim razlikama položaja. Ove metode korisne su kao jeftine referentne konfiguracije jer ne zahtijevaju duboke vizualne značajke.

3.4.2 NvDeepSORT

NvDeepSORT proširuje model praćenja vizualnim značajkama izdvojenima neuralnim ReID modelom. Pri asocijaciji se zato ne promatra samo geometrijska kompatibilnost nego i sličnost reprezentacija izgleda. Takav pristup može smanjiti zamjene identiteta kada se više objekata preklapa, ali ovisi o kvaliteti podataka na kojima je opisnik treniran. U pomorskom okruženju slični trajekti ili plovila u jednakim bojama trupa

mogu imati vrlo bliske opisnike, dok promjena kuta snimanja, osvjetljenja i udjela pozadine u izrezu može narušiti stabilnost značajki.

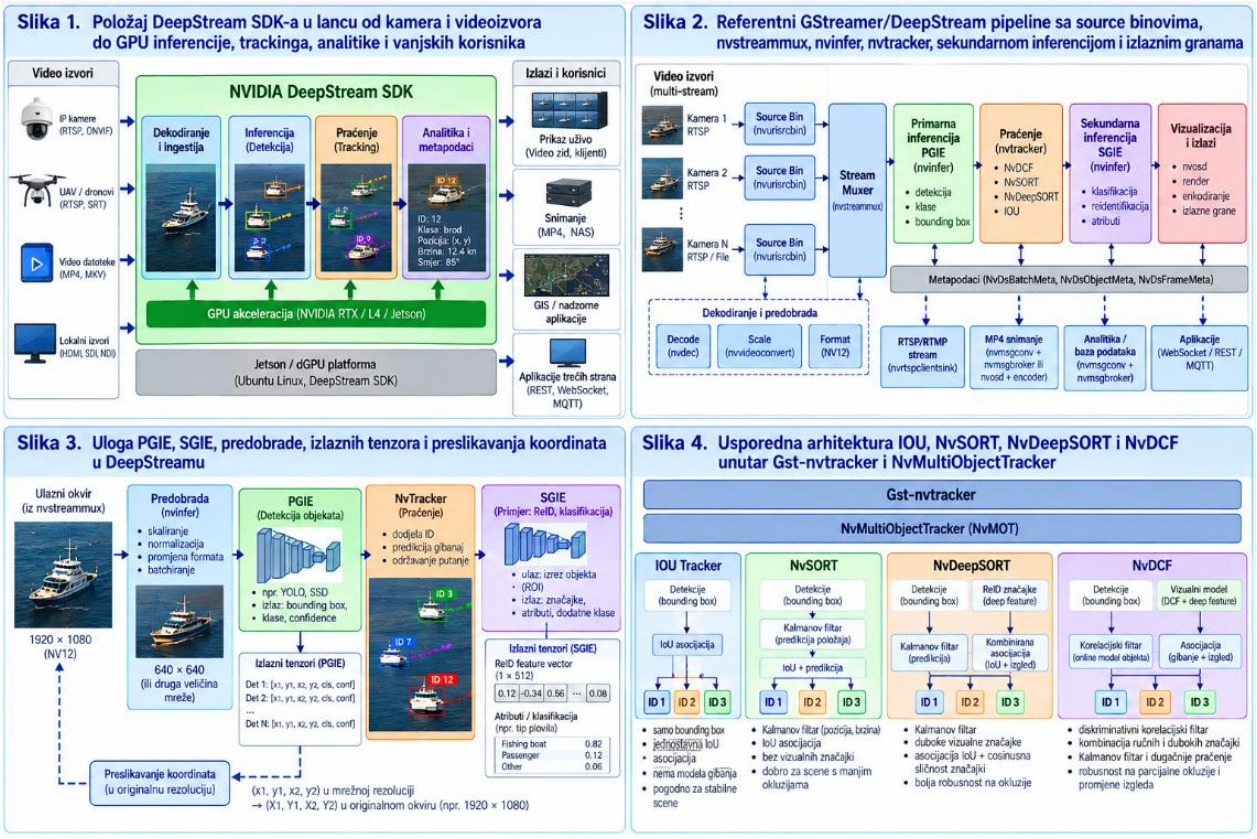
3.4.3 NvDCF

NvDCF temelji vizualno praćenje na diskriminacijskom korelacijskom filtru (Discriminative Correlation Filter, DCF). Za praćeni objekt održava se vizualni predložak koji omogućuje lokalizaciju i između dvaju poziva detektora. U odnosu na geometrijski tracker, DCF može iskoristiti informacije iz slikovnog sadržaja te tako olakšati praćenje objekta kada detekcija kratkotrajno izostane. Referentne konfiguracije nude različite kompromise točnosti i performansi, a detalji poput HOG ili ColorNames značajki, veličine mape značajki, brzine prilagodbe filtra i praga pouzdanosti utječu na ponašanje [20, 21].

DCF pritom nije bez ograničenja. Ako se izgled objekta tijekom djelomične okluzije znatno promijeni, ažuriranje predloška može uključiti okludirajući objekt ili pozadinu te izazvati drift. Praćenje nakon duljeg izostanka relevantnih opažanja stoga nije zajamčeno samom uporabom korelacijskog filtra. Ovisno o konfiguraciji, biblioteka NvMultiObjectTracker može dodati upravljanje sjenovitim stanjima putanje, povezivanje prema vremenskoj povijesti i ReID mehanizme. Takve značajke nisu inherentna svojstva svih konfiguracija NvDCF-a.

3.4.4 Upravljanje životnim ciklusom putanja

Tracker mora razlikovati privremeno opažene objekte od potvrđenih putanja, povezivati nova opažanja sa starim identitetima i odlučivati kada objekt više nije moguće pouzdano pratiti. Pritom su važni pragovi za potvrdu putanje, duljina dopuštenog izostanka detekcije, pouzdanost detektora i trackera te pravila za ponovno pridruživanje. Upravo u ovoj razini nastaju praktične razlike među sustavima koji nominalno koriste isti osnovni tracker. Predugo zadržavanje neaktivnih putanja povećava mogućnost pogrešnog pridruživanja, dok prerano brisanje stvara fragmentaciju i nove identifikatore nakon okluzije.



Slika 3.4. Usporedna arhitektura IOU, NvSORT, NvDeepSORT i NvDCF unutar Gst-nvtracker i NvMultiObjectTracker (Generirano AI).

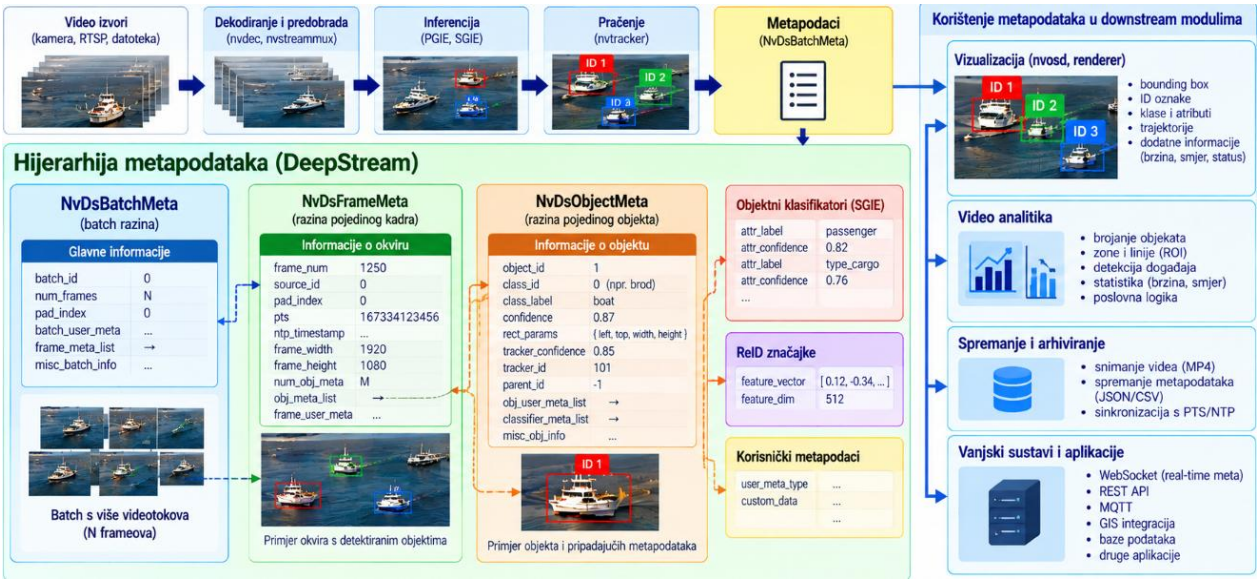
3.5 Reidentifikacija, metapodaci i analitički sloj

Pojam reidentifikacije (ReID) označuje usporedbu prikaza objekta radi ponovnog pronalaženja istog fizičkog objekta nakon promjene položaja ili razdoblja nevidljivosti. Duboke mreže iz slikovnog izreska formiraju vektor značajki (embedding), a sličnost se najčešće mjeri kosinusnom sličnošću ili udaljenošću među normaliziranim vektorima. U DeepStreamu se ReID može ostvariti kroz odgovarajuću tracker konfiguraciju ili odvojenu inferencijsku komponentu, ovisno o odabranoj arhitekturi i verziji SDK-a. Važno je razlikovati lokalni identitet koji dodjeljuje tracker od eventualnog identiteta koji bi povezivao više kamera ili zasebnih vremenskih sesija: ovo drugo zahtijeva dodatne mehanizme i nije automatski posljedica uključenog ReID-a.

DeepStreamov metapodatkovni model organizira podatke u hijerarhiju struktura kao što su NvDsBatchMeta, NvDsFrameMeta i NvDsObjectMeta. Metapodaci o objektima mogu uključivati koordinate pravokutnika, klasu, pouzdanost i identifikator koji vraća tracker. Po potrebi se pridružuju korisnički metapodaci, rezultati sekundarne inferencije ili posebne analitičke oznake. Takav podatkovni model omogućuje da aplikacija izdvaja strukturirani izlaz bez ponovnog analiziranja vizualiziranih okvira na videoslici [16].

Analitičke komponente poput nvdsanalytics mogu izračunavati brojanje objekata, prelazak virtualnih linija ili zadržavanje u definiranim područjima, uz odgovarajuću konfiguraciju koordinatnog sustava. Element nvdsosd služi iscrtavanju pravokutnika, oznaka i grafičkih elemenata na slici. Za slanje poruka vanjskim sustavima koriste se

odgovarajući elementi za pretvaranje metapodataka u poruke i njihov prijenos, primjerice nvmsgconv i nvmsgbroker. Metapodaci i video imaju različite obrasce sinkronizacije, pa se u distribuiranim aplikacijama moraju očuvati vremenske oznake i informacije o izvornom toku.



Slika 3.5. Hijerarhija NvDsBatchMeta – NvDsFrameMeta – NvDsObjectMeta i put podataka prema vizualizaciji, analitici i vanjskim sustavima (Generirano AI).

3.6 Konfiguriranje sustava i ključni parametri

Konfiguracija se u referentnoj aplikaciji deepstream-app često definira preko INI sekcija za izvore, sinkove, streammux, primarnu i sekundarnu inferenciju, tracker i prikaz. Pojedine NVIDIA biblioteke, naročito NvMultiObjectTracker, koriste dodatne YAML datoteke. Vlastita GStreamer aplikacija može iste postavke dodjeljivati programski. Konfiguracija određuje algoritamsko ponašanje jednako kao i mrežnu i računalnu učinkovitost pa treba razlikovati postavke obrade video ulaza, inferencije, trackinga i izlaza [17, 20].

Sljedeći isječak ilustrira način razdvajanja postavki u referentnoj aplikaciji. Vrijednosti su samo primjer i moraju se uskladiti s verzijom DeepStreama, brojem izvora i raspoloživim računalnim resursima [17].

Tablica 3.1 Postavke komponenata sustava

Sloj	**Primjeri važnih postavki**	**Utjecaj na rad**
Izvor / dekoder	URI, RTSP transport, ponovno povezivanje, frekvencija kadrova	Dostupnost i kontinuitet ulaznih kadrova
StreamMux	batch-size, batched-push-timeout, dimenzije	Kompromis propusnosti i latencije
PGIE / SGIE	config-file-path, interval, batch-size, engine, pragovi	Točnost, učestalost i trošak inferencije
Tracker	ll-lib-file, ll-config-file, tracker-width, tracker-height	Vrsta algoritma, skala obrade, utrošak resursa
NvDCF	featureImgSizeLevel, filterLr, parametri asocijacije	Stabilnost vizualnog predloška i trackinga
Izlaz	encoder, bitrate, sink, sinkronizacija	Prijenos, kvaliteta prikaza i kašnjenje

3.6.1 Primjer konfiguracije glavnih elemenata

[streammux]
batch-size=4
width=1280
height=720
batched-push-timeout=40000

[primary-gie]
enable=1
config-file=config_infer_primary.txt

[tracker]
enable=1
tracker-width=640
tracker-height=384
ll-lib-file=/putanja/do/libnvds_nvmultiobjecttracker.so
ll-config-file=config_tracker_NvDCF_accuracy.yml

Parametar batch-size u muxeru definira ciljnu veličinu grupe kadrova. Ujednačenost s batch veličinom inferencije treba provjeriti zasebno jer te postavke ne moraju biti sinonimi. Vrijednost batched-push-timeout prikazana je u mikrosekundama za odgovarajuću inačicu muxera. tracker-width i tracker-height odnose se na dimenzije radne slike trackera, ali iznimke postoje kod specifičnih implementacija i kalibriranih 3D modela. Konfiguracijska biblioteka ll-lib-file određuje niskorazinsku implementaciju praćenja, dok ll-config-file određuje njezine algoritamske parametre [17, 20].

3.6.2 Primjer skupina postavki NvDCF-a

NVIDIA konfiguracijske datoteke NvDCF-a razvrstavaju parametre u skupine za kontrolu trackera, upravljanje ciljevima, asocijaciju, izdvajanje značajki i korelacijski filtar [21]. Pojedine datoteke sadrže i postavke za ReID. Neki parametri utječu na trajanje razdoblja bez detekcije, drugi na brzinu ažuriranja vizualnog predloška, a treći na kriterije pridruživanja. Brže prilagođavanje filtra omogućuje praćenje promjena izgleda, ali može povećati rizik od uključivanja pozadine ili okludirajućeg objekta; sporije prilagođavanje daje stabilniji predložak, ali otežava prilagodbu stvarnoj promjeni perspektive [20, 21].

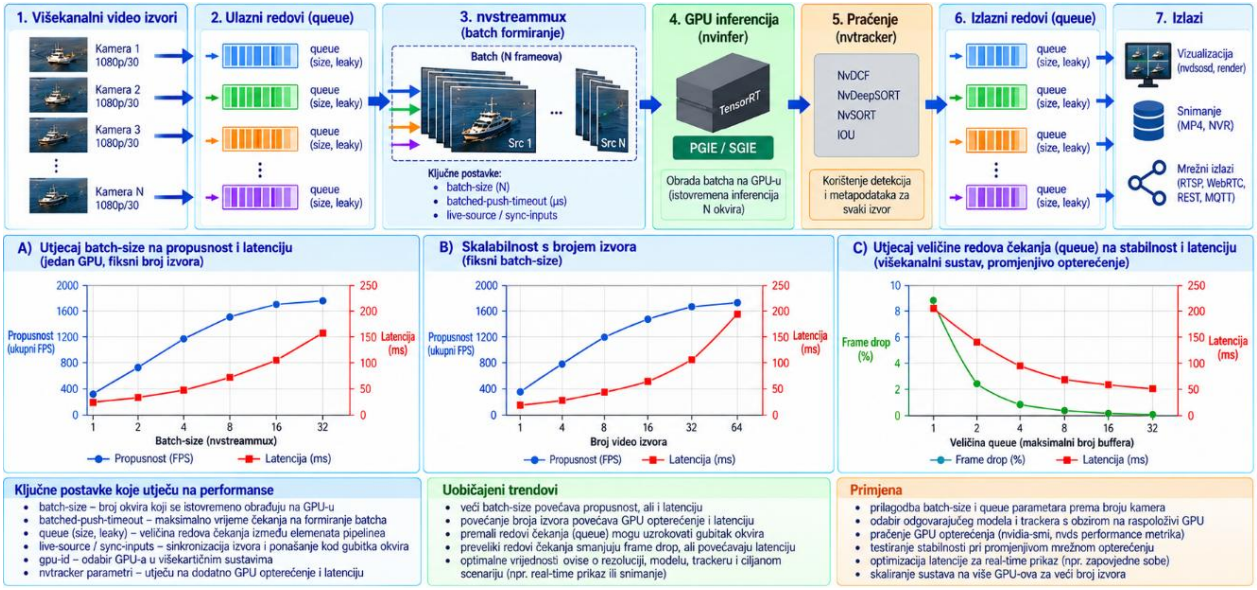
Primjerice, useHog i useColorNames uključuju vrste vizualnih značajki, featureImgSizeLevel utječe na prostornu rezoluciju značajki, filterLr određuje brzinu učenja filtra, a gaussianSigma prostornu širinu željenog odziva. U uporabi ReID-a parametri poput reidFeatureSize, inferDims i reidHistorySize definiraju karakteristike opisnika i njegove povijesti. Vrijednosti nisu neovisne: rezolucija ulaza, vrijeme inferencije, frekvencija detekcije i broj objekata zajednički određuju razinu računskog opterećenja i točnost.

3.7 Performanse, latencija i višestruki videotokovi

Učinak sustava ne može se svesti na brzinu inferencije jednog modela. Potrebno je mjeriti efektivnu brzinu obrade po ulazu (FPS), ukupnu propusnost, end-to-end latenciju, iskorištenost GPU-a i CPU-a, zauzeće memorije te stabilnost vremenskih oznaka. Razlika između latencije i propusnosti ključna je: sustav može obrađivati velik broj kadrova u sekundi, a ipak prikazivati svaki pojedinačni kadar sa znatnim kašnjenjem zbog međuspremnika i čekanja na grupiranje.

Pri skaliranju na više izvora uska grla mogu biti hardverski dekoder, serijalizacija u CPU dodacima, kapacitet GPU memorije, pojedini inferencijski model, enkoder ili mrežna propusnost. Zato povećanje batch-size ne jamči bolje ponašanje. DeepStream podržava dinamičko dodavanje i uklanjanje izvora i odgovarajuće aplikacijske obrasce, ali je potrebno provjeriti ponovnu inicijalizaciju pojedinih modula te očuvanje konzistentnosti identifikatora i resursa [22].

Za analitičke sustave u stvarnom vremenu važna je politika odbacivanja ili zadržavanja zakašnjelih kadrova. Pri preopterećenju se mogu pojaviti rast redova čekanja, povećanje latencije i promjenjiv interval između opažanja. Ako tracker pretpostavlja približno konstantan vremenski korak, neredoviti dolazak kadrova može povećati pogrešku predikcije. Korektna primjena stoga traži provjeru stvarnih vremenskih oznaka, odvajanje mjera trajanja obrade od mjera kašnjenja i testiranje u reprezentativnim uvjetima.



Slika 3.6. Utjecaj batch-size, broja izvora i redova čekanja na propusnost i latenciju višekanalnog DeepStream sustava (Generirano AI).

3.8 Proširivost i razvoj vlastitih komponenti

Arhitektura GStreamera dopušta razvoj vlastitih elemenata ili umetanje obrade preko odgovarajućih pad-probe mehanizama. Time je moguće analizirati rezultate detekcije, dopunjavati NvDsObjectMeta, dodavati vlastite metapodatke ili izdvajati rezultate za vanjske sustave. NVIDIA objavljuje primjere integracije OpenCV-a i izrade vlastitih GStreamer elemenata [15]. Razvoj dodatnih komponenti zahtijeva brigu o životnom ciklusu memorijskih struktura, vlasništvu nad metapodacima, sinkronizaciji paralelnih niti i ponašanju kod dodavanja ili uklanjanja tokova.

Za aplikacije s više kamera važno je upravljati povratnim pritiskom između grana. Dugotrajna operacija u sinkronom callbacku može zaustaviti dio pipelinea, dok neograničeno prikupljanje metapodataka povećava potrošnju memorije. Prikladna arhitektura obično uključuje ograničene redove, jasno razdvajanje zahtjevnih pomoćnih zadataka i definiran postupak oporavka od pogreške. Mogućnost proširivanja zato je važna prednost DeepStreama, ali ona prenosi odgovornost za stabilnost dodatnog aplikacijskog koda na implementatora.

3.9 Prednosti, ograničenja i prikladnost za detekciju i praćenje objekata

Prednosti DeepStreama očituju se u integraciji obrade videosignala i inferencije, uporabi GPU-ubrzanih modula, podršci za više ulaza te dostupnosti različitih tracker biblioteka iza zajedničkog sučelja. Framework omogućuje usporedbu različitih modela detekcije i praćenja unutar relativno jedinstvene infrastrukturne osnove. Njegova modularnost pogoduje sustavima u kojima je potrebno eksperimentirati s modelima, mijenjati broj videoizvora ili izvoziti metapodatke bez preoblikovanja cijelog lanca obrade.

Ipak, DeepStream ne uklanja temeljna ograničenja računalnog vida. Pogreške detektora, vizualna sličnost objekata, djelomične i dugotrajne okluzije, promjene osvjetljenja te neprimjerenost ReID modela konkretnoj domeni nastavljaju utjecati na rezultate. Povećanje učinkovitosti izvođenja ne podrazumijeva automatsko povećanje

točnosti praćenja. Dodatno, značajna ovisnost o NVIDIA softverskom i hardverskom ekosustavu, kompatibilnosti izdanja te složenost konfiguriranja velikog broja dodataka mogu otežati razvoj i održavanje.

Pri procjeni prikladnosti potrebno je razlikovati točnost detekcije, točnost asocijacije i sustavne pokazatelje poput brzine obrade ili dostupnosti ulaza. Za praćenje više objekata relevantne su MOTA, IDF1 i HOTA, dok su za produkcijsku videoanalitiku jednako važne latencija, pouzdanost rada i mogućnost obrade više izvora. Time DeepStream predstavlja tehnološku infrastrukturu pogodnu za implementaciju i evaluaciju MOT metoda, ali ne zamjenjuje metodološki odabir modela i znanstvenu evaluaciju njihova ponašanja.

3.10 Zaključak

NVIDIA DeepStream SDK omogućuje cjelovitu i učinkovitu organizaciju videoanalitičkih zadataka zahvaljujući modularnom GStreamer pipelineu, podršci za hardversko dekodiranje, grupiranje kadrova, TensorRT inferenciju, više algoritama praćenja i strukturirane metapodatke. Razdvajanje infrastrukture od algoritamskih komponenti pruža fleksibilnost: isti okvir može podržavati različite detektore, različite strategije MOT asocijacije i dodatne modele za izdvajanje atributa. Posebno su važni moduli nvstreammux, nvinfer i nvtracker, jer njihova konfiguracija neposredno određuje odnos između računske učinkovitosti i pouzdanosti praćenja.

Pri korištenju DeepStreama potrebno je konfiguracijske odluke donositi na temelju konkretnih zahtjeva, a ne samo preporučenih tvorničkih profila. Interval detekcije, veličina serije, rezolucija i vrsta trackera međusobno su povezani, a dugotrajne okluzije i vizualno slični objekti ostaju otvoreni izazovi neovisno o odabranoj infrastrukturi. DeepStream je stoga primjeren predmet pregleda kao reprezentativna platforma koja omogućuje prijelaz od izoliranog algoritma prema cjelovitom sustavu računalnog vida, uz zadržavanje potrebe za zasebnom evaluacijom detekcije, asocijacije i praćenja u stvarnim uvjetima.

4 SKUPOVI PODATAKA ZA DETEKCIJU, REIDENTIFIKACIJU I PRAĆENJE PLOVILA

4.1 Uloga skupova podataka u istraživanju računalnog vida za pomorski nadzor

Razvoj sustava koji automatski detektiraju i prate plovila temelji se na dostupnosti reprezentativnih podataka. Skup primjeren klasifikaciji tipa plovila ne mora biti primjeren procjeni njegova položaja, a skup neovisnih slika, ma koliko velik, sam po sebi ne može provjeriti zadržavanje identiteta kroz vrijeme. Pri analizi dostupnih baza potrebno je zato razlikovati najmanje tri zadatka: detekciju (gdje se plovilo nalazi i kojoj kategoriji pripada), reidentifikaciju (je li ponovno opaženo plovilo isti fizički objekt) i višeciljno praćenje (može li se zadržati dosljedna putanja istog objekta kroz videozapis).

U pomorskim scenama dodatni izazov predstavljaju velike razlike veličine objekata u pikselima, odražaj svjetlosti na površini mora, valovi, vrlo slični trupovi i nadgrađa, zaklanjanje malih plovila velikima te postojanje statičnih plovila na vezu. Neke baze bilježe kameru na obali, druge se oslanjaju na kamere postavljene na plovilu, a treće na satelitske senzore. Takve se domene ne smiju automatski poistovjećivati: prijenos modela s jednog geometrijskog i radiometrijskog okruženja na drugo može proizvesti pad točnosti.

Ovo poglavlje najprije obrađuje opće skupove podataka i baze iz pomorskih okruženja opisane u disertaciji Ivane Marin [9], potom detaljno prikazuje njezin Split Ship MOT (SSMOT) skup, a završno razmatra novije javno opisane baze iz 2025. i 2026. godine. Izvorne statistike SSMOT-a preuzete su iz disertacije; novije baze imaju zasebno navedene izvore.

4.2 Opći skupovi podataka i njihova ograničenja

CIFAR-10, Caltech-256, ImageNet, PASCAL VOC i MS COCO omogućili su značajan napredak klasifikacije i detekcije objekata. Sadrže i slike plovila, ali u pravilu ne obuhvaćaju dovoljno fino razdvojene kategorije relevantne za luke. U MS COCO-u plovila pripadaju općoj klasi boat, dok zadaci nadzora pomorskog prometa često zahtijevaju razlikovanje trajekta, jedrilice, glisera, ribarske brodice i drugih vizualno različitih objekata. Disertacija [9] ističe i razliku distribucije veličina: u općim skupovima plovilo često zauzima veći dio kadra, dok je u nadzornom videu ono često vrlo maleno.

Tablica 4.1. Karakteristike skupova podataka

Skup	Broj klasa plovila	Slike	Instance	Primarna namjena
CIFAR-10	1	6.000	—	klasifikacija
Caltech-256	4	418	—	klasifikacija
ImageNet	6	525	613	klasifikacija / detekcija
PASCAL VOC	1	363	791	detekcija
MS COCO	1	3.025	10.759	detekcija

Vrijednosti u tablici odnose se na prikaz brojeva primjera plovila u [9], a ne na ukupan broj slika cjelovitih javnih baza. Standardni skupovi za višeciljno praćenje, primjerice MOTChallenge i PETS, dominantno su oblikovani za pješake, a baze autonomne vožnje prvenstveno za cestovna vozila. Njihova metodološka vrijednost ostaje značajna, ali ne mogu zamijeniti ciljano prikupljene podatke iz pomorskih uvjeta.

4.3 Pomorski skupovi podataka prikazani u dosadašnjim istraživanjima

4.3.1 Klasifikacijske baze

VAIS objedinjuje vidljive i infracrvene prikaze plovila; prema disertaciji uključuje 2.865 slika, od kojih 1.623 RGB i 1.242 infracrvene, uključujući 1.088 uparenih prikaza. Takva baza omogućuje istraživanje klasifikacije u različitim spektralnim modalitetima, ali ne sadrži sama po sebi vremenske oznake potrebne za mjerenje MOTA ili IDF1. MARVEL obuhvaća približno dva milijuna internetski prikupljenih slika 109 tipova plovila, sa 26 nadređenih kategorija. Prednost mu je raznolikost prikaza, a ograničenje domenska razlika između internetske fotografije i stalne nadzorne kamere. Game of Deep Learning: Ship Dataset sadrži 6.252 označene slike za učenje i 2.680 neoznačenih testnih slika, raspoređenih u pet klasa [9].

4.3.2 Skupovi za detekciju plovila

SeaShips se usredotočuje na detekciju šest kategorija brodova. McShips obuhvaća civilna i vojna plovila iz heterogenih izvora; takva raznolikost korisna je za generalizaciju, no različita kvaliteta izvora i način nastanka kadrova otežavaju direktnu procjenu vremenskog kontinuiteta. ABOships prikupljen je kamerom na plovilu tijekom plovidbe u Finskoj te zato prikazuje kretanje pozadine i promjenu perspektive. GLSD sadrži znatan broj malih objekata, ali je u Ivaninoj tablici označen kao nejavno dostupan. Sintetički MarSyn sadržava modelirane pomorske prizore, uključujući promjene vremena i refleksije mora; sintetičnost olakšava kontrolu scena, ali zahtijeva provjeru prijenosa na stvarne snimke.

Posebno je važan Split Port Ship Classification Dataset (SPSCD), razvijen na splitskim snimkama uz pomorsku stručnu klasifikaciju. Prema disertaciji obuhvaća 19.337 slika i 27.849 instanci iz 12 kategorija. Njegova je prednost bliskost realnoj splitskoj prometnoj domeni, ali oznake nisu organizirane kao kontinuirane putanje objekata te sam SPSCD nije prikladan za izravnu evaluaciju višeciljnog praćenja. Iz SPSCD-a potekao je važan dio koncepcijskih i anotacijskih polazišta kasnije baze SSMOT [9].

4.3.3 Video baze za praćenje

Singapore Maritime Dataset (SMD) sadrži videozapise snimljene s obale i plovila u pokretu: 40 obalnih i 11 brodskih sekvenci, odnosno ukupno 51 videozapis. Time omogućuje provjeru dva različita tipa gibanja kamere, ali dio objekata nije plovilo, a format .mat zahtijeva dodatnu konverziju za standardne MOT evaluatore. LMD-TShip obuhvaća 191 videozapis, s razdjelom 152 za učenje i 39 za testiranje; prema Ivaninoj analizi često se radi o jednom plovilu po sekvenci, što ograničava mjerenje međusobne asocijacije sličnih plovila. SeaSAw, iako opsežan i važan za situacijsku svjesnost, u disertaciji je označen kao nejavno dostupan. FVessel povezuje videopodatke s AIS-om, čime otvara prostor za multimodalnu analizu, ali ima relativno malu gustoću istodobno opaženih plovila [9].

Tablica 4.2. Baze videa plovila

Skup (god.)	Slike / kadrovi	Objekti	Klase	Glavni zadatak	Javno prema [9]
SMD (2017)	20.367	157.668	10	praćenje	da
SeaShips (2018)	31.455	40.077	6	detekcija	da
Harbor Surveillance (2018)	48.966	70.513	1	detekcija	ne
McShips (2020)	14.709	26.259	13	detekcija	da
ABOships (2021)	9.880	41.967	11	detekcija	da
GLSD (2021)	152.576	212.357	13	detekcija	ne
LMD-TShip (2021)	40.240	nije navedeno	5	praćenje	da
MarSyn (2022)	25.000	34.000	6	detekcija / segmentacija	ne
SeaSAw (2022)	1,9 mil.	14,6 mil.	12	detekcija / praćenje	ne
SPSCD (2023)	19.337	27.849	12	detekcija	da
FVessel (2023)	7.625+	nije navedeno	1	detekcija / praćenje	da
MVDD13 (2024)	35.474	40.839	13	detekcija	da

Podaci i status dostupnosti preuzeti su iz tablice 3.2 disertacije [9] te opisuju stanje u njezinu istraživačkom kontekstu. Javni pristup pojedinim bazama mogao se naknadno promijeniti. Brojeve treba čitati uz način anotiranja: broj slika, broj detekcijskih instanci i broj identiteta nisu iste veličine.



Slika 4.1. Primjeri različitih pomorskih podatkovnih domena: fotografije, obalni video, kamera u pokretu i multimodalna opažanja)Generirano AI).

4.4 Split Ship MOT (SSMOT): struktura, nastanak i znanstvena vrijednost

Split Ship MOT, opisan u četvrtom poglavlju disertacije Ivane Marin [9], oblikovan je kao povezani skup za detekciju, reidentifikaciju i višeciljno praćenje. Njegova ključna vrijednost nije samo količina anotacija, nego mogućnost odvojenog ispitivanja pogrešaka detektora, razlikovanja fizičkih identiteta i očuvanja putanja u istome pomorskom okruženju. Tri podskupa služe različitim eksperimentalnim zadacima, ali koriste međusobno kompatibilne kategorije plovila.

4.4.1 Izvor snimaka i specifičnosti splitske luke

Podaci su nastali na snimkama nadzorne kamere u splitskoj luci. U disertaciji je navedena kamera Dahua DH-TPC-PT8620A-T sa Sony CMOS senzorom 1944 × 1092 piksela, žarišnom duljinom 6–180 mm i optičkim povećanjem 30×, postavljena približno devet metara iznad mora na ulazu u luku. Na snimkama se pojavljuju trajekti, katamarani, manja rekreacijska plovila, turistički brodovi i ribarske brodice. Stalni položaj kamere pogoduje praćenju pozadinski stabilnih scena, ali istodobno ističe problem velikih razlika u udaljenosti objekata od kamere te međusobnog zaklanjanja na pomorskim pravcima [9].



Slika 4.2. Lokacija i vidno polje kamere u splitskoj luci [9].

4.4.2 Kategorizacija na temelju vizualnih obilježja

Za razliku od pomorskih taksonomija temeljenih na stvarnoj duljini ili funkciji plovila, SSMOT je kategorije prilagodio vizualno prepoznatljivim svojstvima na RGB kadrovima. Ivanina disertacija navodi 11 klasa: Small Craft, Small Fishing Boat, Small Passenger Ship, Fishing Trawler, Large Passenger Ship, Sailing Boat, Speed Craft, Motorboat, Pleasure Yacht, Ferry i High-speed Craft. Ujedinjenje prethodnih SPSCD kategorija Medium Ferry i Large Ferry u Ferry odražava činjenicu da ih je iz promatranih kadrova teško pouzdano odijeliti prema duljini. Ta je odluka metodološki važna: manje pogrešaka anotacije klase ne znači nužno lakšu reidentifikaciju fizički različitih brodova koji vizualno izgledaju gotovo jednako.

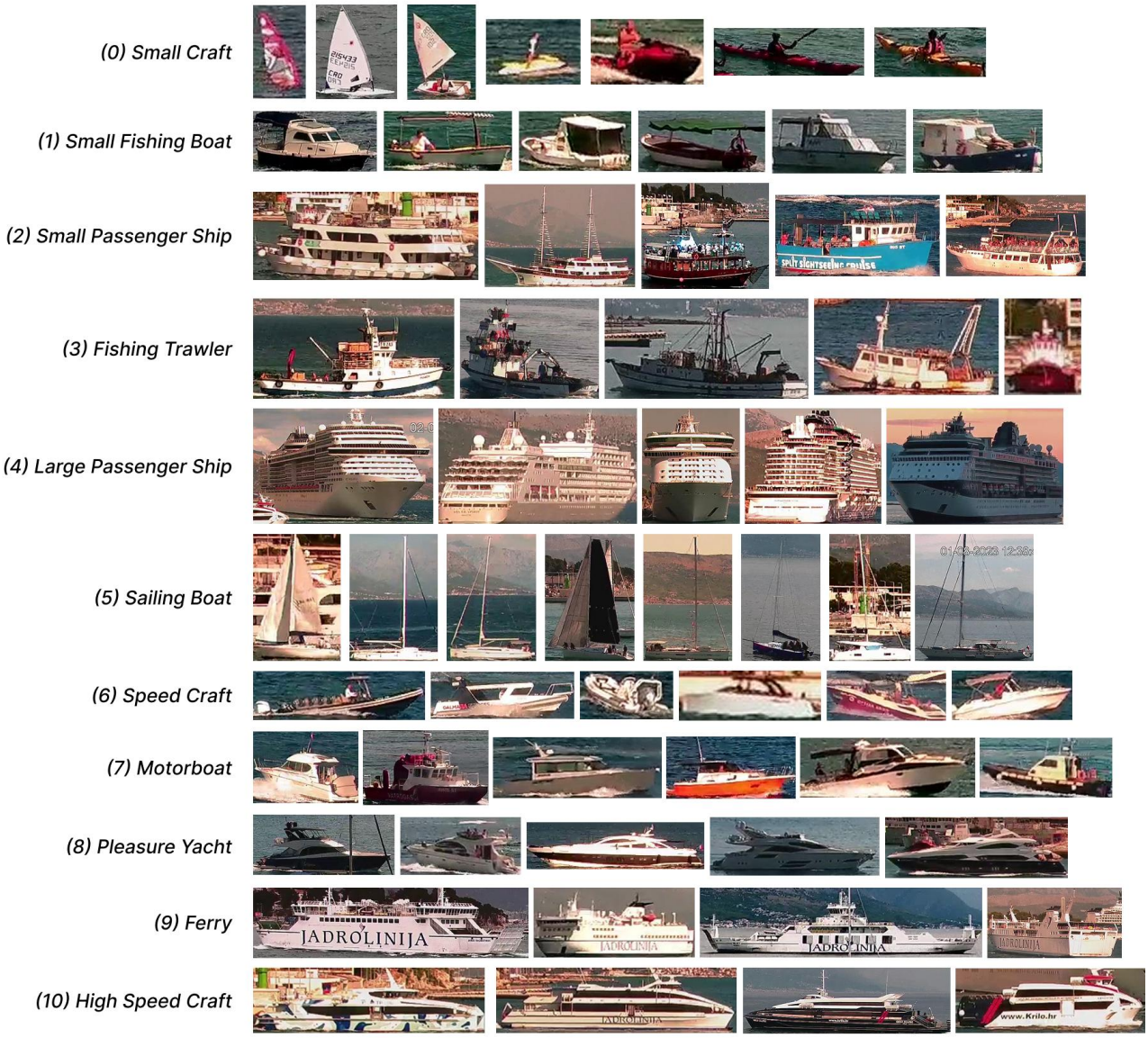
4.4.3 Detekcijski podskup: odabir slika i raspodjela klasa

SSMOT detekcijski podskup obuhvaća 9.827 slika, od kojih su 9.601 razlučivosti 1920 × 1080, a 226 razlučivosti 1280 × 720. Od toga je 7.632 slika namijenjeno učenju, 1.347 validaciji i 848 testiranju; testni podskup potječe iz drugih videozapisa od onih korištenih za učenje i validaciju. Sveukupno je anotirano 50.437 instanci plovila. Posebno je važno da pojedinačni primjeri nisu nekritički uzimani iz uzastopnih vrlo sličnih kadrova, jer bi tada broj slika prividno povećavao veličinu skupa, ali ne i njegovu statističku raznolikost.

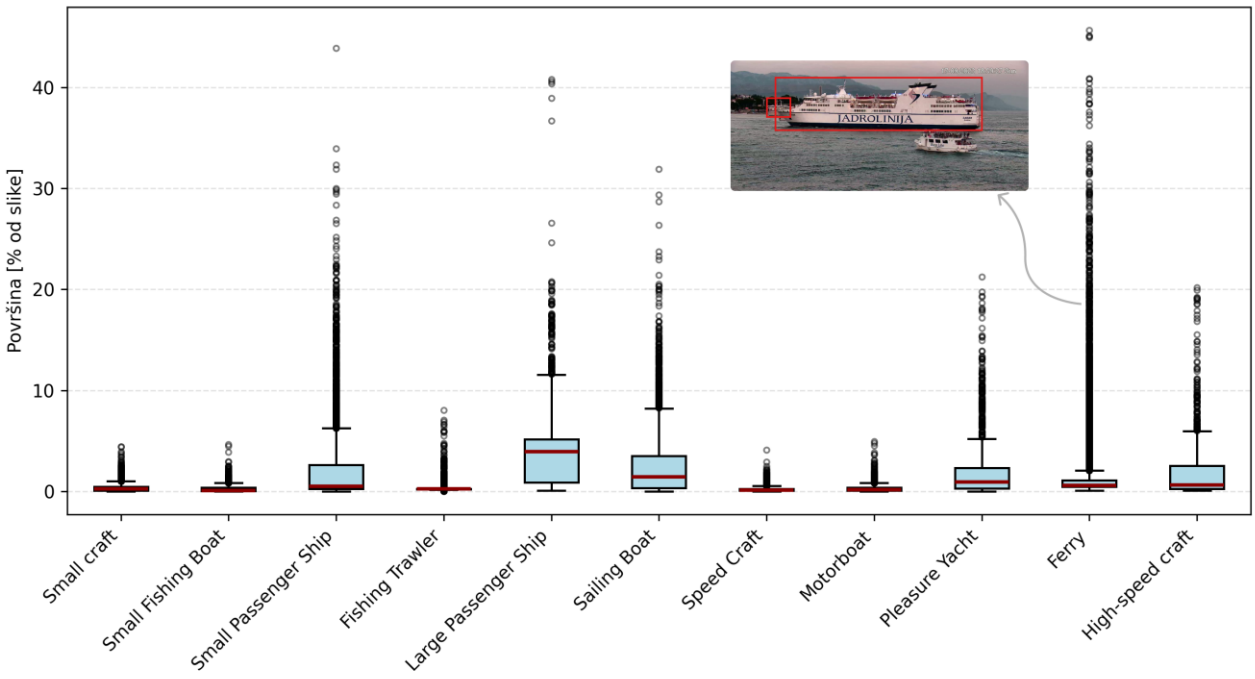
Tablica 4.3. Raspodjela instanci u bazi slika.

Klasa	**Broj instanci SSMOT detekcija**
Small Craft	1.885
Small Fishing Boat	1.372
Small Passenger Ship	4.410
Fishing Trawler	5.401
Large Passenger Ship	5.498
Sailing Boat	4.206
Speed Craft	5.837
Motorboat	2.455
Pleasure Yacht	1.777
Ferry	15.799
High-speed Craft	1.797
UKUPNO	50.437

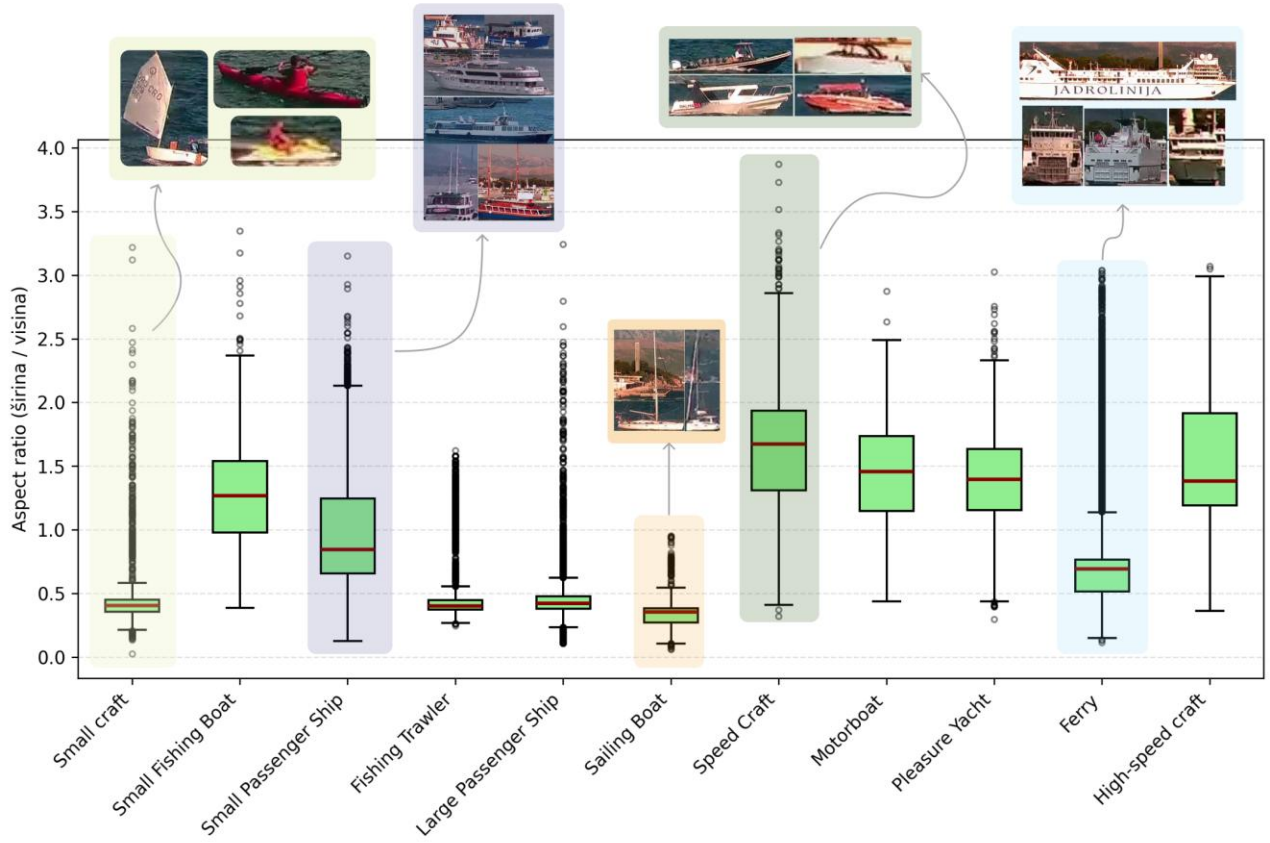
Najzastupljeniji je Ferry s 15.799 instanci. Takva neravnoteža ima stvarno prometno objašnjenje: trajekti redovito prometuju i često dugo ostaju vezani unutar kadra. Međutim, algoritam treniran bez korektivnih strategija mogao bi postati pristran prema dominantnim klasama. Za znanstveno korektnu evaluaciju zato je, uz ukupnu mAP vrijednost, potrebno prikazivati i AP po klasi te uspješnost na malenim plovilima i zaklonjenim objektima. Opisani zaključci proizlaze iz distribucija u disertaciji; ne predstavljaju novo provedene eksperimente [9].



Slika 4.3. Primjeri svih 11 vizualnih kategorija SSMOT-a [9].



Slika 4.4. Box-plot dijagrami površina graničnih okvira po klasama [9].



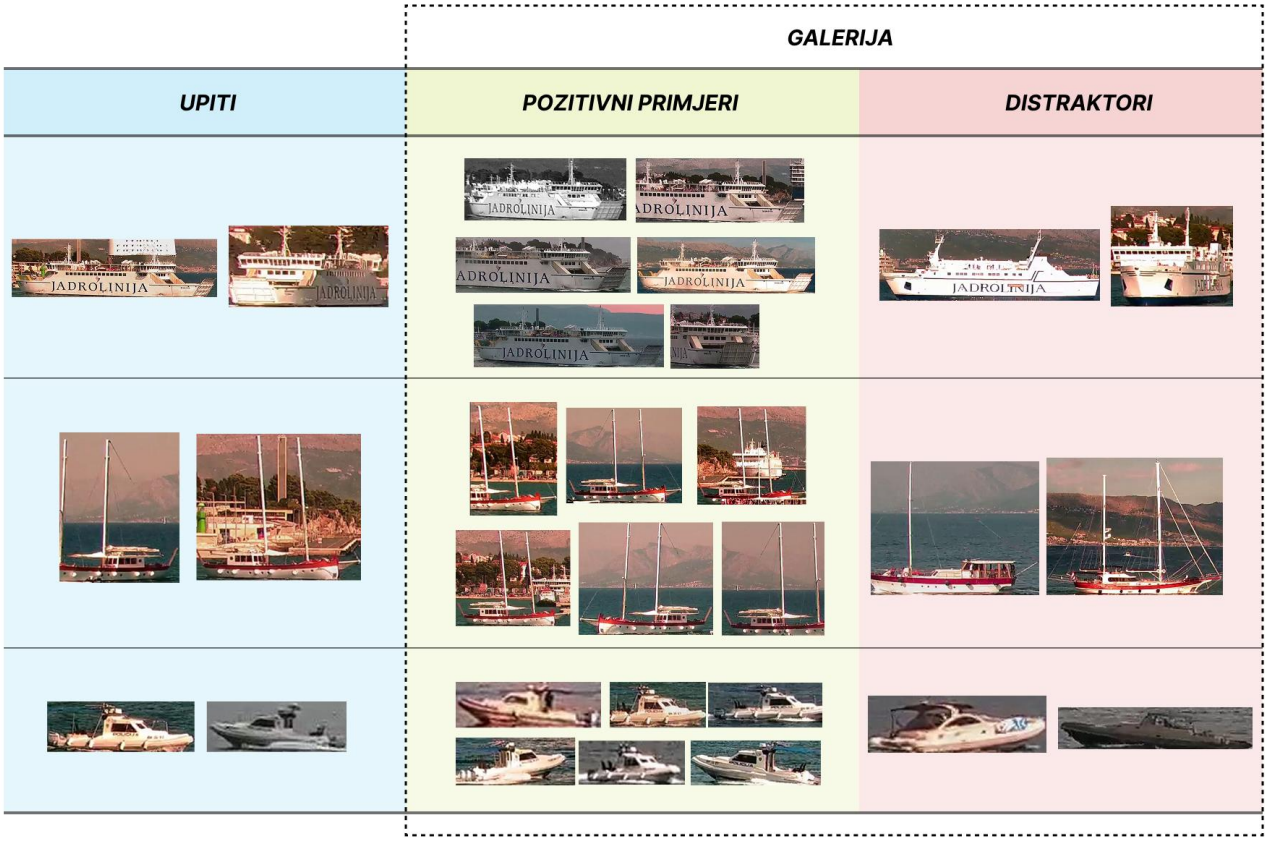
Slika 4.5. Box-plot dijagram omjera širine i visine grani˘cnih okvira po klasama [9].

4.4.4 Reidentifikacijski podskup: fizički identitet iznad kategorije

SSMOT ReID podskup uključuje 281 različitih plovila, pri čemu je za svaki identitet prikupljeno barem osam vizualno različitih prikaza. Ukupno 220 identiteta izdvojeno

je za učenje, a 61 za testiranje. U skupu za učenje nalazi se 5.227 slika; skup upita sastoji se od 122 slike (dvije po testnom identitetu), dok galerija obuhvaća 488 slika (šest pozitivnih i dva distraktora po testnom identitetu). Broj slika pojedinog identiteta u učenju kreće se od 8 do 128; 54 % identiteta ima između 8 i 20 slika. Time je omogućeno provjeravati koliko se opisnik dobro prenosi na različite položaje, osvjetljenje i udaljenost istoga broda [9].

ReID evaluacija ima drukčiju logiku od klasifikacije plovila. Pripadnost istoj klasi, primjerice Ferry, ne podrazumijeva isti identitet. Za zadanu upitnu sliku potrebno je rangirati odgovarajuće slike istoga fizičkog plovila ispred vizualno sličnih distraktora. Usto, uzorci iz jednoga videa ne smiju nehotice procuriti između skupova za učenje i testiranje na način koji olakšava prepoznavanje statične pozadine umjesto broda. Premda je u SSMOT-u provedena podjela identiteta, procjenu generalizacije na potpuno novu kameru ili novu luku treba tretirati kao zaseban istraživački zadatak.



Slika 4.6. Primjer ReID protokola: upitne slike, pozitivni primjeri istog plovila i vizualno slični distraktori u galeriji [9].

4.4.5 Video podskup za praćenje više plovila

SSMOT MOT skup sastoji se od 18 stvarnih videozapisa prikupljenih u kolovozu i rujnu 2023. Traju ukupno 9 minuta i 34 sekunde, sadrže 12.946 kadrova u rezoluciji 1280 × 720, 149 anotiranih putanja plovila i 91.006 anotiranih graničnih okvira, odnosno prosječno 7,0 objekata po kadru. Većina videozapisa snimljena je brzinom od 25 FPS, a pojedini pri 10 ili 30 FPS. Sekvence variraju od jednostavnijih do onih s velikom gustoćom, mimoilaženjima i sličnim plovilima. Primjer SSMOT_12 posebno ilustrira istodobno praćenje više vizualno sličnih objekata [9].

Anotacije su izrađene u alatu CVAT te izvožene u MOT 1.1 formatu: frame, id, x, y, width, height, confidence, class, visibility. Prisutnost oznake identiteta kroz sve kadrove omogućuje računanje standardnih MOT metrika. U disertaciji je visibility zapisan binarno: nula za potpuno zaklonjeni objekt, a jedan za vidljiv objekt. Zbog toga iz samog binarnog polja nije moguće odrediti finu razinu djelomične zaklonjenosti; ako bi novo istraživanje željelo mjeriti progresivan ulazak u okluziju, potrebna je dopuna anotacija razinom vidljivosti, stupnjem preklapanja i/ili maskama objekata.

4.4.6 Okluzije kao najvažniji dio SSMOT-a za analizu identiteta

U videozapisima je zabilježena ukupno 31 potpuna okluzija. Najkraća traje 0,4 s, a najduža 29,05 s. Četrnaest okluzija traje dulje od pet sekundi, a tri dulje od dvadeset sekundi. Ivanina disertacija uvodi tri kategorije prema trajanju: kratka do 2 s (13 događaja; 41,9 %), srednje duga više od 2 do 8 s (8 događaja; 25,8 %) i duga iznad 8 s (10 događaja; 32,3 %). Posebno upečatljiv primjer jest sekvenca SSMOT_14 u kojoj trajekt zaklanja jedrilicu 29,05 s [1, str. 76–78].

Tablica 4.4. Raspodjela događaja unutar videa.

Kategorija potpune okluzije	**Trajanje**	**Broj događaja**	**Udio**
Kratka	do 2 s	13	41,9 %
Srednje duga	više od 2 do 8 s	8	25,8 %
Duga	više od 8 s	10	32,3 %
Ukupno	0,4–29,05 s	31	100 %

Ovakav sadržaj pogoduje evaluaciji kontinuiteta identiteta prije i nakon nestanka vidljivosti, ali je važno razlikovati nekoliko problema. Metrika MOTA računa propuštene detekcije, lažne pozitivne detekcije i zamjene identiteta, dok IDF1 i asocijacijske komponente HOTA-e osjetljivije opisuju kontinuitet uparivanja. U sekvencama sa samo nekoliko dugih okluzija ukupna zbirna MOTA može ostati visoka i kada se nakon jednog dugog nestanka dogodi zamjena identiteta. Zbog toga okluzije valja izvještavati kao zaseban stratificirani podskup, uz broj uspješnih povrataka izvornog ID-a, vrijeme oporavka i broj pogrešnih reidentifikacija.



Slika 4.7. Usporedni primjeri kratke, srednje duge i dugotrajne okluzije iz SSMOT videozapisa [9].

4.5 Noviji relevantni skupovi

4.5.1 ShipsMOT (2025.)

ShipsMOT je novija video baza iz 2025. s 121 sekvencom, od kojih je 81 namijenjena učenju, a 40 testiranju. Prema objavljenom radu sadrži 608 putanja i 237.999 anotiranih graničnih okvira u 15 kategorija. Izvori uključuju sustave pomorskog nadzora i internetski dostupne videozapise, s kadrovima razlučivosti 1920 × 1080 ili 1280 × 720. Autori posebno navode međuklasnu i unutar-klasnu sličnost, promjene gledišta, okluzije i gustoću plovila [25, 26]. Za razliku od SSMOT-a, ShipsMOT je veći skup praćenja, no izvorna diferencijacija između klasifikacije, ReID galerije i kontinuiranog MOT testa nije ista kao u SSMOT-u.

4.5.2 ViV-ReID (2025./2026.)

ViV-ReID je predstavljen kao opsežan video-bazirani skup za reidentifikaciju plovila preko različitih kamera i luka. Prema autorima sadrži 480 identiteta, 7.165 kratkih putanja (tracklets), približno 1,14 milijuna kadrova i 20 pogleda kamera u različitim lukama [27]. Njegova je istraživačka vrijednost činjenica da umjesto pojedinačnih izrezanih slika sadrži vremenski strukturirane sekvence identiteta, zbog čega je prikladan za modele koji iskorištavaju promjene orijentacije, dijelova trupa i dugoročne prostorno-vremenske značajke. Usporedba sa SSMOT-om korisna je za pitanje prijenosa ReID modela iz jedne luke na višekamerni sustav, ali ti skupovi nisu međusobno izravno zamjenjivi bez prilagodbe protokola.

4.5.3 HOSS ReID (2025.) i ostale multimodalne baze

HOSS ReID usmjeren je na prepoznavanje istog broda na optičkim i SAR satelitskim snimkama iz različitih vremenskih trenutaka i kutova snimanja. Dostupan je na Zenodu i ima bitno drugačiju domenu od obalnog RGB videa [28]. Stoga predstavlja zanimljiv primjer razvoja multimodalnih reprezentacija, ali nije izravan benchmark za praćenje brodova iz jedne fiksne lučke kamere. MMDEC (2026.) agregira satelitske SAR detekcije, AIS i kontekstne prostorno-vremenske slojeve na području Engleskog kanala te je prikladniji za istraživanje multisenzorskog povezivanja tragova i rekonstrukciju kretanja nego za evaluaciju pikselnih ReID vektora iz RGB snimaka [27, 29].

Tablica 4.5. Karakteristike HOSS ReID.

Skup	Zadatak	Glavni objavljeni opseg	Prednost za budući rad
SSMOT (2025)	detekcija + ReID + MOT	9.827 slika; 281 ReID ID; 18 videozapisa	kontrolirane lučke okluzije
ShipsMOT (2025)	MOT	121 videa; 237.999 okvira objekata	šira raznolikost plovila
ViV-ReID (2025./26.)	video ReID	480 ID; 7.165 trackleta	višekamerni identiteti
HOSS ReID (2025)	optički/SAR ReID	različiti senzori i vremena	multimodalna robustnost
MMDEC (2026)	SAR + AIS + kontekst	tromjesečni multimodalni period	provjera tragova uz AIS

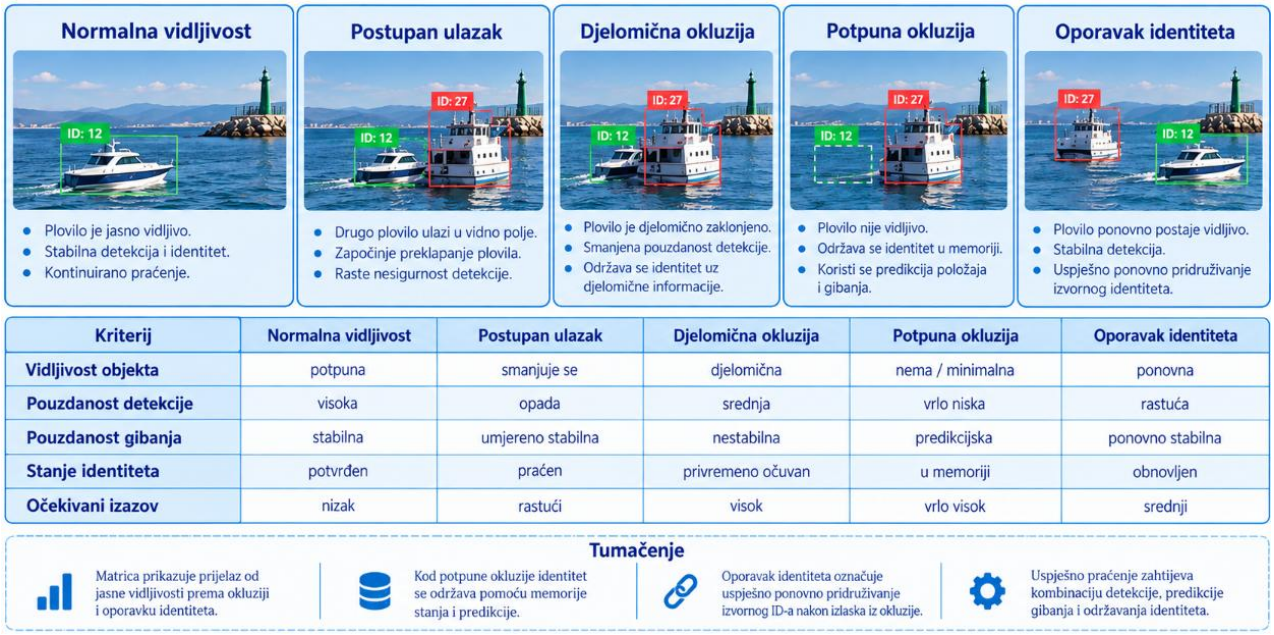
4.6 Kritička usporedba primjenjivosti baza za buduću evaluaciju

Izbor baze ne treba temeljiti samo na njezinoj veličini. Za detektor su važni raznolikost klasa, veličina graničnih okvira, pokrivenost različitih svjetlosnih uvjeta te konzistentne kategorije. Za reidentifikaciju su ključni broj identiteta, broj pozitivnih primjera po identitetu, te broj i težina distraktora. Za praćenje su nužni kontinuirani videozapisi, oznake istoga identiteta kroz vrijeme, dovoljno gusta istodobna pojava objekata te eksplicitno prepoznatljivi događaji okluzije. Kombiniranje različitih baza bez usklađivanja klasa, rezolucija, FPS-a i protokola evaluacije može dati prividno precizne, ali metodološki neusporedive vrijednosti.

Za evaluaciju sustava s ciljanim upravljanjem okluzijom SSMOT je posebno relevantan zbog dokumentiranih potpunih okluzija i jedinstvenih identiteta. Međutim, 31 događaj i jedna dominantna lučka lokacija ograničavaju zaključivanje o svim mogućim pomorskim uvjetima. Uvođenje dodatnih anotacija postupnog ulaska i izlaska iz okluzije, stupnja vidljivosti, pouzdanosti mjerenja položaja i kontaminacije pozadinom omogućilo bi precizniju analizu ponašanja filtriranja gibanja i vizualnih značajki. Takva dopuna bila bi proširenje evaluacijskog protokola, a ne nužno stvaranje nove taksonomije plovila ili novog algoritma.

Predlaže se stratificirano izvještavati rezultate prema trajanju okluzije, broju plovila u kadru, veličini objekata, vizualnoj sličnosti i promjeni osvjetljenja. Važno je odvojiti detekcijske pogreške od asocijacijskih: sustav koji ne vidi plovilo ne može pokazati isti

obrazac pogrešaka kao sustav koji plovilo vidi, ali mu dodjeljuje pogrešan identitet. Zato bi se uz mAP, MOTA, IDF1 i HOTA trebale koristiti metrike uspjeha povratka identiteta nakon okluzije i statistike pogrešnih pridruživanja po događaju. Ova preporuka predstavlja metodološku sintezu, a ne tvrdnju da su takvi eksperimenti već provedeni.



Kriterij	Normalna vidljivost	Postupan ulazak	Djelomična okluzija	Potpuna okluzija	Oporavak identiteta
Vidljivost objekta	potpuna	smanjuje se	djelomična	nema / minimalna	ponovna
Pouzdanost detekcije	visoka	opada	srednja	vrlo niska	rastuća
Pouzdanost gibanja	stabilna	umjereno stabilna	nestabilna	predikcijska	ponovno stabilna
Stanje identiteta	potvrđen	praćen	privremeno očuvan	u memoriji	obnovljen
Očekivani izazov	nizak	rastući	visok	vrlo visok	srednji

Slika 4.8. Predložena usporedna matrica uvjeta: normalna vidljivost, postupan ulazak, djelomična okluzija, potpuna okluzija i oporavak identiteta (Generirano AI).

4.7 Zaključak

Pregled javnih i istraživačkih skupova podataka pokazuje da veliki broj slika nije dovoljan za znanstveno vrednovanje praćenja identiteta. SSMOT je vrijedan jer povezuje označene fotografije za detekciju, provjeru ReID opisnika i kontinuirano praćenje u realnoj splitskoj luci, s posebno odabranim dugotrajnim okluzijama. Noviji ShipsMOT i ViV-ReID donose dodatnu širinu i nove izazove te omogućuju razmatranje generalizacije izvan jednog videoizvora. Upravo razlike među tim podatkovnim strukturama argument su za transparentan eksperimentalni protokol i detaljnu analizu pogrešaka po scenarijima, što predstavlja prirodan prijelaz prema završnom poglavlju o mogućnostima daljnjih istraživanja.

5 EKSPERIMENTALNA ANALIZA VIZUALNE SLIČNOSTI KLASA PLOVILA I EVALUACIJA YOLO DETEKTORA

U ovom poglavlju izdvajaju se i kritički interpretiraju rezultati prethodno provedenoga Istraživačkog seminara „Utjecaj kvalitete anotacije na točnost i odziv detekcije" (Muslim, 2025.). Za razliku od prethodnog pregleda pomorskih skupova podataka, ovdje se ne ponavlja način nastanka i prikupljanja Ivanina skupa, nego se razmatraju poteškoće nastale zbog vizualne sličnosti klasa, posljedice pojednostavljenja klasne taksonomije te utjecaj ulazne rezolucije na YOLO detektor.

5.1 Polazište i istraživačko pitanje

Za praćenje više plovila pouzdana detekcija mora osigurati stabilnu lokalizaciju objekta kroz uzastopne kadrove. Detektor koji pouzdano nalazi plovilo, ali mu često mijenja oznaku vrste, stvara dodatne poteškoće ako se klasna kompatibilnost koristi pri pridruživanju putanja. Problem je izražen kod udaljenih objekata: broj piksela na kojima se zasniva razlikovanje maloga ribarskog broda, motornog čamca, glisera ili jedrilice može biti nedostatan za preciznu vizualnu klasifikaciju.

U Istraživačkom seminaru analizirana je početna shema s 11 klasa plovila, pri čemu su se posebno istaknule zamjene između klasa Ferry i Large/Small Passenger Ship te Speed Craft, Small Craft, Small Fishing Boat i Motorboat. Uočeno je da dio odluka o vrsti plovila ovisi o kontekstu (primjerice valu iza broda), a ne samo o dovoljno razlučivim obilježjima samoga trupa. Takve razlike mogu biti teško uočljive na statičnim kadrovima, osobito pri velikim udaljenostima i smanjenoj rezoluciji.

Središnje pitanje ovoga dijela stoga nije koliko je pomorska klasifikacija semantički detaljna, nego koliko je odabrana taksonomija pouzdano razlikovljiva na slikama koje sustav stvarno prima. Redukcija broja klasa može ukloniti međuklasne zamjene, ali pritom gubi detaljnu semantičku informaciju. Zbog toga se rezultati jednoklasne detekcije moraju tumačiti kao rezultati drugog zadatka, a ne kao izravno nadmoćna klasifikacija istih 11 kategorija.

5.2 Eksperimentalni skup i postupak anotiranja – relevantne razlike

Detaljan opis izvora i nastanka pomorskih slika naveden je u prethodnom poglavlju o skupovima podataka. Za razumijevanje eksperimenata dovoljno je istaknuti da Istraživački seminar navodi 3.131 sliku s ukupno 13.680 označenih objekata te podjelu 2.661 slike za treniranje i 470 za validaciju (85 : 15). U samom eksperimentalnom dijelu naveden je i omjer 90 : 10; ta dva zapisa potrebno je uskladiti s izvornim konfiguracijama izvođenja prije konačne publikacije.

Provjera oznaka provedena je u CVAT-u, a anotacije su zapisane u YOLO formatu kao klasa i normalizirane koordinate graničnog pravokutnika. Za potrebe ove analize važna je dosljednost označavanja vizualno sličnih plovila: čak i točno postavljen pravokutnik ne rješava problem kada dva anotatora istu daleku brodicu mogu razumno pridružiti različitim klasama. Seminar također opisuje korekciju pretjerano širokih pravokutnika i nedostajućih oznaka, što utječe na detekciju neovisno o broju klasa.

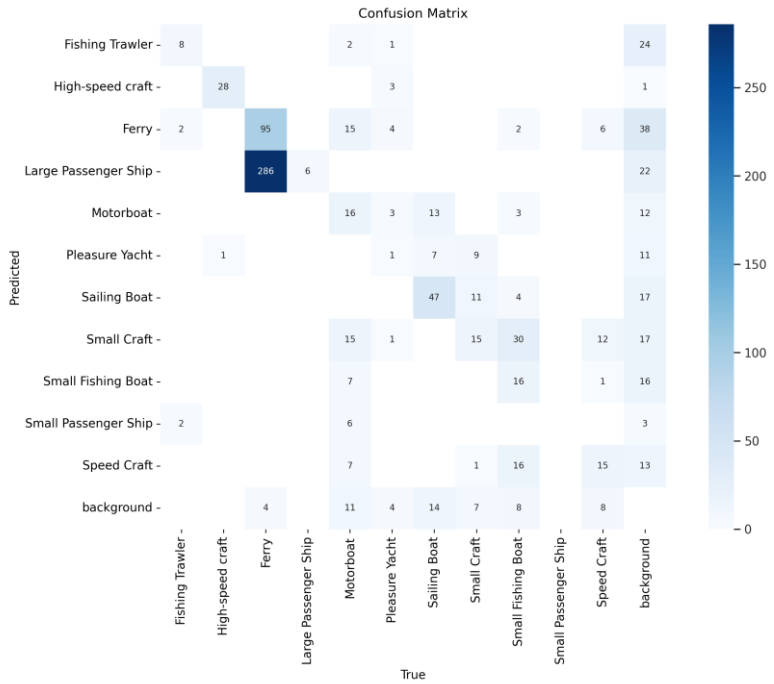
Shematska usporedba vizualno sličnih klasa može se dopuniti u završnoj verziji prikazima brodova iz izvornoga skupa; već provedene evaluacijske slike koje slijede predstavljaju izravniji empirijski dokaz navedenog problema.

5.3 Višeklasna detekcija: YOLO11m na 640 px

Prvi prikazani pokus treniranja modela YOLO11m koristio je 11 klasa, ulaznu veličinu imgsz=640, 100 epoha, batch=16 te više GPU uređaja NVIDIA GeForce GTX 1080 Ti. Analiza konfuzijske matrice u seminaru pokazuje da model bolje izdvaja veće i vizualno karakterističnije brodove (primjerice Large Passenger Ship i Fishing Trawler), dok se male i međusobno slične kategorije često pogrešno klasificiraju ili ih model ne detektira. Razlike između klasa Speed Craft i Small Craft te Ferry i putničkih brodova posebno su problematične.

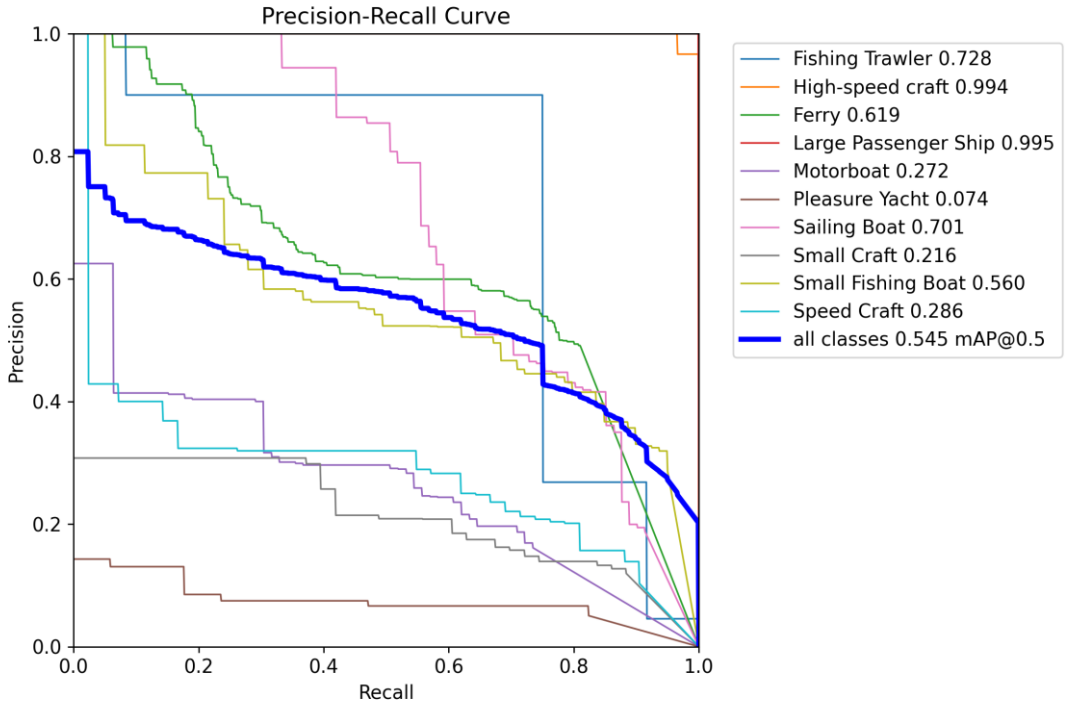
U odjeljku o rezultatima za 640 px navode se približne vrijednosti mAP@0,5 ≈ 0,45 i mAP@0,5:0,95 ≈ 0,35, uz nestabilnost validacijskih krivulja. U kasnijem usporednom odjeljku istog seminara nalaze se drugačije vrijednosti za referentni model na 640 px (mAP@0,5 ≈ 0,24 i mAP@0,5:0,95 ≈ 0,11). Te brojke nisu međusobno konzistentne. Za kvalifikacijski rad zato se naglasak stavlja na dokumentirani kvalitativni zaključak i ne izvodi se numeričko poboljšanje između rezolucija na temelju neusuglašenih zapisa.

Razdvajanje regresijske i klasifikacijske komponente posebno je važno. Seminar navodi uredno smanjenje box loss i DFL loss vrijednosti, dok je klasifikacijski gubitak nestabilniji. To podupire tumačenje da je velik dio ograničenja vezan uz semantičko razdvajanje sličnih klasa, no samo iz tih krivulja nije moguće dokazati da lokalizacijski problemi ne postoje.



Slika 5.1. Konfuzijska matrica višeklasnog modela YOLO11m (11 klasa, imgsz = 640 Izvor: [30].

Matrica pokazuje koncentraciju ispravnih predikcija u dominantnim klasama i brojne propuste te zabune kod vizualno sličnih i slabije zastupljenih malih plovila. Zbog različitih učestalosti klasa apsolutni brojevi nisu izravna mjera njihove uspješnosti: potrebna je interpretacija po retku, odnosno odziva svake klase.



Slika 5.2. Precision–Recall krivulje po klasama za višeklasni model na 640 px. Izvor: [30].

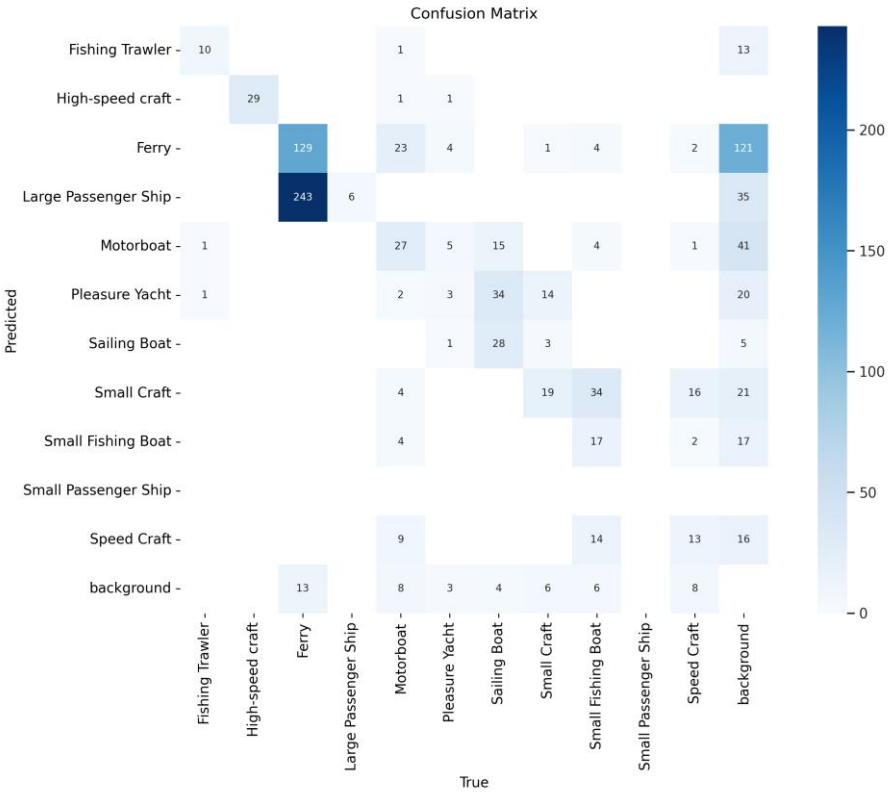
Razdvojene PR krivulje pokazuju neravnomjernu sposobnost modela da istodobno postiže visoku preciznost i odziv za pojedine vrste plovila. Time se dodatno potvrđuje da je ukupna vrijednost mAP-a nedovoljna za razumijevanje pogrešaka između vizualno sličnih klasa.

5.4 Višeklasna detekcija pri povećanoj rezoluciji

U drugom pokusu veličina ulaza povećana je na imgsz=1920 uz smanjenje batch veličine na osam, što je omogućilo izvođenje na dostupnim GPU karticama. U seminaru se navodi približno 11 GB iskorištene memorije te trajanje jedne epohe oko 8 minuta i 40 sekundi. Za razliku od modela na manjoj rezoluciji, model s većim ulaznim slikama zadržava više detalja udaljenih i sitnih objekata. Prema opisnoj analizi konfuzijske matrice, osjetljivost na neke male klase raste, ali se međuklasne zamjene ne uklanjaju.

Za ovaj pokus navedeni su mAP@0,5 ≈ 0,42 (s povremenim vrijednostima do 0,45) i mAP@0,5:0,95 ≈ 0,32–0,34. Zbog prethodno navedenih proturječja u izvještavanju rezultata na 640 px te različite batch veličine zaključak treba ograničiti: povećana rezolucija može olakšati opažanje malih plovila, ali uz znatno viši računarski trošak i bez jamstva razdvajanja vrlo sličnih klasa.

Kod evaluacije se u sljedećim istraživanjima preporučuje zasebno prijaviti rezultate za male, srednje i velike objekte, profiliranje latencije i memorije te provjeru istih podjela na treniranje i validaciju. Time bi se mogao odvojiti učinak veličine slike od promjena u uvjetima učenja.



Slika 5.3. Konfuzijska matrica višeklasnog modela YOLO11m (11 klasa, imgsz = 1920) Izvor: [30] .

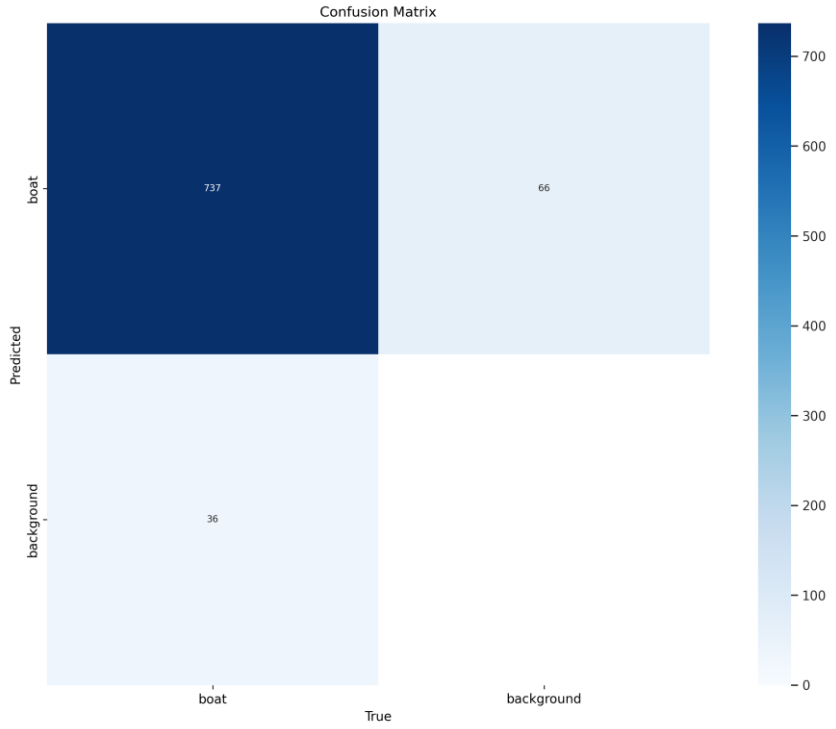
Veća ulazna veličina omogućuje bolju reprezentaciju malih objekata, ali prikazana konfuzijska matrica i dalje pokazuje problem slabije razlikovljivih kategorija. Izravna numerička usporedba s 640 px zahtijeva provjeru izvornog evaluacijskog izvoza zbog neusuglašenih mAP vrijednosti u seminaru.

5.5 Redukcija klasa: eksperiment s jedinstvenom klasom „plovilo”

Treći i četvrti pokus u seminaru predstavljaju radikalnu redukciju taksonomije: svih 11 početnih klasa spojeno je u jednu zajedničku klasu, „brod” (ship/boat). Model tada ne pokušava razlikovati plovila prema tipu, nego samo pronaći svaki brod u sceni i precizno odrediti njegov okvir. Ova promjena nije ekvivalentna uklanjanju dijela slika ili brodova iz skupa; riječ je prvenstveno o preslikavanju klasnih oznaka uz zadržavanje zadatka lokalizacije.

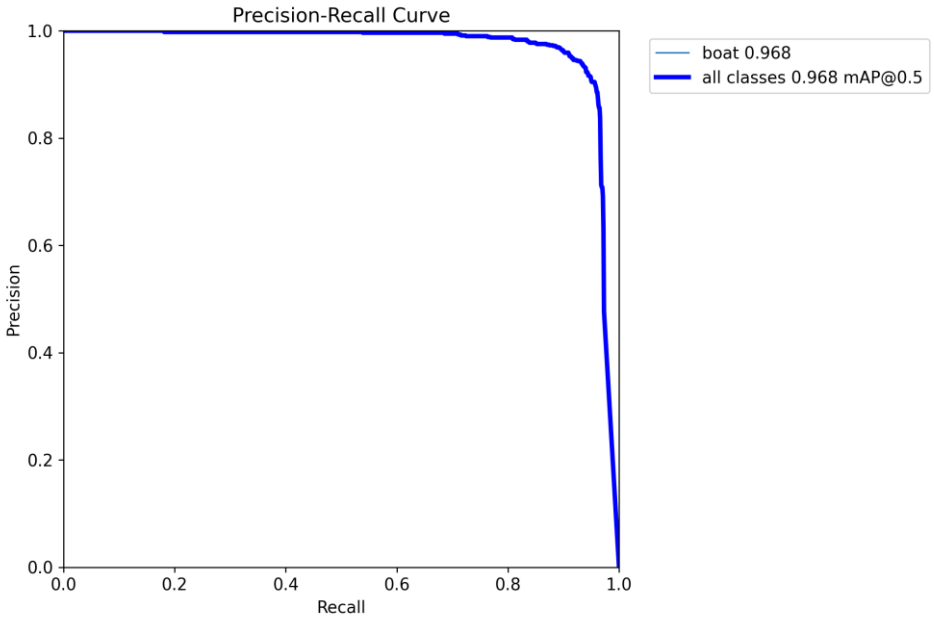
Za jednoklasni pokus pri imgsz=640 u seminaru su navedeni mAP@0,5 ≈ 0,968 i mAP@0,5:0,95 ≈ 0,80. Opisane su i vrijednosti precision ≈ 0,92 i recall ≈ 0,95 iz konfuzijske matrice, uz F1 maksimum oko 0,94. Za jednoklasni pokus pri imgsz=1920 navode se približno mAP@0,5 ≈ 0,85–0,90, mAP@0,5:0,95 ≈ 0,70–0,75, precision ≈ 0,90 i recall ≈ 0,85. Vrijednosti su prenesene kao približne procjene iz prikazanih rezultata, a ne kao ponovno izračunati eksperimentalni podaci.

Viši rezultati jednoklasne konfiguracije jasno upućuju na to da se smanjuje teret semantičkog razdvajanja izrazito sličnih objekata. Međutim, mAP iz jednoklasnog i jedanaestoklasnog zadatka nije istovrsna mjera težine klasifikacije: jednoklasni model više ne može pogriješiti između klasa Ferry i Passenger Ship. Stoga je prikladnije govoriti o poboljšanju kvalitete detekcije prisutnosti plovila za pojednostavljeni zadatak nego o dokazanoj nadmoći u prepoznavanju pojedinačnih tipova.



Slika 5.4. Konfuzijska matrica YOLO11m nakon objedinjavanja oznaka u jednu klasu „brod” (imgsz = 640) Izvor: [30].

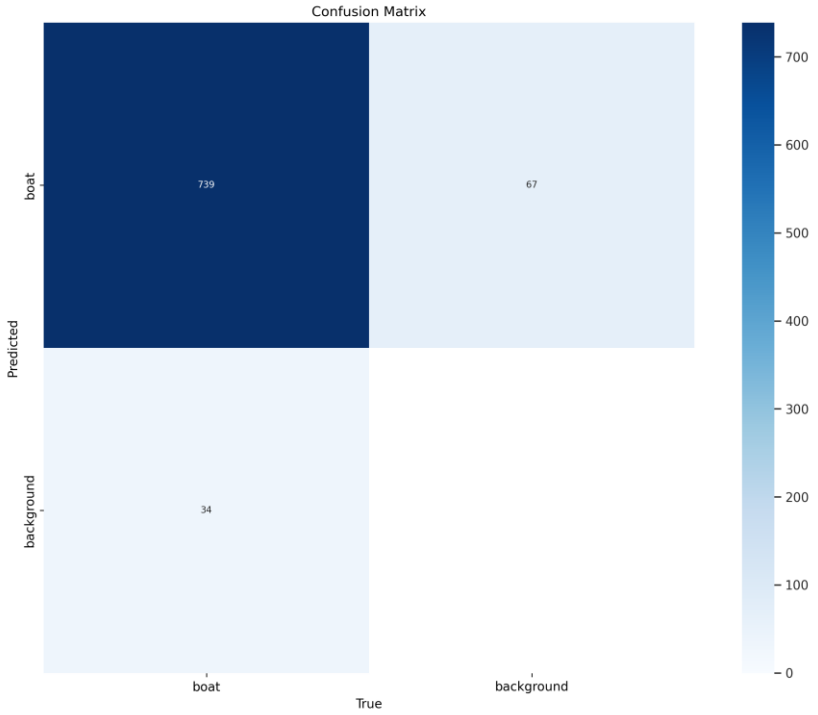
U jednoklasnom zadatku nestaju međuklasne zamjene jer se svako plovilo označava jednakom oznakom. Preostale pogreške odnose se primarno na zamjenu objekta i pozadine, odnosno na propuštene i lažne detekcije. Ova je evaluacija zato informativna za lokalizaciju, ali nije dokaz sposobnosti klasifikacije vrste plovila.



Slika 5.5. Precision–Recall krivulja za jednoklasnu detekciju pri 640 px. Izvor: [30].

Prema legendi izvornog grafa ostvarena je vrijednost AP@0,5 ≈ 0,968. Usporedba sa slikom 5.2 pokazuje koliko je zadatak otkrivanja bilo kojega broda jednostavniji od

klasnog razvrstavanja. Međutim, jednoklasni eksperiment koristi prijenosno učenje iz prethodno treniranih težina, pa učinak redukcije klasa nije eksperimentalno potpuno izoliran.



Slika 5.6. Konfuzijska matrica jednoklasnog modela na 1920 px. Izvor: [30].

Četvrti eksperiment zadržava jedinstvenu oznaku plovila uz veću veličinu ulazne slike. Valja napomenuti da je i ovdje korišteno daljnje treniranje postojećih težina, a ne potpuno neovisno inicijaliziran model; usporedba različitih konfiguracija zato služi prikazu smjera razvoja, ne strogo kontroliranoj ablacijskoj procjeni.

5.6 Usporedni prikaz provedenih pokusa

Klasna shema	**Ulaz**	**mAP@0,5 (seminar)**	**mAP@0,5:0,95 (seminar)**
11 klasa	640	≈ 0,45*	≈ 0,35*
11 klasa	1920	≈ 0,42	≈ 0,32–0,34
1 klasa	640	≈ 0,968	≈ 0,80
1 klasa	1920	≈ 0,85–0,90	≈ 0,70–0,75

* U kasnijoj usporedbi u seminaru za model 11 klasa/640 px navode se druge približne vrijednosti (≈ 0,45 i ≈ 0,35). Prije konačnog kvantitativnog zaključka potrebno je provjeriti izvorne izvozne rezultate i eksperimentalne konfiguracije.

Sva četiri pokusa, prema prikazu seminara, temelje se na varijanti YOLO11m. Iako uvod seminara spominje plan usporedbe YOLO11s i YOLO11m, prikazani i razrađeni rezultati ne daju zaseban, usporediv skup brojčanih pokazatelja za YOLO11s. Zbog toga se iz ovog izvora ne zaključuje o relativnoj točnosti s/m varijanti.

5.7 Metodološka ograničenja i implikacije za praćenje plovila

Pri tumačenju Slika 5.1.–5.6. nužno je razdvojiti dvije različite vrste usporedbe. Promjena ulazne rezolucije može se analizirati samo uz provjeru iste validacijske podjele i postavki inferencije. Spajanje klasa mijenja samu definiciju zadatka i broj izlaznih kategorija, dok su jednoklasni pokusi u seminaru dodatno koristili prethodno naučene težine. Slijedom toga nije opravdano sve razlike AP/mAP vrijednosti pripisati isključivo kvaliteti anotacije ili redukciji broja klasa.

Usporedba je korisna kao prethodno provedena analiza, ali nije još potpuna kontrolirana ablacijska studija: u zapisu postoje različiti omjeri podjele skupa, neusuglašene vrijednosti pojedinih mAP pokazatelja i razlike u batch veličini. Također, za model na 1920 px s jednom klasom seminar upućuje na korištenje prethodno treniranih utega, pa se u interpretaciji mora uzeti u obzir način inicijalizacije. Uočena poboljšanja ne treba automatski pripisati isključivo rezoluciji ili samo broju klasa.

Za MOT sustav posljedice su dvojake. Detektor jedne klase može stabilizirati ulazni skup opažanja jer se isti brod neće naizmjence označavati različitim tipovima; to može olakšati obradu putanja kada tracker koristi klase kao uvjet uparivanja. S druge strane, uklanjanje oznake vrste smanjuje dostupne semantičke informacije pri razdvajanju susjednih vizualno sličnih objekata. Za konačan izbor potrebna je procjena kompletnog lanca detekcije i praćenja, s jednakim testnim videima i metrikama MOTA, IDF1, HOTA, AssA te stopom uspješnog oporavka identiteta nakon okluzije.

Zaključak da bi dvostupanjska arhitektura — opći detektor plovila pa zaseban klasifikator na izrezanim objektima — mogla biti prikladna formuliran je i u završnom dijelu seminara. Valja ga zadržati kao utemeljenu istraživačku smjernicu, a ne kao već verificiran rezultat: seminar ne sadrži zasebnu eksperimentalnu evaluaciju takvog sekundarnog klasifikatora niti mjerenja njegove latencije, robusnosti i doprinosa praćenju.

5.8 Zaključak poglavlja i poveznica s budućim istraživanjima

Eksperimenti prikazani u [30] ukazuju na to da su vizualno slične i slabo zastupljene klase plovila važan izvor pogrešaka višeklasne detekcije. Povećanje ulazne rezolucije povećava dostupnost detalja, ali i računalne zahtjeve, dok jednoklasna detekcija u prikazanim pokusima daje visoke vrijednosti metrika za zadatak lokalizacije svih plovila. Takav rezultat opravdava razmatranje jednostavnije klasne sheme ili hijerarhijskog postupka detekcije i naknadne klasifikacije, ali ne ukida potrebu za provjerom očuvanja identiteta kroz vrijeme.

6 ZAKLJUČAK

Pregled literature i tehnoloških rješenja pokazuje da pouzdano praćenje plovila predstavlja složeniji zadatak od povezivanja uzastopnih detekcija. Geometrija, predikcija gibanja i vizualne značajke daju komplementarne informacije, ali njihova pouzdanost nije jednaka u svim okolnostima. U normalnoj vidljivosti i pri stabilnom gibanju jednostavniji pristupi mogu biti učinkoviti, dok u uvjetima djelomične i dugotrajne okluzije dolazi do odstupanja detekcijskih pravokutnika, akumulacije pogrešaka modela gibanja i moguće kontaminacije ReID opisa. Posebno su zahtjevne scene u kojima se međusobno zaklanjaju brodovi sličnih silueta, boja i dimenzija.

Analizirani algoritmi, od SORT-a i DeepSORT-a do OC-SORT-a, StrongSORT-a i novijih metoda namijenjenih okluzijama, nude različite mehanizme predikcije i ponovnog pridruživanja. Nijedan se pristup, međutim, ne može smatrati univerzalno najboljim za pomorsku domenu bez evaluacije na reprezentativnim snimkama. Stoga uz MOTA metriku treba promatrati IDF1, HOTA i asocijacijske pokazatelje, broj zamjena identiteta te stopu ispravnog povratka izvornog identiteta nakon okluzije. Usporedba mora razlikovati kvalitetu same detekcije od kontinuiteta identiteta i učinka cijelog sustava.

Pregled NVIDIA DeepStream SDK-a pokazuje da je riječ o prikladnoj infrastrukturi za GPU-ubrzanu videoanalitiku i integraciju primarne i sekundarne inferencije, praćenja objekata te obrade metapodataka. Njegova stvarna primjenjivost ipak ovisi o konfiguraciji, broju izvora, veličini serija, latenciji, hardverskim resursima i svojstvima pojedinog trackera. Analiza pomorskih skupova podataka dodatno potvrđuje važnost odgovarajućih anotacija, raznolikih perspektiva i posebnih scenarija okluzije, pri čemu SSMOT predstavlja relevantno polazište za sustavno vrednovanje praćenja i reidentifikacije plovila.

Provedena istraživanja prikazana u Istraživačkom seminaru ukazala su na izraženu zabunu između vizualno sličnih kategorija plovila te na povoljnije detekcijske rezultate nakon objedinjavanja klasa u pojednostavljeni zadatak. To ne dokazuje da je jedna klasa optimalna za sve primjene niti da se uočene razlike mogu pripisati isključivo promjeni oznaka; potrebno je uzeti u obzir i konfiguracije treniranja te razlike u evaluacijskim zadacima. Ipak, nalazi opravdavaju daljnje ispitivanje vizualno utemeljene taksonomije i odvajanja pouzdane detekcije plovila od naknadnog određivanja njegovih atributa.

Provedena analiza potvrđuje da pouzdanost praćenja plovila ovisi o međusobnoj povezanosti kvalitete detekcije, modela gibanja, postupka pridruživanja i vizualne reidentifikacije. Posebno zahtjevne okolnosti predstavljaju djelomične i dugotrajne okluzije, tijekom kojih nepouzdana opažanja mogu narušiti procjenu položaja i očuvanje identiteta objekta. Usporedba postojećih algoritama pokazuje različite pristupe ublažavanju tih problema, ali i njihova ograničenja pri primjeni u pomorskom okruženju. Pregled dostupnih tehnologija i skupova podataka, zajedno s provedenom eksperimentalnom analizom detekcije plovila, pruža cjelovit uvid u teorijske, algoritamske i praktične aspekte područja. Time su obuhvaćene temeljne metode i otvoreni problemi detekcije, praćenja i reidentifikacije plovila u složenim pomorskim scenama.

LITERATURA

[1] Bewley, A. i sur. (2016). Simple Online and Realtime Tracking. ICIP.

[2] Wojke, N., Bewley, A., Paulus, D. (2017). Simple Online and Realtime Tracking with a Deep Association Metric. ICIP.

[3] Zhang, Y. i sur. (2021). FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking. IJCV.

[4] Zhang, Y. i sur. (2022). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. ECCV.

[5] Du, Y. i sur. (2023). StrongSORT: Make DeepSORT Great Again. IEEE TMM. Preprint 2022.

[6] Aharon, N., Orfaig, R., Bobrovsky, B.-Z. (2022). BoT-SORT: Robust Associations Multi-Pedestrian Tracking. arXiv:2206.14651.

[7] Cao, J. i sur. (2023). Observation-Centric SORT: Rethinking SORT for Robust Multi- Object Tracking. CVPR.

[8] Maggiolino, G. i sur. (2023). Deep OC-SORT: Multi-Pedestrian Tracking by Adaptive Re-Identification. arXiv:2302.11813.

[9] Marin, I. (2025). Model dubokog učenja za praćenje plovila s fokusom na rješavanje problema dugotrajnih okluzija. Doktorska disertacija, FESB, Split.

[10] Chen, Y., Meng, F. i Chen, Z., OcclusionTrack: Multi-Object Tracking in Dense Scenes (2025). Applied Sciences, 15(24), 13030. https://www.mdpi.com/2076-3417/15/24/13030

[11] Li, C. i sur. (2026). Occlusion-Aware SORT: Observing Occlusion for Robust Multi- Object Tracking. CVPR, 42560–42570. https://openaccess.thecvf.com/content/CVPR2026/html/Li_Occlusion-Aware_SORT_Observing_Occlusion_for_Robust_Multi-Object_Tracking_CVPR_2026_paper.html

[12] Bernardin, K., Stiefelhagen, R. (2008). Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. EURASIP JIVP.

[13] Luiten, J. i sur. (2021). HOTA: A Higher Order Metric for Evaluating Multi-Object Tracking. IJCV.

[14] Ristani, E. i sur. (2016). Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking. ECCV Workshops.

[15] NVIDIA, DeepStream SDK Developer Guide, izdanje 9.1, https://docs.nvidia.com/metropolis/deepstream/dev-guide/ (pristupljeno: 2026-10-25).

[16] NVIDIA, GStreamer Plugin Overview; Metadata in the DeepStream SDK, službena DeepStream dokumentacija.

[17] NVIDIA, DeepStream Reference Application – deepstream-app, Application Architecture and Configuration, službena dokumentacija.

[18] NVIDIA, Gst-nvstreammux i Gst-nvstreammux New: konfiguracija grupiranja i latencije, službena dokumentacija.

[19] NVIDIA, Gst-nvinfer, Gst-nvinferserver i Gst-nvdspreprocess: plugin specifications, službena dokumentacija.

[20] NVIDIA, Gst-nvtracker: NvMultiObjectTracker, tracker algorithms and configuration, https://docs.nvidia.com/deepstream/text/DS_plugin_gst-nvtracker (pristupljeno: 2026-10-25).

[21] NVIDIA, DeepStream repository: config_tracker_NvDCF_accuracy_3D.yml, GitHub, dokumentirani primjer parametara NvDCF-a i ReID-a.

[22] NVIDIA Developer Blog, Managing Video Streams in Runtime with the NVIDIA DeepStream SDK, 2022.

[23] GStreamer Project, Application Development Manual, https://gstreamer.freedesktop.org/documentation/ .
[24] NVIDIA, DeepStream SDK – Get Started and Release Information, https://developer.nvidia.com/deepstream-getting-started (pristupljeno: 2026-10-25).
[25] Luo, F., Jiang, P., Ho, G. T. S. i Zeng, W. (2025). ShipsMOT: A Comprehensive Benchmark and Framework for Multiobject Tracking of Ships. IET Computer Vision, 19(1), e70042. https://doi.org/10.1049/cvi2.70042
[26] ShipsMOT, javni repozitorij i opis podataka. https://github.com/jpj0916/ShipsMOT
[27] ViV-ReID: Bidirectional Structural-Aware Spatial-Temporal Graph Networks on Large-Scale Video-Based Vessel Re-Identification Dataset. Sažetak i povezani izvor: https://pubmed.ncbi.nlm.nih.gov/41411358/
[28] Wang, H. i sur. HOSS ReID dataset, Zenodo, 2025. https://doi.org/10.5281/zenodo.15860212
[29] MMDEC: Multimodal maritime dataset on the English channel. Data in Brief 65 (2026), 112629. https://doi.org/10.1016/j.dib.2026.112629
[30] Muslim, P. (2025). Utjecaj kvalitete anotacije na točnost i odziv detekcije. Istraživački seminar, Fakultet elektrotehnike, strojarstva i brodogradnje, Sveučilište u Splitu.