

UNIVERZITET SINGIDUNUM

BEOGRAD

DEPARTMAN ZA POSLEDIPLOMSKE STUDIJE

DOKTORSKA DISERTACIJA

PRECIZNA SEGMENTACIJA INSTANCI NA VAZDUŠNIM SNIMCIMA VISOKE REZOLUCIJE KORIŠĆENJEM TILED INFERENCIJE I GLOBALNE KONSOLIDACIJE MASKI

MENTOR:

Prof. dr Marina Marjanović

STUDENT:

Marko Mihajlović

Broj indeksa:
2023/480218

Beograd, 2026.

UNIVERZITET SINGIDUNUM
DEPARTMAN ZA POSLEDIPLOMSKE STUDIJE
Danijelova 32, Beograd

Kandidat:	Marko Mihajlović, master inženjer elektrotehnike i računarstva
Broj indeksa:	2023/480218
Studijski program:	Inteligentno softversko inženjerstvo
Tema:	PRECIZNA SEGMENTACIJA INSTANCI NA VAZDUŠNIM SNIMCIMA VISOKE REZOLUCIJE KORIŠĆENJEM TILED INFERENCIJE I GLOBALNE KONSOLIDACIJE MASKI

Mentor:

Ime i prezime:	Prof. dr Marina Marjanović
Zvanje:	redovni profesor
Institucija:	Univerzitet Singidunum, Beograd

Članovi komisije:

Ime i prezime:	Prof. dr Marko Tanasković
Zvanje:	vanredni profesor
Institucija:	Univerzitet Singidunum, Beograd

Ime i prezime:	Doc. dr Milan Čabarkapa
Zvanje:	docent
Institucija:	Fakultet inženjerskih nauka, Univerzitet u Kragujevcu

Beograd, 2026. godine

ZAHVALNICA

Zahvaljujem se svim profesorima Univerziteta Singidunum koji su svojim predavanjima, savetima, diskusijama i akademskim usmerenjem doprineli mom stručnom i istraživačkom razvoju tokom doktorskih studija. Njihovo znanje, iskustvo i posvećenost nastavi predstavljali su važan oslonac u oblikovanju mog istraživačkog puta i u razumevanju savremenih pravaca razvoja veštačke inteligencije i softverskog inženjerstva.

Posebnu zahvalnost dugujem prof. dr Nebojši Bačaninu–Džakuli za podršku, stručne savete i doprinos u razvoju mog naučno-istraživačkog rada, posebno kroz saradnju na pripremi i objavljivanju radova na naučnim konferencijama i u akademskim publikacijama. Njegovo iskustvo, istraživačka energija i spremnost da usmeri istraživače imali su značajan uticaj na moje razumevanje naučnog rada, metodologije istraživanja i značaja međunarodne akademske saradnje.

Zahvaljujem se prof. dr Angelini Njeguš i prof. dr Milošu Antonijeviću na podršci, saradnji, savetima i doprinosu tokom mog akademskog i istraživačkog rada. Posebno sam im zahvalan na pomoći u razmeni ideja, pripremi i objavljivanju radova na različitim naučnim konferencijama, kao i na podršci u tehničkoj pripremi i oblikovanju istraživačkih rezultata koji su prethodili izradi ove disertacije. Njihova spremnost da pomognu i doprinesu razvoju istraživačkog pristupa predstavljala je važnu podršku u procesu mog profesionalnog i naučnog sazrevanja.

Posebnu i iskrenu zahvalnost dugujem svojoj mentorki, prof. dr Marini Marjanović, na poverenju, strpljenju, stručnom vođenju i usmeravanju tokom izrade doktorske disertacije. Zahvalan sam joj što me je usmerila ka oblasti računarskog vida, u kojoj sam pronašao svoj istraživački interes i prostor za dalji profesionalni razvoj. Posebno joj zahvaljujem na pomoći i podršci u pripremi i objavljivanju naučnih radova u međunarodnim časopisima i publikacijama, što je predstavljalo važan deo ispunjavanja akademskih uslova doktorskih studija i značajno doprinelo razvoju rezultata prikazanih u ovoj disertaciji. Njeni saveti, sugestije i podrška bili su od presudnog značaja za oblikovanje teme, metodologije i konačne strukture ovog rada.

Na kraju, najveću zahvalnost dugujem svojoj porodici, koja je imala razumevanja za vreme koje sam provodio u učenju, istraživanju, pisanju radova i izradi ove disertacije. Njihovo strpljenje, podrška i poverenje bili su mi najvažniji oslonac tokom celog procesa.

Sadržaj

Rezime IX

1 UVOD 1
1.1 Kontekst istraživanja 1
1.2 Motivacija i značaj teme 2
1.3 Izazovi analize snimaka visoke rezolucije 3
1.4 Obim realizovanog istraživanja 5
1.5 Doprinos istraživanja 6

2 TEORIJSKI OKVIR I PREGLED LITERATURE 8
2.1 Konvolucione neuronske mreže 8
2.1.1 Lokalna obrada i hijerarhijske reprezentacije 9
2.1.2 Receptivno polje i granice lokalnog konteksta 9
2.1.3 Značaj za segmentaciju instanci na vazdušnim snimcima 10
2.2 Transformerske arhitekture i mehanizam samopažnje u računarskom vidu . 11
2.2.1 Pažnja, upiti objekata i predikcija bez NMS postprocesiranja 12
2.2.2 Višeskalna i deformabilna pažnja 13
2.2.3 Segmentacija zasnovana na upitima i RF-DETR kontekst 13
2.2.4 Ograničenja pri visokim rezolucijama 14
2.3 Relevantni pristupi segmentaciji i konsolidaciji maski 14
2.3.1 Postprocesna deduplikacija predikcija 14
2.3.2 End-to-end i NMS-free pristupi 15
2.3.3 Lokalna deduplikacija i konsolidacija nezavisnih predikcija 16
2.3.4 Od IoU-zasnovanog spajanja ka prostornom grupisanju maski . . . 16
2.3.5 Spatial Mask Merging kao polazni mask-level pristup 17
2.4 Selekcija i oporavak predikcija niže pouzdanosti 18
2.4.1 Confidence, kalibracija i kvalitet predikcije 18
2.4.2 Selektivno odlučivanje i naknadna procena kvaliteta 18
2.4.3 Teški negativni primeri i prostorni kontekst 19
2.4.4 Nezavisan signal drugog modela 20
2.4.5 Selektivni oporavak kao sekundarno odlučivanje 20
2.5 Ograničenja postojećih pristupa i istraživački prostor 21

3 METODOLOŠKI OKVIR 25
3.1 Modeli i skupovi podataka 25
3.1.1 Korišćeni modeli segmentacije instanci 25
3.1.2 Izvršni ONNX artefakti i zajednički izlazni format 26
3.1.3 Skup iSAID 27
3.1.4 Skup BONAI 28
3.2 Inferencija nad pločicama 30
3.2.1 Dekompozicija slike na pločice 30
3.2.2 Model-specifična predobrada i lokalna inferencija 31
3.2.3 Rekonstrukcija lokalnih predikcija u globalnom koordinatnom sistemu 32
3.2.4 Formiranje početnog globalnog skupa predikcija 32
3.2.5 Evidencija, reproduktivnost i merenje vremena 33
3.3 Integrisani cevovod obrade 34
3.4 Selektivni oporavak predikcija niže pouzdanosti 35
3.4.1 Formiranje problema selektivnog oporavka 36
3.4.2 Overlap kao izvor dodatnog prostornog signala 36
3.4.3 Reprezentacija low-confidence kandidata 37
3.4.4 Nezavisni cross-model signal 39
3.4.5 Sekundarni klasifikator 39
3.4.6 Optimizacija i finalna konfiguracija XGBoost klasifikatora 39
3.4.7 Izbor operativnog praga 40
3.4.8 Mere kvaliteta rangiranja 41
3.4.9 Izolacija razvojnog i konačnog evaluacionog skupa 41
3.4.10 Integracija u globalni pipeline 42
3.5 BC-SMM: granično ograničeno prostorno spajanje maski za globalnu konsolidaciju 42
3.5.1 Kriterijum kompatibilnosti predikcija 43
3.5.2 Koridori i fizički kontakti između pločica 44
3.5.3 Formiranje kandidatnih parova 45
3.5.4 Lokalna kontaktna validacija marginalnih mostova 45
3.5.5 Topološko grupisanje predikcija 46
3.5.6 Izbor lokalne particije 46
3.5.7 Sažeti postupak BC-SMM 47
3.5.8 Parametri BC-SMM 48
3.5.9 Izlaz BC-SMM 49
3.6 Funkcija fuzije maski 49
3.6.1 Fuzija višečlanih grupa 49
3.6.2 Agregacija skora pouzdanosti 50
3.6.3 Postupanje sa jednočlanim grupama 51

3.6.4 Izlaz funkcije . . . 51
3.7 Evaluacioni protokol i predložene tile-boundary metrike . . . 52
3.7.1 Standardni evaluacioni protokol . . . 52
3.7.2 Motivacija za specijalizovanu tile-boundary evaluaciju . . . 53
3.7.3 Definicija crossing populacije . . . 53
3.7.4 Pridruživanje predikcija i uklanjanje duplikata . . . 54
3.7.5 Definicija predloženih tile-boundary metrika . . . 55
3.7.5.1 Tile-Boundary Miss Rate . . . 55
3.7.5.2 Tile-Boundary Fragmentation Rate . . . 55
3.7.5.3 Tile-Boundary Excess Fragmentation . . . 55
3.7.5.4 Tile-Boundary Duplicate Rate . . . 56
3.7.5.5 Reconstruction IoU . . . 56
3.7.5.6 Merge Rate . . . 57
3.7.6 Zajednička interpretacija metrika . . . 58
3.7.7 Agregacija i statistički protokol . . . 58
3.8 Eksperimentalne konfiguracije . . . 59
3.9 Dodatno ispitane metodološke varijante . . . 60
3.9.1 Eksperimentalno morfološko prečišćavanje . . . 60
3.9.2 Naknadna analiza i korekcija semantičke klase . . . 60

4 EKSPERIMENTALNI REZULTATI I ANALIZA PREDLOŽENOG SISTEMA . 62
4.1 Eksperimentalni protokol i pregled poređenih konfiguracija . . . 62
4.2 Uticaj preklapanja pločica pri tiled inferenciji . . . 63
4.2.1 Ukupni rezultati bez globalne konsolidacije . . . 63
4.2.2 Uticaj na COCO metrike i kvalitet lokalizacije . . . 64
4.2.3 Analiza uzroka povećanja false-positive predikcija . . . 65
4.2.4 Analiza smanjenja false-negative slučajeva . . . 65
4.2.5 Uticaj na reprezentativne klase . . . 66
4.3 SMM i efekat globalne konsolidacije . . . 67
4.3.1 Efekat SMM konsolidacije pri 8 procenata overlap-a . . . 67
4.3.2 Kontrolno poređenje SMM-a pri 0 i 8 procenata overlap-a . . . 68
4.3.3 Struktura grupa i efekat overlap-a . . . 68
4.3.4 Promena kandidatnog grafa . . . 69
4.3.5 Uticaj anti-chaining korekcije . . . 69
4.3.6 Efekat overlap-a nakon globalne konsolidacije . . . 69
4.4 Rezultati BC-SMM globalne konsolidacije . . . 70
4.4.1 Ukupni evaluacioni rezultat . . . 71
4.4.2 Poboljšanje COCO mask metrika . . . 71
4.4.3 Promena strukture globalne particije . . . 72

4.4.4 Ograničavanje kandidatnog prostora . . . 72
4.4.5 Rezultati po klasama . . . 73
4.4.6 Kontrolna analiza BC-SMM bez sistematskog preklapanja . . . 74
4.5 Analiza ponašanja sistema na granicama pločica . . . 75
4.5.1 Rezultati bez preklapanja pločica . . . 75
4.5.2 Rezultati pri 8 procenata preklapanju . . . 76
4.5.3 Dodatni doprinos BC-SMM . . . 78
4.5.4 Razdvajanje geometrijske rekonstrukcije i identiteta instance . . . 78
4.5.5 Kontrola rizika od prekomernog spajanja . . . 78
4.5.6 Odnos tile-boundary i standardnih metrika . . . 78
4.5.7 Uloga preklapanja u globalnoj konsolidaciji . . . 79
4.5.8 Robusnost rezultata: uparena bootstrap analiza . . . 79
4.5.9 Sažetak nalaza . . . 80
4.6 Eksperimentalna evaluacija selektivnog oporavka predikcija niže pouzdanosti 80
4.6.1 Informativnost i progresija reprezentacije karakteristika . . . 81
4.6.2 Razvojna reintegracija i izbor operativnog režima . . . 82
4.6.3 Zamrznuta finalna evaluacija na official validation skupu . . . 83
4.6.4 Razvojni i finalni rezultat . . . 84
4.7 Analiza rezultata po klasama i veličini objekata . . . 84
4.7.1 Lokalna dijagnostika segmentacionih modela po klasama . . . 84
4.7.2 Uticaj veličine objekta . . . 85
4.7.3 Analiza karakterističnih klasa . . . 86
4.7.4 Guste klase vozila i efekat selektivnog oporavka . . . 88
4.8 Kvalitativna analiza i vizuelni primeri . . . 88
4.8.1 Uticaj preklapanja pločica i globalne konsolidacije . . . 89
4.8.2 Redundantne predikcije pre i posle konsolidacije . . . 89
4.8.3 SMM naspram BC-SMM . . . 89
4.8.4 Uspešan primer selektivnog oporavka predikcija . . . 90
4.8.5 Ograničavajući primer . . . 91
4.8.6 Kvalitativna analiza prenosivosti na BONAI skup . . . 91
4.9 Računske performanse . . . 93
4.9.1 Cena tiled inferencije . . . 93
4.9.2 Vreme formiranja grupa . . . 94
4.9.3 Redukcija kandidatnog prostora . . . 94
4.9.4 Cena fuzije maski . . . 95
4.9.5 Računski uticaj rescue grane . . . 96
4.10 Ablacione i dodatne analize . . . 96
4.10.1 Razvojna ablacija BC-SMM pristupa . . . 97
4.10.2 Morfološko prečišćavanje maski . . . 98

4.10.3 Naknadna analiza semantičke korekcije klase 98
4.10.4 Ablacija reprezentacije rescue kandidata 100
4.10.5 Zaključak ablacionih i dodatnih analiza 100
4.11 Sinteza glavnih eksperimentalnih nalaza 100

5 DISKUSIJA 103
5.1 Interpretacija glavnih nalaza 103
5.2 Overlap i globalna konsolidacija 104
5.3 Prostorno indeksiranje i grafovska konsolidacija 105
5.4 Granice dodatnog postprocesiranja 106
5.5 Generalizacija i prenosivost 106
5.6 Računska i praktična primenljivost 108
5.7 Ograničenja predloženog pristupa 109
5.8 Pravci budućeg razvoja 110

6 ZAKLJUČAK 112

SPISAK SLIKA 118

SPISAK TABELA 120

SPISAK AKRONIMA 121

SPISAK ENGLESKIH TERMINA 122

LITERATURA 123

Rezime

Ova disertacija istražuje preciznu segmentaciju instanci na vazdušnim snimcima visoke rezolucije korišćenjem inferencije nad pločicama, globalne konsolidacije maski i selektivnog oporavka predikcija niže pouzdanosti. Podela velikih slika na pločice omogućava očuvanje lokalnih detalja i ublažava memorijska ograničenja savremenih segmentacionih modela, ali istovremeno uvodi fragmentaciju, višestruka lokalna opažanja iste fizičke instance i redundantne predikcije. Istraživački fokus rada je formiranje globalno konzistentnog skupa instanci na nivou originalne slike i korišćenje ograničenog prostora kandidata niže pouzdanosti bez nekritičkog snižavanja osnovnog praga pouzdanosti.

Glavni metodološki doprinos je Boundary-Constrained Spatial Mask Merging (BC-SMM), algoritam globalne grafovske konsolidacije koji koristi geometriju i topologiju prostora organizovanog pločicama za ograničavanje kandidatnih veza između lokalnih maski. U odnosu na referentni Spatial Mask Merging (SMM) pristup, BC-SMM smanjuje kandidatni prostor za približno 54%, uz gotovo nepromenjen broj kompatibilnih veza. Predloženi evaluacioni okvir usmeren na granice pločica dodatno razdvaja propuštanje instanci, fragmentaciju, višak fragmenata, geometrijsku rekonstrukciju i prekomerno spajanje. Ove metrike ukazuju da je glavni detektovani efekat BC-SMM-a smanjenje identitetske fragmentacije na granicama pločica, a ne samo promena agregatnih referentnih evaluacionih metrika.

Dodatni metodološki doprinos predstavlja selektivni oporavak predikcija ispod osnovnog praga pouzdanosti. Rezultati pokazuju da sirovi skor pouzdanosti nije dovoljan za rangiranje takvih kandidata, dok lokalno-geometrijske karakteristike, karakteristike između pločica i duž njihove zajedničke granice i karakteristike zasnovane na signalu drugog modela povećavaju površinu ispod krive preciznost–odziv (*Precision–Recall Area Under the Curve*, PR-AUC) sa približno 0,23 na 0,45. Finalna konfiguracija selektivnog oporavka, pri definisanom ograničenju odnosa lažno pozitivnih i tačno pozitivnih kandidata ($FP/TP \leq 0{,}25$), ostvaruje približno 78% preciznosti pri vraćanju predikcija niže pouzdanosti u završni cevovod. Dodatni eksperimenti pokazali su da se predloženi metodološki okvir može preneti i na drugi skup podataka i drugu segmentacionu arhitekturu, što pruža dodatnu evidenciju o njegovoj prenosivosti. Ukupni rezultati ukazuju da redundansa nastala obradom po pločicama nije samo problem deduplikacije, već i dodatna prostorna evidencija koja se može iskoristiti za rekonstrukciju globalnih instanci.

Ključne reči: segmentacija instanci; snimci visoke rezolucije; inferencija nad pločicama; globalna konsolidacija maski; selektivni oporavak predikcija

Naučna oblast: Tehničko-tehnološke nauke
Uža naučna oblast: Elektrotehnika i računarstvo

1. UVOD

1.1 Kontekst istraživanja

Snimci daljinske detekcije visoke prostorne rezolucije, prikupljeni satelitskim, avionskim ili platformama bespilotnih letelica (engl. *Unmanned Aerial Vehicle*, UAV), omogućavaju detaljnu analizu pojedinačnih objekata u urbanim i infrastrukturnim scenama[1, 2]. U takvim snimcima od interesa nisu samo semantičke klase, već i zasebne fizičke instance: pojedinačna vozila, zgrade, plovila, sportski tereni, rezervoari ili drugi objekti koji se mogu pojaviti u velikom broju i u međusobno bliskim prostornim odnosima. Zbog toga segmentacija instanci ima posebnu ulogu, jer za svaki detektovani objekat određuje zasebnu masku, klasu i prostorni položaj[3].

Takve informacije su važne u zadacima mapiranja objekata, praćenja promena u prostoru, analize gustine i ažuriranja geoprostornih baza podataka. Za razliku od graničnih pravougaonika, maske instanci omogućavaju preciznije merenje oblika i zauzete površine, što je posebno značajno kada su objekti mali, gusto raspoređeni ili delimično zaklonjeni[3].

Visoka prostorna rezolucija istovremeno povećava informativnost snimka i otežava njegovu obradu. Direktno smanjivanje cele slike na ulaznu rezoluciju neuronske mreže može ukloniti sitne detalje potrebne za detekciju malih objekata, dok obrada cele slike u izvornoj rezoluciji često prevazilazi memorijska i računska ograničenja[4, 5]. Zbog toga se velike slike u praksi dele na manje pločice koje se nezavisno prosleđuju segmentacionom modelu. Inferencija nad pločicama (engl. *tiled inference*) tako omogućava očuvanje lokalnih detalja i primenu savremenih modela nad snimcima koji su znatno veći od njihovog standardnog ulaza[6, 7].

Ovaj postupak pomera problem sa lokalne segmentacije na rekonstrukciju rezultata. Model više ne vidi celu sliku odjednom, već niz lokalnih regiona koji se naknadno moraju povezati u jedinstven skup instanci u koordinatnom sistemu originalne slike.

Podela na pločice, međutim, uvodi novi problem. Objekti mogu biti presečeni granicama pločica, delimično vidljivi ili opaženi više puta ako se koriste preklopljene pločice[7, 6]. Preklapanje smanjuje rizik da objekat bude nepovoljno pozicioniran u samo jednom lokalnom prozoru, ali istovremeno proizvodi parcijalne, redundantne ili geometrijski neusaglašene lokalne predikcije iste fizičke instance[8]. Nakon lokalne inferencije potrebno je sve predikcije vratiti u globalni koordinatni sistem originalne slike i odlučiti koje maske predstavljaju isti objekat.

Dodatni izazov nastaje pri filtriranju predikcija prema skoru pouzdanosti. Fiksni prag pouzdanosti (engl. *confidence threshold*) uklanja veliki broj nepouzdanih lokalnih izlaza, ali može odbaciti i validne instance koje su delimično vidljive, nalaze se blizu granice ploči-

ce ili imaju nepovoljan lokalni kontekst. Zato konačan kvalitet sistema ne zavisi samo od osnovnog segmentacionog modela, već i od načina na koji se lokalne, parcijalne i redundantne predikcije selektuju, rekonstruišu i konsoliduju u pouzdan skup fizičkih instanci na nivou originalne slike.

1.2 Motivacija i značaj teme

Motivacija ovog istraživanja proizlazi iz činjenice da lokalno dobra maska ne garantuje ispravan globalni skup instanci. Segmentacioni model može precizno obraditi pojedinačnu pločicu, ali konačan rezultat na originalnoj slici može i dalje sadržati duplikate, fragmente iste instance ili pogrešno spojene objekte. Za primene u kojima se maske koriste za brojanje, merenje površine, procenu gustine ili izgradnju geoprostornih baza, takve greške nisu samo vizuelni nedostatak, već utiču na kasniju kvantitativnu analizu.

Posebno je važno pravilno tumačiti redundansu nastalu preklapanjem pločica. Ona nije isključivo problem koji treba ukloniti. Više lokalnih opažanja može doneti korisnu evidenciju o objektu, naročito kada je u jednoj pločici delimično presečen ili prikazan bez dovoljno konteksta. Istraživački problem zato nije samo deduplikacija, već globalna rekonstrukcija fizičke instance iz više lokalnih predikcija. Potrebno je razlikovati kada više maski opisuje isti objekat, a kada su u pitanju različite susedne instance.

Ova razlika je važna zato što standardno uklanjanje duplikata pretpostavlja da su višestruke predikcije već dovoljno slične i preklopljene. Kod tiled inferencije isti objekat može biti predstavljen komplementarnim fragmentima, dok susedni objekti iste klase mogu biti veoma blizu. Zbog toga odluka o spajanju mora koristiti širi prostorni i geometrijski kontekst od samog lokalnog preklapanja maski.

Drugi motivacioni aspekt odnosi se na filtriranje prema skoru pouzdanosti (engl. *confidence*). Jedan fiksni prag pouzdanosti pojednostavljuje lokalni izlaz modela, ali sve predikcije ispod praga tretira kao neupotrebljive. U tiled scenariju to može biti pregrubo: deo validnih objekata može dobiti niži skor zato što je delimično vidljiv, nalazi se blizu granice pločice ili je opažen u nepovoljnom lokalnom kontekstu. Zbog toga prostor neposredno ispod osnovnog confidence praga predstavlja kandidatni skup za selektivnu sekundarnu procenu, a ne prostor za nekritičko snižavanje praga.

Metodološki značaj rada proizlazi iz razdvajanja dve vrste odluka. Selekcija predikcija niže pouzdanosti može koristiti naučeni model koji procenjuje ograničen skup kandidata za selektivni oporavak (engl. *rescue*), dok se globalna konsolidacija višestrukih lokalnih opažanja može zasnivati na eksplicitnim prostornim i geometrijskim odnosima između maski. Takvo razdvajanje omogućava da se odvojeno analizira da li poboljšanje potiče od oporavka dodatnih validnih predikcija, od boljeg grupisanja redundantnih maski ili od njihove kombinacije.

Na taj način se problem ne svodi na izbor jednog agresivnijeg ili konzervativnijeg

confidence praga. Predikcije koje su već prihvaćene i kandidati koji zahtevaju dodatnu proveru ulaze u jasno razdvojene faze, nakon čega se njihov doprinos meri na nivou konačnih globalnih instanci.

Praktični značaj pristupa je u tome što postprocesni deo nije vezan za unutrašnju arhitekturu jednog konkretnog segmentacionog modela. Osnovni princip konsolidacije može se primeniti nad predikcijama različitih segmentacionih arhitektura, što je dodatno ispitano na drugom modelu i drugom skupu podataka. Time se istraživanje usmerava na pouzdanije formiranje konačnog globalnog skupa instanci, a ne samo na poboljšanje pojedinačne lokalne predikcije.

1.3 Izazovi analize snimaka visoke rezolucije

Analiza snimaka visoke rezolucije obuhvata više povezanih izazova koji proizlaze iz prostorne skale podataka, velikog broja objekata, lokalnog konteksta i računarskih ograničenja. Kod vazdušnih i satelitskih snimaka urbanih područja na jednoj slici mogu se istovremeno pojaviti veoma mali objekti, velike strukture, guste grupe instanci i složena pozadina.

Prvi izazov predstavlja velika varijacija prostorne skale. Male instance mogu biti predstavljene sa svega nekoliko desetina piksela, dok velike strukture zauzimaju značajan deo snimka. Smanjivanje cele slike na standardnu ulaznu rezoluciju neuronskog modela smanjuje memorijske i računske zahteve, ali može ukloniti sitne detalje potrebne za detekciju malih objekata [9, 10]. Direktna obrada izvorne rezolucije, sa druge strane, često je ograničena raspoloživom memorijom i računskim kapacitetima.

Računarski kompromis je zato prisutan već na početku obrade. Manji lokalni prozor lakše čuva rezoluciju sitnih objekata, ali smanjuje kontekst u kome model posmatra scenu. Veći prozor daje više konteksta, ali povećava memorijske zahteve i broj detalja koje model mora obraditi u jednom prolazu.

Podela velikih snimaka na manje regione rešava deo tog problema, jer čuva lokalnu prostornu rezoluciju, ali uvodi ograničen lokalni kontekst. Objekat u unutrašnjosti pločice može biti prikazan zajedno sa okolinom, dok objekat uz granicu može biti samo delimično vidljiv. Posledica mogu biti nepotpune maske, niže vrednosti pouzdanosti ili različite klasne odluke za različita lokalna posmatranja istog fizičkog objekta.

Granični efekti posebno su važni kod segmentacije instanci. Ne zahteva se samo detekcija prisustva objekta, već i očuvanje njegovog identiteta i prostorne konture. Jedna fizička instanca može zato biti predstavljena sa više parcijalnih maski čiji se oblici, confidence vrednosti i konture razlikuju. Čak i kada više lokalnih predikcija pripada istom objektu, njihovo međusobno preklapanje može biti malo, pa sama vrednost preseka nad unijom (engl. *Intersection over Union*, IoU) nije dovoljan pokazatelj njihove pripadnosti istoj instanci.

Na nivou maski to znači da treba očuvati geometrijsku informaciju iz lokalnih predikcija, ali bez slepog spajanja svih bliskih kontura. Parcijalne maske mogu biti korisni delovi

iste instance, dok nepravilne ili neusaglašene konture mogu istovremeno povećati rizik od pogrešnog spajanja susednih objekata.

Suprotan problem javlja se u scenama sa velikom gustinom objekata. Vozila, plovila, rezervoari, zgrade i druge instance iste klase mogu biti međusobno veoma blizu, pa prostorna blizina nije dokaz da dve maske pripadaju istom objektu. Pouzdana konsolidacija mora razlikovati fragmente jedne fizičke instance od različitih objekata koji se nalaze na malom rastojanju. Ovi izazovi mogu se istovremeno javiti u jednoj sceni, kao što je prikazano na Slici 1.1.



Slika 1.1. Primer karakterističnih izazova na visokorezolucionom iSAID snimku: (a) veoma male instance, (b) velika lokalna gustina objekata, (c) objekat velikih prostornih dimenzija i (d) instanca koja preseca unutrašnju granicu pločice.

Primer ilustruje da ista scena može zahtevati očuvanje sitnih detalja, razdvajanje gusto raspoređenih instanci, obradu objekata različitih skala i pouzdano postupanje sa objektima koji prelaze granice lokalnih pločica. Dodatnu neizvesnost mogu uneti osvetljenje, senke, orijentacija i ugao snimanja, ali ključni problem ove disertacije ostaje način na koji se lokalne predikcije prevode u globalni skup instanci.

U takvom okruženju greške se često ne javljaju izolovano. Mali objekat može istovremeno biti blizu druge instance i presečen granicom pločice; veliki objekat može biti posmatran kroz više lokalnih regiona; gusta grupa objekata može proizvesti veliki broj maski koje su prostorno bliske, ali ne pripadaju istoj fizičkoj instanci. Zato globalna odluka mora koristiti informacije o položaju predikcije u pločici, odnosu susednih pločica i geometriji same maske.

Confidence skor predstavlja procenu pouzdanosti lokalne predikcije, ali njegova vrednost zavisi i od uslova u kojima je objekat posmatran. Parcijalna vidljivost, nedostatak kon-

teksta ili položaj uz granicu pločice mogu sniziti skor i kada predikcija odgovara stvarnoj instanci. Jednostavno smanjivanje praga prihvatanja povećava broj lažno pozitivnih predikcija, pa izbor kandidata za dalju obradu mora biti selektivan.

Računska složenost dodatno ograničava globalnu obradu. Ako se svaka lokalna predikcija poredi sa svakom drugom predikcijom iste slike, broj parova raste kvadratno. Ovaj problem je posebno izražen kod gustih scena i kod preklopljenih pločica koje proizvode više opažanja istog prostora. Skalabilna obrada zato zahteva ograničavanje poređenja na prostorno relevantne parove.

Navedeni problemi pokazuju da precizna segmentacija instanci na snimcima visoke rezolucije nije samo problem generisanja kvalitetnih lokalnih maski. Potrebno je očuvati koristan lokalni signal, rekonstruisati identitet fizičke instance i efikasno konsolidovati veliki broj parcijalnih i potencijalno redundantnih predikcija.

1.4 Obim realizovanog istraživanja

Realizovano istraživanje obuhvata razvoj, implementaciju i evaluaciju metodološkog okvira za segmentaciju instanci na snimcima daljinske detekcije visoke rezolucije. Obuhvaćeni su priprema tiled skupova podataka, korišćenje segmentacionih modela obučenih za lokalnu segmentaciju, inferencija nad preklopljenim pločicama, rekonstrukcija lokalnih predikcija u koordinatnom sistemu originalne slike i njihova globalna konsolidacija.

Poseban deo istraživanja odnosi se na selektivni oporavak predikcija niže pouzdanosti. Umesto automatskog odbacivanja primenom jednog confidence praga, ispituje se mogućnost da se pomoću posebno obučenog modela mašinskog učenja izdvoji deo potencijalno validnih predikcija i uključi u dalju obradu. Globalna konsolidacija zadržanih predikcija realizuje se unapređenim Boundary-Constrained Spatial Mask Merging (BC-SMM) algoritmom, nakon čega se formiraju konačne instance na nivou originalnog snimka.

Eksperimentalni deo obuhvata poređenje osnovnog tiled pristupa, prethodnog Spatial Mask Merging (SMM) postupka, unapređene BC-SMM konsolidacije i kompletnog toka obrade sa uključenom rescue granom. Evaluacija se sprovodi standardnim metrikama segmentacije instanci i dodatnim tile-boundary pokazateljima, odnosno pokazateljima usmerenim na granice pločica (engl. *tile-boundary*), za analizu fragmentacije, duplikacije, propuštenih instanci i efekata granica pločica. Primarni eksperimentalni domen predstavlja skup Instance Segmentation in Aerial Images Dataset (iSAID), dok se dodatnom evaluacijom ispituje prenosivost pristupa na drugi skup podataka i drugu segmentacionu arhitekturu.

Obim istraživanja ograničen je na segmentaciju instanci u dvodimenzionalnim snimcima daljinske detekcije i obradu predikcija generisanih tiled inferencijom. Istraživanje ne obuhvata trodimenzionalnu rekonstrukciju, vremensko praćenje objekata, multimodalnu fuziju senzora niti optimizaciju procesa prikupljanja snimaka. Fokus je na povećanju pouzdanosti konačnog skupa instanci kroz selektivno očuvanje korisnih lokalnih predikcija i njihovu

prostorno konzistentnu globalnu konsolidaciju.

1.5 Doprinos istraživanja

Doprinos ove disertacije ogleda se u razvoju, formalizaciji, implementaciji i eksperimentalnoj evaluaciji metodološkog okvira za segmentaciju instanci na snimcima daljinske detekcije visoke rezolucije. Predloženi pristup lokalne izlaze segmentacionog modela tretira kao predikcije koje se selektuju, preslikavaju u globalni koordinatni sistem i konsoliduju u konačne fizičke instance.

Za razliku od pristupa koji tiled izlaz posmatra prvenstveno kao skup duplikata za uklanjanje, ovde se redundantna lokalna opažanja razmatraju kao izvor dodatne prostorne evidencije. Doprinosi su zato definisani na nivou metoda, evaluacionog protokola i kontrolisane eksperimentalne provere, a ne kao pojedinačne vrednosti rezultatnih metrika.

Osnovni doprinosi rada su:

- razvoj i formalizacija Boundary-Constrained Spatial Mask Merging (BC-SMM) algoritma, koji koristi geometriju i topologiju tiled prostora za ograničavanje kandidatnih veza i globalnu konsolidaciju lokalnih maski;
- formalizacija celovitog metodološkog okvira od ulaza do izlaza koji povezuje tiled inferenciju, globalnu rekonstrukciju lokalnih predikcija, selekciju prema pouzdanosti, selektivni oporavak, BC-SMM grupisanje i fuziju maski;
- razvoj selektivnog oporavka predikcija ispod osnovnog praga pouzdanosti korišćenjem lokalno-geometrijskih karakteristika, karakteristika između susednih pločica i duž njihove zajedničke granice (engl. *cross-tile/seam*) i karakteristika zasnovanih na signalu drugog modela (engl. *cross-model*);
- uvođenje specijalizovanog tile-boundary evaluacionog okvira, uključujući Tile-Boundary Miss Rate (TBMR), Tile-Boundary Fragmentation Rate (TBFR), Tile-Boundary Excess Fragmentation (TBEF), Tile-Boundary Duplicate Rate (TBDR), Reconstruction IoU (RIoU) i Merge Rate (MR), za merenje propuštanja, fragmentacije, viška fragmenata, geometrijske rekonstrukcije i prekomernog spajanja na granicama pločica;
- sistematska eksperimentalna evaluacija predloženog okvira kroz kontrolisana poređenja, tile-boundary analizu, ablacije i dodatnu proveru prenosivosti na drugom skupu podataka i sa drugom segmentacionom arhitekturom.

Disertacija tretira redundansu nastalu obradom slike po pločicama ne samo kao problem deduplikacije, već kao dodatnu prostornu evidenciju koja se može koristiti za rekonstrukciju globalnih instanci. Time se odvojeno razmatraju gubitak potencijalno validnih kandidata niže pouzdanosti (engl. *low-confidence*) i konsolidacija višestrukih lokalnih opažanja

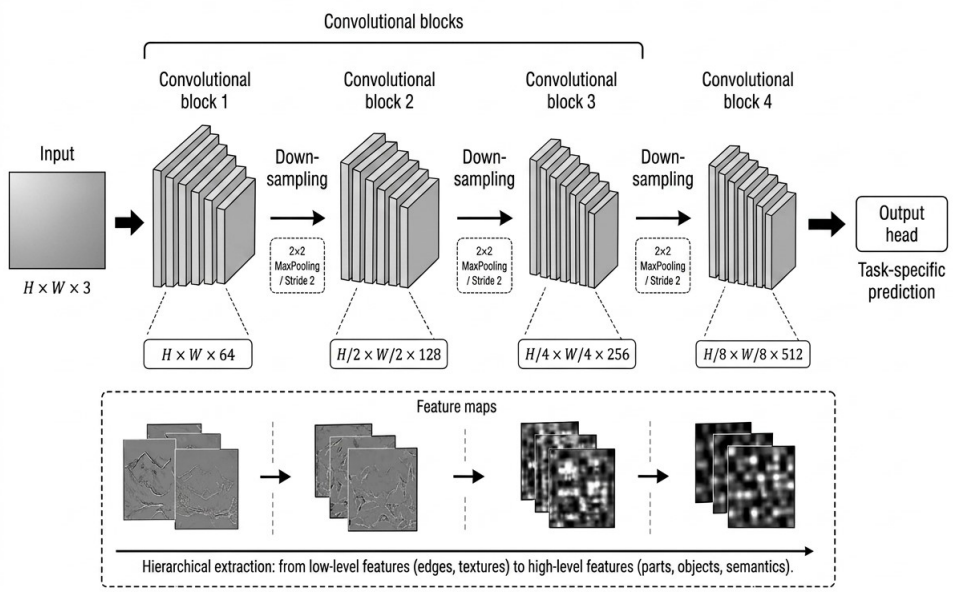
iste fizičke instance, što omogućava precizniju analizu i kontrolu različitih izvora grešaka u segmentaciji instanci nad pločicama. Rad polazi od stanovišta da ovu redundansu nije dovoljno samo ukloniti, već je treba strukturisano iskoristiti pri formiranju konačnog skupa globalnih instanci.

2. TEORIJSKI OKVIR I PREGLED LITERATURE

2.1 Konvolucione neuronske mreže

Konvolucione neuronske mreže (engl. *Convolutional Neural Networks*, CNN) predstavljaju jednu od osnovnih arhitektonskih porodica savremenog računarskog vida. Njihova relevantnost za ovu disertaciju ne proizlazi iz opšte sposobnosti klasifikacije slika, već iz načina na koji lokalno obrađuju prostor, grade hijerarhijske reprezentacije i kombinuju semantički bogate osobine sa mapama veće prostorne rezolucije potrebnim za segmentaciju instanci [11, 12, 13].

U konvolucionom sloju isti naučeni filter primenjuje se na različitim prostornim pozicijama ulazne mape. Time se uvode dve važne induktivne pristrasnosti: lokalna povezanost i deljenje parametara. Lokalna povezanost odgovara činjenici da se ivice, teksture, uglovi i drugi osnovni obrasci najčešće prepoznaju iz ograničenog susedstva, dok deljenje parametara omogućava da se isti obrazac prepozna nezavisno od njegove apsolutne pozicije u slici [14, 15]. Za visokorezolucione snimke ova osobina je posebno praktična, jer omogućava efikasnu obradu velikog broja prostornih lokacija bez potpunog povezivanja svakog piksela sa svakim izlazom.



Slika 2.1. Opšta struktura konvolucione neuronske mreže i hijerarhijska ekstrakcija vizuelnih osobina. Sa povećanjem dubine mreže prostorna rezolucija mapa osobina se smanjuje, dok se broj kanala i nivo apstrakcije reprezentacije povećavaju.

2.1.1 Lokalna obrada i hijerarhijske reprezentacije

Slaganjem konvolucionih slojeva mreža postepeno prelazi od lokalnih vizuelnih obrazaca ka semantički bogatijim reprezentacijama. Rani slojevi tipično reaguju na ivice, gradijente i teksture, srednji slojevi kombinuju ove elemente u motive i delove objekata, dok dublji slojevi kodiraju šire konfiguracije i odnose među delovima scene [13, 14]. Ova hijerarhija je važna za segmentaciju instanci jer konačna odluka mora istovremeno da razlikuje klasu objekta, njegovu lokalizaciju i preciznu konturu.

Cena takve hijerarhije često je smanjenje prostorne rezolucije. Poduzorkovanje, konvolucije sa većim korakom i agregiranje povećavaju efikasnost i šire kontekst, ali mogu ukloniti sitne strukture i oslabiti informaciju o preciznom položaju granica. Zato segmentacione arhitekture obično kombinuju osobine iz više nivoa: dublje mape daju semantički kontekst, dok ranije ili bočne mape više rezolucije pomažu pri rekonstrukciji malih objekata i kontura.

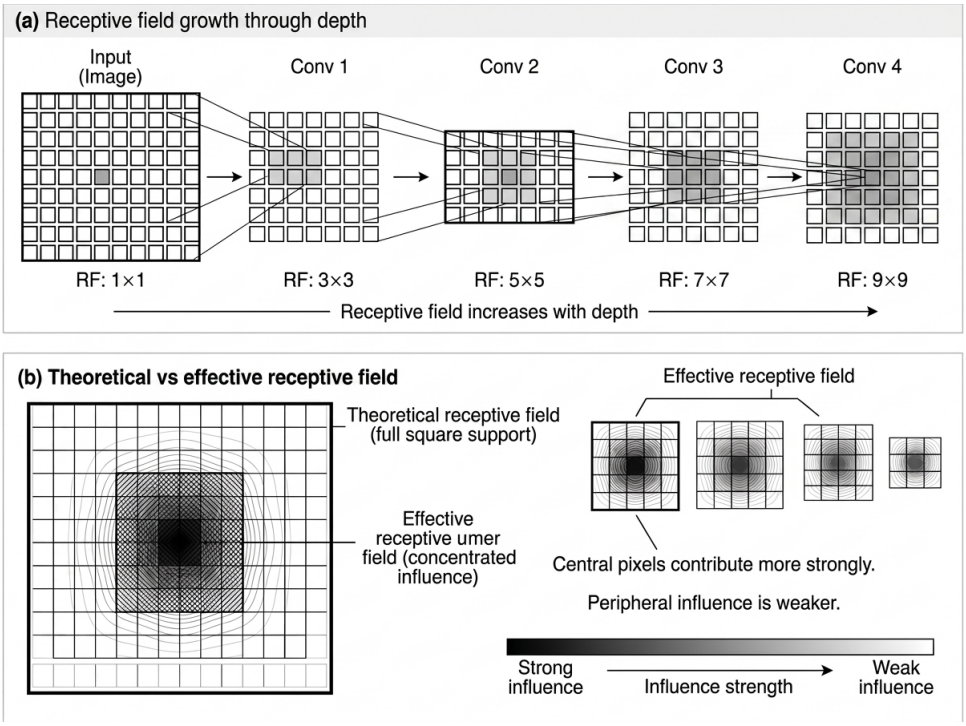
Kod vazdušnih snimaka ovaj kompromis je posebno izražen. Vozila, plovila ili mali elementi infrastrukture mogu zauzimati mali broj piksela, dok veliki objekti zahtevaju širi kontekst da bi bili pravilno razdvojeni od pozadine i susednih struktura. Zbog toga su višeskalne karakteristike i očuvanje prostorne rezolucije centralni zahtevi za modele koji se koriste u instancnoj segmentaciji takvih podataka.

2.1.2 Receptivno polje i granice lokalnog konteksta

Receptivno polje opisuje deo ulazne slike koji može uticati na vrednost određene aktivacije. U plitkim slojevima ono je malo i neposredno lokalno, dok se u dubljim slojevima povećava sukcesivnim sastavljanjem lokalnih operacija. Teorijski veliko receptivno polje ipak ne znači da mreža ravnomerno koristi ceo raspoloživi kontekst: efektivno receptivno polje često je koncentrisano oko centralnog dela teorijskog regiona [16]. To ograničenje je važno kod segmentacije objekata koji prelaze veliki prostorni opseg ili su presečeni granicom lokalnog ulaza.

Lokalnost konvolucije uvodi i specifično ponašanje uz granice slike ili pločice. Dopunjavanje ivica stvara veštački kontekst, a objekat koji se nalazi uz granicu lokalnog ulaza može biti posmatran bez dela svoje stvarne okoline. Kod tiled inferencije ista fizička instanca može zato biti prikazana kao centralni objekat u jednoj pločici, kao delimičan objekat u susednoj pločici ili kao nekoliko lokalnih fragmenata. Ove razlike ne moraju biti posledica slabog modela, već posledica ograničenog lokalnog pogleda.

Standardne konvolucione mreže takođe nisu prirodno robusne na sve geometrijske transformacije. Deljenje parametara daje korisnu translacionu strukturu, ali proizvoljne rotacije, promene skale i perspektivne deformacije zahtevaju podatkovnu augmentaciju, višeskalne reprezentacije ili dodatne arhitektonske mehanizme [17, 15]. Vazdušni snimci upravo često sadrže objekte sa proizvoljnom orijentacijom i velikim rasponom veličina.



Slika 2.2. Povećavanje teorijskog receptivnog polja kroz dubinu mreže i koncentracija efektivnog uticaja oko centralnog regiona. Slaganjem lokalnih konvolucionih operacija receptivno polje raste, dok efektivni doprinos ulaznih piksela ostaje izraženiji u centralnom delu teorijskog receptivnog polja.

2.1.3 Značaj za segmentaciju instanci na vazdušnim snimcima

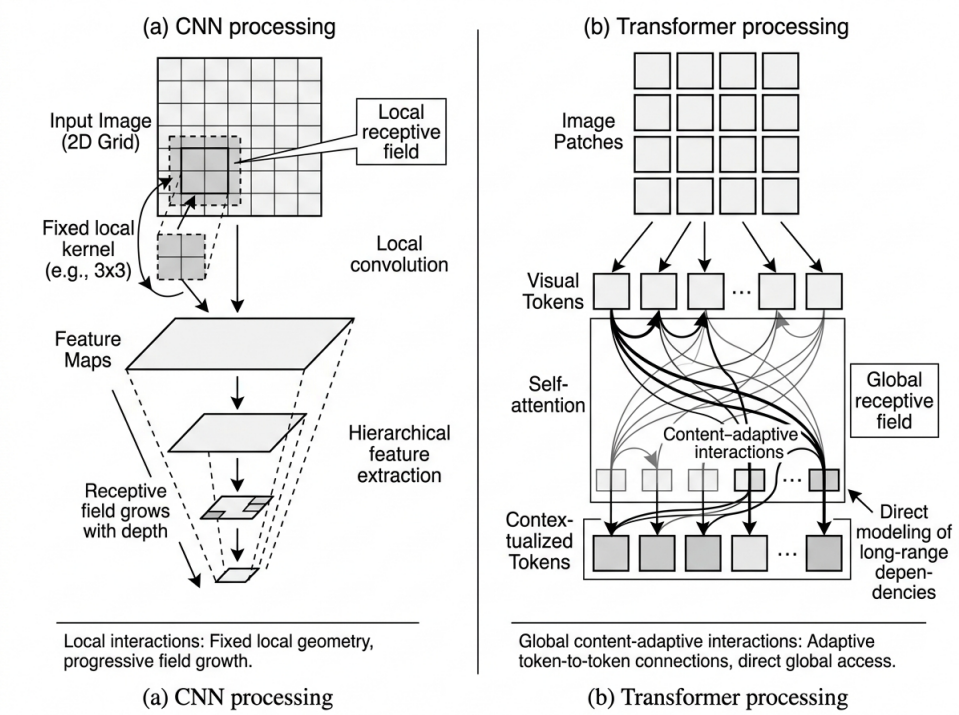
Segmentacija instanci zahteva spoj semantičkog razumevanja i precizne prostorne lokalizacije. Model mora da razlikuje objekte iste klase koji su blizu jedan drugom, da očuva njihove pojedinačne maske i da ne izgubi male instance tokom smanjivanja rezolucije. CNN komponente su zato i dalje prisutne u mnogim savremenim segmentacionim sistemima, bilo kao samostalni backbone, bilo kao deo hibridnih konvoluciono-transformerskih arhitektura.

Visokorezolucioni vazdušni snimci dodatno uvode računsko ograničenje: obrada cele slike u originalnoj rezoluciji često nije praktična zbog memorije i veličine ulaza. Podela slike na preklopljene pločice omogućava da se lokalni detalji zadrže i da mali objekti ostanu dovoljno veliki u odnosu na ulaz modela, ali istovremeno ograničava kontekst na jednu pločicu. Zbog toga ista lokalna i višeskalna obrada koja omogućava preciznu segmentaciju malih objekata može uz granice pločica proizvesti redundantne, fragmentisane ili međusobno neusaglašene maske, što zahteva vraćanje lokalnih predikcija u globalni koordinatni sistem i njihovu naknadnu konsolidaciju.

2.2 Transformerske arhitekture i mehanizam samopažnje u računarskom vidu

Transformerske arhitekture uvode drugačiji princip obrade od lokalne konvolucije. Umesto da unapred definisano jezgro povezuje susedne piksele ili tokene, mehanizam pažnje dinamički određuje koje delove reprezentacije treba povezati na osnovu njihovog sadržaja [18]. Time se omogućava neposrednije modelovanje udaljenih odnosa, što je posebno korisno kada lokalno slični regioni imaju različito značenje u zavisnosti od šireg konteksta scene.

U računarskom vidu ovaj princip se najpre široko afirmisao kroz Vision Transformer, koji sliku posmatra kao sekvencu vizuelnih tokena dobijenih podelom na zakrpe [19]. Za segmentaciju instanci važniji je razvoj modela zasnovanih na upitima (engl. *query-based*), počev od Detection Transformer (DETR) modela, u kome se detekcija formuliše kao direktna predikcija neuređenog skupa objekata pomoću naučivih upita [20]. Kasniji modeli proširili su ovaj okvir na višeskalnu detekciju, semantičku, instancnu i panoptičku segmentaciju [21, 22, 23].



Slika 2.3. Poređenje lokalne konvolucione obrade i sadržajno adaptivnog povezivanja vizuelnih tokena pomoću mehanizma samopažnje. Kod CNN-a receptivno polje raste kroz dubinu mreže, dok Transformer omogućava neposredno modelovanje odnosa između prostorno udaljenih tokena.

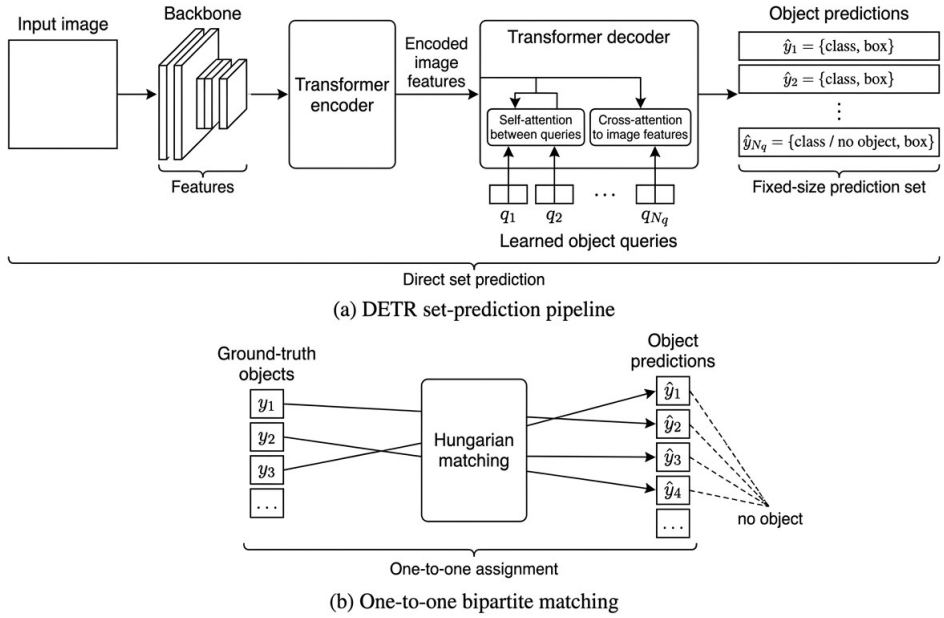
U kontekstu detekcije i segmentacije, klasični postprocesni korak za uklanjanje redundantnih predikcija jeste Non-Maximum Suppression (NMS). Zato su query-based modeli relevantni i zbog nastojanja da se deo te deduplikacije prebaci u samu formulaciju predikcije.

2.2.1 Pažnja, upiti objekata i predikcija bez NMS postprocesiranja

U sloju samopažnje (engl. *self-attention*) svaki token ažurira svoju reprezentaciju agregiranjem informacija iz drugih tokena istog skupa. U sloju unakrsne pažnje (engl. *cross-attention*) upiti potiču iz jednog skupa, dok ključevi i vrednosti potiču iz drugog. U DETR porodici modela to znači da latentni upiti objekata preuzimaju informacije iz vizuelnih karakteristika slike i postepeno se specijalizuju za pojedinačne instance.

Za razliku od konvolucionog filtera, težine pažnje zavise od konkretnog ulaza. Model zato može povezati prostorno udaljene, ali semantički ili geometrijski relevantne regione. U praksi se pažnja koristi u više glava i slojeva, uz pozicione informacije koje tokenima daju prostorni poredak. Za ovu disertaciju nije presudan formalni zapis operatora pažnje, već posledica: lokalna reprezentacija može biti dopunjena sadržajno izabranim kontekstom.

DETR formuliše detekciju kao predikciju konačnog skupa objekata [20]. Model koristi fiksan broj naučivih upita, a tokom obučavanja bipartitno uparivanje povezuje anotirane objekte sa predikcijama jedan-prema-jedan (engl. *one-to-one*). Višak upita uči se kao klasa "bez objekta". Ovakva set-prediction formulacija smanjuje potrebu za klasičnim NMS postprocesiranjem, jer se redundantnost pokušava rešiti u samom učenju modela.



Slika 2.4. DETR formulacija detekcije kao direktne predikcije skupa objekata uz naučive upite i bipartitno uparivanje jedan-prema-jedan. Transformer dekoder generiše fiksan skup predikcija, dok se tokom obučavanja Hungarian algoritmom svaka anotirana instanca uparuje sa najviše jednom predikcijom, a preostali upiti označavaju kao "bez objekta".

Važno je razlikovati lokalni izlaz bez NMS postprocesiranja (engl. *NMS-free*) od problema koji nastaje u tiled inferenciji. Čak i ako model u jednoj pločici proizvede neredundantan skup instanci, isti fizički objekat može biti nezavisno predviđen u više susednih pločica. U tom slučaju redundantnost nije posledica višestrukih izlaza jednog modelskog prolaza, već

posledica više lokalnih pogleda na isti globalni region. Zato se NMS-free svojstvo modela mora dopuniti globalnom konsolidacijom maski.

2.2.2 Višeskalna i deformabilna pažnja

Puna pažnja nad svim tokenima ima kvadratni memorijski i vremenski rast po broju tokena. To je ozbiljno ograničenje za slike visoke rezolucije, naročito ako se žele očuvati sitni objekti i fine granice. Zbog toga se vizuelni transformeri u detekciji i segmentaciji najčešće koriste u efikasnijim oblicima: kroz hijerarhijske reprezentacije, prozorsku pažnju, retko uzorkovanje ili kombinovanje sa konvolucionim mapama osobina.

Swin Transformer, na primer, koristi pažnju unutar lokalnih prozora i njihovo pomeranje kroz slojeve, čime zadržava hijerarhijsku višeskalnu strukturu uz manji trošak od pune globalne pažnje [24]. Deformable DETR uvodi višeskalnu deformabilnu pažnju: svaki upit ne poredi se sa svim prostornim položajima, već uzorkuje mali broj relevantnih tačaka oko referentnih lokacija na više nivoa osobina [21]. Time se smanjuje broj poređenja i olakšava lokalizacija objekata različitih veličina.

Ove osobine su neposredno relevantne za vazdušne snimke. Scene mogu sadržati veliki broj malih objekata, objekte vrlo različitih razmera i ponavljajuće teksture koje zahtevaju i lokalnu preciznost i širi kontekst. Višeskalne mape osobina i deformabilna pažnja omogućavaju da query-based model selektivno koristi prostorne regione koji su informativni za određenu instancu, bez pune pažnje nad celom visokorezolucionom slikom.

2.2.3 Segmentacija zasnovana na upitima i RF-DETR kontekst

Kod segmentacije instanci query-based pristupi povezuju reprezentaciju objekta sa predikcijom klase i binarne maske. Mask2Former formuliše semantičku, instancnu i panoptičku segmentaciju kao klasifikaciju maski i koristi maskiranu pažnju da svaki upit agregira karakteristike iz regiona relevantnog za njegovu trenutnu masku [22]. Mask DINO dodatno povezuje DINO/DETR detekcioni okvir sa predikcijom maski i objedinjuje više segmentacionih zadataka u jednoj arhitekturi [23].

U ovoj disertaciji značaj ovih modela nije u detaljima njihovog gubitka ili unutrašnjeg dekodera, već u tri svojstva. Prvo, upiti objekata daju prirodan mehanizam za predikciju skupa instanci. Drugo, višeskalna i deformabilna pažnja čine ih pogodnim za objekte različitih dimenzija. Treće, lokalni izlaz modela može biti NMS-free, pa se kasnija redundantnost u tiled režimu mora tumačiti kao globalni problem usklađivanja nezavisnih lokalnih predikcija.

RF-DETR i srodni savremeni modeli koriste upravo ovakvu kombinaciju efikasne ekstrakcije vizuelnih karakteristika, objektnih upita i pažnje usmerene na relevantne prostorne regione. To ih čini pogodnim kandidatima za segmentaciju instanci na pločicama visoke rezolucije, ali ne uklanja potrebu za posebnim tretmanom predikcija koje nastaju na granicama i preklopima pločica.

2.2.4 Ograničenja pri visokim rezolucijama

Transformeri olakšavaju modelovanje udaljenih odnosa, ali ne mogu neograničeno zameniti globalnu obradu cele slike. Kod vrlo velikih vazdušnih snimaka puna pažnja nad svim pikselima ili sitnim zakrpama ostaje nepraktična, a smanjenje rezolucije može ukloniti upravo one male objekte zbog kojih se koristi visokorezolucioni ulaz. Zbog toga se i transformerski modeli često primenjuju nad preklopljenim pločicama.

Tiled obrada vraća problem konteksta na nivo inferencionog režima. Model u jednoj pločici vidi samo lokalni prozor, bez neposrednog znanja o tome kako se njegova predikcija odnosi prema predikcijama iz susednih pločica. Objekat presečen granicom može biti predstavljen različitim query predikcijama u različitim lokalnim prolazima, sa različitim maskama i confidence skorovima. Te lokalne predikcije čine ulaz za kasniju globalnu konsolidaciju, a ne problem koji se može rešiti samo izborom CNN ili Transformer backbone-a.

2.3 Relevantni pristupi segmentaciji i konsolidaciji maski

U prethodnim odeljcima predstavljene su konvolucione i transformerske arhitekture kao osnova savremenih modela za segmentaciju instanci. Za ovu disertaciju jednako je važno kako se od lokalnih izlaza formira konačan skup instanci: tradicionalni detektori često generišu više kandidata koji se naknadno redukuju, dok end-to-end modeli nastoje da već tokom obučavanja nauče preslikavanje slike u skup razdvojenih objekata.

Za ovu disertaciju važno je razlikovati dva nivoa obrade. Prvi se odnosi na formiranje konačnog skupa predikcija unutar jednog modelskog ulaza, gde se koriste postupci kao što je NMS, ili arhitekture projektovane tako da takav korak nije potreban. Drugi nivo odnosi se na povezivanje predikcija iz više nezavisnih modelskih prolaza. Taj problem nastaje kod preklopljenih pločica i ne rešava se uklanjanjem NMS-a iz lokalnog modela.

2.3.1 Postprocesna deduplikacija predikcija

Tradicionalni objekatski detektori često proizvode više predikcija za isti fizički objekat. Standardni NMS ovu redundantnost rešava sortiranjem predikcija prema skoru pouzdanosti: najpouzdaniji kandidat se zadržava, a slabije predikcije iste klase uklanjaju se ako njihovo preklapanje sa već prihvaćenom predikcijom prelazi zadati prag [25].

Postupak je jednostavan i efikasan, ali odluka zavisi od praga i poretka confidence skorova. U gustim scenama veliki IoU može nastati između različitih instanci, dok fragmenti istog objekta mogu imati mali ili nulti IoU.

Soft-NMS ublažava ovaj problem tako što preklopljene kandidate ne uklanja odmah, već postepeno umanjuje njihove confidence vrednosti prema stepenu preklapanja [26]. Time se smanjuje mogućnost da validna predikcija bude eliminisana, ali odluka i dalje prvenstveno zavisi od odnosa među graničnim pravougaonicima.

Kod segmentacije instanci dostupna je dodatna informacija u obliku binarne ili probabilističke maske. Dve instance mogu imati preklopljene granične pravougaonike, a prostorno razdvojene maske. Zato box-level deduplikacija ne koristi punu geometrijsku informaciju koju segmentacioni model proizvodi.

Syncretic-NMS, odnosno Non-Maximum Merging (NMM), uvodi mogućnost da se odgovarajuće predikcije ne samo potisnu već i objedine u zajedničku reprezentaciju [27]. Takvi postupci predstavljaju prelaz od uklanjanja kandidata ka konsolidaciji. Isti princip proširen je na maske nastale tiled inferencijom kroz *Enhanced Syncretic Mask Merging*, gde se prostorni i geometrijski kriterijumi koriste za objedinjavanje fragmenata iste instance na granicama pločica [28].

2.3.2 End-to-end i NMS-free pristupi

Alternativni pravac pokušava da potrebu za postprocesnom deduplikacijom ukloni iz inferencionog postupka. Ključna ideja jeste da tokom obučavanja jedan stvarni objekat bude povezan sa ograničenim, idealno jednim dominantnim modelskim izlazom, tako da mreža ne generiše veliki skup redundantnih predikcija.

DETR formuliše detekciju kao direktnu predikciju neuređenog skupa objekata [20]. Model koristi naučene objektne upite, a tokom obučavanja bipartitno uparivanje uspostavlja odnos između referentnih objekata i predikcija. Hungarian matching omogućava da svaka anotirana instanca bude povezana sa najviše jednom predikcijom, dok se preostali izlazi uče kao predikcije bez objekta. Deduplikacija se time ugrađuje u formulaciju učenja.

Slična ideja razvijena je kod potpuno konvolucionih detektora. OneNet pokazuje da je one-to-one dodeljivanje pozitivnih uzoraka ključan uslov za konačan skup predikcija bez NMS-a [29], dok DeFCN uvodi Prediction-aware One-To-One (POTO) dodeljivanje za end-to-end inferenciju bez klasičnog NMS postprocesiranja [30].

Pošto one-to-one dodeljivanje smanjuje broj pozitivnih uzoraka koji doprinose optimizaciji, razvijeni su pristupi koji kombinuju bogatiji signal jedan-prema-više (engl. *one-to-many*) tokom treninga sa one-to-one granom za konačnu inferenciju. Co-DETR kombinuje one-to-one dodeljivanje sa pomoćnim granama obučenim one-to-many strategijama, koje pružaju bogatiji signal supervizije tokom treninga, dok se pomoćne grane odbacuju tokom inferencije [31]. Sličan princip doslednog dualnog dodeljivanja primenjen je u YOLOv10 radi NMS-free end-to-end inferencije u okviru You Only Look Once (YOLO) porodice modela [32].

Kod segmentacije instanci isti koncept proširuje se sa detekcionih pravougaonika na maske. Mask R-CNN je klasičan region-based pristup u kome se maska instance predviđa paralelno sa klasifikacijom i regresijom graničnog pravougaonika [33]. Nasuprot tome, savremeni query-based pristupi neposredno povezuju objektnu reprezentaciju sa klasom i maskom. Mask2Former, na primer, formuliše segmentaciju kao problem klasifikacije maski i koristi masked attention za povezivanje objektnih upita sa prostornim regionima [22].

Termin *NMS-free* ne znači odsustvo inferencionih operacija. Model i dalje može primenjivati confidence prag, izbor ograničenog broja predikcija, dekodiranje maski i skaliranje na originalnu rezoluciju; u užem smislu, lokalni skup samo ne zavisi od pohlepnog potiskivanja preklopljenih predikcija.

2.3.3 Lokalna deduplikacija i konsolidacija nezavisnih predikcija

Važno ograničenje NMS-free formulacije jeste njen domen delovanja. One-to-one dodeljivanje i set-prediction regulišu odnos između predikcija nastalih u istom modelskom prolazu, ali ne uspostavljaju vezu između rezultata više nezavisnih inferencija.

Ova razlika je značajna kod metoda zasnovanih na podeli velikih slika. SAHI koristi manje isečke kako bi povećao relativnu veličinu malih objekata i omogućio efikasniju inferenciju nad visokorezolucionim snimcima [34]. Kod preklopljenih isečaka isti deo slike može biti prosleđen modelu više puta, pa dve lokalne predikcije predstavljaju nezavisne procene iste oblasti, a ne redundantne izlaze jednog prolaza.

Zato se razlikuju intra-tile redundantnost, odnosno više predikcija istog objekta unutar jednog modelskog ulaza, i inter-tile redundantnost ili fragmentacija, koje nastaju kada se ista fizička instanca nezavisno predvidi u više delimično preklopljenih regiona. NMS-free arhitekture prvenstveno utiču na prvi problem; drugi ostaje prisutan jer predikcije iz različitih pločica nemaju zajednički modelski kontekst.

Nakon vraćanja lokalnih maski u koordinatni sistem originalne slike mogu se pojaviti duplikati, delimično preklopljene maske, komplementarni fragmenti ili geometrijski različite reprezentacije istog objekta. Njihova konsolidacija je zato zaseban prostorni problem i zahteva kriterijume koji prevazilaze lokalni mehanizam formiranja predikcija.

2.3.4 Od IoU-zasnovanog spajanja ka prostornom grupisanju maski

Najjednostavniji postupci globalne deduplikacije mogu primeniti NMS ili merging varijantu nakon transformacije predikcija u zajednički koordinatni sistem. Međutim, to pretpostavlja da predikcije iste instance imaju dovoljno veliki međusobni presek, što kod lokalno generisanih maski često nije ispunjeno.

Dva komplementarna fragmenta jednog objekta mogu se dodirivati ili biti blizu, a imati nizak ili nulti IoU. Nasuprot tome, maske različitih bliskih objekata mogu imati preklopljene obuhvatne pravougaonike. Odluka o pripadnosti istoj fizičkoj instanci zato se bolje posmatra kao problem prostorne kompatibilnosti nego kao problem maksimalnog preklapanja.

Jedan način formalizacije jeste predstavljanje predikcija grafom: lokalna predikcija je čvor, a grana označava dovoljno jaku prostornu ili semantičku vezu. Grupisanje tada nije sekvencijalno uklanjanje kandidata, već određivanje skupova međusobno kompatibilnih objekata. Korelaciono klasterovanje daje opšti okvir za takvo grupisanje na osnovu pozitiv-

nih i negativnih odnosa sličnosti, bez unapred poznatog broja grupa [35].

Kod velikog broja lokalnih predikcija poseban problem je generisanje kandidatnih parova. Potpuno poređenje svih parova ima kvadratni broj relacija, iako je većina maski prostorno udaljena. R-stablo (engl. *R-tree*) omogućava organizaciju obuhvatnih pravougaonika i efikasnije izdvajanje relevantnih suseda [36]. Indeks ne odlučuje o pripadnosti istoj instanci, već izbegava detaljnu analizu očigledno nerelevantnih parova.

2.3.5 Spatial Mask Merging kao polazni mask-level pristup

Spatial Mask Merging (SMM) razvijen je kao postupak globalne konsolidacije predikcija segmentacije instanci na snimcima visoke rezolucije [37]. Za razliku od postupaka zasnovanih na IoU graničnih pravougaonika, SMM odluku o povezivanju formuliše na nivou maski i kombinuje preklapanje, udaljenost granica i pouzdanost predikcija.

Predikcije se predstavljaju čvorovima kompatibilnosnog grafa, dok se kandidatni susedi najpre izdvajaju pomoću R-tree prostornog indeksa. Detaljna procena se time ograničava na prostorno relevantne parove. Kompatibilne predikcije zatim se grupišu, a maske iz iste grupe objedinjuju u jednu globalnu instancu.

Ovakav pristup predstavlja pomak u odnosu na klasičnu deduplikaciju. Cilj nije samo izabrati najbolju među redundantnim predikcijama, već rekonstruisati fizičku instancu iz više lokalno validnih opažanja, uključujući predikcije sa malim međusobnim preklapanjem kada njihova prostorna geometrija ukazuje na isti objekat.

Eksperimentalna evaluacija SMM-a na skupu iSAID pokazala je da mask-level prostorno grupisanje može poboljšati kvalitet konačnih predikcija u odnosu na standardne NMS i merging postupke, uz korišćenje prostornog indeksiranja za ograničavanje broja kandidatnih relacija [37]. SMM zato predstavlja neposrednu metodološku osnovu za dalje istraživanje globalne konsolidacije u ovoj disertaciji.

Istovremeno, opšti SMM prostornu organizaciju predikcija razmatra prvenstveno kroz njihov položaj u globalnom koordinatnom sistemu. Kod tiled inferencije dostupna je i dodatna strukturna informacija: iz koje je pločice predikcija nastala, koje pločice se preklapaju i u kom delu njihovih granica može nastati fragmentacija iste instance. Takva informacija otvara prostor za preciznije ograničavanje kandidatnih veza.

Pregled pristupa pokazuje da se konačno formiranje instanci rešava na više nivoa. NMS redukuje redundantne lokalne kandidate, NMS-free arhitekture nastoje da redundantnost uklone kroz učenje, merging pristupi objedinjuju relevantne predikcije, a prostorno i grafovsko grupisanje rekonstruiše instance iz više nezavisnih lokalnih opažanja. Za visokorezolucione snimke obrađene preklopljenim pločicama poslednji nivo ostaje potreban i kada osnovni segmentacioni model proizvodi NMS-free lokalni izlaz.

2.4 Selekcija i oporavak predikcija niže pouzdanosti

Savremeni modeli detekcije i segmentacije instanci uz svaku predikciju generišu skor pouzdanosti koji se koristi za rangiranje i selekciju izlaza. U praktičnim sistemima najčešće se primenjuje unapred definisan confidence prag: predikcije iznad praga se zadržavaju, a one ispod njega odbacuju. Ovaj postupak je jednostavan i računski efikasan, ali pretpostavlja da confidence skor dovoljno verno opisuje ispravnost predikcije.

Ta pretpostavka nije uvek opravdana. Confidence vrednost nije nužno jednaka verovatnoći da je predikcija tačna, jer modeli mogu biti nedovoljno ili prekomerno samouvereni [38]. Kod detekcije i segmentacije instanci problem je složeniji nego kod klasifikacije cele slike: kvalitet izlaza zavisi od klasne oznake, lokalizacije i geometrije maske [39, 40]. Zato confidence prag treba posmatrati kao operativnu radnu tačku koja kontroliše kompromis između zadržavanja validnih instanci i uklanjanja lažno pozitivnih predikcija.

2.4.1 Confidence, kalibracija i kvalitet predikcije

Confidence skor je koristan za rangiranje kandidata, ali ne opisuje u potpunosti njihov kvalitet. Predikcija može imati visok confidence, a ipak nepreciznu masku ili granični pravougaonik; obrnuto, predikcija nešto niže pouzdanosti može odgovarati stvarnom objektu. Izbor praga zato neposredno utiče na odnos preciznosti i odziva. U neuravnoteženim problemima ovaj odnos se pogodnije analizira Precision–Recall krivom nego ROC krivom [41], a optimalna radna tačka može zavisiti od modela, klase i kvaliteta lokalizacije [42].

Dobro kalibrisan model trebalo bi da zadovolji princip da među predikcijama kojima je dodeljena određena pouzdanost približno isti udeo zaista bude ispravan. Kod instancne segmentacije ovaj odnos zavisi od semantičke klase, položaja, skale, veličine predikovanog pravougaonika i geometrijskog poklapanja sa referentnim objektom. Zbog toga su razvijeni multivarijantni pristupi kalibraciji koji, pored confidence vrednosti, uključuju prostorne i geometrijske informacije [39, 40]. Kalibracija ipak nije isto što i procena neizvesnosti: prva meri slaganje numeričkog skora sa empirijskom verovatnoćom ispravnosti, dok druga procenjuje sigurnost modela u konkretnu odluku. U segmentaciji instanci jedna vrednost dodatno sažima više aspekata kvaliteta, od klase do tačnosti konture, pa je confidence prag praktična radna tačka, a ne prirodna granica između tačnih i pogrešnih predikcija.

2.4.2 Selektivno odlučivanje i naknadna procena kvaliteta

Problem donošenja odluke samo nad podskupom dovoljno pouzdanih primera razmatra se kroz *selective prediction*, *selective classification* i pristupe sa opcijom odbacivanja (*reject option*). Osnovna ideja je da model ne mora prihvatiti svaki primer, već može odbiti slučajeve za koje procenjuje da je rizik greške previsok. Povećavanjem selektivnosti moguće je smanjiti rizik nad prihvaćenim primerima, ali uz manju pokrivenost. Geifman i El-Yaniv

pokazuju da se ovaj kompromis može eksplicitno kontrolisati radi dostizanja zadatog nivoa rizika [43], dok SelectiveNet integriše selekcionu funkciju u sam proces učenja uz ograničenje minimalne pokrivenosti [44].

Selektivni oporavak low-confidence predikcija predstavlja obrnut, ali konceptualno srodan problem. Polazni confidence filter je već odbacio deo izlaza, a sekundarni postupak pokušava da iz tog prostora vrati samo ograničen broj kandidata čiji je procenjeni rizik dovoljno nizak. Cilj nije maksimalna pokrivenost low-confidence prostora, već izdvajanje dovoljno čistog podskupa, naročito kada je cena lažno pozitivnog oporavka velika.

Jedan način da se prevaziđu ograničenja osnovnog confidence skora jeste naknadna procena kvaliteta već generisanih predikcija. Predikcija osnovnog modela tada se posmatra kao novi uzorak opisan dodatnim karakteristikama, a sekundarni model procenjuje da li ona odgovara validnom objektu. MetaSeg koristi karakteristike izvedene iz predikovanih segmenata radi procene njihove validnosti i očekivanog IoU kvaliteta [45]. Slična ideja primenjena je u MetaDetect pristupu, gde se dodatni model koristi za razlikovanje tačnih i lažno pozitivnih detekcija i procenu kvaliteta lokalizacije [46]. Zajednička ideja ovih pristupa jeste da confidence vrednost osnovnog modela ne mora biti jedini kriterijum odluke: oblik, veličina, geometrija i odnos predikcije prema drugim kandidatima mogu nositi dodatnu informaciju.

Ovakva formulacija je pogodna za postprocesne sisteme zato što ne zahteva promenu osnovne segmentacione mreže. Sekundarni model može raditi nad već formiranim predikcijama, koristeći njihove skorove, geometriju i lokalni kontekst, dok se rizičan deo izlaznog prostora analizira posebnim kriterijumom.

2.4.3 Teški negativni primeri i prostorni kontekst

Detekcioni problemi često sadrže veliki broj jednostavnih negativnih primera i manji broj negativa koje je teško razlikovati od pravih objekata. U takvim uslovima za kvalitet selekcije nije presudan ukupan broj negativa, već negativni primeri blizu granice odluke ili oni kojima model dodeljuje visok skor. Online Hard Example Mining uveden je za obučavanje region-based detektora upravo da bi se učenje usmerilo na primere sa najvećim trenutnim gubitkom, umesto na veliki broj trivijalnih background regiona [47].

Sličan princip važi i za sekundarnu verifikaciju predikcija. Ako je cilj izdvojiti čist podskup validnih kandidata, najvažniji negativni primeri nisu očigledno pogrešne predikcije niskog skora, već oni kojima sekundarni klasifikator dodeljuje skor sličan pozitivnim primerima. Takvi slučajevi određuju kvalitet vrha rang-liste i ograničavaju koliko se odziv može povećati uz kontrolisan broj lažno pozitivnih prihvatanja.

Procena kvaliteta može se dodatno poboljšati korišćenjem prostornog konteksta. Noviji pristupi pokazuju da karakteristike susednih segmenata mogu doprineti proceni kvaliteta ciljne predikcije [48]. Ovaj princip je posebno relevantan kod tiled inferencije sa preklapanjem, gde isti deo originalne slike može biti posmatran u više susednih pločica. Položaj kandidata, njegova udaljenost od unutrašnje granice pločice, geometrijska saglasnost sa su-

sednim predikcijama i ponovljivost kroz više lokalnih pogleda mogu pružiti signal koji nije sadržan u pojedinačnom confidence skoru. Prostorna potvrda ipak nije dokaz ispravnosti, jer se ista sistematska greška može ponoviti u više lokalnih prikaza; zato je najkorisnija kada se kombinuje sa confidence vrednošću i drugim karakteristikama kandidata.

Posebno su važni kandidati u blizini unutrašnjih granica pločica. Oni mogu biti predviđeni na osnovu nepotpunog lokalnog konteksta, pa standardni confidence skor ne mora odražavati činjenicu da se odgovarajući objekat nastavlja u susednom lokalnom prikazu. Ako se slična maska javlja u preklopljenoj pločici, ta ponovljivost može biti dodatni argument za zadržavanje kandidata; ako se takva podrška ne javlja, kandidat i dalje može biti validan, ali odluka mora biti opreznija.

2.4.4 Nezavisan signal drugog modela

Jedan model može proizvoditi sistematske greške koje se ponavljaju kroz više lokalnih prikaza istog regiona, pa ponovljena predikcija istog modela ne predstavlja uvek potpuno nezavisan dokaz. U literaturi o proceni neizvesnosti ovaj problem se često ublažava korišćenjem više modela ili više stohastičkih realizacija istog modela. Deep Ensembles kombinuju više nezavisno obučenih neuronskih mreža i koriste raspodelu njihovih izlaza kao praktičnu procenu predictive uncertainty [49]. Srodan princip primenjen je kod detekcije objekata pomoću dropout uzorkovanja, pri čemu je za agregaciju uncertainty karakteristika potrebno povezati detekcije iz različitih uzoraka [50, 51].

Kod sistema koji raspolaže izlazima dva različita segmentaciona modela moguće je koristiti sličnu ideju bez formiranja formalnog ansambla. Predikcija drugog modela može predstavljati nezavisnu evidenciju o istom prostornom regionu: saglasnost pruža dodatnu podršku kandidatu, dok neslaganje može ukazivati na povećanu neizvesnost. Takav signal nije kalibrisana verovatnoća ispravnosti, jer različiti modeli mogu deliti trening skup, pristrasnosti i teške primere, ali kod različitih arhitektura i nezavisnih inferencionih mehanizama može biti korisna karakteristika koja nije dostupna iz confidence skora jednog modela.

U kontekstu tiled inferencije posebno je korisna kombinacija dve vrste redundantne evidencije: cross-tile evidencije, nastale preklapanjem pločica, i cross-model evidencije, nastale nezavisnom analizom istog regiona drugim modelom. Njihova kombinacija omogućava da sekundarni postupak procene kandidata koristi širi informacioni kontekst od onog koji je sadržan u pojedinačnoj lokalnoj predikciji, uz zadržavanje opreza prema slučajevima u kojima oba modela dele istu vizuelnu dvosmislenost ili odsustvo potvrde ne znači nužno grešku prvog modela.

2.4.5 Selektivni oporavak kao sekundarno odlučivanje

Sa predikcijama niže pouzdanosti može se postupati na tri načina. Prvi je neposredno odbacivanje fiksnim pragom. Drugi je promena kalibracije ili načina obučavanja osnovnog

modela kako bi izlazne vrednosti bolje odgovarale kvalitetu predikcije. Treći je zadržavanje osnovnog modela i uvođenje sekundarnog postupka koji dodatno procenjuje samo neizvesne predikcije.

Treća formulacija je pogodna kada osnovni segmentacioni model već daje pouzdane rezultate za predikcije visoke pouzdanosti. Umesto promene ponašanja celog modela, dodatna analiza se ograničava na granični deo izlaznog prostora. Problem se tada ne formuliše kao ponovno prepoznavanje objekta iz slike, već kao procena da li već generisana predikcija ima dovoljno dokaza da bude zadržana. Kod tiled inferencije tom opisu se mogu pridružiti položaj kandidata u pločici, odnos prema granicama, podrška iz drugih lokalnih posmatranja i nezavisni signal drugog modela.

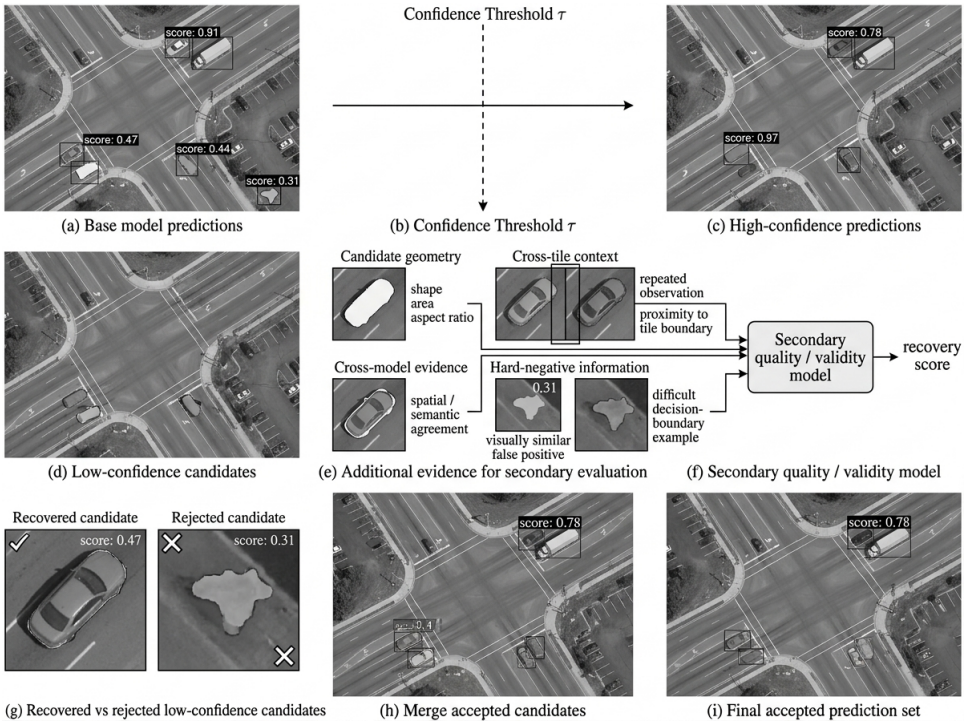
Cilj selektivnog oporavka nije nekritičko snižavanje confidence praga. Jednostavno snižavanje praga povećava broj razmatranih kandidata, ali povećava i broj lažno pozitivnih predikcija. Cilj sekundarne procene jeste da u prostoru niže pouzdanosti pronađe podskup kandidata za koje postoje dodatni dokazi o validnosti, uz kontrolu odnosa između dobijenog odziva i uvedenih lažno pozitivnih instanci. Time se selektivni oporavak predikcija niže pouzdanosti pozicionira kao zaseban problem: osnovni modelski skor ostaje važna karakteristika, ali se konačna odluka zasniva na širem skupu kalibracionih, geometrijskih, kontekstualnih i modelskih informacija.

Pregled literature zato daje više komplementarnih osnova za ovakvu formulaciju: kalibracija objašnjava zašto confidence skor nije savršena procena ispravnosti, selektivno odlučivanje uvodi kontrolu rizika, meta-klasifikacija omogućava naknadnu procenu već generisanih predikcija, rudarenje teških negativa (engl. *hard-negative mining*) naglašava značaj teških negativa, a prostorni i modelski kontekst pružaju dodatne signale posebno dostupne u tiled inferenciji.

2.5 Ograničenja postojećih pristupa i istraživački prostor

Pregled postojećih pristupa pokazuje da su segmentacija instanci na slikama visoke rezolucije, tiled inferencija, uklanjanje redundantnih predikcija, spajanje maski i naknadna procena kvaliteta modelskih izlaza razvijani kao povezani, ali ne potpuno objedinjeni pravci. Otvoreni problemi postaju najizraženiji kada se velika slika obrađuje pomoću preklopljenih pločica: lokalni detalji se čuvaju, ali se istovremeno proizvode nezavisne, delimično redundantne i često fragmentisane predikcije.

Prvi istraživački prostor odnosi se na način globalne konsolidacije lokalnih predikcija. Standardni IoU, NMS i NMM postupci korisni su za uklanjanje ili spajanje očigledno preklopljenih kandidata, ali nisu dovoljni za sve oblike fragmentacije koje proizvodi tiled inferencija. Predikcije iste fizičke instance mogu imati mali međusobni presek, a bliske instance mogu imati slične obuhvatne pravougaonike. Inferencija nad pločicama pritom pruža dodatnu strukturnu informaciju: poznato je iz kog lokalnog regiona predikcija potiče, koje se



Slika 2.5. Koncept selektivnog oporavka predikcija niže pouzdanosti. Predikcije iznad osnovnog praga pouzdanosti zadržavaju se neposredno, dok se kandidati niže pouzdanosti dodatno procenjuju korišćenjem geometrijskih karakteristika, prostornog konteksta između susednih pločica i nezavisne modelske evidencije. Sekundarni postupak prihvata samo podskup kandidata za koje postoji dovoljno dodatnih pokazatelja validnosti, dok se preostali kandidati odbacuju.

pločice međusobno preklapaju i u kojim delovima slike je moguće očekivati višestruko opažanje iste instance. Ova informacija je posebno značajna kod modela koji već unutar jednog lokalnog inferencionog prolaza proizvode NMS-free izlaz, jer tada glavni izvor redundantnosti više nije lokalno umnožavanje kandidata, već nezavisno posmatranje istog objekta iz više pločica.

Postojeći mask-level pristupi, uključujući Spatial Mask Merging, pokazuju da prostorno indeksiranje, rastojanje granica i preklapanje maski mogu biti efikasnija osnova za globalnu konsolidaciju od oslanjanja isključivo na IoU graničnih pravougaonika [37]. Međutim, SMM ne koristi eksplicitno poznatu topologiju tiled prostora i fizičke odnose između susednih pločica. Umesto jednakog tretiranja svih prostorno bliskih predikcija, moguće je prioritet dati regionima u kojima se susedne pločice zaista preklapaju, unutrašnjim šavovima između njih i topološki opravdanim odnosima između pločica. Takav pristup potencijalno smanjuje broj nepotrebnih poređenja i rizik od spajanja bliskih, ali fizički različitih instanci.

Dodatni prostor za unapređenje postoji kod složenijih slučajeva u kojima ista instanca nije predstavljena samo jednim parom lokalnih maski. Objekat može biti prisutan u tri ili više susednih pločica, naročito u blizini preseka njihovih granica. Tada parovno spajanje ili strogo zahtevana direktna kompatibilnost između svih članova grupe može ostaviti deo lokalnih opažanja nekonsolidovanim. Poznata topologija tiled rasporeda zato može predstavljati dodatni

izvor ograničenja pri formiranju grupa, bez potrebe da se dozvoli nekontrolisano prostorno povezivanje udaljenih predikcija.

Drugi istraživački prostor odnosi se na predikcije koje se nalaze ispod osnovnog confidence praga. Kao što je prethodno pokazano, confidence vrednost nije savršena procena ispravnosti predikcije, a sekundarni modeli mogu koristiti dodatne karakteristike kako bi pouzdanije procenili kvalitet detekcije ili segmenta [46, 48]. Kod tiled inferencije postoji dodatna informacija koja u klasičnoj inferenciji nije dostupna u istom obliku: isti deo scene može biti posmatran iz više preklopljenih lokalnih regiona.

Preklapanje zato ne predstavlja samo mehanizam za smanjenje graničnih artefakata, već i potencijalni izvor dodatnog kontekstualnog signala. Položaj kandidata u odnosu na unutrašnju granicu pločice, prisustvo kompatibilne predikcije u susednoj pločici, međusobni geometrijski odnos maski i relativna stabilnost predikcije kroz više lokalnih opažanja mogu pružiti dodatne informacije o kvalitetu kandidata. Takav signal može biti dopunjen i nezavisnim dokazom drugog segmentacionog ili detekcionog modela, čime se odluka o zadržavanju predikcije ne zasniva samo na confidence vrednosti jednog modela.

Ovakva formulacija ima vezu sa pristupima za procenu predictive uncertainty pomoću više modela ili više stohastičkih predikcija, kod kojih međusobna varijabilnost izlaza predstavlja dodatnu informaciju o stabilnosti odluke [49, 50]. Međutim, u kontekstu tiled inferencije drugi model ne mora biti deo formalnog ansambla. Njegova uloga može biti ograničena na obezbeđivanje nezavisnog prostornog i semantičkog dokaza za low-confidence kandidat osnovnog modela. Time se cross-tile redundansa nastala preklapanjem može dopuniti cross-model evidencijom.

Važno je, međutim, da ovakav pristup ne bude poistovećen sa jednostavnim snižavanjem confidence praga. Njegov smisao je selektivna procena ograničenog prostora nesigurnih kandidata, pri čemu se dodatne prostorne, geometrijske i modelske informacije koriste za razlikovanje potencijalno validnih instanci od lažno pozitivnih predikcija. Time se otvara istraživački prostor između dve krajnosti: konzervativnog odbacivanja svih low-confidence izlaza i nekontrolisanog povećanja odziva prihvatanjem velikog broja nepouzdanih kandidata.

Treći istraživački prostor odnosi se na evaluaciju efekata tiled inferencije. Standardne globalne metrike segmentacije daju ukupnu ocenu kvaliteta, ali ne izdvajaju specifično greške koje nastaju na unutrašnjim granicama pločica. U pregledanoj literaturi nije identifikovana jedinstvena, opšteprihvaćena metrika namenjena isključivo evaluaciji instanci koje dodiruju ili presecaju te granice. Istraživanja najčešće koriste standardne metrike nad celim skupom podataka, analiziraju uticaj veličine i stepena preklapanja pločica ili posebno razmatraju degradaciju kvaliteta u blizini njihovih granica [7, 34].

Metrike usmerene na kvalitet konture, kao što su *Boundary IoU*, *Trimap IoU* i granična F-mera, pružaju detaljniju procenu geometrijske tačnosti segmentacije [52]. Međutim, one ne mere direktno da li je jedna fizička instanca zbog obrade slike po pločicama potpuno

propuštena, ostala podeljena na više predikovanih instanci ili proizvela više redundantnih lokalnih izlaza. Zbog toga je za potpuniju evaluaciju globalne konsolidacije potrebno standardne metrike segmentacije dopuniti pokazateljima koji eksplicitno opisuju ove oblike grešaka na objektima pogođenim granicama pločica.

Iz ovih ograničenja proizlaze tri povezana pravca za nastavak disertacije: globalna konsolidacija koja eksplicitnije koristi topologiju i geometriju preklopljenih pločica, evaluacioni pokazatelji usmereni na instance pogođene granicama pločica i selektivna procena ograničenog low-confidence opsega pomoću dodatne prostorne, geometrijske i cross-model evidencije. Ti pravci odgovaraju metodološkim komponentama BC-SMM, tile-boundary evaluacije i selektivnog oporavka predikcija niže pouzdanosti, bez pretpostavke da se rezultati mogu dobiti prostim snižavanjem confidence praga ili lokalnim NMS/NMM pravilom.

3. METODOLOŠKI OKVIR

3.1 Modeli i skupovi podataka

U eksperimentalnom delu istraživanja korišćena su dva modela za segmentaciju instanci: Roboflow Detection Transformer (RF-DETR) XLarge i You Only Look Once (YOLO) 26 Large. Oba modela su korišćena preko Open Neural Network Exchange (ONNX) izvršnih artefakata i obučena su za lokalnu segmentaciju na pločicama veličine 512×512 piksela. RF-DETR XLarge predstavlja primarni segmentacioni model u glavnom iSAID eksperimentalnom toku, dok se YOLO26 Large koristi kao druga arhitektura za poređenje lokalnih modela, za BONAI proveru i kao nezavisan modelski signal u rescue grani.

Teorijske razlike između DETR/query-based i YOLO/konvolucionih pristupa razmatrane su u Poglavlju 2. U ovom odeljku navode se samo one osobine konkretnih modela, ONNX izlaza i skupova podataka koje utiču na reprodukciju inferencije, dekodiranja lokalnih predikcija i kasnije globalne obrade.

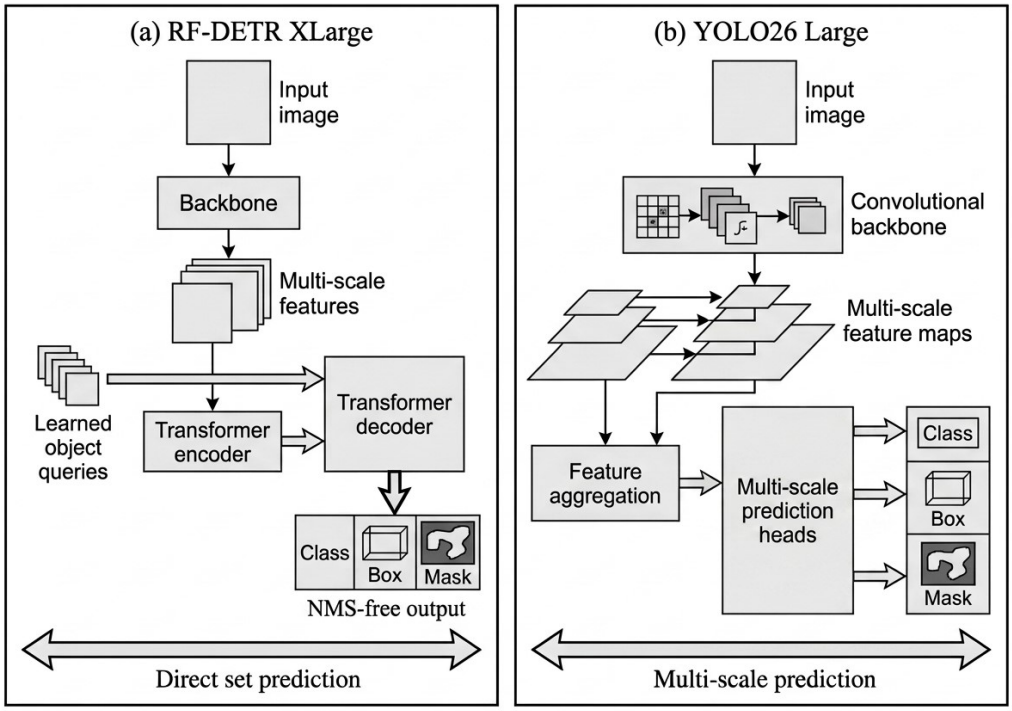
3.1.1 Korišćeni modeli segmentacije instanci

Korišćena RF-DETR konfiguracija odgovara varijanti *RF-DETR Instance Segmentation X Large*. Model pripada DETR paradigmi direktne predikcije skupa objekata i koristi transformersku obradu vizuelnih reprezentacija. Njegova izvršna reprezentacija prima RGB ulaz rezolucije 624×624 piksela i formira najviše 300 lokalnih objekatskih kandidata po jednom modelskom prolazu.

Drugi korišćeni model predstavlja *YOLO26 Large Instance Segmentation* konfiguraciju. Model koristi višeskalnu obradu i posebnu segmentacionu granu za formiranje maski instanci. Kanonska rezolucija tokom izvršavanja iznosi 640×640 piksela, a završni izlaz takođe sadrži najviše 300 lokalnih objekatskih kandidata.

Za oba modela korišćeni ONNX grafovi ne sadrže ugrađenu `NonMaxSuppression` operaciju. Time se dekodiranje lokalnih kandidata odvaja od kasnije globalne konsolidacije predikcija dobijenih iz različitih pločica. Kod YOLO26 je korišćen Roboflow deployment runtime čija konfiguracija definiše nefuzionisani spoljašnji postprocesni NMS korak (`type=nms`, `fused=false`); NMS zato nije deo ONNX grafa. Eksperimentalni kod nije dodavao poseban klijentski, tile-level ili cross-tile NMS nakon dobijanja runtime predikcija, niti je u zahtevima slao dodatni NMS IoU prag ili drugi NMS parametar.

Početna lokalna evaluacija oba modela sprovedena je nad pločicama izvedenim iz zvaničnog iSAID validation skupa, prema jedinstvenom protokolu za lokalne ulaze. Ovi rezultati predstavljaju opisni referentni kontekst za lokalne segmentacione sposobnosti korišćenih modela i prikazani su u Poglavlju 4. Lokalna evaluacija na pločicama ne predstavlja



Slika 3.1. Poređenje dve arhitektonske paradigme segmentacije instanci korišćene u eksperimentima. RF-DETR XLarge koristi transformersku obradu sa naučivim upitima objekata i direktnim formiranjem skupa predikcija, dok YOLO26 Large koristi višeskalnu konvolucionu obradu i agregaciju mapa osobina. Oba modela kao završni izlaz generišu klasu, granični pravougaonik i segmentacionu masku instance.

Tabela 3.1. Osnovne karakteristike modela korišćenih u eksperimentima.

Karakteristika	**RF-DETR XLarge**	**YOLO26 Large**
Zadatak	Segmentacija instanci	Segmentacija instanci
Arhitektonska paradigma	DETR / Transformer	YOLO / konvoluciona
Primarna uloga u radu	Glavni iSAID model; izvor osnovnih i rescue kandidata	BONAI model; nezavisan cross-model signal
Broj ciljnih klasa na iSAID	15	15
Veličina pločice za obučavanje	512×512	512×512
Ulaz izvršnog modela	624×624	640×640
Maksimalan broj kandidata	300	300
Format izvršavanja	ONNX	ONNX
Ugrađeni NMS u ONNX grafu	Ne	Ne

direktnu pre/posle referentnu osnovu za end-to-end konfiguracije evaluirane na originalnim slikama.

3.1.2 Izvršni ONNX artefakti i zajednički izlazni format

ONNX analiza korišćena je da dokumentuje stvarne izvršne artefakte dovoljno precizno da bude jasno kako se lokalne predikcije formiraju i dekodiraju pre ulaska u zajednički tok obrade. Kompletan reverse-engineering svih čvorova i inicijalizatora nije deo metodologije;

izdvojene su samo informacije koje utiču na ulaznu rezoluciju, broj kandidata, koordinatnu konvenciju, formiranje maski i prisustvo ili odsustvo NMS-a.

Tabela 3.2. ONNX i runtime informacije relevantne za dekodiranje lokalnih predikcija.

Svojstvo	**RF-DETR XLarge**	**YOLO26 Large**
Ulazni tenzor	Fiksno $1 \times 3 \times 624 \times 624$	Dinamički `images` $[B, 3, H, W]$; u eksperimentima 640×640
Maksimalan broj kandidata	300, izabranih pomoću `TopK` iz encoder lokacija	300, izabranih završnom one-to-one glavom sa dva `TopK` koraka
BBox izlaz	`dets` oblika $[1, 300, 4]$; normalizovani c_x, c_y, w, h	Deo `output0` vektora; x_1, y_1, x_2, y_2 u koordinatama ulaza
Klasni izlaz	`labels` oblika $[1, 300, 16]$; 15 ciljnih klasa i kanal bez objekta	Deo `output0` vektora; confidence skor i identifikator klase za 15 klasa
Mask izlaz	`masks` oblika $[1, 300, 156, 156]$; direktna mask-logit mapa po kandidatu	32 mask koeficijenta po kandidatu i prototype izlaz $[B, 32, 160, 160]$
Formiranje lokalne maske	Mask-logit mapa kandidata se skalira na lokalnu pločicu i binarizuje prema segmentacionom postprocesu	32 koeficijenta se linearno kombinuju sa 32 prototype mape, zatim se rezultat skalira na lokalnu pločicu i binarizuje
NMS u ONNX grafu	`NonMaxSuppression` nije prisutan	`NonMaxSuppression` nije prisutan; deployment runtime koristi nefuzionisani spoljašnji NMS, bez dodatnog klijentskog NMS-a u eksperimentalnom kodu

Nakon model-specifičnog dekodiranja oba modela daju lokalne instance koje se za potrebe narednih faza svode na zajedničku reprezentaciju

$$o_i = (M_i, b_i, s_i, \ell_i)\,, \tag{3.1}$$

gde M_i označava masku instance u koordinatama lokalne pločice, b_i njen granični pravougaonik, s_i skor pouzdanosti, a ℓ_i klasnu oznaku. Informacije o pločici porekla, njenom položaju, preklapanju i koordinatnom pomeraju ne pripadaju samom ONNX izlazu, već se dodaju tokom tiled inferencije. Time se uspostavlja model-nezavisan interfejs nad kojim se definišu rekonstrukcija globalnih koordinata, selektivni oporavak, BC-SMM grupisanje i fuzija maski.

3.1.3 Skup iSAID

Primarni skup podataka korišćen u eksperimentalnom delu istraživanja je *Instance Segmentation in Aerial Images Dataset* (iSAID) [3]. Reč je o referentnom skupu za segmentaciju instanci na vazdušnim snimcima visoke prostorne rezolucije, formiranom na osnovu

slika iz skupa *Dataset for Object deTection in Aerial images* [53]. iSAID sadrži 2.806 visokorezolucionih slika sa 655.451 anotiranom instancom raspoređenom u 15 kategorija.

Skup obuhvata klase *plane*, *ship*, *storage tank*, *baseball diamond*, *tennis court*, *basketball court*, *track field*, *harbor*, *bridge*, *Large Vehicle*, *Small Vehicle*, *helicopter*, *roundabout*, *soccer field* i *swimming pool*. Instance su anotirane na nivou piksela, pa referentna oznaka predstavlja masku instance, a ne samo granični pravougaonik.

Zvanična podela iSAID skupa obuhvata 1.411 trening, 458 validacionih i 937 test slika. U ovom istraživanju osnovni segmentacioni modeli posmatraju se kao unapred obučeni izvori lokalnih predikcija; predmet metodologije je inferencija nad slikama visoke rezolucije i naknadna obrada dobijenih maski. Glavni eksperimenti nad skupom iSAID izvode se nad zvaničnim validacionim skupom od 458 originalnih slika, za koji su dostupne referentne anotacije instanci.

Tabela 3.3. Osnovne karakteristike skupa iSAID i njegova zvanična podela.

Karakteristika	**Vrednost**
Ukupan broj originalnih slika	2.806
Broj anotiranih instanci	655.451
Broj klasa	15
Trening skup	1.411 slika
Validacioni skup	458 slika
Test skup	937 slika
Eksperimentalni evaluacioni skup	458 validacionih slika

Važno je razlikovati zvaničnu podelu originalnog skupa od pločica koje se generišu tokom inferencije. Eksperimentalna jedinica ostaje originalna validaciona slika. Podela na lokalne pločice predstavlja deo inferencionog postupka i ne formira novi evaluacioni skup. Nakon lokalne inferencije sve predikcije se vraćaju u koordinatni sistem iste originalne slike i evaluiraju u odnosu na njene referentne anotacije.

3.1.4 Skup BONAI

Kao sekundarni skup podataka korišćen je *Buildings in Off-Nadir Aerial Images* (BONAI) [54]. BONAI je razvijen za analizu zgrada na vazdušnim snimcima načinjenim pod kosim uglovima posmatranja i sadrži 3.300 slika sa 268.958 anotiranih instanci zgrada. Za svaku zgradu dostupne su anotacije krova, otiska objekta i vektora pomeraja između krova i otiska; u ovom radu kao segmentaciona instanca koristi se vidljiva zgrada, odnosno njena *roof* maska.

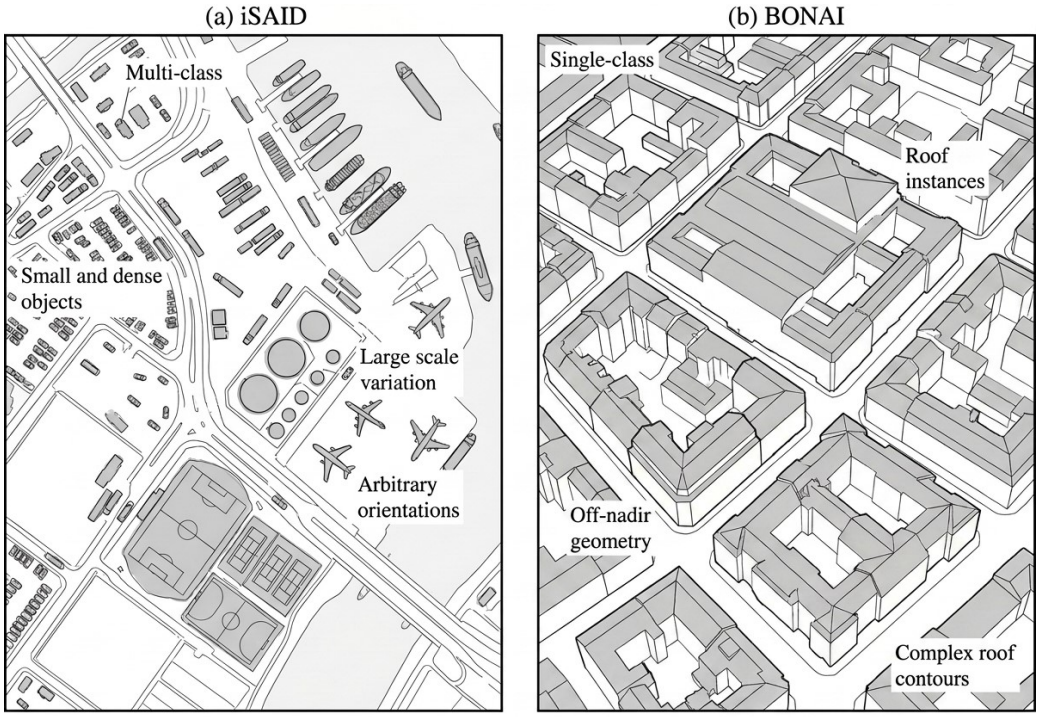
BONAI je jednoklasni domen u odnosu na višeklasni iSAID. Sadrži veće objekte složenih kontura, izraženu perspektivnu geometriju i okluzije, pa se koristi kao kontrolni domen za proveru istog postupka globalne obrade u drugačijim geometrijskim uslovima.

Zvanični BONAI protokol deli skup na 2.700 trening, 300 validacionih i 300 test slika.

Trening skup sadrži slike iz Šangaja, Pekinga, Čengdua i Harbina. Validacioni skup čine 228 slika iz Šangaja i 72 slike iz Đinana, dok test skup sadrži 200 slika iz Šangaja i 100 slika iz Sijana. Za eksperimentalnu proveru metodologije koristi se zvanični BONAI validacioni skup od 300 originalnih slika.

Tabela 3.4. Osnovne karakteristike skupa BONAI i njegova zvanična podela.

Karakteristika	**Vrednost**
Ukupan broj originalnih slika	3.300
Broj anotiranih instanci	268.958
Broj ciljnih klasa	1
Trening skup	2.700 slika
Validacioni skup	300 slika
Test skup	300 slika
Eksperimentalni evaluacioni skup	300 validacionih slika



Slika 3.2. Ilustrativno poređenje domena skupova iSAID i BONAI. iSAID predstavlja višeklasni domen sa velikim brojem malih, gusto raspoređenih i proizvoljno orijentisanih objekata različitih veličina, dok BONAI u ovom istraživanju predstavlja jednoklasni domen segmentacije vidljivih krovova, sa većim instancama, složenijim konturama i izraženom perspektivnom geometrijom.

Kao i kod iSAID-a, eksperimentalna jedinica ostaje originalna slika. Sve pločice generisane iz jedne BONAI slike koriste se isključivo kao lokalni inferencioni ulazi, a konačne predikcije se evaluiraju nakon vraćanja u koordinatni sistem originalne slike. Time tiled generisanje ne menja zvanični evaluation split ni na jednom od korišćenih skupova.

3.2 Inferencija nad pločicama

Inferencija nad originalnim slikama visoke rezolucije izvodi se njihovom dekompozicijom na manje prostorne regione, odnosno pločice. Svaka pločica obrađuje se nezavisnim prolazom segmentacionog modela, a dobijene lokalne predikcije zatim se preslikavaju u koordinatni sistem originalne slike.

U eksperimentima su korišćeni RF-DETR XLarge i YOLO26 Large modeli opisani u prethodnom odeljku. Oba modela primenjuju se nad istim fizičkim pločicama, ali sa različitom internom rezolucijom i model-specifičnom predobradom, čime tilovanje ostaje nezavisno od segmentacione arhitekture.

Preklapanje susednih pločica omogućava više lokalnih posmatranja istog fizičkog objekta. Time se objekti presečeni granicom jedne pločice mogu potpunije predstaviti u susednoj pločici, ali nastaju i redundantne lokalne predikcije koje nakon globalne rekonstrukcije ulaze u selekciju i konsolidaciju.

3.2.1 Dekompozicija slike na pločice

Neka je visokorezolutna slika I visine H i širine W. Slika se deli na kvadratne pločice fizičkog obuhvata $T \times T$, pri čemu je u svim eksperimentima korišćeno $T = 512$ piksela. Razmatrana su dva režima inferencije: osnovna konfiguracija bez sistematskog preklapanja, za koju je $r = 0$, i konfiguracija sa relativnim preklapanjem $r = 0{,}08$.

Za relativno preklapanje r, širina preklopnog regiona izražena u pikselima i korak pomeranja prozora određuju se kao

$$o = \text{round}(rT), \qquad S = T - o. \tag{3.2}$$

Za $T = 512$ i $r = 0{,}08$ dobija se preklapanje od 41 piksela i korak od 471 piksel. Broj pločica po horizontalnoj i vertikalnoj osi određuje se izrazima

$$N_x = 1 + \left\lceil \frac{\max(0, W - T)}{S} \right\rceil, \qquad N_y = 1 + \left\lceil \frac{\max(0, H - T)}{S} \right\rceil. \tag{3.3}$$

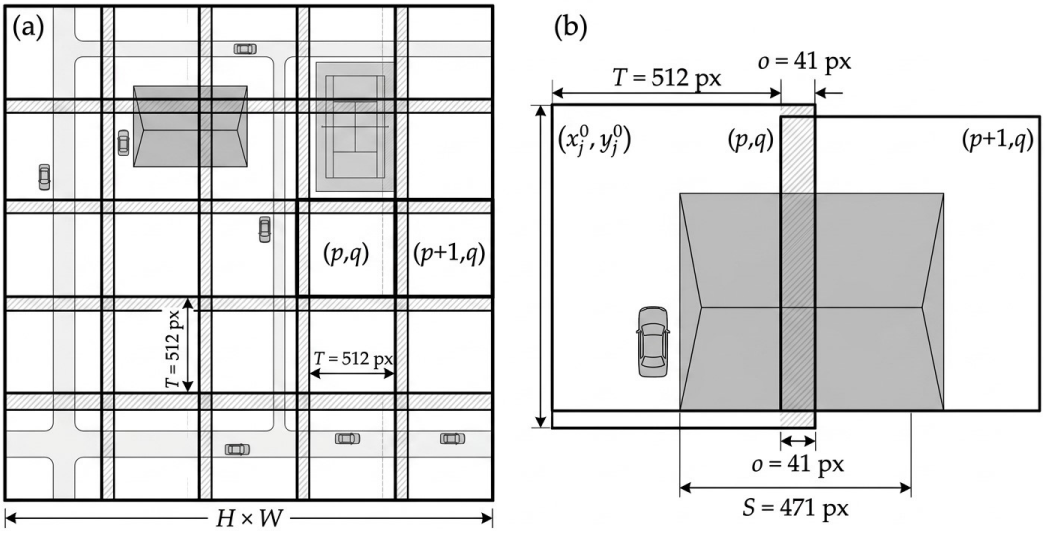
Za pločicu sa indeksima (p, q) početna koordinata određena je kao

$$x_p = \min(pS, \max(0, W - T)), \qquad y_q = \min(qS, \max(0, H - T)). \tag{3.4}$$

Ovim pravilom poslednja pločica po svakoj osi poravnava se sa desnom, odnosno donjom granicom slike, čime se obezbeđuje potpuna pokrivenost. U režimu bez sistematskog preklapanja koristi se $S = T = 512$. Ako dimenzija slike nije deljiva sa 512, lokalno preklapanje može nastati samo u poslednjem redu ili koloni kao posledica potpune pokrivenosti, a

ne kao poseban režim preklapanja (engl. *overlap*).

Za svaku pločicu $\mathcal{T}_j$ čuvaju se početna koordinata (x_j^0, y_j^0), red i kolona u mreži pločica i dimenzije originalne slike. Poreklo svake lokalne predikcije zato ostaje poznato i nakon njenog preslikavanja u globalni koordinatni sistem.



Slika 3.3. Dekompozicija visokorezolucione slike na preklopljene pločice tokom inferencije nad pločicama. Prikazani su fizički obuhvat pločice $T = 512$ piksela, preklapanje $o = 41$ piksel za $r = 0{,}08$ i korak pomeranja $S = 471$ piksel. Uvećani prikaz ilustruje slučaj objekta koji prelazi granicu dve susedne pločice i zato može biti nezavisno posmatran iz oba lokalna regiona.

3.2.2 Model-specifična predobrada i lokalna inferencija

Potrebno je razlikovati fizičku veličinu pločice izdvojene iz originalne slike od rezolucije tenzora koji prima konkretan segmentacioni model. Klijentski deo inferencionog sistema u oba slučaja izdvaja fizičku RGB pločicu dimenzija 512×512 piksela i prosleđuje je lokalnom inferencionom serveru. Promena veličine i ostala model-specifična predobrada izvršavaju se na serverskoj strani.

Inferencija je izvršavana pomoću lokalnog sistema *Roboflow Inference Server*, verzije `1.3.7`, uz GPU akceleraciju. Modeli su pozivani preko endpoint-a za segmentaciju instanci i odgovarajućih ONNX artefakata.

RF-DETR XLarge koristi fiksni interni RGB ulaz rezolucije 624×624 piksela, pa se fizička pločica 512×512 u model-specifičnoj predobradi skalira faktorom $1{,}21875$. YOLO26 Large koristi kanonsku ulaznu rezoluciju 640×640 piksela, uz podršku za dinamičke prostorne dimenzije u izvezenom ONNX artefaktu; u korišćenom inferencionom toku referentni faktor skaliranja iznosi $1{,}25$.

Za opšti slučaj fizičke pločice dimenzija $T_w \times T_h$ i internog modelskog ulaza $W_m \times H_m$, faktori transformacije zapisani su kao

$$\alpha_x = \frac{W_m}{T_w}, \qquad \alpha_y = \frac{H_m}{T_h}. \tag{3.5}$$

Promena modelske rezolucije ne menja prostorni obuhvat scene, već samo internu reprezentaciju slike. Serverski tok obuhvata dekodiranje ulaza, model-specifičnu predobradu, izvršavanje ONNX grafa i pretvaranje izlaza u predikcije segmentacije instanci. Klijent dobija klasu, skor pouzdanosti, granični pravougaonik i geometriju maske u koordinatnom sistemu fizičke pločice.

Kontrolnim lokalnim inferencijama potvrđeno je da se geometrija vraćenih predikcija odnosi na domen pločice 512×512, a ne na internu rezoluciju modelskog tenzora. Zbog toga se prilikom globalne rekonstrukcije ne primenjuje dodatno inverzno skaliranje između modelskog i fizičkog koordinatnog sistema. Višeskalna inferencija i *test-time augmentation* nisu primenjivane, tako da svi kasniji postprocesni postupci polaze od istog definisanog lokalnog inferencionog režima.

3.2.3 Rekonstrukcija lokalnih predikcija u globalnom koordinatnom sistemu

Neka je $\mathcal{O}_j$ skup predikcija dobijen inferencijom nad pločicom $\mathcal{T}_j$ čija je početna koordinata u originalnoj slici (x_j^0, y_j^0). Svaka lokalna predikcija sadrži masku instance, granični pravougaonik, skor pouzdanosti, predikovanu klasu i identifikator pločice iz koje potiče.

Lokalni granični pravougaonik i geometrija maske preslikavaju se u globalni sistem translacijom za koordinatni ofset pločice (x_j^0, y_j^0). Nakon toga globalno rekonstruisana geometrija ograničava se na validan domen originalne slike, odnosno na koordinate $0 \leq x < W$ i $0 \leq y < H$.

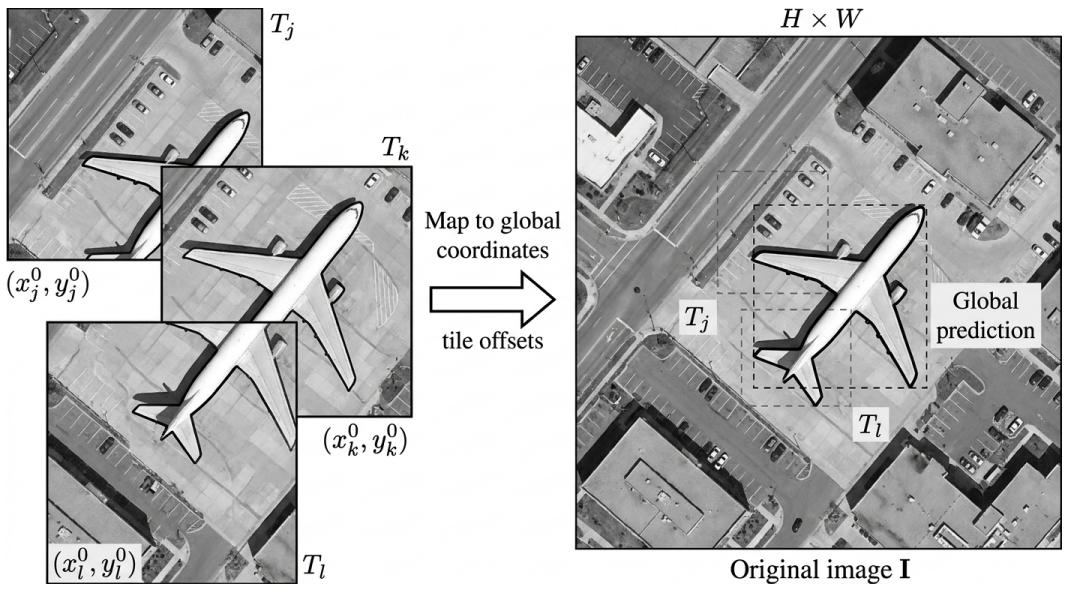
Sve lokalne predikcije time postaju izražene u jedinstvenom koordinatnom sistemu originalne slike. Informacija o izvornoj pločici se ne odbacuje: uz globalnu geometriju zadržavaju se identifikator pločice i njen položaj u mreži, što kasnije omogućava razlikovanje predikcija iz iste pločice od predikcija istog fizičkog regiona iz susednih preklopljenih pločica.

3.2.4 Formiranje početnog globalnog skupa predikcija

Nakon obrade svih pločica lokalni skupovi predikcija preslikavaju se u globalni koordinatni sistem. Početni globalni skup zapisuje se kao

$$\mathcal{O}^{\text{glob}} = \biguplus_{j=1}^{N} \mathcal{R}_j(\mathcal{O}_j), \tag{3.6}$$

gde $\mathcal{R}_j$ predstavlja transformaciju lokalne geometrije iz koordinatnog sistema pločice $\mathcal{T}_j$ u koordinatni sistem originalne slike. Operator $\biguplus$ naglašava da se u ovoj fazi ne vrši uklanjanje ili spajanje prostorno sličnih predikcija. Dve predikcije koje su nastale u različi-



Slika 3.4. Rekonstrukcija lokalnih predikcija u koordinatnom sistemu originalne slike. Predikcije dobijene iz više preklopljenih pločica preslikavaju se iz lokalnih koordinata u jedinstven globalni koordinatni sistem korišćenjem poznatih koordinatnih ofseta pločica. Time predikcije koje potiču iz različitih lokalnih pogleda istog fizičkog regiona zauzimaju odgovarajuće položaje u originalnoj slici.

tim pločicama zadržavaju zaseban identitet čak i kada predstavljaju isti fizički objekat i imaju veliko prostorno preklapanje.

Ovo svojstvo predstavlja osnovu za kasniju globalnu konsolidaciju. U overlap režimu isti fizički objekat može proizvesti više lokalnih predikcija, svaku sa sopstvenom maskom, skorom, klasom i informacijom o pločici porekla. Njihovo grupisanje ne vrši se tokom tiled inferencije, već u posebnoj fazi globalne konsolidacije.

3.2.5 Evidencija, reproduktivnost i merenje vremena

Globalno rekonstruisane predikcije čuvaju se pre primene algoritama za njihovu selekciju, grupisanje ili spajanje. Za svaku predikciju evidentiraju se identifikator originalne slike, izvorna pločica, lokalna i globalna geometrija, klasa, skor pouzdanosti, maska i odgovarajući granični pravougaonik.

Uz predikcije se čuvaju i podaci za analizu njihovog položaja u odnosu na mrežu pločica, uključujući koordinatu početka pločice i odnos prema njenim granicama. Time se ista modelska inferencija može koristiti za više nezavisnih eksperimenata postprocesiranja bez ponovnog pokretanja modela.

Ovakvo razdvajanje modelskog izlaza od naknadne obrade važno je za kontrolisano poređenje konfiguracija. SMM, BC-SMM, različiti režimi selekcije predikcija i druge ispitane varijante primenjuju se nad identičnim početnim skupom lokalnih predikcija, tako da razlike između njih ne potiču iz ponovljene inferencije.

Vreme izvršavanja prati se odvojeno od vremena kasnijih faza globalnog postprocesiranja. Za svaku pločicu evidentira se serverski prijavljeno vreme inferencije, dok se na nivou

kompletnog pokretanja beleži i ukupno proteklo vreme potrebno za obradu slike ili skupa slika.

Posebno se razlikuje modelsko vreme od ukupnog vremena inferencionog toka, koje uključuje i pripremu pločice, lokalnu komunikaciju sa serverom, dekodiranje rezultata, rekonstrukciju geometrije i upis predikcija.

3.3 Integrisani cevovod obrade

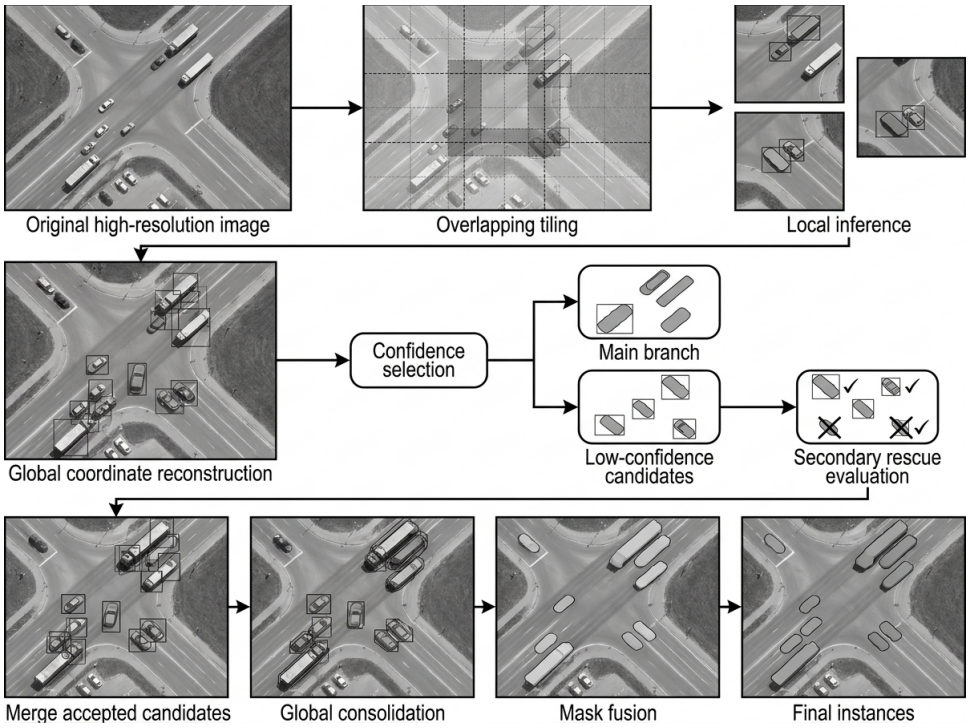
Predloženi metodološki okvir posmatra segmentaciju instanci na snimcima visoke rezolucije kao višefazni cevovod obrade (engl. *pipeline*) u kome konačan rezultat ne nastaje neposredno nakon jednog modelskog prolaza, već nakon niza međusobno povezanih koraka. Polazni segmentacioni modeli proizvode lokalne predikcije nad pojedinačnim pločicama, dok se konačne instance formiraju tek nakon njihove globalne rekonstrukcije, selekcije i konsolidacije.

Osnovna ideja pipeline-a jeste da se lokalne predikcije ne tretiraju kao konačan izlaz, već kao početni skup kandidata koji se dalje obrađuje na dva komplementarna načina. Sa jedne strane, predikcije visoke pouzdanosti zadržavaju se u glavnoj grani obrade. Sa druge strane, predikcije iz ograničenog intervala niže pouzdanosti izdvajaju se u posebnu *rescue* granu, u kojoj se dodatno procenjuje da li predstavljaju potencijalno validne instance. Nakon toga se izabrane predikcije iz obe grane objedinjavaju u zajednički skup kandidata za globalnu konsolidaciju.

Nad tako formiranim skupom zatim se primenjuje postupak prostorno ograničenog grupisanja, kojim se lokalne predikcije dovode u vezu sa odgovarajućim fizičkim instancama u originalnoj slici. Poslednji korak pipeline-a predstavlja fuzija grupisanih maski u konačne instance segmentacije na nivou originalnog snimka. Na taj način ceo postupak eksplicitno razdvaja tri problema: lokalno generisanje predikcija, selektivno zadržavanje kandidata niže pouzdanosti i globalnu rekonstrukciju identiteta instanci.

Sa stanovišta toka obrade, integrisani pipeline može se sažeti u sledeće korake:

1. dekompozicija originalne slike visoke rezolucije na pločice, pri čemu glavne overlap konfiguracije koriste delimično preklapanje susednih pločica;
2. lokalna inferencija segmentacionog modela nad svakom pločicom;
3. rekonstrukcija lokalnih predikcija u globalnom koordinatnom sistemu;
4. razdvajanje predikcija na glavnu confidence granu i *rescue* granu;
5. objedinjavanje prihvaćenih kandidata iz obe grane;
6. prostorno grupisanje i globalna konsolidacija lokalnih predikcija;



Slika 3.5. Integrisani pipeline za segmentaciju instanci na snimcima visoke rezolucije. Originalna slika deli se na pločice; u glavnim overlap konfiguracijama susedne pločice se delimično preklapaju. Nakon lokalne inferencije predikcije se rekonstruišu u globalnom koordinatnom sistemu. Predikcije više pouzdanosti zadržavaju se u glavnoj grani, dok kandidati niže pouzdanosti prolaze kroz sekundarnu procenu. Prihvaćeni kandidati se zatim objedinjuju, globalno konsoliduju i njihove maske se spajaju u konačne instance na nivou originalne slike.

7. fuzija maski u konačne instance segmentacije.

Slika 3.5 služi kao mapa postupka. Detalji tiled inferencije dati su u prethodnom odeljku, selektivni oporavak i tačno mesto njegove reintegracije u Odeljku 3.4, BC-SMM grupisanje u Odeljku 3.5, a fuzija maski u Odeljku 3.6.

3.4 Selektivni oporavak predikcija niže pouzdanosti

Standardna inferenciona putanja zadržava predikcije čiji skor pouzdanosti dostiže unapred definisani prag. Međutim, takav prag ne predstavlja apsolutnu granicu između ispravnih i pogrešnih predikcija. Neposredno ispod njega mogu se nalaziti stvarne instance koje osnovni segmentacioni model nije ocenio dovoljno visokom pouzdanošću da bi prošle kroz glavnu inferencionu granu.

U ovom istraživanju zato je uvedena zasebna grana za *selektivni oporavak predikcija niže pouzdanosti*. Njena uloga nije promena klase predikcije niti korekcija njene maske, već binarna odluka da li određenu predikciju, koja bi inače bila odbačena pragom pouzdanosti, treba vratiti u skup kandidata pre globalne BC-SMM konsolidacije.

U finalno razmatranom režimu rescue grana primenjuje se na predikcije za koje važi

$$0{,}40 \leq s_i < 0{,}50. \tag{3.7}$$

Predikcije sa skorom od najmanje $0{,}50$ nastavljaju regularnom putanjom, dok se predikcije iz intervala (3.7) prosleđuju sekundarnom klasifikatoru. Samo kandidati koje sekundarni model prihvati vraćaju se u glavni skup predikcija i zajedno sa regularnim predikcijama prosleđuju BC-SMM algoritmu.

Metodološki, ulaz u BC-SMM čine regularno prihvaćene predikcije i podskup kandidata iz intervala $[0{,}40, 0{,}50)$ koje prihvati naučena funkcija selektivnog oporavka.

3.4.1 Formiranje problema selektivnog oporavka

Rescue klasifikator se ne trenira da razlikuje proizvoljne ispravne i pogrešne detekcije. Cilj je uži: pronaći low-confidence predikcije koje odgovaraju stvarnim instancama koje nisu već uspešno obuhvaćene regularnom putanjom visoke pouzdanosti (engl. *high-confidence*).

Za potrebe formiranja razvojnog skupa, referentne anotacije (engl. *ground truth*, GT) koriste se isključivo za generisanje ciljnih oznaka. Low-confidence kandidat može dobiti pozitivnu oznaku ako sa stvarnom instancom dostiže IoU od najmanje $0{,}50$ i ako ta GT instanca nije već pokrivena regularnom high-confidence putanjom i njenom BC-SMM konsolidacijom.

Pozitivna klasa zato označava kandidata koji može vratiti prethodno propuštenu fizičku instancu, a ne svaku ispravnu low-confidence predikciju. Kada više kandidata odgovara istoj GT instanci, bira se primarni kandidat kako jedna stvarna instanca ne bi proizvodila više pozitivnih primera.

GT informacije ne predstavljaju ulazne karakteristike modela. One se koriste isključivo za formiranje ciljnih oznaka, razvojnu evaluaciju i dijagnostiku. Sve karakteristike dostupne klasifikatoru moraju biti izračunljive i tokom stvarne inferencije.

3.4.2 Overlap kao izvor dodatnog prostornog signala

Ključna osobina rescue problema proizlazi iz same tiled inferencije sa preklapanjem. Bez overlap-a jedna lokalna predikcija u velikoj meri mora biti procenjena na osnovu sopstvenog skora, maske i lokalnog sadržaja pločice. Uvođenjem preklapanja isti prostorni region može biti posmatran iz više susednih pločica.

Time nastaje dodatna prostorna redundansa: jedna fizička instanca može imati više lokalnih posmatranja. Za low-confidence predikciju zato više nije poznata samo njena sopstvena geometrija. Moguće je analizirati i njen odnos prema unutrašnjoj granici pločice, zoni preklapanja, susednoj pločici i drugim predikcijama koje posmatraju isti ili neposredno susedni prostor. Posebno je važan odnos između kandidata, šava i odgovarajuće susedne pločice.

Na primer, ako se kandidat nalazi blizu desne unutrašnje ivice pločice, poznata geometrija tiling mreže omogućava da se analizira odgovarajući levi region susedne pločice. Analogno se formiraju odnosi levo–desno, desno–levo, gore–dole i dole–gore.

Na taj način overlap nema samo ulogu da smanji gubitak objekata presečenih granicama pločica. On proizvodi dodatne informacije o prostornoj konzistentnosti predikcija koje nisu dostupne pri izolovanom posmatranju jedne pločice (engl. *tile*).

3.4.3 Reprezentacija low-confidence kandidata

Za svaki kandidat formira se tabularna reprezentacija sastavljena isključivo od karakteristika dostupnih pri primeni. Razvoj reprezentacije sproveden je postepeno, od osobina same predikcije prema eksplicitnom modelovanju cross-tile i cross-model konteksta.

Radi metodološki čistijeg poređenja informacionog doprinosa različitih reprezentacija, u njihovoj konačnoj komparativnoj evaluaciji korišćena je ista familija sekundarnog klasifikatora, Extreme Gradient Boosting (XGBoost).

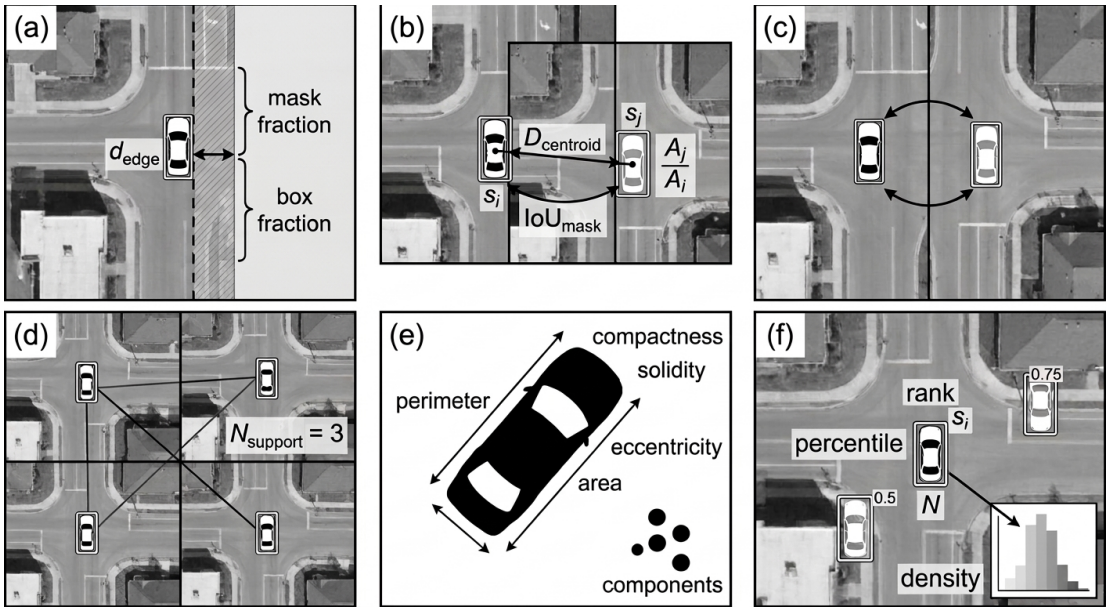
Pojedinačna karakteristika (engl. *feature*) predstavlja ulaznu veličinu sekundarnog klasifikatora. Osnovni skup karakteristika obuhvata podatke kao što su:

- skor pouzdanosti i semantička klasa kandidata;
- površina i dimenzije maske i obuhvatnog pravougaonika;
- relativni položaj predikcije unutar pločice;
- osnovne geometrijske karakteristike maske;
- lokalna gustina i susedske predikcije iste klase;
- relativni rang confidence vrednosti unutar pločice.

Proširena reprezentacija uvodi šest grupa karakteristika koje eksplicitno koriste strukturu preklopljene tiled inferencije, a njihov grafički prikaz dat je na Slici 3.6.

Položaj u odnosu na šav i zonu preklapanja. Za kandidata se određuje da li dodiruje unutrašnju ivicu pločice, koliko je udaljen od nje i koji deo njegove maske ili obuhvatnog pravougaonika leži u overlap regionu. Time se razlikuju predikcije nastale u unutrašnjosti pločice od onih čija geometrija može biti neposredno povezana sa tiled granicom.

Podrška iz susedne pločice. Za prostorno odgovarajuće susedne pločice analiziraju se predikcije iste klase koje mogu predstavljati drugo lokalno posmatranje kandidata. Relevantni odnosi uključuju preklapanje maski i pravougaonika, rastojanje granica i centroida, odnos površina, confidence susedne predikcije i razliku confidence vrednosti.



Slika 3.6. Grafički prikaz proširene reprezentacije kandidata za selektivni oporavak predikcija niže pouzdanosti: (a) položaj kandidata u odnosu na unutrašnju granicu pločice i zonu preklapanja, uključujući udaljenost od granice i udeo maske i obuhvatnog pravougaonika u preklopnom regionu; (b) geometrijska i modelska podrška predikcije iz susedne pločice, opisana preklapanjem maski, rastojanjem centroida, odnosom površina i skorovima pouzdanosti; (c) simetrična prostorna podrška sa odgovarajuće strane iste fizičke granice između susednih pločica; (d) višestruka cross-tile podrška, odnosno prisustvo odgovarajućih lokalnih posmatranja u više susednih pločica; (e) karakteristike oblika maske, uključujući površinu, obim, kompaktnost, solidity, ekscentricitet i broj komponenti; (f) lokalni kontekst na nivou pločice, uključujući rang i percentil confidence vrednosti, broj predikcija, njihovu gustinu i lokalnu raspodelu skorova.

Simetrična podrška preko šava. Geometrija mreže omogućava da se za svaku unutrašnju ivicu odredi njen odgovarajući region u susednoj pločici. Time se ne meri samo postojanje prostorno bliske predikcije, već i da li se ona nalazi na očekivanoj strani istog fizičkog šava.

Ovakav opis omogućava modelu da razlikuje proizvoljnog suseda od geometrijski konzistentnog cross-tile posmatranja.

Višestruka cross-tile podrška. Kandidat se ne opisuje samo najboljim susedom. U reprezentaciju se uključuje i broj susednih pločica sa odgovarajućom podrškom, kao i agregirane karakteristike takvih odnosa. Ovaj broj može se označiti sa N_{support}; na primer, $N_{\text{support}} = 3$ znači da su za posmatrani kandidat pronađene tri prostorno i semantički kompatibilne predikcije iz susednih pločica. Ovo je posebno relevantno u regionima u kojima se ista fizička instanca pojavljuje u više od dve pločice.

Osobine oblika maske. Iz maske se izvode karakteristike oblika, kao što su kompaktnost, solidity, ekscentricitet, odnos granice i površine, broj komponenti i slične mere. Ove karakteristike predstavljaju samo opis kandidata; u ovoj grani se ne primenjuju morfološke transformacije nad izlaznim maskama.

Kontekst pločice. Pouzdanost kandidata posmatra se i relativno u odnosu na ostale predikcije sa iste pločice. Uključuju se rang i položaj skora pouzdanosti u raspodeli, broj predikcija na pločici, raspodela njihovih skorova i lokalna gustina predikcija.

Time ista apsolutna vrednost skora pouzdanosti može imati različito značenje na jednostavnoj pločici i na pločici sa velikim brojem gusto raspoređenih objekata.

3.4.4 Nezavisni cross-model signal

Prostorni kontekst izveden iz overlap-a proširuje informacije dostupne sekundarnom klasifikatoru, ali je i dalje u velikoj meri zasnovan na predikcijama istog osnovnog modela. To znači da više lokalnih posmatranja mogu deliti slične sistematske greške.

Zbog toga je reprezentacija dodatno proširena nezavisnim dokazom drugog segmentacionog modela. U ovom istraživanju kao sekundarni izvor korišćen je YOLO26, dok low-confidence kandidat potiče iz RF-DETR inferencije.

YOLO26 se u ovoj grani ne koristi da direktno zameni RF-DETR predikciju niti da određuje njenu novu klasu. Njegova uloga je da obezbedi dodatni *cross-model evidence*: informaciju o tome da li drugi model, sa drugačijom arhitektonskom i inferencionom formulacijom, pruža prostorno i semantički kompatibilnu podršku posmatranom RF-DETR kandidatu.

Konceptualno se za kandidat o_i zato posmatra kombinacija lokalnih, overlap, cross-tile, kontekstualnih i cross-model karakteristika. Prve četiri komponente izvedene su iz RF-DETR predikcija i geometrije tiled inferencije, dok poslednja predstavlja nezavisnu potvrdu drugog modela.

Time konačna odluka ne zavisi samo od RF-DETR podrške kroz više pločica, već i od toga da li nezavisan model u istom prostoru prepoznaje kompatibilnu instancu.

3.4.5 Sekundarni klasifikator

Problem selektivnog oporavka formuliše se kao binarna klasifikacija koja za reprezentaciju kandidata $\mathbf{x}_i$ vraća skor $p_i \in [0, 1]$, odnosno procenu da kandidata treba vratiti u glavni pipeline.

Za komparativnu evaluaciju razvijenih reprezentacija i za finalnu rescue formulaciju korišćen je XGBoost. Ovaj model je izabran zbog pogodnosti za veliki tabularni skup heterogenih karakteristika, sposobnosti modelovanja nelinearnih interakcija između prostornih, geometrijskih i cross-model signala i mogućnosti efikasnog GPU izvršavanja.

3.4.6 Optimizacija i finalna konfiguracija XGBoost klasifikatora

Hiperparametri finalnog XGBoost klasifikatora određivani su isključivo na razvojnom skupu, pomoću Optuna optimizacije sa TPE (*Tree-structured Parzen Estimator*) postupkom uzorkovanja i 200 pokušaja. Pretraga je obuhvatala broj stabala, brzinu učenja, maksimalnu dubinu, minimalnu težinu deteta, `gamma`, uzorkovanje redova i karakteristika, L_1 i L_2

regularizaciju i težinu pozitivne klase. Fiksni parametri bili su `objective=binary:logistic`, `eval_metric=aucpr`, `tree_method=hist` i `random_state=42`.

Primarni kriterijum izbora nije bio najveći globalni PR-AUC, već maksimalan broj tačno oporavljenih DEV kandidata uz ograničenje $FP/TP \leq 0{,}25$ i $TP > 0$. Ako je više konfiguracija zadovoljavalo isti broj TP, prednost je dobijala konfiguracija sa manjim brojem FP, zatim većom preciznošću, višim PR-AUC i višim pragom. Official validation skup nije korišćen za izbor hiperparametara niti za izbor operativnog praga; nakon izbora, model i prag su zamrznuti pre finalne evaluacije.

Tabela 3.5. Finalni hiperparametri XGBoost klasifikatora za selektivni oporavak.

Parametar	iSAID	BONAI
`n_estimators`	1450	950
`learning_rate`	0,011674	0,012474
`max_depth`	10	5
`min_child_weight`	0,738591	26,438969
`gamma`	5,626422	7,575891
`subsample`	0,995147	0,827281
`colsample_bytree`	0,847250	0,524174
`colsample_bylevel`	0,941797	0,690176
`colsample_bynode`	0,662997	0,728294
`reg_alpha`	0,000001	0,000000
`reg_lambda`	0,317969	0,001400
`scale_pos_weight`	0,895977	0,313976
Operativni prag θ_R	0,624776	0,170566

BONAI klasifikator nije koristio iSAID model niti njegove hiperparametre, već je nezavisno optimizovan istom procedurom nad BONAI TRAIN/DEV podelom. U oba slučaja grana selektivnog oporavka ima izrazito asimetričan cilj: mali broj dodatnih lažno pozitivnih predikcija može poništiti korist ostvarenu vraćanjem propuštenih instanci.

3.4.7 Izbor operativnog praga

XGBoost za svaki kandidat vraća kontinuirani skor p_i. Kandidat se prihvata kada taj skor dostigne operativni prag θ_R, koji se ne postavlja unapred na standardnu vrednost $0{,}5$, već se određuje na razvojnom skupu.

Za svaki razmatrani prag računaju se preciznost i odziv, ali cilj nije maksimizovanje odziva samo po sebi, već oporavak što većeg broja stvarnih instanci uz unapred ograničen broj pogrešno prihvaćenih kandidata.

Operativni prag zato se bira tako da maksimizuje broj tačno oporavljenih kandidata uz ograničenje. U nastavku se TP i FP odnose na tačno pozitivne i lažno pozitivne prihvaćene rescue kandidate:

$$\frac{FP(\theta_R)}{TP(\theta_R)} \leq 0{,}25, \qquad TP(\theta_R) > 0. \tag{3.8}$$

Ograničenje (3.8) odgovara zahtevu da precision bude najmanje $0{,}80$. Drugim rečima, cilj sekundarnog modela nije agresivno vraćanje velikog broja kandidata, već pronalaženje malog, dovoljno čistog podskupa koji može biti bezbedno vraćen u glavni pipeline.

Ako više pragova daje isti broj TP, prednost se daje rešenju sa manjim brojem FP, zatim većom precision vrednošću i, konačno, višim pragom.

3.4.8 Mere kvaliteta rangiranja

Pored primarne operativne tačke određene zadatim ograničenjem, kvalitet kontinuiranog skora klasifikatora za selektivni oporavak prati se pomoću površine ispod krive preciznost–odziv (PR-AUC) i površine ispod ROC krive (ROC-AUC). Obe krive dobijaju se promenom operativnog praga θ_R: PR-AUC opisuje kvalitet rangiranja pozitivnih kandidata u odnosu na kompromis između preciznosti i odziva, dok ROC-AUC meri ukupnu sposobnost skora da razdvoji pozitivne i negativne kandidate nezavisno od jednog fiksnog praga.

U posmatranom rescue problemu PR-AUC ima veću praktičnu važnost od ROC-AUC zbog izražene neuravnoteženosti klasa: broj low-confidence kandidata koje ne treba vratiti znatno je veći od broja stvarno korisnih rescue kandidata. Kada je negativna klasa velika, relativno mali FPR može odgovarati značajnom apsolutnom broju false-positive kandidata. Zbog toga ROC-AUC može ostati relativno visok i kada model u praktično važnom gornjem delu rang-liste proizvodi više pogrešno prihvaćenih predikcija nego što rescue grana može bezbedno tolerisati.

PR-AUC se zato koristi kao glavna sekundarna mera kvaliteta rangiranja, dok ROC-AUC služi kao komplementarna mera ukupne diskriminativne sposobnosti. Nijedna od ove dve metrike ne određuje neposredno operativni prag rescue grane. Primarni metodološki kriterijum ostaje maksimizacija broja tačno oporavljenih instanci uz ograničenje

$$\frac{FP}{TP} \leq 0{,}25. \tag{3.9}$$

Model zato može imati dobar globalni PR-AUC ili ROC-AUC, ali nedovoljno čist gornji deo distribucije za bezbednu primenu rescue grane. Praktična upotrebljivost se zato ocenjuje u unapred definisanom constrained operating point-u.

3.4.9 Izolacija razvojnog i konačnog evaluacionog skupa

Razvoj rescue klasifikatora sproveden je isključivo na podacima izvedenim iz trening dela skupa. Oni su podeljeni na zaseban deo za obučavanje (TRAIN) i razvojni deo (DEV), pri čemu se podela vrši na nivou originalnih slika kako predikcije nastale iz iste slike ne bi dospele u oba skupa.

DEV deo koristi se za:

- izbor karakteristika;
- poređenje različitih reprezentacija i rescue konfiguracija;
- optimizaciju hiperparametara;
- izbor operativnog praga θ_R;

Zvanični validacioni skup ne koristi se ni u jednoj od ovih odluka. Konačni klasifikator, njegov prostor karakteristika (engl. *feature space*) i operativni prag zamrzavaju se pre primene na nezavisni validacioni skup.

Ovakav protokol sprečava prilagođavanje rescue grane konačnim evaluacionim anotacijama i omogućava nezavisnu procenu njene generalizacije.

3.4.10 Integracija u globalni pipeline

Nakon zamrzavanja klasifikatora i operativnog praga, rescue grana se izvršava pre BC-SMM konsolidacije.

Neka je

$$\mathcal{R}_{\theta_R} = \left\{ o_i \in \mathcal{O}^{\text{low}} : f_\theta(x_i) \geq \theta_R \right\} \tag{3.10}$$

skup prihvaćenih rescue kandidata. Ulaz globalne konsolidacije tada postaje

$$\mathcal{O}^{\text{BC}} = \mathcal{O}^{\text{high}} \cup \mathcal{R}_{\theta_R}. \tag{3.11}$$

Od ovog trenutka prihvaćeni kandidat tretira se kao regularna predikcija. BC-SMM ne dobija informaciju da je predikcija prošla kroz ML rescue granu i ne koristi njen ML skor kao kriterijum grupisanja. Konsolidacija se i dalje zasniva na sopstvenim prostornim, geometrijskim i topološkim pravilima.

Time su funkcije dve komponente jasno razdvojene: rescue klasifikator procenjuje da li pojedinačnu low-confidence predikciju treba vratiti, dok BC-SMM rešava kojim lokalnim predikcijama ona pripada u globalnoj konsolidaciji. Funkcija fuzije zatim, kao i u regularnoj putanji, pretvara svaku konačnu višečlanu grupu u jednu izlaznu segmentacionu instancu.

3.5 BC-SMM: granično ograničeno prostorno spajanje maski za globalnu konsolidaciju

Nakon tiled inferencije i rekonstrukcije lokalnih predikcija u globalnom koordinatnom sistemu raspoloživ je globalni skup predikcija $\mathcal{O}^{\text{glob}}$. Svaka predikcija sadrži binarnu

masku instance M_i, obuhvatni pravougaonik b_i u globalnim koordinatama, skor pouzdanosti s_i, klasnu oznaku ℓ_i i skup identifikatora pločica $\mathcal{S}_i$ iz kojih predikcija potiče.

Zbog preklapanja susednih pločica ista fizička instanca može biti nezavisno predviđena u više lokalnih modelskih prolaza. Nakon globalne rekonstrukcije takve predikcije mogu biti gotovo identične, delimično preklopljene ili prostorno susedne sa malim, pa čak i bez značajnog površinskog preseka. Problem globalne konsolidacije zato nije samo uklanjanje duplikata, već određivanje koje lokalne predikcije nastale u različitim pločicama predstavljaju istu fizičku instancu.

Polaznu osnovu predstavlja Spatial Mask Merging (SMM), u kome se kandidatni parovi pronalaze prostornim indeksiranjem, a njihova kompatibilnost procenjuje na osnovu rastojanja granica maski, njihovog preklapanja i skorova pouzdanosti [8]. BC-SMM zadržava ovaj osnovni princip, ali kandidatnu pretragu prilagođava poznatoj geometriji i topologiji mreže pločica.

Umesto globalnog razmatranja svih prostorno bliskih predikcija, početna pretraga ograničava se na vertikalne i horizontalne kontakte susednih pločica. Za svaki unutrašnji šav formira se lokalni **koridor**, a razmatraju se samo predikcije čiji obuhvatni pravougaonici ulaze u taj region. Na taj način se već u početnoj fazi eliminiše veliki broj predikcija koje se nalaze u unutrašnjosti pločica i za koje ne postoji prostorno opravdanje da predstavljaju isto fizičko pojavljivanje kao predikcija iz susedne pločice.

Nad preostalim kandidatnim parovima primenjuje se kriterijum kompatibilnosti preuzet iz SMM algoritma, zasnovan na klasnoj saglasnosti, rastojanju granica, preklapanju maski i skorovima pouzdanosti. Tako formirane početne grupe predstavljaju osnovu za dodatnu topološku analizu karakterističnu za BC-SMM.

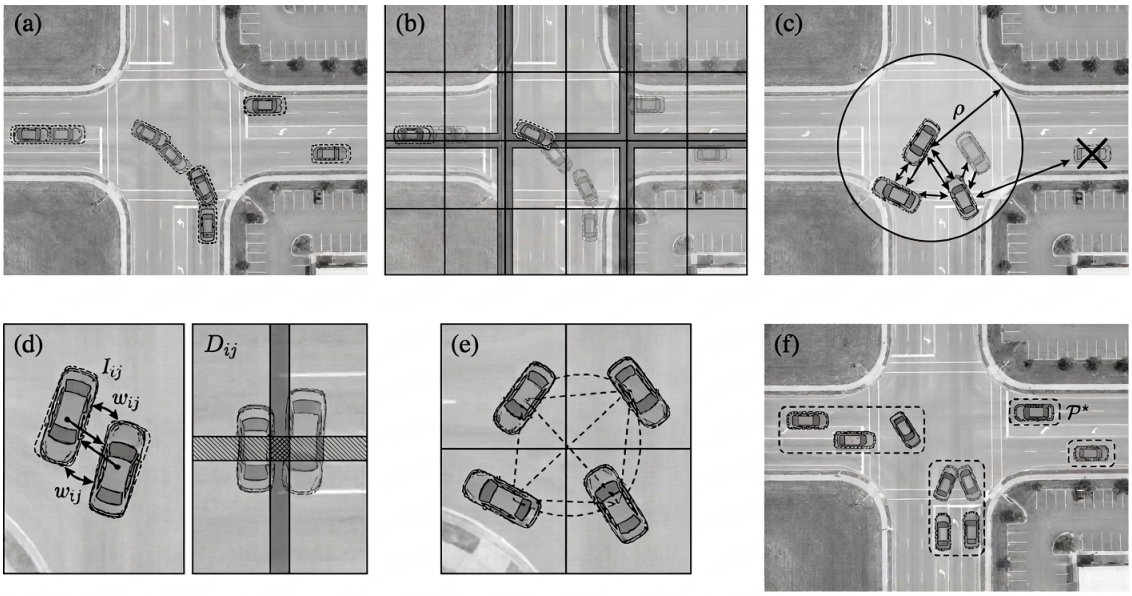
Važno je naglasiti da je izlaz BC-SMM *particija predikcija*, a ne nova segmentaciona maska. Algoritam određuje koje lokalne predikcije pripadaju istoj fizičkoj instanci, dok se njihovo geometrijsko objedinjavanje izvodi naknadno, zasebnom funkcijom fuzije.

Celokupan tok BC-SMM postupka prikazan je na Slici 3.7, dok su pojedinačne faze algoritma detaljno razmotrene u narednim podsekcijama.

3.5.1 Kriterijum kompatibilnosti predikcija

Za dve predikcije iste klase procenjuju se minimalno rastojanje njihovih granica D_{ij}, mask IoU I_{ij} i težina veze w_{ij}. Težina kombinuje normalizovanu blizinu granica, globalno preklapanje maski i manji skor pouzdanosti posmatranog para, sa koeficijentima β_1, β_2 i β_3.

Standardna kompatibilnost zahteva klasnu saglasnost i barem jedan jak geometrijski signal: dovoljno malo rastojanje granica ili dovoljno veliki mask IoU. Par se prihvata samo ako ukupna težina veze prelazi prag γ. Na ovaj način par može biti podržan blizinom granica, značajnim preklapanjem maski ili njihovom kombinacijom, dok skorovi pouzdanosti dodatno utiču na ukupnu težinu veze. Konkretne vrednosti svih pragova i težinskih koeficijenata date su u Tabeli 3.6.



Slika 3.7. Grafički prikaz BC-SMM postupka: (a) globalno rekonstruisane lokalne predikcije sa redundantnim posmatranjima istih fizičkih instanci; (b) ograničavanje kandidatne pretrage korišćenjem topologije pločica i kontaktnih koridora; (c) formiranje kandidatnih parova u prostornom dometu ρ; (d) procena kompatibilnosti parova na osnovu rastojanja granica, preklapanja maski i težine veze, uz lokalnu kontaktnu validaciju marginalnih parova; (e) topološki ograničeno grupisanje predikcija iz susednih pločica; (f) izlazna particija $\mathcal{P}^*$, čije se grupe u narednoj fazi prosleđuju funkciji fuzije maski.

3.5.2 Koridori i fizički kontakti između pločica

Neka je $\mathcal{T}$ mreža pločica jedne slike. Za svaku pločicu poznati su njen red, kolona i validni region u globalnom koordinatnom sistemu.

Ključna razlika u odnosu na globalnu SMM pretragu jeste to što BC-SMM najpre određuje gde između pločica uopšte postoji fizički prostor u kome se mogu pojaviti redundantne predikcije.

Osnovu kandidatne pretrage predstavljaju horizontalni i vertikalni šavovi između susednih pločica. Oko svakog takvog šava formira se **kontaktni koridor** širine δ_{corr}, koja je u korišćenoj konfiguraciji jednaka prostornom dometu kandidatne pretrage ρ.

Predikcija učestvuje u međutile poređenju samo ako njen obuhvatni pravougaonik ulazi u odgovarajući koridor. Predikcije udaljene od granice pločice ne ulaze u ovu kandidatnu pretragu.

Pored ortogonalnih šavova, BC-SMM razmatra i geometrijske situacije koje nastaju oko uglova i preklopljenih validnih regiona pločica. Zato se razlikuju:

1. horizontalni i vertikalni kontakti susednih pločica;
2. dijagonalni ugaoni kontakti sa fizičkim preklapanjem;
3. dijagonalni ugaoni kontakti u kojima se pločice dodiruju u uglu;

4. druga stvarna fizička preklapanja validnih regiona pločica.

Ovi kontakti zajedno određuju lokalne regione u kojima BC-SMM traži moguće parove predikcija istog fizičkog objekta.

3.5.3 Formiranje kandidatnih parova

Za kontakt između pločica T_a i T_b posmatraju se predikcije koje potiču iz odgovarajućih pločica i nalaze se u blizini kontaktnog regiona. Prostorno pretraživanje realizuje se pomoću R-tree indeksa, čime se izbegava eksplicitno poređenje svih mogućih parova.

Za dodatni fizički kontakt C_{ab} lokalni upitni region dobija se proširivanjem obuhvatnog pravougaonika tog kontakta za ρ. Kandidatni par se zadržava samo ako predikcije imaju istu klasu, potiču iz različitih pločica i njihovi obuhvatni pravougaonici nisu udaljeni više od ρ. Prvi uslov sprečava povezivanje različitih semantičkih klasa. Drugi je neophodan zato što BC-SMM rešava redundansu nastalu višestrukim posmatranjem objekta u tiled inferenciji. Treći ograničava detaljnu geometrijsku analizu na prostorno bliske instance.

Ako je isti par pronađen preko više fizičkih kontakata, evaluira se samo jednom, dok se informacije o svim relevantnim kontaktima zadržavaju uz kandidatni par.

Za novouvedene parove najpre se računa rastojanje granica. Globalni mask IoU računa se samo kada obuhvatni pravougaonici imaju stvarni površinski presek; u suprotnom se tretira kao nula. Nakon izračunavanja težine veze primenjuje se standardni kriterijum kompatibilnosti. Prihvaćeni parovi postaju veze koje mogu učestvovati u grupisanju predikcija.

3.5.4 Lokalna kontaktna validacija marginalnih mostova

Standardni SMM kriterijum namerno je restriktivan, ali kod objekata presečenih granicom pločice mogu postojati parovi čiji je globalni geometrijski signal slabiji iako se dve maske lokalno nastavljaju upravo u regionu njihovog fizičkog kontakta.

BC-SMM zato za takve marginalne parove primenjuje dodatnu **lokalnu kontaktnu validaciju marginalnih mostova**. Ona se primenjuje samo na par koji nije zadovoljio standardni kriterijum, ali i dalje ima dovoljno veliku težinu veze i barem jedan osnovni geometrijski signal: dovoljno malo rastojanje granica ili minimalni globalni IoU. Korišćeni pragovi dati su u Tabeli 3.6.

Za takav par ne posmatra se ponovo kompletna geometrija maski, već njihov odnos neposredno oko fizičkog kontakta pločica. Kontaktni region se širi za δ_C, a u tom lokalnom pojasu računa se kontaktni IoU kao preklapanje maski ograničeno na kontaktni region. Marginalni par prihvata se kao validan most kada lokalni kontaktni IoU dostigne prag τ_C.

Ovaj mehanizam primenjuje se kod ortogonalnih kontakata i dijagonalnih kontakata sa stvarnim površinskim preklapanjem. Njegova namena nije globalno ublažavanje kriterijuma kompatibilnosti, već lokalna provera kontinuiteta maski upravo u regionu u kome je isti objekat posmatran iz dve pločice.

Termin lokalna kontaktna validacija koristi se namerno kako bi se ovaj geometrijski mehanizam jasno razlikovao od zasebne ML grane za selektivni oporavak predikcija niže pouzdanosti, opisane u kasnijem odeljku.

3.5.5 Topološko grupisanje predikcija

Veze pronađene preko horizontalnih i vertikalnih koridora najpre formiraju početnu particiju $\mathcal{P}_0$. Početne grupe zatim se posmatraju kao čvorovi meta-grafa. Između dve početne grupe postoji grana ako između njihovih članova postoji prihvaćena međutile veza, bilo standardna ili potvrđena lokalnom kontaktnom validacijom.

Topološka analiza sprovodi se nezavisno za svaku povezanu komponentu ovog meta-grafa. Na taj način se globalni problem grupisanja razlaže na male lokalne konfiguracije.

Kandidatna finalna grupa G mora zadovoljiti sledeće uslove:

1. svi članovi pripadaju istoj semantičkoj klasi;
2. različiti članovi potiču iz različitih izvorišnih pločica;
3. broj članova ne prelazi $g_{\max} = 4$;
4. pločice zastupljene u grupi formiraju povezan graf fizičkih kontakata;
5. prihvaćene veze između predikcija formiraju povezan graf;
6. isti fizički kontakt ili isti smer jedne predikcije ne može biti iskorišćen za povezivanje sa više različitih partnera.

Posebno ograničenje primenjuje se na grupu veličine četiri. Četiri predikcije tada moraju poticati iz četiri pločice koje formiraju lokalni 2×2 blok susednih pločica. Takva grupa mora biti povezana sa najmanje tri međusobne veze, pri čemu najmanje dve moraju zadovoljavati standardni SMM kriterijum. Lokalna kontaktna validacija marginalnih mostova može podržati najviše jednu vezu u jednoj finalnoj grupi.

Ovo ograničenje sprečava da nekoliko pojedinačno prihvatljivih parova tranzitivno proizvede geometrijski neopravdanu veću grupu.

3.5.6 Izbor lokalne particije

Za svaku povezanu komponentu meta-grafa formiraju se moguća lokalna grupisanja koja zadovoljavaju prethodno definisana pravila. Za kandidatnu grupu bira se skup najjačih prihvaćenih veza koji povezuje sve njene članove.

Veze se prioritetno rangiraju prema većoj težini w_{ij}, većem globalnom IoU-u I_{ij} i manjem rastojanju granica D_{ij}.

Za komponentu sa najviše deset predikcija razmatraju se validne lokalne particije, a njihov kvalitet određuje se funkcijom

$$J = \sum_{(i,j)\in E_{\text{sel}}} w_{ij} - 0{,}05\, r_C + 0{,}001\, \Delta_q, \tag{3.12}$$

gde je E_{sel} skup izabranih povezujućih ivica, r_C broj marginalnih mostova prihvaćenih lokalnom kontaktnom validacijom, a Δ_q smanjenje broja početnih grupa.

Dominantan kriterijum je ukupna kompatibilnost izabranih veza. Član $-0{,}05 r_C$ daje prednost rešenjima zasnovanim na standardnoj kompatibilnosti, dok veoma mali član $0{,}001 \Delta_q$ daje prednost opravdanoj konsolidaciji kada su konkurentna rešenja veoma slična.

Kod većih lokalnih komponenti veze se obrađuju redosledom njihove kompatibilnosti i prihvataju samo kada nastala grupa i dalje zadovoljava prethodno definisana topološka ograničenja. Time se kombinatorna pretraga zadržava lokalnom, bez promene osnovnih pravila formiranja grupe.

3.5.7 Sažeti postupak BC-SMM

BC-SMM prima globalno rekonstruisani skup predikcija $\mathcal{O}^{\text{glob}}$, mrežu pločica $\mathcal{T}$ i parametre kompatibilnosti i topološkog grupisanja. Izlaz algoritma je particija $\mathcal{P}^*$, odnosno skup grupa lokalnih predikcija koje predstavljaju iste fizičke instance. Postupak se izvodi kroz sledeće korake:

1. određuju se fizički kontakti pločica i formiraju koridori oko horizontalnih i vertikalnih šavova;
2. u kontaktnim regionima R-tree indeksiranjem pronalaze se parovi predikcija iste klase, iz različitih pločica i unutar prostornog dometa ρ;
3. za kandidatne parove procenjuju se rastojanje granica, mask IoU i težina veze, a zadržavaju se parovi koji zadovoljavaju standardni SMM kriterijum;
4. kandidatni skup dopunjuje se relacijama preko dijagonalnih i drugih fizičkih kontakata, dok se nad marginalnim parovima primenjuje lokalna kontaktna validacija;
5. na osnovu ortogonalnih međutile veza formira se početna particija $\mathcal{P}_0$, a njene grupe postaju čvorovi meta-grafa;
6. početne grupe povezuju se validnim međutile vezama i formiraju se kandidatne grupe veličine najviše $g_{\max} = 4$, uz ograničenja klase, porekla iz pločica i povezanosti fizičkog i predikcionog grafa;
7. za svaku lokalnu komponentu bira se topološki validna particija sa najvećom ukupnom kompatibilnošću, a lokalna rešenja objedinjuju se u $\mathcal{P}^*$.

3.5.8 Parametri BC-SMM

BC-SMM koristi tri grupe vrednosti: nasleđene parametre SMM kriterijuma kompatibilnosti, nove konfiguracione parametre uvedene u finalni BC-SMM postupak i nekoliko fiksnih internih konstanti konkretne BC-SMM realizacije. Njihove vrednosti i uloge prikazane su u Tabeli 3.6.

Tabela 3.6. Parametri i fiksne konstante finalnog BC-SMM postupka.

Oznaka	Vrednost	Tip	Uloga
ρ	39 px	SMM	Maksimalni prostorni domet kandidatne pretrage.
τ_d	6 px	SMM	Granična vrednost rastojanja granica u standardnoj kompatibilnosti.
τ_i	0,859	SMM	Minimalni globalni mask IoU u standardnom kriterijumu.
β_1	0,362	SMM	Težina komponente zasnovane na rastojanju granica.
β_2	0,461	SMM	Težina komponente globalnog mask IoU-a.
β_3	0,120	SMM	Težina komponente pouzdanosti predikcija.
γ	0,476	SMM	Minimalna ukupna težina standardno kompatibilnog para.
λ	0,777	SMM	Kazneni parametar osnovnog SMM grupisanja.
$g_{\max}$	4	BC-SMM	Maksimalan broj lokalnih predikcija koje mogu pripadati jednoj finalnoj grupi.
$w^C_{\min}$	0,44	BC-SMM	Minimalna težina standardno odbijenog marginalnog para za ulazak u lokalnu kontaktnu validaciju marginalnih mostova.
$D^C_{\max}$	8 px	BC-SMM	Maksimalno rastojanje granica koje može pružiti geometrijsku podršku marginalnom paru.
$I^C_{\min}$	0,10	BC-SMM	Alternativni minimalni globalni IoU za ulazak marginalnog para u lokalnu kontaktnu validaciju marginalnih mostova.
τ_C	0,10	BC-SMM	Minimalni lokalni kontaktni IoU potreban da marginalni par bude prihvaćen kao validan most.
δ_C	8 px	fiksna konstanta BC-SMM	Širina lokalnog pojasa oko fizičkog kontakta u kome se računa kontaktni IoU.
n_{LS}	10	fiksna konstanta BC-SMM	Granični broj predikcija u lokalnoj komponenti do kog se koristi lokalna iscrpnija pretraga; veće komponente prelaze na deterministički greedy postupak.
η_C	0,05	fiksna konstanta BC-SMM	Koeficijent kazne za marginalni most u lokalnoj ciljnoj funkciji.
ε_q	0,001	fiksna konstanta BC-SMM	Mali koeficijent kojim se, kod vrlo sličnih rešenja, favorizuje opravdana redukcija broja grupa.

Prva grupa vrednosti nasleđena je iz SMM formulacije i određuje standardni kriterijum kompatibilnosti. Druga grupa uvodi novo topološko ograničavanje i lokalnu kontaktnu validaciju marginalnih mostova u BC-SMM. Poslednja grupa predstavlja fiksne interne konstante konkretne konačne realizacije algoritma, a ne zasebne izložene konfiguracione para-

metre.

3.5.9 Izlaz BC-SMM

Rezultat algoritma je particija $\mathcal{P}^*$ sastavljena od grupa predikcija. Grupe su međusobno disjunktne i zajedno obuhvataju početni globalni skup predikcija $\mathcal{O}^{\text{glob}}$.

Grupa sa jednim članom predstavlja predikciju za koju nije pronađeno drugo opravdano lokalno posmatranje iste instance. Višečlana grupa predstavlja skup predikcija iz različitih pločica za koje je, na osnovu prostorne, geometrijske i topološke analize, zaključeno da pripadaju istoj fizičkoj instanci.

BC-SMM u ovoj fazi ne obavlja dodatno filtriranje predikcija na osnovu toga da li je za njih pronađen odgovarajući par u drugim pločicama. Jednočlane grupe zato se zadržavaju i neposredno prenose u narednu fazu kao pojedinačne instance, dok se kod višečlanih grupa njihove članice prosleđuju funkciji fuzije radi formiranja jedinstvene izlazne instance.

BC-SMM se završava formiranjem ovih grupa. Njihove maske u ovoj fazi još uvek ostaju zasebne. Pretvaranje svake grupe G_k u jednu konačnu masku, određivanje njenog obuhvatnog pravougaonika i atributa rekonstruisane instance predstavlja narednu, odvojenu fazu funkcije fuzije.

3.6 Funkcija fuzije maski

BC-SMM kao izlaz formira particiju lokalnih predikcija na grupe koje pripadaju istoj fizičkoj instanci, dok se tek u narednoj fazi funkcijom fuzije Φ svaka takva grupa prevodi u jednu konačnu masku, kao što je ilustrovano na Slici 3.8.

Time su globalna konsolidacija i geometrijsko objedinjavanje jasno razdvojeni: BC-SMM određuje *koje* predikcije pripadaju zajedno, dok funkcija Φ određuje *kako* se njihove maske i atributi objedinjuju u konačan izlaz.

3.6.1 Fuzija višečlanih grupa

Neka je

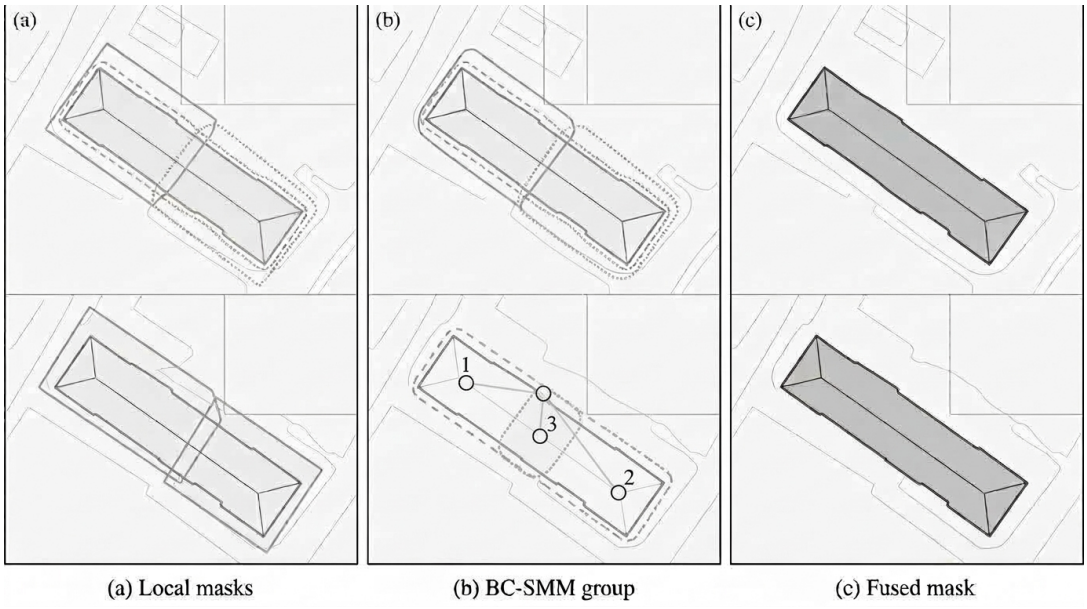
$$G_k = \{o_{i_1}, o_{i_2}, \ldots, o_{i_{m_k}}\}, \qquad m_k > 1, \tag{3.13}$$

jedna višečlana grupa dobijena BC-SMM algoritmom. Pošto je klasna konzistentnost već obezbeđena tokom grupisanja, svi njeni članovi imaju zajedničku semantičku klasu.

Konačna maska grupe dobija se piksel-po-piksel logičkom unijom svih maski:

$$M_k^{\text{con}} = \bigvee_{t=1}^{m_k} M_{i_t}. \tag{3.14}$$

Ovakva operacija zadržava prostornu podršku svih ulaznih predikcija i omogućava



Slika 3.8. Odnos između BC-SMM grupisanja i naknadne fuzije maski: (a) više lokalnih predikcija iste fizičke instance nakon globalne rekonstrukcije; (b) iste predikcije povezane u jednu BC-SMM grupu, bez promene njihove geometrije; (c) konačna maska formirana fuzijom članova grupe.

objedinjavanje kako gotovo potpunih duplikata, tako i delimično preklopljenih ili komplementarnih fragmenata iste instance. Fuzija ne generiše nove piksele izvan unije ulaznih maski i ne primenjuje dodatne morfološke transformacije.

Praktično, maske članova najpre se postavljaju u zajednički lokalni koordinatni prostor određen njihovim globalnim pomerajima. Nakon izvođenja logičke unije uklanjaju se prazni spoljašnji redovi i kolone, tako da se čuva samo minimalni lokalni region koji sadrži rezultujuću masku.

Obuhvatni pravougaonik zatim se ponovo određuje neposredno iz spojene maske:

$$b_k^{\text{con}} = \text{bbox}\left(M_k^{\text{con}}\right). \tag{3.15}$$

Na taj način obuhvatni pravougaonik nije agregat ulaznih pravougaonika, već geometrijski opis stvarne rezultujuće maske.

3.6.2 Agregacija skora pouzdanosti

Skor višečlane konsolidovane instance određuje se površinski ponderisanom sredinom skorova njenih članova. Ako je

$$a_{i_t} = |M_{i_t}| \tag{3.16}$$

površina maske člana o_{i_t}, tada je

$$s_k^{\text{con}} = \frac{\sum_{t=1}^{m_k} a_{i_t} s_{i_t}}{\sum_{t=1}^{m_k} a_{i_t}}. \tag{3.17}$$

Ovakvim ponderisanjem veće lokalne maske imaju proporcionalno veći uticaj na konačni skor, dok veoma mali fragmenti ne mogu nesrazmerno promeniti pouzdanost konsolidovane instance.

Pošto su ponderi nenegativni, konačni skor ostaje unutar opsega skorova članova grupe:

$$\min_t(s_{i_t}) \leq s_k^{\text{con}} \leq \max_t(s_{i_t}). \tag{3.18}$$

Semantička klasa konsolidovane instance nasleđuje zajedničku klasu članova grupe, dok se informacija o izvorišnim pločicama i pojedinačnim predikcijama može sačuvati kao provenijencija rezultujuće instance.

3.6.3 Postupanje sa jednočlanim grupama

Za grupu sa jednim članom

$$G_k = \{o_i\} \tag{3.19}$$

fuzija nije potrebna. Takva predikcija predstavlja već kompletnu izlaznu instancu i prenosi se bez izmene geometrije, maske i skora pouzdanosti:

$$\Phi(\{o_i\}) = o_i. \tag{3.20}$$

U realizovanom postupku za singleton se zato ne dekodira i ponovo ne kodira maska, ne računa se novi obuhvatni pravougaonik i ne preračunava se confidence skor. Zadržavaju se originalna maska, originalni bbox i originalna vrednost pouzdanosti.

Ovo svojstvo direktno prati način na koji BC-SMM formira particiju: predikcija za koju nije pronađeno opravdano drugo lokalno posmatranje ne odbacuje se, već ostaje jednočlana grupa i neizmenjena ulazi u konačni skup instanci.

3.6.4 Izlaz funkcije fuzije

Za višečlanu grupu funkcija fuzije može se zapisati kao

$$\Phi(G_k) = \left(M_k^{\text{con}}, b_k^{\text{con}}, s_k^{\text{con}}, \ell_k, \mathcal{S}_k\right), \tag{3.21}$$

gde je M_k^{con} logička unija maski, b_k^{con} obuhvatni pravougaonik ponovo izračunat iz

te maske, s_k^{con} površinski ponderisan skor, ℓ_k zajednička semantička klasa, a $\mathcal{S}_k$ objedinjena informacija o poreklu članova grupe.

Primenom funkcije Φ na sve grupe finalne particije dobija se skup

$$\mathcal{O}^{\text{con}} = \{\Phi(G_k) : G_k \in \mathcal{P}^*\}\,. \tag{3.22}$$

Broj izlaznih instanci zato je jednak broju grupa u particiji $\mathcal{P}^*$. Više lokalnih predikcija koje pripadaju istoj grupi postaju jedna konsolidovana instanca, dok jednočlane grupe ostaju nepromenjene.

Funkcija fuzije time predstavlja završni geometrijski korak globalne konsolidacije. Ona ne menja prethodno određenu strukturu grupa, ne vrši dodatno filtriranje predikcija i ne primenjuje morfološke operacije. Njena uloga ograničena je na deterministično pretvaranje svake grupe iz BC-SMM particije u jednu izlaznu segmentacionu instancu.

3.7 Evaluacioni protokol i predložene tile-boundary metrike

Evaluacija predloženog sistema obuhvata dva komplementarna nivoa. Prvi je standardna procena kvaliteta segmentacije instanci, koja omogućava poređenje modela i postprocesnih postupaka sa uobičajenim referentnim protokolima. Drugi je specijalizovana analiza grešaka nastalih zbog tiled inferencije, posebno u slučajevima kada granica pločice preseca fizičku instancu.

Predloženi skup tile-boundary metrika predstavlja deo metodološkog doprinosa ove disertacije. Njegova svrha je da razdvoji mehanizme greške koji kod tiled inferencije ostaju skriveni iza agregatnih standardnih metrika: potpuno propuštanje crossing instance, fragmentaciju njenog identiteta, redundantna opažanja, kvalitet prostorne rekonstrukcije i prekomerno spajanje različitih fizičkih objekata.

3.7.1 Standardni evaluacioni protokol

Osnovna evaluacija zasniva se na Common Objects in Context (COCO) mask AP protokolu [55]. Prikazuju se prosečna preciznost (engl. *Average Precision*, AP), AP_{50} i AP_{75}, kao i prosečni odziv (engl. *Average Recall*, AR), rezultati prema klasama i veličini objekata kada je to primenljivo. Za analize pri jednom pragu podudaranja koriste se tačno pozitivni (TP), lažno pozitivni (FP), lažno negativni (FN), preciznost, odziv i F_1, uz jedan-prema-jedan uparivanje predikcija i referentnih instanci iste klase na osnovu mask IoU. Prag podudaranja iznosi $0{,}5$.

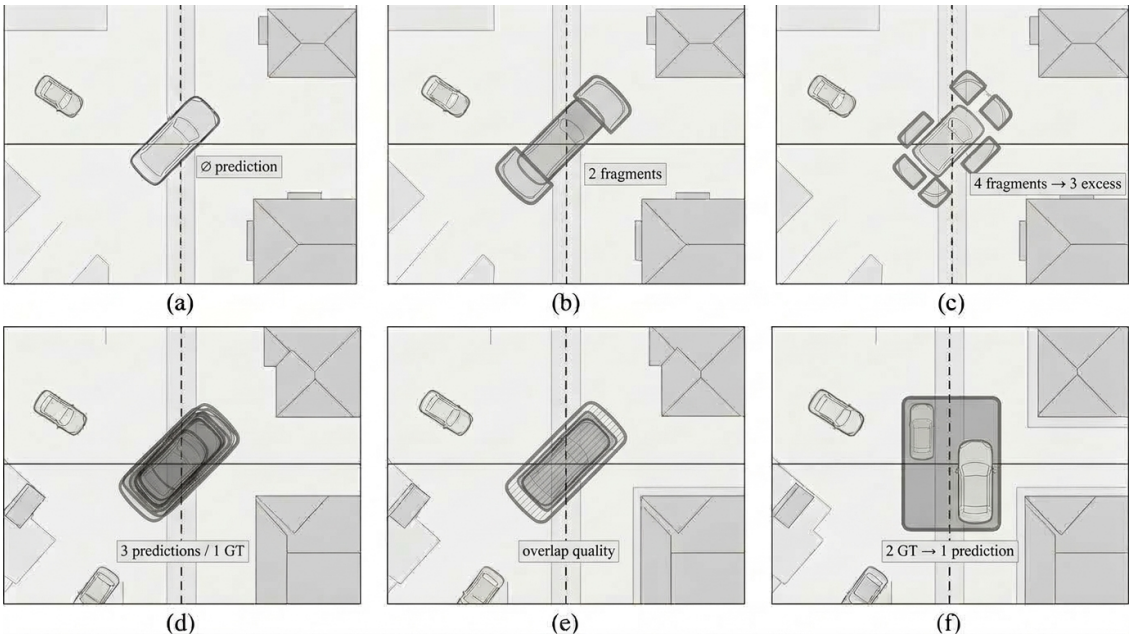
Za geometrijsku analizu uspešno uparenih instanci dodatno se prikazuju prosečan mask IoU, Dice i Boundary IoU [52]. Ove metrike opisuju kvalitet maske i konture, ali ne određuju da li je isti fizički objekat nakon tiled inferencije ostao predstavljen kao jedna ili više izlaznih instanci.

3.7.2 Motivacija za specijalizovanu tile-boundary evaluaciju

Inferencija nad pločicama je praktičan način obrade slika visoke rezolucije, jer sliku deli na lokalne ulaze koje model može neposredno obraditi [7, 6]. Međutim, granice pločica uvode diskontinuitete: ista fizička instanca može biti posmatrana kroz više lokalnih prozora, delimično prekrivena, višestruko predviđena ili podeljena na fragmente koji se u finalnom izlazu ne objedinjuju.

Standardni AP, IoU ili Boundary IoU ne razlikuju da li je greška nastala zbog potpunog propuštanja objekta, zbog podele identiteta instance, zbog redundantnih lokalnih predikcija ili zbog prekomernog spajanja susednih objekata. Zbog toga se uvodi poseban evaluacioni okvir koji posmatra samo objekte pogođene stvarnim unutrašnjim granicama pločica i razdvaja detekcionu, geometrijsku i identitetsku komponentu greške.

Specijalizovane metrike uvedene u ovom radu mere različite oblike grešaka koji mogu nastati kada fizička instanca dodiruje ili preseca unutrašnju granicu između susednih pločica. Njihova osnovna interpretacija prikazana je na Slici 3.9, dok su formalne definicije date u narednim podsekcijama.



Slika 3.9. Vizuelna interpretacija predloženih tile-boundary metrika: (a) TBMR meri propuštene instance koje presecaju granicu pločica; (b) TBFR registruje pojavu fragmentacije jedne fizičke instance; (c) TBEF meri višak fragmenata nastao takvom fragmentacijom; (d) TBDR meri redundantne predikcije iste fizičke instance; (e) $\mathrm{RIoU}_{\mathrm{cross}}$ *opisuje kvalitet prostorne rekonstrukcije instance koja preseca granicu; (f) MR registruje slučajeve prekomernog spajanja različitih fizičkih instanci.*

3.7.3 Definicija crossing populacije

Specijalizovane metrike računaju se nad referentnim instancama koje stvarna unutrašnja granica tiled mreže deli na delove pozitivne površine. Neka je $\mathcal{S}$ skup svih unutrašnjih

horizontalnih i vertikalnih granica pločica u globalnom koordinatnom sistemu. Referentna instanca G_i pripada skupu crossing instanci ako postoji granica $\sigma \in \mathcal{S}$ koja njenu masku deli na dva dela sa strogo pozitivnom površinom:

$$\mathcal{G}_{\text{cross}} = \left\{ G_i : \exists \sigma \in \mathcal{S},\ |G_i \cap \Omega_\sigma^-| > 0 \wedge |G_i \cap \Omega_\sigma^+| > 0 \right\}. \tag{3.23}$$

Samo dodirivanje granične linije ne smatra se crossing slučajem. Time se analiza usmerava na instance čija je vidljivost zaista podeljena između najmanje dve pločice. Granice se određuju iz stvarne tiled geometrije, odnosno iz manifesta generisanog tokom inferencije, a ne iz nominalne procene koraka. Zbog toga promena tiled geometrije, uključujući promenu preklapanja ili položaja poslednjih pločica, menja i skup $\mathcal{G}_{\text{cross}}$.

3.7.4 Pridruživanje predikcija i uklanjanje duplikata

Za tile-boundary dijagnostiku potrebno je odrediti koje predikcije predstavljaju istu crossing instancu. Standardni IoU nije dovoljan za parcijalne maske, jer mali fragment može gotovo potpuno pripadati referentnoj instanci, a ipak imati nizak IoU zbog velike razlike površina. Zato se koristi *Intersection over Minimum Area*:

$$\text{IoM}(G_i, P_j) = \frac{|G_i \cap P_j|}{\min(|G_i|, |P_j|)}. \tag{3.24}$$

Predikcija P_j pridružuje se najviše jednoj crossing instanci iste klase. Kriterijum je najveći IoM, uz prag $\tau_f = 0{,}50$. Ako više referentnih instanci zadovoljava prag, primenjuje se stabilan redosled razrešavanja: veći IoM, zatim veća površina preseka, a zatim stabilni identifikator referentne instance. Ovo pridruživanje koristi se samo za tile-boundary dijagnostiku i ne menja standardni COCO evaluacioni protokol.

Nakon inicijalnog pridruživanja razdvajaju se duplikati od stvarnih fragmenata. Predikcije dodeljene istoj instanci sortiraju se prema opadajućoj pouzdanosti, a kasnija predikcija uklanja se samo ako je duplikat neke već zadržane predikcije. Duplikat se prepoznaje pomoću međusobnog IoU praga $0{,}70$, containment/IoM praga $0{,}90$ i odnosa površina $0{,}67$. Ovi pragovi su fiksirani pre finalnog poređenja kako bi prag za pridruživanje obezbedio dokaz pripadnosti referentnoj instanci, a stroži duplikacioni pragovi sprečili da komplementarni fragmenti budu pogrešno uklonjeni kao redundantne predikcije. Skup inicijalno pridruženih predikcija označava se sa $\mathcal{A}_i$, a skup nakon uklanjanja duplikata sa $\widetilde{\mathcal{A}}_i$. Broj nereundantnih fragmenata reference G_i je

$$n_i^{\text{frag}} = |\widetilde{\mathcal{A}}_i|. \tag{3.25}$$

Duplikacija i fragmentacija se zato ne tretiraju kao ista pojava. Gotovo identične predikcije iste oblasti povećavaju duplikaciju, dok komplementarne predikcije različitih delova istog objekta predstavljaju fragmentaciju identiteta instance.

3.7.5 Definicija predloženih tile-boundary metrika

3.7.5.1 Tile-Boundary Miss Rate

Tile-Boundary Miss Rate meri udeo crossing instanci za koje nije pronađena nijedna odgovarajuća predikcija:

$$\text{TBMR} = \frac{\sum_{G_i \in \mathcal{G}_{\text{cross}}} \mathbf{1}\left[n_i^{\text{frag}} = 0\right]}{|\mathcal{G}_{\text{cross}}|}. \tag{3.26}$$

U ovoj formuli $\mathcal{G}_{\text{cross}}$ predstavlja skup svih referentnih instanci koje prelaze stvarnu unutrašnju granicu pločice. Veličina n_i^{frag} označava broj nereundantnih predikcija koje su pridružene instanci G_i. Indikator $\mathbf{1}[n_i^{\text{frag}} = 0]$ ima vrednost jedan samo kada za tu instancu ne postoji nijedan pridruženi fragment. Brojilac zato broji promašene crossing instance, a imenilac ukupan broj crossing instanci.

Poželjna je niža vrednost, jer TBMR meri potpuno odsustvo odgovarajuće predikcije za objekat koji je pogođen granicom pločice. Ova metrika ima kontrolnu ulogu u interpretaciji ostalih metrika: bez nje bi metoda mogla prividno smanjiti fragmentaciju tako što bi izgubila problematične objekte.

3.7.5.2 Tile-Boundary Fragmentation Rate

Tile-Boundary Fragmentation Rate računa se nad detektovanim crossing instancama, označenim sa $\mathcal{G}_{\text{cross}}^{+}$:

$$\text{TBFR} = \frac{\sum_{G_i \in \mathcal{G}_{\text{cross}}^{+}} \mathbf{1}\left[n_i^{\text{frag}} > 1\right]}{|\mathcal{G}_{\text{cross}}^{+}|}. \tag{3.27}$$

U ovoj formuli $\mathcal{G}_{\text{cross}}^{+}$ označava samo detektovane crossing instance, odnosno one za koje postoji najmanje jedna pridružena predikcija. Indikator $\mathbf{1}[n_i^{\text{frag}} > 1]$ izdvaja slučajeve u kojima je jedna fizička instanca ostala predstavljena sa više od jednog nereundantnog dela. Brojilac zato broji fragmentisane detektovane crossing instance, a imenilac sve detektovane crossing instance.

TBFR meri koliko često je detektovana crossing instanca u izlazu ostala predstavljena sa najmanje dve nereundantne predikcije. Niža vrednost je poželjna, jer označava da se parcijalna lokalna opažanja uspešnije organizuju u jedinstven identitet fizičke instance.

3.7.5.3 Tile-Boundary Excess Fragmentation

TBFR ne razlikuje slučaj sa dva fragmenta od slučaja sa više fragmenata. Intenzitet fragmentacije meri *Tile-Boundary Excess Fragmentation*:

$$\text{TBEF} = \frac{1}{|\mathcal{G}^{+}_{\text{cross}}|} \sum_{G_i \in \mathcal{G}^{+}_{\text{cross}}} \max\left(0, n_i^{\text{frag}} - 1\right). \tag{3.28}$$

U ovoj formuli član $\max(0, n_i^{\text{frag}} - 1)$ meri koliko dodatnih fragmenata postoji za instancu G_i nakon prvog, neophodnog fragmenta. Ako je instanca predstavljena jednom predikcijom, doprinos je nula; ako je predstavljena sa dve predikcije, doprinos je jedan; ako je predstavljena sa tri predikcije, doprinos je dva. Zbir u brojiocu sabira višak fragmenata za sve detektovane crossing instance, a imenilac taj zbir normalizuje njihovim brojem.

Ispravno rekonstruisana instanca doprinosi nulom; instanca predstavljena sa dva nereundantna dela doprinosi jednim viškom fragmenta, a sa tri dela sa dva viška fragmenta. Niža vrednost je poželjna, jer ne meri samo učestalost fragmentacije nego i njenu jačinu kod detektovanih crossing objekata.

3.7.5.4 Tile-Boundary Duplicate Rate

Tile-Boundary Duplicate Rate meri redundantno višestruko pokrivanje iste oblasti crossing objekta:

$$\text{TBDR} = \frac{\sum_{G_i \in \mathcal{G}_{\text{cross}}} \left(|\mathcal{A}_i| - |\widetilde{\mathcal{A}}_i|\right)}{\sum_{G_i \in \mathcal{G}_{\text{cross}}} |\mathcal{A}_i|}. \tag{3.29}$$

U ovoj formuli $|\mathcal{A}_i|$ označava broj svih predikcija koje su inicijalno pridružene instanci G_i, dok $|\widetilde{\mathcal{A}}_i|$ označava broj predikcija koje ostaju nakon uklanjanja duplikata. Razlika između ove dve vrednosti predstavlja broj uklonjenih redundantnih predikcija za tu instancu. Brojilac sabira sve uklonjene duplikate, a imenilac ukupan broj inicijalno pridruženih predikcija.

Visok TBDR ukazuje na redundantna opažanja iste oblasti, dok visok TBFR ukazuje na više različitih delova iste fizičke instance. Niža vrednost TBDR je poželjna jer znači da izlaz ne sadrži više skoro istih predikcija za istu oblast objekta. Vrednost nula znači da, prema zadatim duplikacionim kriterijumima, među predikcijama pridruženim crossing instancama nije ostao detektovan redundantni par.

3.7.5.5 Reconstruction IoU

Da bi se prostorna rekonstrukcija razdvojila od očuvanja identiteta, za svaku crossing instancu formira se unija njenih nereundantnih predikcija:

$$U_i = \bigcup_{P_j \in \widetilde{\mathcal{A}}_i} P_j. \tag{3.30}$$

Reconstruction IoU (RIoU) za pojedinačnu instancu definiše se kao

$$\mathrm{RIoU}_i = \begin{cases} \dfrac{|G_i \cap U_i|}{|G_i \cup U_i|}, & U_i \neq \varnothing, \\ 0, & U_i = \varnothing. \end{cases} \tag{3.31}$$

U ovoj formuli U_i predstavlja rekonstruisanu masku instance G_i dobijenu unijom svih njenih nereundantnih predikcija. Presek $G_i \cap U_i$ meri deo referentne instance koji je rekonstruisan, dok unija $G_i \cup U_i$ uzima u obzir i promašene delove reference i višak površine u predikciji. Ako za instancu ne postoji nijedan pridruženi fragment, U_i je prazan skup i RIoU se definiše kao nula.

Agregirana vrednost računa se preko svih crossing instanci:

$$\overline{\mathrm{RIoU}}_{\mathrm{cross}} = \frac{1}{|\mathcal{G}_{\mathrm{cross}}|} \sum_{G_i \in \mathcal{G}_{\mathrm{cross}}} \mathrm{RIoU}_i \,. \tag{3.32}$$

U agregiranoj formuli sabiraju se pojedinačne RIoU vrednosti svih crossing instanci i dele njihovim ukupnim brojem. Za razliku od TBFR i TBEF, ovde u imenilac ulaze i propuštene crossing instance, jer se za njih RIoU računa kao nula. Time metrika istovremeno kažnjava lošu rekonstrukciju i potpuno odsustvo predikcije.

Za razliku od prethodnih metrika, viša vrednost RIoU je poželjna, jer označava bolju prostornu pokrivenost crossing instance nakon objedinjavanja njenih nereundantnih delova. Visok RIoU uz visok TBFR znači da je objekat prostorno dobro pokriven, ali je njegov identitet podeljen. Visok RIoU uz nizak TBFR označava slučaj u kojem je objekat i geometrijski rekonstruisan i predstavljen kao jedinstvena instanca.

3.7.5.6 Merge Rate

Smanjenje fragmentacije treba tumačiti zajedno sa rizikom prekomernog spajanja. Za predikciju P_j definiše se skup referentnih instanci iste klase koje ona pokriva najmanje zadatim udelom:

$$\mathcal{Q}_j(\tau_m) = \left\{ G_i \in \mathcal{G} \;\middle|\; \frac{|G_i \cap P_j|}{|G_i|} \geq \tau_m \right\}. \tag{3.33}$$

U ovoj formuli $\mathcal{Q}_j(\tau_m)$ predstavlja skup referentnih instanci iste klase koje predikcija P_j pokriva najmanje u zadatom udelu τ_m. Ako je taj skup veći od jedan, ista izlazna maska obuhvata više fizičkih objekata i tretira se kao overmerge slučaj.

Za proveru prekomernog spajanja koristi se Merge Rate (MR) uz prag $\tau_m = 0{,}50$, usklađen sa osnovnim pragom podudaranja u evaluaciji. Predikcija se smatra overmerge slučajem ako pokriva više od jedne referentne instance. Za M izlaznih predikcija važi:

$$\mathrm{MR} = \frac{\sum_{j=1}^{M} \mathbf{1}\left[|\mathcal{Q}_j(\tau_m)| > 1\right]}{M}. \tag{3.34}$$

U ovoj formuli indikator ima vrednost jedan za svaku izlaznu predikciju koja pokriva više od jedne referentne instance. Brojilac je ukupan broj takvih overmerge predikcija, a imenilac M ukupan broj izlaznih predikcija.

MR se računa nad svim finalnim predikcijama, a ne samo nad crossing populacijom. Niža vrednost je poželjna, jer ukazuje na manji udeo izlaznih maski koje pokrivaju više različitih fizičkih objekata. Smanjenje TBFR predstavlja poboljšanje samo ako nije ostvareno detektabilnim povećanjem MR.

3.7.6 Zajednička interpretacija metrika

Predložene metrike su komplementarne i zajedno opisuju različite mehanizme greške povezane sa tiled inferencijom i naknadnom globalnom konsolidacijom.

Tabela 3.7. Uloge specijalizovanih metrika za evaluaciju tiled inferencije.

Metrika	**Meri**	**Poželjno ponašanje**
TBMR	Promašene crossing instance.	Niže.
TBFR	Učestalost fragmentacije identiteta instance.	Niže.
TBEF	Intenzitet fragmentacije, odnosno višak fragmenata.	Niže.
TBDR	Redundantna predviđanja iste oblasti.	Niže.
$\mathrm{RIoU}_{\mathrm{cross}}$	Kvalitet prostorne rekonstrukcije crossing instance.	Više.
MR	Prekomerno spajanje različitih fizičkih instanci.	Niže.

TBMR sprečava da propušteni objekti veštački poboljšaju TBFR i TBEF. Kombinacija visokog RIoU i visokog TBFR označava dobru geometrijsku pokrivenost uz podeljen identitet instance. Nizak TBFR uz stabilan MR ukazuje na smanjenje fragmentacije bez prekomernog spajanja različitih objekata.

3.7.7 Agregacija i statistički protokol

Tile-boundary metrike agregiraju se mikroagregacijom, odnosno sabiranjem odgovarajućih brojnika i imenilaca pre konačnog izračunavanja odnosa. Ne koristi se prost prosek vrednosti po slikama, jer bi time isti značaj bio dodeljen slikama sa različitim brojem crossing instanci. Slike bez crossing referentnih instanci ne doprinose odgovarajućem brojniku ni imeniocu.Ako crossing instance postoje, ali nijedna nije detektovana, slika doprinosi izračunavanju TBMR metrike, dok TBFR i TBEF za tu sliku nisu definisani. Nedefinisane vrednosti ne tretiraju se kao nula.

Statistička pouzdanost razlika procenjuje se uparenim bootstrap postupkom na nivou originalnih slika. U svakoj od 1000 iteracija uzorkuje se, sa vraćanjem, skup od 458 originalnih slika. Isti bootstrap uzorak image ID vrednosti koristi se za sve metode unutar iste tiled geometrije, čime se čuva uparena struktura poređenja.

U svakoj bootstrap iteraciji metrike se ponovo računaju iz agregiranih numerator/denominator countera, a ne prosečenjem per-image odnosa. Na primer, ako $I^{(b)}$ označava skup image indeksa u iteraciji b, bootstrap vrednost TBFR računa se kao

$$\text{TBFR}^{(b)} = \frac{\sum_{k \in I^{(b)}} N_{\text{frag},k}}{\sum_{k \in I^{(b)}} N_{\text{det},k}}. \tag{3.35}$$

Za dve metode A i B u svakoj iteraciji računa se uparena razlika

$$\Delta^{(b)} = M_B^{(b)} - M_A^{(b)}. \tag{3.36}$$

Interval poverenja dobija se percentile postupkom:

$$CI_{95\%} = \left[P_{2.5}(\Delta^{(b)}), P_{97.5}(\Delta^{(b)}) \right]. \tag{3.37}$$

Za metrike kod kojih je niža vrednost poželjna, kao što su TBMR, TBFR, TBEF, TBDR i MR, negativna delta označava poboljšanje metode B. Za RIoU pozitivna delta označava poboljšanje.

Predloženi evaluacioni okvir omogućava da se ukupna promena kvaliteta segmentacije razloži na tri odvojene komponente: detekciju crossing objekta, geometrijsku rekonstrukciju njegove maske i očuvanje identiteta instance. Time se efekat globalne konsolidacije ne ocenjuje samo kroz promenu standardnih referentnih metrika, već i kroz mehanizam kojim je ta promena ostvarena. Ovaj skup tile-boundary metrika predstavlja specijalizovani evaluacioni doprinos disertacije.

3.8 Eksperimentalne konfiguracije

Eksperimentalni dizajn formiran je tako da se odvojeno porede efekti preklapanja pločica, referentne globalne konsolidacije, predloženog BC-SMM postupka i selektivnog oporavka predikcija niže pouzdanosti. Sve glavne konfiguracije koriste pločice veličine 512×512 piksela, isti evaluacioni skup i isti evaluacioni protokol; razlikuju se samo po komponentama uključenim u pojedine faze obrade.

Konfiguracija B0 predstavlja tiled inferenciju bez sistematskog preklapanja (0%) i bez globalne konsolidacije, pri čemu se lokalne predikcije samo transformišu u globalni koordinatni sistem originalne slike. B1 koristi 8% preklapanja, ali i dalje bez globalne konsolidacije. SMM koristi isti 8% overlap ulaz, primenjuje originalni Spatial Mask Merging postupak i zatim funkciju fuzije Φ. B2 zamenjuje SMM predloženim BC-SMM algoritmom,

uz nepromenjen ulaz, confidence režim i funkciju fuzije. B3 dodaje rescue granu pre BC-SMM konsolidacije, pri čemu se selektivno prihvaćeni kandidati iz opsega $0{,}40 \leq s_i < 0{,}50$ dalje tretiraju kao regularne predikcije u istom BC-SMM postupku i funkciji fuzije.

Ove konfiguracije čine osnovu eksperimentalnih poređenja u Poglavlju 4. Pregled konfiguracija i njihove uloge u kontrolisanim poređenjima dat je u Tabeli 4.1.

Prenosivost kompletnog B3 postupka proverava se i na BONAI skupu podataka, pri čemu se kao primarni segmentacioni model koristi YOLO26 Large. Ta provera zadržava istu logiku konfiguracije, ali se izvodi nad drugim skupom podataka i jednoklasnim domenom.

3.9 Dodatno ispitane metodološke varijante

Pored komponenti uključenih u finalni metodološki tok, tokom razvoja sistema ispitane su dodatne postprocesne varijante. One nisu deo finalnog pipeline-a, već se evaluiraju kao zasebne ablacione analize u Poglavlju 4. Ovaj odeljak navodi šta je testirano i kako su varijante bile ograničene, bez tumačenja njihovih eksperimentalnih ishoda.

3.9.1 Eksperimentalno morfološko prečišćavanje

Nakon funkcije fuzije razmatrana je mogućnost lokalnog morfološkog prečišćavanja konačne maske operatorom Ψ sa parametrima θ_Ψ. Ova grana nije menjala pripadnost predikcija grupama, broj konsolidovanih instanci niti BC-SMM particiju; delovala je isključivo nad već formiranom izlaznom maskom.

Ispitivane su standardne morfološke operacije: otvaranje, zatvaranje, dilatacija, popunjavanje manjih rupa i uklanjanje malih povezanih komponenti. Za otvaranje i zatvaranje razmatrani su disk i kvadrat kao strukturni elementi, uz fiksne i adaptivne veličine kernela. Da bi se sprečila velika promena geometrije, uvedena su zaštitna ograničenja: obrađena maska morala je zadržati visoku IoU sličnost sa izvornom maskom i ograničenu relativnu promenu površine.

Morfološka obrada nije uključena u finalni metodološki tok i ne primenjuje se u konfiguracijama B0–B3.

3.9.2 Naknadna analiza i korekcija semantičke klase

Druga dodatna grana odnosila se na situacije u kojima je fizička instanca geometrijski detektovana i konsolidovana, ali joj je dodeljena pogrešna semantička klasa. Analiza je sprovedena nad klasama *Small Vehicle* i *Large Vehicle*, kao vizuelno i geometrijski srodnim kategorijama.

Za ovu svrhu formiran je zaseban, naknadni skup podataka iz konačnih BC-SMM instanci koje se mogu pouzdano upariti sa referentnim objektom. Kao ulazna obeležja korišćene su samo informacije dostupne u trenutku inferencije: postojeća klasa i confidence

konačne instance, statistike confidence vrednosti članova grupe, broj izvornih predikcija i pločica, raspodela klasa među članovima fusion grupe, površina maske i obuhvatnog pravougaonika, širina, visina, odnos stranica, položaj prema granicama pločica i geometrijske osobine konsolidovane instance.

Ispitane su dve formulacije problema. U prvoj model procenjuje da li je postojeća klasa pogrešna. U drugoj model neposredno procenjuje konačnu klasu između kategorija *Small Vehicle* i *Large Vehicle*. Testirano je više tabularnih klasifikatora, uključujući linearne i ansambl modele. Pored direktne zamene postojeće klase analiziran je i konzervativni režim u kojem se klasa menja samo kada ML model predloži drugu klasu sa dovoljno visokom pouzdanošću.

Tokom razvoja rescue grane ispitano je više tabularnih klasifikacionih modela i skupova karakteristika. Za finalnu konfiguraciju izabran je XGBoost, dok su njegovi hiperparametri optimizovani na razvojnom skupu pomoću Optuna okvira i TPE (*Tree-structured Parzen Estimator*) postupka za pretragu prostora hiperparametara. Nakon izbora konfiguracije, parametri modela su zamrznuti pre evaluacije na finalnom skupu. Ovaj postupak optimizacije odnosi se isključivo na sekundarni rescue klasifikator i nezavisan je od parametara BC-SMM algoritma.

4. EKSPERIMENTALNI REZULTATI I ANALIZA PREDLOŽENOG SISTEMA

4.1 Eksperimentalni protokol i pregled poređenih konfiguracija

Eksperimentalna evaluacija organizovana je tako da se odvojeno ispita doprinos preklapanja pločica, globalne konsolidacije maski i selektivnog oporavka predikcija niže pouzdanosti. Svih pet glavnih konfiguracija koristi pločice veličine 512×512, isti evaluacioni skup i isti evaluacioni protokol, a razlikuju se samo u komponentama navedenim u Tabeli 4.1. Poređenja su organizovana tako da izoluju efekat preklapanja (B0–B1), referentne globalne konsolidacije (B1–SMM), predloženog BC-SMM postupka (SMM–B2) i selektivnog oporavka (B2–B3).

Ovakvo razdvajanje je važno zato što iste lokalne predikcije mogu drugačije doprineti konačnom rezultatu u zavisnosti od toga da li se posmatraju neposredno, konsoliduju u globalnom prostoru ili dopunjavaju kandidatima iz rescue grane.

Tabela 4.1. Pregled glavnih eksperimentalnih konfiguracija.

Oznaka	**Preklapanje**	**Rescue**	**Globalna konsolidacija**	**Fuzija**	**Uloga u poređenju**
B0	0%	ne	ne	ne	Kontrolna tiled konfiguracija bez sistematskog preklapanja i bez globalne konsolidacije.
B1	8%	ne	ne	ne	Izolovanje efekta 8% preklapanja bez konsolidacije.
SMM	8%	ne	SMM	Φ	Referentna globalna konsolidacija nad istim 8% overlap ulazom.
B2	8%	ne	BC-SMM	Φ	Izolovanje efekta BC-SMM-a u odnosu na referentni SMM.
B3	8%	da	BC-SMM	Φ	Kompletna konfiguracija sa selektivnim oporavkom kandidata iz opsega $0{,}40 \leq s_i < 0{,}50$ pre BC-SMM konsolidacije.

SMM dodatno je analiziran i nad 0% overlap konfiguracijom, kako bi se proverilo kako preklapanje menja broj i strukturu redundantnih lokalnih predikcija kada algoritam konsolidacije ostaje isti.

Rezultati se prikazuju kroz tri grupe pokazatelja. Prvu čine standardne referentne evaluacione metrike segmentacije: AP, AP_{50}, AP_{75}, Precision, Recall i F1. Drugu čine predložene

tile-boundary metrike: TBMR, TBFR, TBEF, TBDR, prosečni RIoU nad graničnim instancama i MR. One razdvajaju propuštanje, fragmentaciju, redundantnu duplikaciju, kvalitet rekonstrukcije i pogrešno spajanje objekata pogođenih granicama pločica. Treću grupu čine dijagnostičke veličine: broj ulaznih predikcija, broj i veličina grupa, struktura kandidatnog grafa, vreme izvršavanja, a za rescue granu PR-AUC, ROC-AUC i broj prihvaćenih TP i FP kandidata pri ograničenju $FP/TP \leq 0{,}25$.

Standardne i specijalizovane metrike ne posmatraju se kao zamene. COCO AP opisuje ukupan kvalitet segmentacije instanci, dok tile-boundary metrike izdvajaju ponašanje nad objektima za koje inferencija nad pločicama uvodi dodatni rizik fragmentacije i duplikacije. Zbog toga se rezultati prikazuju postupno: najpre overlap, zatim originalna i predložena konsolidacija, potom tile-boundary analiza i doprinos selektivnog oporavka.

4.2 Uticaj preklapanja pločica pri tiled inferenciji

Prvi eksperimentalni korak ispituje uticaj sistematskog preklapanja pločica nezavisno od bilo kog postupka globalne konsolidacije. Porede se konfiguracije B0 i B1. B0 koristi tiled inferenciju bez sistematskog preklapanja, dok B1 koristi preklapanje od 8%. U oba slučaja lokalne predikcije vraćene su u globalni koordinatni sistem originalne slike, ali nad njima nisu primenjeni SMM, BC-SMM niti funkcija fuzije maski.

Na ovaj način moguće je neposredno ispitati dve suprotstavljene posledice preklapanja: povećanje pokrivenosti objekata zahvaljujući dodatnim lokalnim pogledima i povećanje broja redundantnih predikcija nastalih višestrukim posmatranjem istog prostora.

Obe konfiguracije evaluirane su nad istih 458 slika i 116 648 validnih referentnih instanci. Evaluacioni protokol i kriterijum podudaranja ostali su nepromenjeni, pa razlike potiču iz geometrije tiled inferencije.

4.2.1 Ukupni rezultati bez globalne konsolidacije

Konfiguracija bez sistematskog preklapanja proizvela je ukupno $102\,886$ predikcija, dok je uvođenje 8% preklapanja povećalo taj broj na $115\,140$. Preklapanje je time generisalo $12\,254$ dodatne lokalne predikcije, odnosno približno $11{,}9\%$ više nego režim bez sistematskog overlap-a. Glavni rezultati su u Tabeli 4.2.

Tabela 4.2. Poređenje tiled inferencije bez sistematskog preklapanja i sa 8% *preklapanja, pre globalne konsolidacije.*

Metrika	**B0:** 0%	**B1:** 8%	**Promena**
Broj predikcija	102 886	115 140	+12 254
TP	59 582	61 508	+1 926
FP	43 304	53 632	+10 328
FN	57 066	55 140	-1 926
Precision	0,5791	0,5342	-0,0449
Recall	0,5108	0,5273	+0,0165
F1	0,5428	0,5307	-0,0121
AP	0,1830	0,1903	+0,0073
AP_{50}	0,3338	0,3373	+0,0035
AP_{75}	0,1789	0,1892	+0,0103
High-density AP_{50}	0,4232	0,4253	+0,0021
High-density AP_{75}	0,2176	0,2331	+0,0156
AR_{6000}	0,3967	0,4291	+0,0325

Rezultati pokazuju jasan kompromis. Uvođenjem overlap-a pronađeno je $1\,926$ dodatnih tačnih instanci, pa se broj FN smanjuje za isti iznos. Recall raste sa $0{,}5108$ na $0{,}5273$, dok se AR_{6000} povećava sa $0{,}3967$ na $0{,}4291$. Istovremeno, broj FP raste za čak $10\,328$. Kao posledica toga precision opada sa $0{,}5791$ na $0{,}5342$, a F1 sa $0{,}5428$ na $0{,}5307$.

Prema tome, sam overlap ne daje jednoznačno poboljšanje ukoliko se lokalne predikcije neposredno evaluiraju kao konačne instance. On povećava pokrivenost scene, ali dodatnu informaciju proizvodi u redundantnom obliku.

4.2.2 Uticaj na COCO metrike i kvalitet lokalizacije

Iako fixed-IoU precision i F1 opadaju, standardne COCO metrike pokazuju drugačiji obrazac. AP raste sa $0{,}1830$ na $0{,}1903$, dok AP_{75} raste sa $0{,}1789$ na $0{,}1892$. Raste i high-density AP_{75}, sa $0{,}2176$ na $0{,}2331$, kao i AR_{6000}, sa $0{,}3967$ na $0{,}4291$.

To pokazuje da overlap ne proizvodi samo veći broj predikcija, već povećava i dostupnost kvalitetno lokalizovanih maski. Predikcije koje nastaju u susednoj pločici mogu imati povoljniji lokalni kontekst i bolju geometriju od odgovarajuće predikcije iz pločice u kojoj je objekat presečen granicom.

Prosečne geometrijske vrednosti uspešno uparenih maski pokazuju mali, ali konzistentan dobitak nakon uvođenja preklapanja. Srednji mask IoU raste sa $0{,}7416$ na $0{,}7466$, dok Dice raste sa $0{,}8469$ na $0{,}8503$. To ukazuje da dodatni lokalni pogledi ne povećavaju samo broj pronađenih objekata, već kod dela uspešno detektovanih instanci omogućavaju i nešto kvalitetnije geometrijsko poklapanje sa referentnom maskom.

Za scene sa veoma velikim brojem predikcija dodatno je korišćen high-density eva-

luacioni režim sa maksimalno 6000 detekcija po slici. Pri interpretaciji rezultata posebno se koristi AR_{6000}, koji direktno meri odziv u ovom proširenom režimu. Njegova vrednost raste sa $0{,}3967$ pri 0% overlap-a na $0{,}4291$ pri 8% overlap-a, što dodatno potvrđuje povećanu pokrivenost objekata.

4.2.3 Analiza uzroka povećanja false-positive predikcija

Da bi se razjasnio rast FP, slučajevi su razvrstani prema dominantnom uzroku. Rezultati su prikazani u Tabeli 4.3.

Tabela 4.3. Promena strukture false-positive predikcija nakon uvođenja 8% overlap-a.

Uzrok FP	0%	8%	**Promena**
Duplikat već uparenog GT objekta	14 369	21 207	+6 838
Lokalizacija ispod IoU praga	17 374	19 833	+2 459
Pozadina / bez odgovarajućeg preklapanja	10 485	11 490	+1 005
Pogrešna semantička klasa	1 076	1 102	+26
Ukupno	**43 304**	**53 632**	**+10 328**

Najveći deo povećanja FP predstavlja upravo porast duplikata. Od ukupno $10\,328$ dodatnih FP, njih $6\,838$, odnosno približno $66{,}2\%$, pripada ovoj kategoriji.

Dodatnih $2\,459$ FP, odnosno približno $23{,}8\%$, predstavlja predikcije koje prostorno odgovaraju objektu, ali ne dostižu zadati IoU prag. Background predikcije objašnjavaju približno $9{,}7\%$ porasta, dok je dodatnih klasnih konfuzija svega 26.

Prema tome, približno 90% povećanja FP potiče od duplikacije i nedovoljno dobre lokalizacije. Pad precision-a zato nije posledica naglog povećanja pogrešnih klasnih odluka, već načina na koji su višestruka lokalna posmatranja istih objekata predstavljena pre globalne konsolidacije.

4.2.4 Analiza smanjenja false-negative slučajeva

Komplementarna FN analiza pokazuje koje je objekte overlap povratio.

Tabela 4.4. Promena strukture false-negative slučajeva nakon uvođenja 8% overlap-a.

Uzrok FN	0%	8%	**Promena**
Bez preklapajuće predikcije	44 703	43 079	-1 624
Lokalizacija ispod IoU praga	11 312	10 995	-317
Postoji podudaranje pogrešne klase	1 050	1 066	+16
Konkurencija pri pridruživanju	1	0	-1
Ukupno	**57 066**	**55 140**	**-1 926**

Najveći deo dobitka potiče od objekata za koje u režimu bez overlap-a nije postojala odgovarajuća preklapajuća predikcija. Njihov broj smanjen je za 1 624, što predstavlja približno 84% ukupnog neto smanjenja FN.

Dodatnih 317 slučajeva prelazi iz nedovoljno dobro lokalizovanih u uspešno detektovane instance. Broj FN povezanih sa pogrešnom klasom praktično ostaje nepromenjen, što potvrđuje da je glavni doprinos preklapanja prostorne, a ne semantičke prirode.

4.2.5 Uticaj na reprezentativne klase

Isti obrazac može se uočiti i na nivou pojedinačnih kategorija. U Tabeli 4.5 prikazane su četiri reprezentativne klase sa velikim brojem instanci ili izraženim uticajem položaja objekta unutar pločice.

Tabela 4.5. Promena precision i recall vrednosti za reprezentativne klase.

	Precision		**Recall**	
Klasa	0%	8%	0%	8%
Plane	0,5843	0,5236	0,8977	0,9325
Ship	0,6617	0,6077	0,8127	0,8300
Small Vehicle	0,5864	0,5434	0,4284	0,4424
Large Vehicle	0,5059	0,4641	0,5818	0,6025

Za sve četiri klase recall raste, dok precision opada. Kod klase *Plane*, na primer, recall raste sa 0,8977 na 0,9325, dok precision opada sa 0,5843 na 0,5236. Sličan obrazac postoji i kod obe kategorije vozila.

Kod klase *Small Vehicle*, najzastupljenije u evaluacionom skupu, TP raste sa 36 025 na 37 203, dok broj predikcija raste sa 61 432 na 68 461. Overlap zato pronalazi dodatne validne instance, ali povećava i broj izlaza koji se odnose na iste ili prostorno bliske objekte.

Detaljnija FP analiza za ovu klasu pokazuje da broj duplikata raste sa 7 006 na 10 926, odnosno za 3 920 predikcija. Kod *Large Vehicle* odgovarajući rast iznosi 692 duplikata. Samo ove dve klase zato objašnjavaju 4 612 od ukupno 6 838 dodatnih duplicate FP nastalih prelaskom sa 0% na 8% overlap.

Rezultati B0 i B1 zato definišu problem za narednu fazu: preklapanje uvodi dodatna lokalna opažanja i smanjuje FN, ali bez globalne konsolidacije istovremeno povećava redundantne FP. Sledeći eksperiment meri koliko SMM može da zadrži dobitak u pokrivenosti uz uklanjanje duplikata.

4.3 SMM i efekat globalne konsolidacije

Nakon analize tiled inferencije bez globalnog postprocesiranja, ispituje se koliko originalni SMM konsoliduje redundantne predikcije nastale nezavisnim posmatranjem iste scene kroz različite pločice. Glavno poređenje odnosi se na prelazak sa B1 konfiguracije, koja koristi 8% preklapanja bez konsolidacije, na $\text{SMM}_{8\%}$. Kao kontrola analizira se isti SMM postupak i pri 0% overlap-a. U oba slučaja korišćen je isti skup parametara originalnog SMM-a, pa razlike odražavaju promenu geometrije tiled inferencije, a ne promenu algoritma.

4.3.1 Efekat SMM konsolidacije pri 8% overlap-u

Pre globalne konsolidacije konfiguracija B1 sadržala je $115\,140$ predikcija. SMM ih grupiše u $82\,919$ grupa, od kojih su $57\,495$ singleton grupe, dok $25\,424$ grupe sadrže dve ili više predikcija. Time se broj izlaznih predikcija smanjuje za $32\,221$, odnosno za približno $27{,}98\%$. Nakon primene funkcije fuzije Φ, svaka grupa generiše jednu konačnu instancu. Direktni efekat na metrike prikazan je u Tabeli 4.6.

Tabela 4.6. Poređenje 8% tiled inferencije bez konsolidacije i nakon SMM konsolidacije.

Metrika	**B1**	**$\text{SMM}_{8\%}$**	**Promena**
Broj predikcija	115 140	82 919	-32 221
TP	61 508	61 249	-259
FP	53 632	21 670	-31 962
FN	55 140	55 399	+259
Precision	0,5342	0,7387	+0,2045
Recall	0,5273	0,5251	-0,0022
F1	0,5307	0,6138	+0,0831
AP	0,1903	0,3123	+0,1220
AP_{50}	0,3373	0,4950	+0,1577
AP_{75}	0,1892	0,3329	+0,1437
AR_{6000}	0,4291	0,4489	+0,0198

Najizraženiji efekat konsolidacije jeste smanjenje broja FP sa $53\,632$ na $21\,670$, uz gubitak od samo 259 TP, odnosno približno $0{,}42\%$ početnog broja tačnih detekcija. Precision zato raste za više od 20 procentnih poena, recall ostaje praktično nepromenjen, a F1 raste sa $0{,}5307$ na $0{,}6138$. Još veći dobitak vidi se u COCO metrikama: standardni AP raste sa $0{,}1903$ na $0{,}3123$, a AP_{75} sa $0{,}1892$ na $0{,}3329$. To pokazuje da SMM ne uklanja samo redundantne izlaze, već nakon fuzije proizvodi kvalitetniji rangirani skup konačnih instanci.

4.3.2 Kontrolno poređenje SMM-a pri 0% i 8% overlap-u

Posebno informativan rezultat dobija se kada se isti SMM postupak primeni na tiled inferenciju bez i sa sistematskim overlap-om. Rezultati particionisanja prikazani su u Tabeli 4.7.

Tabela 4.7. Poređenje strukture SMM particija za 0% i 8% overlap-a.

Pokazatelj	**SMM$_{0\%}$**	**SMM$_{8\%}$**	**Razlika**
Ulazne predikcije	102 886	115 140	+12 254
Izlazne grupe	82 923	82 919	-4
Singleton grupe	66 334	57 495	-8 839
Spojene grupe	16 589	25 424	+8 835
Predikcije u spojenim grupama	36 552	57 645	+21 093
Potencijalna redukcija	19 963	32 221	+12 258
Relativna redukcija	19,40%	27,98%	+8,58 p.p.
Kandidatni parovi	509 657	626 437	+116 780
Kompatibilni parovi	25 108	43 193	+18 085

Najvažniji rezultat je gotovo identičan broj konačnih grupa: 82 923 pri 0% i 82 919 pri 8% overlap-a. Drugim rečima, 8% overlap povećava broj ulaznih predikcija za 12 254, ali nakon SMM konsolidacije broj očekivanih fizičkih instanci ostaje praktično nepromenjen. Razlika u potencijalnoj redukciji iznosi 12 258, što je gotovo jednako broju dodatnih predikcija nastalih overlap-om. Ovaj rezultat pokazuje da SMM apsorbuje gotovo celokupnu dodatnu redundansu uvedenu preklapanjem pločica.

4.3.3 Struktura grupa i efekat overlap-a

Raspodela grupa prema broju članova dodatno pokazuje kako overlap menja strukturu globalnog problema konsolidacije.

Tabela 4.8. Raspodela veličina SMM grupa.

Veličina grupe	0%	8%
1	66 334	57 495
2	14 814	21 755
3	184	613
4	1 584	3 010
5	6	20
6	1	26

Pri 0% overlap-a singleton grupe čine približno 80% svih grupa, dok se njihov udeo

pri 8% overlap-a smanjuje na približno 69%. Udeo ulaznih predikcija u grupama sa najmanje dva člana istovremeno raste sa približno 35,5% na 50,1%, pa pri 8% overlap-a oko polovine lokalnih predikcija učestvuje u globalnoj konsolidaciji. Posebno raste broj grupa veličine četiri, što je očekivano kada isti region može biti obuhvaćen sa četiri susedne pločice.

4.3.4 Promena kandidatnog grafa

Overlap ne povećava samo broj ulaznih predikcija, već menja i strukturu kandidatnog grafa. Pri 0% overlap-a SMM generiše 509 657 kandidatnih parova, od kojih je 25 108 kompatibilno. Pri 8% overlap-a broj kandidata raste na 626 437, a broj kompatibilnih parova na 43 193. Relativno, broj kandidata raste za približno 22,9%, dok broj kompatibilnih relacija raste za približno 72%.

Odnos kompatibilnih i ukupnih kandidatnih parova povećava se sa približno 4,93% na 6,90%. Prema tome, dodatni parovi nastali overlap-om nisu samo slučajni prostorni susedi. Znatno veći deo njih zadovoljava geometrijske i semantičke uslove kompatibilnosti originalnog SMM-a, što potvrđuje da overlap proizvodi korisnu cross-tile evidenciju o istim fizičkim objektima.

4.3.5 Uticaj anti-chaining korekcije

Gušći kandidatni graf pri 8% overlap-a povećava broj situacija u kojima indirektne veze između predikcija mogu formirati nedovoljno konzervativne grupe. Pri 0% overlap-a anti-chaining korekcija zahvatila je 30 predikcija i povećala broj grupa za 10. Pri 8% overlap-a isti korak zahvatio je 646 predikcija i povećao broj grupa za 186.

Iako je apsolutni broj intervencija znatno veći, korekcija i dalje obuhvata mali deo ukupnog skupa predikcija. Rezultat pokazuje da preklapanje stvara gušću mrežu kompatibilnih relacija i time povećava verovatnoću indirektnih lanaca između lokalnih predikcija.

4.3.6 Efekat overlap-a nakon globalne konsolidacije

Najvažniji dokaz korisnosti overlap-a dobija se tek nakon primene globalne konsolidacije. Bez konsolidacije prelazak sa 0% na 8% povećavao je recall, ali je smanjivao precision i F1 zbog velikog broja redundantnih predikcija. Nakon primene SMM-a obrazac se menja.

Tabela 4.9. Poređenje konačnih SMM rezultata pri 0% i 8% overlap-a.

Metrika	**SMM$_{0\%}$**	**SMM$_{8\%}$**	**Promena**
Predikcije	82 923	82 919	-4
TP	59 328	61 249	+1 921
FP	23 595	21 670	-1 925
FN	57 320	55 399	-1 921
Precision	0,7155	0,7387	+0,0232
Recall	0,5086	0,5251	+0,0165
F1	0,5946	0,6138	+0,0193
AP	0,2396	0,3123	+0,0727
AP$_{50}$	0,4296	0,4950	+0,0654
AP$_{75}$	0,2352	0,3329	+0,0977
AR$_{6000}$	0,3958	0,4489	+0,0530

Za praktično isti broj konačnih instanci, konfiguracija sa 8% overlap-a sadrži 1 921 dodatni TP i 1 925 manje FP. Precision, recall i F1 zato istovremeno rastu. Posebno je izražen rast AP$_{75}$ sa 0,2352 na 0,3329, što pokazuje da dodatni lokalni pogledi iz preklopljenih pločica, nakon konsolidacije, daju kvalitetniji skup precizno lokalizovanih instanci. Sam overlap uvodi i korisnu informaciju i redundansu; globalna konsolidacija uklanja redundantni deo, dok se dobitak u pokrivenosti zadržava.

Ovi rezultati uspostavljaju SMM kao referentnu globalnu konsolidaciju za naredni eksperiment. BC-SMM se zatim poredi sa istim ulaznim predikcijama i istom tiled geometrijom, ali sa eksplicitnijim korišćenjem fizičkih odnosa između pločica.

4.4 Rezultati BC-SMM globalne konsolidacije

Centralni eksperiment ovog poglavlja ispituje doprinos predloženog BC-SMM postupka u odnosu na originalni Spatial Mask Merging. Obe metode primenjene su nad identičnim skupom od 115 140 lokalnih predikcija, dobijenih tiled inferencijom sa 8% preklapanja, uz isti osnovni segmentacioni model i isti evaluacioni protokol. Zato se razlika između SMM$_{8\%}$ i BC-SMM može pripisati promeni postupka globalne konsolidacije, a ne promeni ulaznih predikcija.

BC-SMM je glavni algoritamski doprinos ove disertacije. Za razliku od originalnog SMM-a, on koristi topologiju pločica, tipove njihovih fizičkih kontakata i lokalno ograničene koridore, uz lokalnu kontaktnu validaciju marginalnih mostova i topološki kontrolisano formiranje grupa.

4.4.1 Ukupni evaluacioni rezultat

Glavni rezultati prikazani su u Tabeli 4.10.

Tabela 4.10. Poređenje SMM i BC-SMM globalne konsolidacije pri 8% *overlap-u.*

Metrika	**SMM**	**BC-SMM**	**Promena**
Broj izlaznih instanci	82 919	**82 166**	-753
TP	61 249	**61 304**	+55
FP	21 670	**20 862**	-808
FN	55 399	**55 344**	-55
Precision	0,7387	**0,7461**	+0,0074
Recall	0,5251	**0,5255**	+0,0005
F1	0,6138	**0,6167**	+0,0029
COCO AP	0,3123	**0,3219**	+0,0097
Standard AP_{50}	0,4950	**0,5078**	+0,0128
Standard AP_{75}	0,3329	**0,3447**	+0,0118
High-density AP_{50}	0,6022	**0,6158**	+0,0136
High-density AP_{75}	0,3868	**0,3982**	+0,0114
AR_{6000}	0,4489	**0,4521**	+0,0032
Mean mask IoU	0,7459	**0,7461**	+0,0002
Mean Dice	0,8498	**0,8499**	+0,0001

BC-SMM smanjuje broj konačnih instanci za dodatnih 753 u odnosu na SMM. Međutim, ova dodatna konsolidacija nije praćena smanjenjem broja tačnih detekcija. Naprotiv, broj TP raste sa $61\,249$ na $61\,304$, dok se FP smanjuje sa $21\,670$ na $20\,862$, a FN za 55. Ovaj rezultat pokazuje da dodatna konsolidacija nije na agregatnom nivou dovela do izraženog over-merging efekta. Kada bi dodatno spajanje prvenstveno povezivalo različite fizičke objekte, očekivao bi se pad TP i recall-a, odnosno rast FN. Dobijeni obrazac je suprotan: precision raste sa $0{,}7387$ na $0{,}7461$, dok je recall blago povećan sa $0{,}5251$ na $0{,}5255$.

F1 mera zato raste sa $0{,}6138$ na $0{,}6167$. Iako je apsolutna promena F1 relativno mala, ona nastaje u odnosu na već snažno konsolidovan Original SMM izlaz i praćena je istovremenim poboljšanjem precision-a i recall-a.

4.4.2 Poboljšanje COCO mask metrika

Razlika između dve metode izraženija je u COCO mask evaluaciji. Ukupni AP raste sa $0{,}3123$ na $0{,}3219$, što odgovara apsolutnom povećanju od približno $0{,}97$ procentnih poena, odnosno relativnom poboljšanju od približno $3{,}1\%$ u odnosu na SMM.

Standardni AP_{50} raste sa $0{,}4950$ na $0{,}5078$, a AP_{75} sa $0{,}3329$ na $0{,}3447$. U high-density režimu odgovarajuće vrednosti rastu sa $0{,}6022$ na $0{,}6158$ za AP_{50} i sa $0{,}3868$ na $0{,}3982$ za AP_{75}.

Posebno je značajno da poboljšanje nije ograničeno samo na blaži IoU kriterijum. Rast AP_{75} pokazuje da dodatna konsolidacija nije ostvarena po cenu degradacije prostorne preciznosti konačnih maski.

Fixed-IoU geometrijski pokazatelji podržavaju isti zaključak: srednji mask IoU ostaje praktično stabilan, uz mali rast sa $0{,}7459$ na $0{,}7461$, dok Dice raste sa $0{,}8498$ na $0{,}8499$. Dobitak zato potiče pre svega iz preciznijeg odlučivanja koje lokalne predikcije pripadaju istoj fizičkoj instanci, a ne iz velike naknadne promene geometrije maski.

4.4.3 Promena strukture globalne particije

Poboljšanje evaluacionih rezultata prati i promena strukture formiranih grupa. SMM proizvodi $82\,919$ grupa, dok BC-SMM daje $82\,166$ grupa.

Tabela 4.11. Promena raspodele veličina grupa između SMM i BC-SMM.

Veličina grupe	SMM	BC-SMM	Promena
1	57 495	56 602	-893
2	21 755	21 383	-372
3	613	952	+339
4	3 010	3 229	+219
5	20	0	-20
6	26	0	-26
Ukupno	**82 919**	**82 166**	**-753**

BC-SMM formira $25\,564$ višečlanih grupa koje obuhvataju ukupno $58\,538$ lokalnih predikcija. U odnosu na početnih $115\,140$ predikcija, nakon fuzije ostaje $82\,166$ konačnih instanci, što predstavlja redukciju od $32\,974$ predikcije, odnosno približno $28{,}64\%$ ulaznog skupa.

BC-SMM ovu redukciju ostvaruje uz maksimalnu veličinu grupe četiri. Iako SMM sadrži mali broj grupa veličine pet i šest, BC-SMM ukupno proizvodi manje grupa i bolji evaluacioni rezultat, pa rezultat ne ukazuje na potrebu za povećanjem maksimalne veličine BC-SMM grupe.

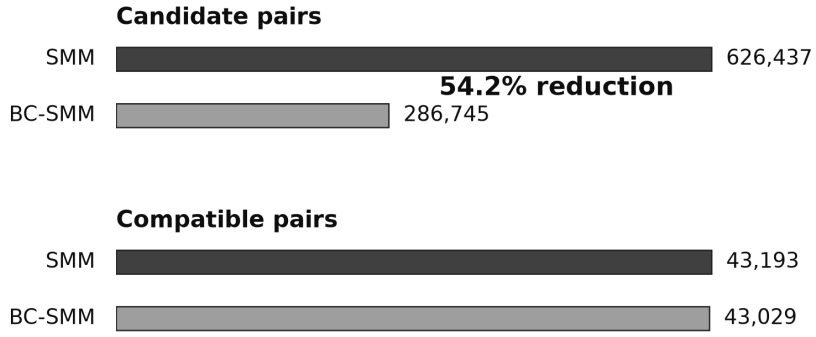
4.4.4 Ograničavanje kandidatnog prostora

Druga važna razlika između metoda odnosi se na broj parova nad kojima se izvršavaju detaljne provere kompatibilnosti.

Tabela 4.12. Struktura kandidatnog prostora SMM i BC-SMM.

Pokazatelj	SMM	BC-SMM	Promena
Ulazne predikcije	115 140	115 140	0
Kandidatni parovi	626 437	**286 745**	-339 692
Kompatibilni standardni parovi	43 193	43 029	-164
Solver komponente	10 623	**7 583**	-3 040
Najveća grupa	6	**4**	–

BC-SMM ispituje 286 745 kandidatnih parova, naspram 626 437 kod SMM-a, pa je broj parnih poređenja smanjen za približno 54,2%. Istovremeno je zadržano 43 029 standardno kompatibilnih relacija, naspram 43 193 kod SMM-a. To pokazuje da smanjenje kandidatnog prostora nije ostvareno agresivnim odbacivanjem potencijalnih veza, već korišćenjem geometrije i topologije tiled inferencije za usmeravanje pretrage prema fizički relevantnim kontaktima pločica.



Slika 4.1. Redukcija kandidatnog prostora primenom BC-SMM pristupa u odnosu na referentni SMM. Broj kandidatnih parova smanjuje se sa 626 437 na 286 745, odnosno za približno 54,2%, dok broj identifikovanih kompatibilnih parova ostaje gotovo nepromenjen (43 193 naspram 43 029).

Ovakav odnos pokazuje da BC-SMM uklanja veliki deo prostorno nerelevantnih poređenja, uz očuvanje gotovo celog skupa parova koji zadovoljavaju kriterijum kompatibilnosti. Dodatne relacije koje ne zadovoljavaju standardnu SMM kompatibilnost mogu biti prihvaćene samo kroz lokalnu kontaktnu validaciju marginalnih mostova. Takve relacije predstavljaju manjinski deo ukupnog skupa veza i podležu dodatnim geometrijskim i topološkim ograničenjima, čime se sprečava nekontrolisano relaksiranje kriterijuma spajanja.

4.4.5 Rezultati po klasama

Ukupno poboljšanje nije koncentrisano na jednu kategoriju. BC-SMM ostvaruje veću F1 vrednost kod 14 od 15 evaluiranih klasa.

Najveća poboljšanja prikazana su u Tabeli 4.13.

Tabela 4.13. Klase sa najizraženijom promenom F1 vrednosti između Original SMM i BC-SMM.

Klasa	SMM F1	BC-SMM F1	ΔF1
Soccer ball field	0,5871	**0,6361**	+0,0490
Ground track field	0,5534	**0,5985**	+0,0451
Harbor	0,7122	**0,7537**	+0,0415
Plane	0,8637	**0,8951**	+0,0314
Tennis court	0,8844	**0,9136**	+0,0292
Helicopter	**0,4351**	0,4298	-0,0054

Kod klasa *Plane*, *Harbor* i *Tennis court* poboljšanje prvenstveno prati povećanje precision-a uz očuvan ili blago povećan recall. Za *Plane* precision raste sa $0{,}8047$ na $0{,}8603$, a za *Harbor* sa $0{,}6091$ na $0{,}6680$.

Jedini pad F1 zabeležen je kod klase *Helicopter*, sa $0{,}4351$ na $0{,}4298$. Ova kategorija sadrži samo 86 referentnih instanci, pa mali broj promenjenih predikcija ima relativno veliki uticaj na vrednost agregirane metrike.

Poseban slučaj predstavlja *Small Vehicle*. BC-SMM za ovu klasu ostvaruje $36\,953$ TP, $15\,749$ FP i $47\,142$ FN, uz $F1 = 0{,}5403$. Ova kategorija sama sadrži $84\,095$ referentnih instanci i ostaje dominantan izvor ukupnih FP i FN grešaka sistema. Zbog toga ograničenje ukupnog rezultata nakon konsolidacije u velikoj meri potiče iz ponašanja osnovnog segmentacionog modela nad veoma sitnim objektima, a ne samo iz postupka globalnog spajanja.

Direktno poređenje sa SMM pokazuje da BC-SMM smanjuje kandidatni prostor za približno $54{,}2\%$, zadržava gotovo isti broj kompatibilnih parova i pri tom ostvaruje manji broj FP uz praktično očuvan recall. Time je pripremljen osnov za narednu proveru: da li se razlika između metoda posebno vidi na instancama koje dodiruju ili presecaju granice pločica.

4.4.6 Kontrolna analiza BC-SMM bez sistematskog preklapanja

Radi odvajanja doprinosa globalne konsolidacije od doprinosa višestrukog posmatranja objekata, BC-SMM je evaluiran i u kontrolnom režimu bez sistematskog preklapanja pločica.

U ovom režimu SMM proizvodi $82\,923$ izlazne instance, a BC-SMM $82\,913$. Uprkos gotovo identičnom broju izlaza, BC-SMM povećava TP sa $59\,328$ na $59\,388$, smanjuje FP sa $23\,595$ na $23\,525$ i FN sa $57\,320$ na $57\,260$. Precision, recall i F1 zato blago rastu.

Ovaj kontrolni rezultat pokazuje da doprinos BC-SMM nije posledica samo većeg broja spajanja. Čak i kada se broj konačnih grupa gotovo ne promeni, drugačiji izbor topološki opravdanih relacija može proizvesti povoljniju globalnu particiju predikcija.

Još izraženiji rezultat dobija se poređenjem BC-SMM izlaza za 0% i 8% preklapanja. Sa uvođenjem 8% preklapanja TP raste sa $59\,388$ na $61\,304$, FP opada sa $23\,525$ na $20\,862$, a FN sa $57\,260$ na $55\,344$. Precision raste sa $0{,}7163$ na $0{,}7461$, recall sa $0{,}5091$ na $0{,}5255$, a

F1 sa $0{,}5952$ na $0{,}6167$. Kod strožih mask kriterijuma high-density AP_{75} raste sa $0{,}2770$ na $0{,}3982$, dok AR_{6000} raste sa $0{,}3956$ na $0{,}4521$.

Time se dobija važan eksperimentalni zaključak: sistematsko preklapanje je naročito korisno kada se dodatna lokalna opažanja obrade globalnom konsolidacijom. BC-SMM višestruka posmatranja pretvara iz izvora redundantnih predikcija u koristan signal, pa konfiguracija sa 8% preklapanja istovremeno ostvaruje veći recall i veći precision od režima bez preklapanja.

4.5 Analiza ponašanja sistema na granicama pločica

U Sekciji 3.7 definisan je skup specijalizovanih tile-boundary metrika TBMR, TBFR, TBEF, TBDR, RIoU i MR. U ovoj sekciji te metrike primenjuju se na nekonsolidovani tiled izlaz, SMM i BC-SMM. Rezultati se prikazuju odvojeno za 0% i 8% preklapanja, a poređenja se vrše unutar iste tiled geometrije, jer se promenom geometrije menja i skup $\mathcal{G}_{\text{cross}}$.

Radi lakše interpretacije specijalizovanih pokazatelja, na Slici 4.2 prikazana su četiri karakteristična tipa događaja koji se javljaju kod instanci koje presecaju unutrašnje granice pločica. U prikazu su referentne GT konture označene crvenom bojom sa belim kontrastnim obrubom, predikovane konture cijan/plavom bojom sa crnim obrubom, dok je unutrašnja granica pločice prikazana žutom isprekidanom linijom sa crnim kontrastom.

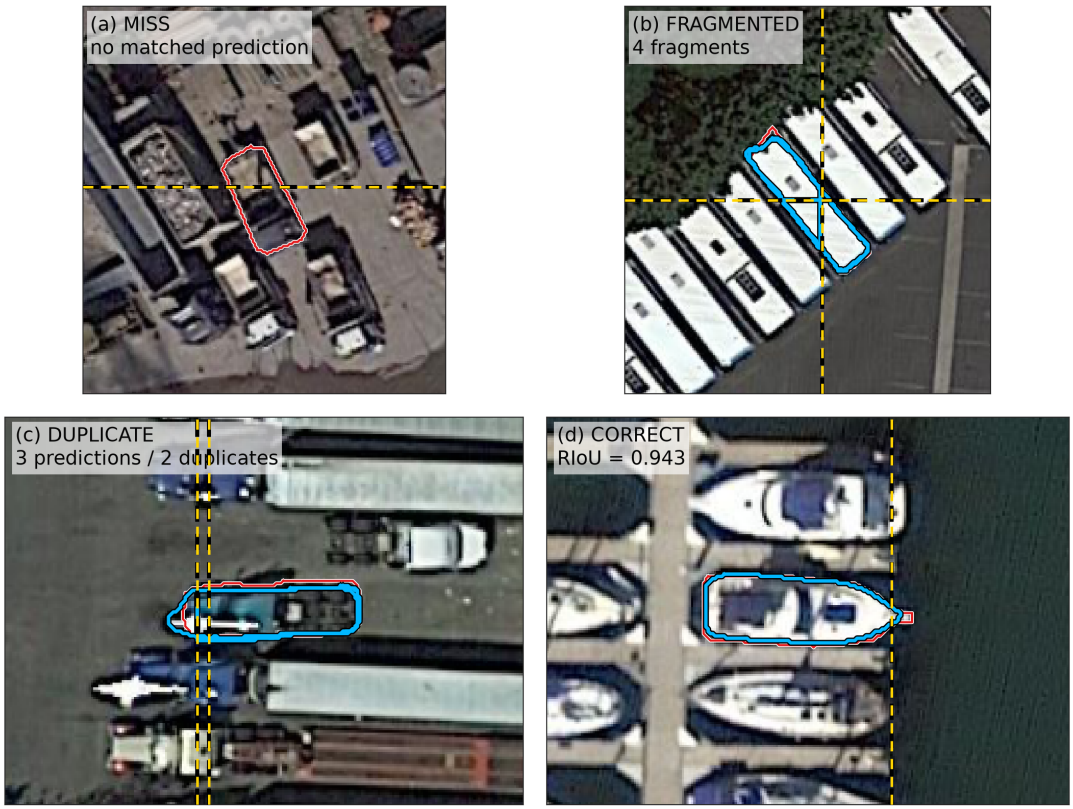
4.5.1 Rezultati bez preklapanja pločica

Konfiguracija bez preklapanja služi kao kontrolni eksperiment. Sve metode u Tabeli 4.14 koriste istu tiled geometriju, pa je crossing populacija ista i iznosi 10 287 instanci.

Tabela 4.14. Tile-boundary metrike za tiled inferenciju bez preklapanja pločica. Strelice označavaju poželjan smer promene, a podebljane vrednosti predstavljaju numerički najbolji rezultat po pojedinačnoj metrici. Kontrolne metrike TBMR i MR ne tumače se izolovano.

Metoda	$\lvert\mathcal{G}_{\text{cross}}\rvert$	TBMR ↓	TBFR ↓	TBEF ↓	TBDR ↓	$\overline{\text{RIoU}}$ ↑	MR ↓
Bez konsolidacije	10 287	**0.2262**	0.5317	0.6563	0.1785	0.5826	**0.00204**
SMM	10 287	0.2271	0.3700	0.4447	**0.0000**	0.5858	0.00288
BC-SMM	10 287	0.2271	**0.3587**	**0.4299**	0.00105	**0.5859**	0.00256

Bez globalne konsolidacije 53.17% detektovanih crossing instanci ostaje fragmentisano. SMM smanjuje TBFR na 0.3700, a TBEF sa 0.6563 na 0.4447. BC-SMM ostvaruje manji dodatni dobitak u odnosu na SMM: TBFR pada na 0.3587, a TBEF na 0.4299. Broj promašenih objekata ostaje praktično nepromenjen, što potvrđuje da SMM i BC-SMM u ovoj fazi konsoliduju postojeće lokalne predikcije, a ne uvode novu detekcionu granu. Vrednost TBDR od 0.0000 kod originalnog SMM-a znači da, prema korišćenom kriterijumu duplikacije, među predikcijama pridruženim crossing instancama nije identifikovana redundantna predikcija koja bi bila uklonjena postupkom deduplikacije.



Slika 4.2. Primeri karakterističnih događaja na granicama pločica: (a) propuštena instanca koja preseca granicu pločice, bez uparene predikcije; (b) fragmentisana instanca predstavljena sa više lokalnih fragmenata; (c) redundantne predikcije iste fizičke instance; (d) pravilno rekonstruisana instanca koja preseca granicu pločice, sa visokim $\mathrm{RIoU}_{\mathrm{cross}}$.

Zato se TBDR posmatra zajedno sa TBFR i TBEF: prva metrika meri redundantna, međusobno veoma slična opažanja iste instance, dok druge dve mere da li je isti fizički objekat i nakon uklanjanja duplikata ostao podeljen na više prostorno različitih fragmenata.

Tabela 4.15. Apsolutni brojevi crossing instanci za režim bez preklapanja.

Metoda	Crossing GT	Detektovano	Promašeno	Fragmentisano	Višak fragmenata	Broj izlaznih predikcija
Bez konsolidacije	10 287	7 960	2 327	4 232	5 224	102 886
SMM	10 287	7 951	2 336	2 942	3 536	82 923
BC-SMM	10 287	7 951	2 336	**2 852**	**3 418**	82 913

Reconstruction IoU se menja samo sa 0.5826 na približno 0.586. To pokazuje da smanjenje fragmentacije nije posledica velikog poboljšanja prostorne pokrivenosti objekata, već bolje organizacije parcijalnih predikcija u koherentnije instance. Bez preklapanja je prostor za korekciju koja uvažava granice pločica ograničen, jer ne postoje dodatna preklapajuća lokalna opažanja istog objekta koja bi mogla biti iskorišćena tokom globalne konsolidacije.

4.5.2 Rezultati pri 8% preklapanju

Glavni eksperiment prikazan je u Tabeli 4.16. Sve metode koriste istu tiled geometriju, pa je crossing populacija jednaka 15 915 instanci.

Tabela 4.16. Tile-boundary metrike za tiled inferenciju sa 8% preklapanja pločica. Strelice označavaju poželjan smer promene, a podebljane vrednosti predstavljaju numerički najbolji rezultat po pojedinačnoj metrici. Kontrolne metrike TBMR i MR ne tumače se izolovano.

Metoda	$\lvert\mathcal{G}_{\text{cross}}\rvert$	TBMR ↓	TBFR ↓	TBEF ↓	TBDR ↓	$\overline{\text{RIoU}}$ ↑	MR ↓
Bez konsolidacije	15 915	**0.2049**	0.4907	0.6128	0.2370	0.6016	**0.00201**
SMM	15 915	0.2063	0.0866	0.1136	**0.00014**	**0.6068**	0.00287
BC-SMM	15 915	0.2063	**0.0356**	**0.0529**	0.00195	**0.6068**	0.00276

Nekonsolidovani izlaz pokazuje obrazac grešaka koji nastaje kada se preklapajuća inferencija nad pločicama koristi bez naknadne globalne konsolidacije. Od 12 654 detektovane crossing instance, 6 209 ostaje fragmentisano, odnosno približno svaki drugi detektovani objekat koji prelazi granicu pločice ostaje predstavljen kroz više finalnih instanci. TBEF iznosi 0.6128, uz ukupno 7 755 viška fragmenata.

TBDR od 0.2370 pokazuje da nekonsolidovani izlaz sa preklapanjem sadrži izraženu redundansu: od 26 750 predikcija dodeljenih crossing objektima, 6 341 je identifikovana kao redundantna prema korišćenim pragovima za duplikaciju. Bez globalne konsolidacije redundantna opažanja ostaju prisutna kao višestruke predikcije, dok deo objekata istovremeno ostaje podeljen na prostorno različite fragmente. Najniži TBMR u ovoj tabeli ima nekonsolidovani izlaz, ali ta vrednost se ne tumači kao bolji ukupan rezultat: isti izlaz istovremeno ima znatno veće TBFR, TBEF i TBDR vrednosti.

TBMR i MR imaju pre svega kontrolnu ulogu i ne tumače se izolovano: TBMR proverava da li je smanjenje fragmentacije ostvareno gubitkom crossing objekata, dok MR proverava da li je ono ostvareno prekomernim spajanjem različitih fizičkih instanci; zato se njihovo značenje procenjuje zajedno sa TBFR, TBEF i RIoU.

Tabela 4.17. Apsolutni brojevi crossing instanci i fragmenata pri 8% preklapanju.

Metoda	Detektovano	Promašeno	Fragmentisano	Višak fragmenata	Dodeljene predikcije	Uklonjeni duplikati	Izlazne predikcije
Bez konsolidacije	12 654	3 261	6 209	7 755	26 750	6 341	115 140
SMM	12 632	3 283	1 094	1 435	14 069	2	82 919
BC-SMM	12 632	3 283	**450**	**668**	13 326	26	82 166

SMM uklanja dominantan deo problema: TBFR pada sa 0.4907 na 0.0866, a broj fragmentisanih crossing instanci sa 6 209 na 1 094. TBEF opada sa 0.6128 na 0.1136, uz smanjenje broja viška fragmenata sa 7 755 na 1 435. Istovremeno se TBDR smanjuje sa 0.2370 na 0.00014, odnosno na samo dve redundantne predikcije među 14 069 predikcija dodeljenih crossing instancama.

Reconstruction IoU raste sa 0.6016 na 0.6068. Nekonsolidovani tiled izlaz zato u velikom broju slučajeva već sadrži dovoljno prostorne informacije za rekonstrukciju objekta, ali je ta informacija raspodeljena kroz više parcijalnih predikcija. SMM prvenstveno menja njihovu instance-level organizaciju.

4.5.3 Dodatni doprinos BC-SMM

BC-SMM dodatno smanjuje fragmentaciju koja ostaje nakon originalnog SMM-a. Pri 8% preklapanju TBFR se smanjuje sa 0.086605 na 0.035624, što odgovara relativnom smanjenju od približno 58.9%, dok broj fragmentisanih crossing instanci pada sa 1 094 na 450. TBEF se istovremeno smanjuje sa 0.113600 na 0.052882, odnosno za približno 53.4%, uz pad broja viška fragmenata sa 1 435 na 668.

Agregatni TBMR ostaje isti, a RIoU se praktično ne menja. Statistička pouzdanost ovih razlika analizira se u Sekciji 4.5.8.

4.5.4 Razdvajanje geometrijske rekonstrukcije i identiteta instance

Zajednička interpretacija TBFR, TBEF i RIoU pokazuje da glavni doprinos BC-SMM nije povećanje pokrivenosti maske, već smanjenje identitetske fragmentacije. Pri prelasku sa originalnog SMM-a na BC-SMM prosečni RIoU ostaje praktično nepromenjen, dok se TBFR više nego prepolovljava.

Drugim rečima, objekat može biti prostorno dobro opisan unijom lokalnih maski, a istovremeno pogrešno predstavljen kao više nezavisnih instanci. BC-SMM deluje upravo na taj drugi problem: povezuje parcijalna lokalna opažanja istog fizičkog objekta u konzistentniji globalni identitet.

4.5.5 Kontrola rizika od prekomernog spajanja

Snažno smanjenje fragmentacije bilo bi metodološki problematično ako bi bilo ostvareno agresivnim spajanjem različitih objekata. Zato se TBFR i TBEF tumače zajedno sa MR.

Pri 8% preklapanju MR iznosi 0.00201 za nekonsolidovani izlaz, 0.00287 za SMM i 0.00276 za BC-SMM. Point estimate pri prelasku sa SMM-a na BC-SMM pada sa 0.002870 na 0.002763, ali 95% bootstrap interval za razliku obuhvata nulu. Prema tome, nema dokaza da BC-SMM povećava overmerge, niti je snažno smanjenje fragmentacije ostvareno detektabilnim porastom prekomernog spajanja.

4.5.6 Odnos tile-boundary i standardnih metrika

Tile-boundary rezultati treba da se posmatraju zajedno sa standardnim COCO metrikama. Tabela 4.18 prikazuje standardne rezultate za tri konfiguracije sa 8% preklapanja.

Tabela 4.18. Standardne metrike segmentacije za konfiguracije sa 8% preklapanja.

Metoda	AP	AP_{50}	AP_{75}	Precision	Recall	F1	AR_{6000}
Bez konsolidacije	0.1903	0.3373	0.1892	0.5342	**0.5273**	0.5307	0.4291
SMM	0.3123	0.4950	0.3329	0.7387	0.5251	0.6138	0.4489
BC-SMM	**0.3219**	**0.5078**	**0.3447**	**0.7461**	0.5255	**0.6167**	**0.4521**

Prelazak sa nekonsolidovanog izlaza na SMM povećava AP sa 0.1903 na 0.3123, dok BC-SMM dodatno povećava AP na 0.3219. AP_{50} raste sa 0.3373 na 0.5078, a AP_{75} sa 0.1892 na 0.3447. Precision raste sa 0.5342 na 0.7461 uz praktično stabilan recall, što je konzistentno sa funkcijom globalne konsolidacije: metoda prvenstveno uklanja redundantne i pogrešno razdvojene predikcije, dok osnovna sposobnost modela da pronađe objekte ostaje približno nepromenjena.

Standardne metrike pokazuju ukupno poboljšanje kvaliteta segmentacije, dok tile-boundary metrike objašnjavaju mehanizam tog poboljšanja: redukciju fragmentacije identiteta i redundanse nastale višestrukim lokalnim posmatranjem istog objekta.

4.5.7 Uloga preklapanja u globalnoj konsolidaciji

Rezultati režima sa 0% i 8% preklapanja ne porede se direktno kao ista crossing populacija, jer promena tiled geometrije menja $\mathcal{G}_{\text{cross}}$. Informativna su poređenja unutar iste geometrije i obrazac promena između nekonsolidovanog izlaza, SMM-a i BC-SMM.

inferencija nad pločicama stvara uslove za boundary fragmentaciju, jer isti fizički objekat može biti podeljen između više lokalnih modelskih ulaza. Preklapanje obezbeđuje više lokalnih, delimično redundantnih opažanja istog objekta. Bez globalne konsolidacije redundantna lokalna opažanja ostaju kao višestruke predikcije, dok parcijalna opažanja pojedinih objekata mogu ostati organizovana kao više prostornih fragmenata.

To se vidi iz razlike u obrascu poboljšanja. Bez preklapanja BC-SMM dodatno smanjuje TBFR sa 0.3700 na 0.3587. Pri 8% preklapanju ista razlika je znatno veća: TBFR pada sa 0.0866 na 0.0356. Ovaj nalaz pokazuje da konsolidacija koja uvažava granice pločica najviše dobija na značaju kada inferencija nad pločicama obezbedi redundantna opažanja koja se mogu prostorno uskladiti.

4.5.8 Robusnost rezultata: uparena bootstrap analiza

Statistička pouzdanost razlika procenjena je uparenim bootstrap postupkom opisanim u Sekciji 3.7, sa 1000 ponavljanja na nivou originalnih slika. Tabela 4.19 prikazuje procenjene razlike i njihove 95% percentile bootstrap intervale za glavna poređenja pri 8% preklapanju.

U poređenju nekonsolidovanog izlaza sa originalnim SMM-om, 95% intervali za TBFR i TBEF u potpunosti su negativni, što potvrđuje stabilno smanjenje dominantnog dela fragmentacije prisutne u nekonsolidovanom tiled izlazu. Za SMM prema BC-SMM, intervali za TBFR i TBEF takođe su u potpunosti negativni: $[-0.06070, -0.04341]$ za TBFR i $[-0.07183, -0.05133]$ za TBEF. Dakle, dodatni doprinos BC-SMM stabilan je upravo na metrikama koje mere učestalost i intenzitet fragmentacije.

Za TBMR interval obuhvata nulu. SMM i BC-SMM imaju identičnu agregatnu TBMR vrednost i isti ukupan broj detektovanih i promašenih crossing instanci, iako raspodela pojedinačnih slučajeva po slikama nije potpuno identična. Za RIoU interval takođe obuhvata

Tabela 4.19. Uparene bootstrap razlike tile-boundary metrika pri 8% preklapanju pločica. Prikazani su point estimate razlike i 95% percentile bootstrap intervali na osnovu 1000 iteracija.

Poređenje	Metrika	M_A	M_B	$\Delta = B - A$	95% CI za Δ
Bez konsolidacije → SMM	TBFR	0.49067	0.08661	-0.40407	[-0.43018, -0.38155]
Bez konsolidacije → SMM	TBEF	0.61285	0.11360	-0.49925	[-0.53883, -0.46661]
Bez konsolidacije → SMM	TBMR	0.20490	0.20628	+0.00138	[0.00083, 0.00201]
Bez konsolidacije → SMM	RIoU	0.60160	0.60679	+0.00519	[0.00399, 0.00638]
SMM → BC-SMM	TBFR	0.08661	0.03562	**-0.05098**	**[-0.06070, -0.04341]**
SMM → BC-SMM	TBEF	0.11360	0.05288	**-0.06072**	**[-0.07183, -0.05133]**
SMM → BC-SMM	TBMR	0.20628	0.20628	0.00000	[-0.00036, 0.00036]
SMM → BC-SMM	RIoU	0.60679	0.60678	-0.00002	[-0.00036, 0.00031]
SMM → BC-SMM	MR	0.00287	0.00276	-0.00011	[-0.00024, 0.00003]
Bez konsolidacije → BC-SMM	TBFR	0.49067	0.03562	**-0.45505**	**[-0.48522, -0.42815]**
Bez konsolidacije → BC-SMM	TBEF	0.61285	0.05288	**-0.55997**	**[-0.60814, -0.52120]**
Bez konsolidacije → BC-SMM	RIoU	0.60160	0.60678	+0.00517	[0.00391, 0.00640]

nulu, pa nije detektovana pouzdana promena geometrijske rekonstrukcije. Interval za MR obuhvata nulu, pa nema dokaza da BC-SMM povećava overmerge.

Kada se BC-SMM direktno poredi sa nekonsolidovanim 8% izlazom, efekat je najizraženiji: TBFR se smanjuje za približno 92.7%, a TBEF za približno 91.4%. Bootstrap intervali za obe razlike ostaju u potpunosti u domenu poboljšanja.

Kontrolno, pri 0% preklapanju BC-SMM takođe dodatno smanjuje TBFR i TBEF u odnosu na SMM, ali je apsolutna korist manja: razlika TBFR iznosi -0.01132 uz interval $[-0.01519, -0.00833]$, dok je za TBEF razlika -0.01484 uz interval $[-0.01918, -0.01126]$. Ovaj obrazac je u skladu sa interpretacijom da se najveći doprinos konsolidacije koja uvažava granice pločica javlja kada overlap obezbedi više redundantnih opažanja istog objekta.

4.5.9 Sažetak nalaza

Rezultati pokazuju da glavni merljivi doprinos BC-SMM-a nije u dodatnom otkrivanju objekata niti u promeni geometrijskog kvaliteta rekonstruisanih maski, već u pouzdanijem očuvanju identiteta instance preko granica pločica, odnosno u smanjenju fragmentacije već detektovanih objekata. Na primer, avion koji prelazi granicu dve susedne pločice može biti ispravno detektovan u obe pločice, ali predstavljen sa dve odvojene maske; BC-SMM u takvom slučaju češće rekonstruiše jednu globalnu instancu bez nužne promene geometrijskog kvaliteta same maske.

4.6 Eksperimentalna evaluacija selektivnog oporavka predikcija niže pouzdanosti

Selektivni oporavak ispituje da li se među RF-DETR predikcijama ispod regularnog inferencionog praga može pronaći mali, dovoljno pouzdan podskup stvarnih instanci i vratiti u pipeline pre BC-SMM konsolidacije. Razvoj modela, izbor karakteristika, hiperparametri i

operativni prag određeni su isključivo na podacima izvedenim iz trening dela iSAID skupa. Zvanični iSAID validation skup nije korišćen ni u jednom razvojnom koraku.

4.6.1 Informativnost i progresija reprezentacije karakteristika

U ovoj sekciji izraz reprezentacija karakteristika (engl. *feature representation*) označava skup ulaznih atributa koje sekundarni klasifikator koristi za rangiranje rescue kandidata.

Početna analiza obuhvatila je RF-DETR predikcije sa skorom pouzdanosti u intervalu $[0{,}30, 0{,}50)$. Na 300 originalnih slika izdvojenih iz trening dela iSAID skupa dobijeno je 102 452 low-confidence kandidata, od čega 16 304 pozitivna i 86 148 negativna. Kandidat je pozitivan samo ako odgovara stvarnoj instanci koja nije već pokrivena regularnom high-confidence putanjom.

Ukupan broj instanci koje nisu bile pokrivene regularnom putanjom iznosio je 126 327. U posmatranom low-confidence prostoru pronađeno je 16 304 potencijalno oporavljivih instanci, što odgovara oracle potencijalu od približno 12,91%. Pozitivna stopa raste sa približavanjem regularnom pragu, ali validni kandidati postoje u celom posmatranom opsegu.

Tabela 4.20. Udeo pozitivnih rescue kandidata u različitim confidence intervalima.

Confidence interval	Udeo pozitivnih kandidata
$[0{,}30, 0{,}35)$	10,59%
$[0{,}35, 0{,}40)$	13,91%
$[0{,}40, 0{,}45)$	17,91%
$[0{,}45, 0{,}50)$	22,09%

Za razvoj modela odabran je interval $[0{,}40, 0{,}50)$. Na internom DEV skupu on sadrži 11 068 kandidata, od kojih je 2 245 pozitivno i 8 823 negativno. Isti kandidati, oznake i TRAIN–DEV split korišćeni su u svim poređenjima reprezentacija. Raw confidence rangiranje daje PR-AUC od 0,23008.

Tri naučene reprezentacije ispitane su istim XGBoost klasifikatorom. Lokalno-geometrijska reprezentacija sadrži confidence, klasu, geometriju maske i bbox-a, položaj unutar pločice, lokalnu gustinu i prostorno bliske predikcije. Prostorno proširena overlap–seam reprezentacija dodaje odnos prema overlap zoni, unutrašnjim granicama pločica i susednim tile-ovima. Cross-model reprezentacija dopunjuje ove informacije kompatibilnim signalom nezavisnog YOLO26 segmentacionog modela; YOLO26 se ne koristi za dodavanje svojih instanci, promenu klase ili zamenu maske RF-DETR kandidata.

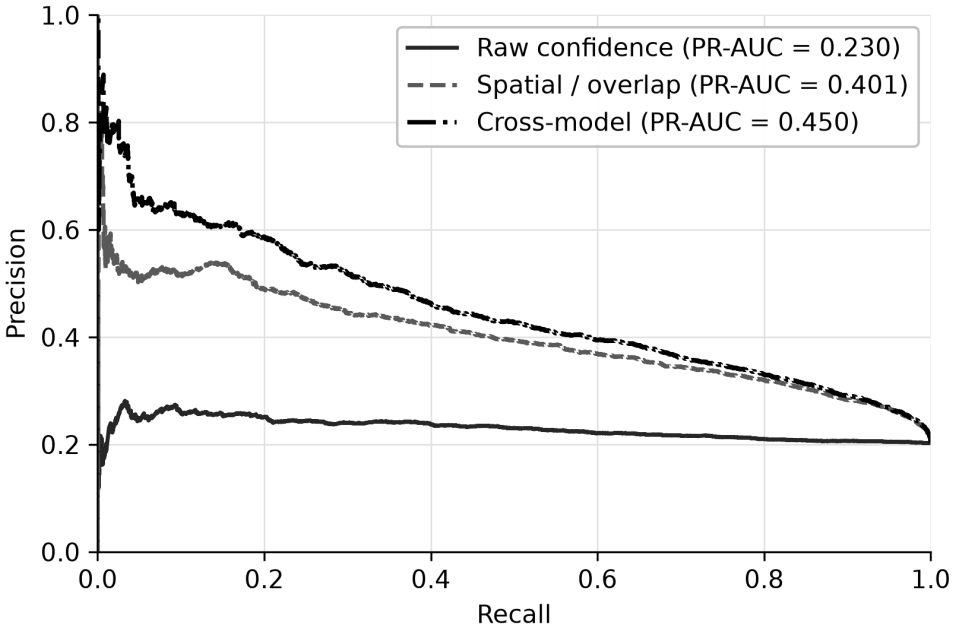
Rezultati progresije prikazani su u Tabeli 4.21. TP i FP odnose se na najbolju operativnu tačku uz unapred definisano ograničenje $FP/TP \leq 0{,}25$.

Najveći početni dobitak nastaje prelaskom sa raw confidence skora na lokalno-geometrijsku reprezentaciju. Overlap–seam kontekst daje manji dodatni pomak, dok cross-model signal povećava PR-AUC na 0,45048 i omogućava 57 TP uz 14 FP u konzervativnom režimu. Zamrznuti rescue prag za finalnu reprezentaciju iznosi $\theta_R = 0{,}6247759461$. Od 57 pozitivnih

Tabela 4.21. Progresija kvaliteta reprezentacije low-confidence kandidata na TRAIN izvedenom DEV skupu za interval $[0{,}40, 0{,}50)$.

Pristup	Model	PR-AUC	ROC-AUC	TP	FP	Precision
Raw confidence	–	0,23008	–	–	–	–
Lokalno-geometrijska	XGBoost	0,39462	0,74457	12	2	85,71%
Prostorno proširena	XGBoost	0,40113	0,75009	12	3	80,00%
Cross-model proširena	XGBoost	**0,45048**	**0,76894**	**57**	14	80,28%

DEV kandidata, 17 potiče iz intervala $[0{,}40, 0{,}45)$, a 40 iz intervala $[0{,}45, 0{,}50)$.



Slika 4.3. Precision–recall krive za tri reprezentativna nivoa opisa kandidata za selektivni oporavak. Proširivanje osnovnog confidence signala prostornim karakteristikama i nezavisnim cross-model signalom povećava PR-AUC sa 23,01% na 40,11% i 45,05%.

Pri približno 50% rescue recall-a lokalno-geometrijska reprezentacija daje 1122 TP i 1760 FP, overlap–seam reprezentacija 1123 TP i 1729 FP, a cross-model reprezentacija 1123 TP i 1511 FP. Ove tačke pokazuju zašto je za finalni sistem izabran konzervativan prag: agresivno povećavanje rescue recall-a brzo povećava broj pogrešno prihvaćenih kandidata.

4.6.2 Razvojna reintegracija i izbor operativnog režima

Rezultat na nivou kandidata proverava se kroz ceo cevovod, jer prihvaćene predikcije ulaze u sistem pre BC-SMM grupisanja. Na internom razvojnom (DEV) skupu zamrznuta konfiguracija prihvatila je 71 kandidata, od kojih 57 stvarnih i 14 pogrešnih. Nakon njihove reintegracije pre BC-SMM grupisanja dobijeno je 54 dodatnih tačno pozitivnih predikcija, uz 17 dodatnih lažno pozitivnih i 54 manje lažno negativnih; 54 od 57 ispravno prihvaćenih kandidata zadržalo je pozitivan doprinos i u konačnom rezultatu.

Tabela 4.22. Razvojna end-to-end evaluacija selektivnog oporavka nakon BC-SMM konsolidacije.

Metrika	BC-SMM	BC-SMM + rescue	Promena
TP	17 909	**17 963**	+54
FP	10 028	10 045	+17
FN	27 009	**26 955**	-54
Precision	64,105%	**64,135%**	+0,030 pp
Recall	39,870%	**39,991%**	+0,120 pp
F1	49,163%	**49,264%**	+0,100 pp

Marginalni odnos nakon reintegracije iznosi 17/54, odnosno 0,3148, a marginalni precision približno 76,06%. Taj odnos se ne poklapa potpuno sa candidate-level odnosom, jer BC-SMM određuje kako se ponovo uvedene lokalne predikcije prostorno konsoliduju u globalne instance.

4.6.3 Zamrznuta finalna evaluacija na official validation skupu

Pre finalne evaluacije zamrznuti su skup i redosled 191 ulazne karakteristike, XGBoost model i hiperparametri, interval $[0{,}40, 0{,}50)$, prag $\theta_R = 0{,}6247759461$, RF-DETR i YOLO26 inferencioni parametri, BC-SMM konfiguracija, fusion procedura i score politika. Referentne anotacije validation skupa nisu korišćene tokom generisanja karakteristika, selekcije kandidata niti određivanja rescue odluke.

Parity provera je pokazala da RF-DETR inferencija sa akvizicionim pragom $s \geq 0{,}30$, nakon filtriranja na $s \geq 0{,}50$, daje isti skup od 115 140 high-confidence predikcija kao prethodni regularni run sa pragom 0,50. Na zvaničnom validation skupu zatim je pronađeno 24 200 kandidata u intervalu $[0{,}40, 0{,}50)$. Zamrznuti klasifikator prihvatio je 159 kandidata, odnosno 0,657%; 139 kandidata pripadalo je klasi Small Vehicle, sedam klasi Large Vehicle, šest klasi Ship, četiri klasi Storage Tank i tri klasi Swimming Pool.

Nakon dodavanja 159 kandidata regularnim RF-DETR predikcijama broj pre-BC-SMM predikcija povećan je sa 115 140 na 115 299. BC-SMM i fusion izvršeni su bez promene parametara. Rezultati fixed-IoU evaluacije prikazani su u Tabeli 4.23.

Tabela 4.23. Konačna nezavisna evaluacija selektivnog oporavka na zvaničnom iSAID validation skupu.

Metrika	BC-SMM	BC-SMM + rescue	Promena
TP	61 304	**61 428**	+124
FP	20 862	20 897	+35
FN	55 344	**55 220**	-124
Precision	74,6099%	**74,6165%**	+0,0065 pp
Recall	52,5547%	**52,6610%**	+0,1063 pp
F1	61,6697%	**61,7451%**	+0,0754 pp

Finalna rescue grana dodaje 124 TP uz 35 FP i smanjuje FN za 124. Marginalni precision dodatka iznosi 77,99%, pa ukupni precision ne opada uprkos rastu recall-a. Prihvaćeno je

samo 159 od 24 200 raspoloživih kandidata, što odgovara izrazito selektivnom operativnom režimu.

COCO evaluacija završne konfiguracije daje AP=0,32199, $AP_{50} = 0{,}50784$ i $AP_{75} = 0{,}34466$ u standardnom režimu sa maxDets $= 100$. Po veličini objekta dobijeno je $AP_{\text{small}} = 0{,}18691$, $AP_{\text{medium}} = 0{,}40748$ i $AP_{\text{large}} = 0{,}50313$. U high-density režimu ostvareni su $AP_{50} = 0{,}61624$, $AP_{75} = 0{,}39819$ i $AR_{6000} = 0{,}45229$. Standardni COCO AP ostaje praktično nepromenjen u odnosu na osnovni BC-SMM sistem, što je u skladu sa malim brojem dodatih predikcija.

4.6.4 Razvojni i finalni rezultat

Na DEV skupu nakon BC-SMM dobijeno je +54 TP i +17 FP, dok je na zvaničnom validation skupu ostvareno +124 TP i +35 FP. Marginalni precision dodatih predikcija iznosi približno 76,06% na DEV skupu i 77,99% na validation skupu. Ovo poređenje pokazuje isti smer promene i sličan odnos dodatnih TP i FP predikcija posle zamrzavanja konfiguracije.

Eksperimenti selektivnog oporavka zato daju dva odvojena rezultata: feature reprezentacije sa lokalnim, prostornim i cross-model signalima bolje rangiraju low-confidence kandidate od raw confidence skora, a konzervativna operativna tačka se može reintegrisati u kompletan RF-DETR + BC-SMM pipeline uz mali, ali merljiv dobitak u fixed-IoU odzivu i kontrolisan FP trošak.

4.7 Analiza rezultata po klasama i veličini objekata

Agregatne metrike iz prethodnih odeljaka potvrđuju ukupan efekat globalne konsolidacije i selektivnog oporavka, ali ne pokazuju kako dobitak zavisi od veličine i semantičke klase objekta. Zato je završna analiza sprovedena prema COCO kategorijama veličine objekata i prema nekoliko karakterističnih iSAID klasa.

Analiza obuhvata četiri reprezentativne konfiguracije: tiled inferenciju bez konsolidacije, SMM, predloženi BC-SMM i konačni BC-SMM sistem sa selektivnim oporavkom predikcija niže pouzdanosti. Kompletni rezultati za svih 15 iSAID klasa mogu se prikazati u dodatku, dok su u glavnom tekstu izdvojene klase koje najbolje ilustruju karakteristične efekte predloženog pipeline-a.

4.7.1 Lokalna dijagnostika segmentacionih modela po klasama

Početno poređenje lokalnih RF-DETR XLarge i YOLO26 Large modela po klasama prikazano je u Tabeli 4.24. Ova tabela predstavlja dijagnostiku modela na validation pločicama veličine 512×512 piksela i ne predstavlja end-to-end poređenje kompletnog B3 sistema nad originalnim slikama.

Tabela 4.24. Početne lokalne performanse RF-DETR XLarge i YOLO26 Large po klasama na iSAID validation pločicama.

Klasa	**RF-DETR XLarge**				**YOLO26 Large**			
	AP$_{50}$	**P**	**R**	F_1	**AP**$_{50}$	**P**	**R**	F_1
Baseball diamond	70,78	87,92	72,26	79,32	65,87	79,77	70,21	74,68
Basketball court	41,32	70,47	47,95	57,07	41,78	67,11	45,66	54,35
Bridge	38,64	79,67	41,54	54,61	35,26	59,62	43,42	50,25
Ground track field	55,96	85,65	58,36	69,42	54,43	72,39	61,20	66,32
Harbor	63,89	71,70	73,83	72,75	62,14	72,59	72,91	72,75
Helicopter	26,19	33,33	44,23	38,02	29,21	31,06	39,42	34,75
Large Vehicle	53,25	81,01	56,82	66,79	47,63	81,37	50,01	61,95
Plane	86,22	91,10	87,54	89,28	82,10	91,05	83,08	86,88
Roundabout	48,65	86,15	50,68	63,82	55,09	82,80	58,82	68,78
Ship	73,74	95,37	78,53	86,13	70,41	93,20	74,76	82,97
Small Vehicle	35,60	73,75	41,53	53,13	24,27	69,54	29,87	41,79
Soccer field	56,99	85,56	59,84	70,43	45,07	80,26	47,41	59,61
Storage tank	61,44	86,06	63,32	72,96	54,52	84,78	56,84	68,06
Swimming pool	63,61	88,98	66,55	76,14	49,83	78,43	52,00	62,54
Tennis court	85,49	95,06	86,21	90,42	85,73	91,71	86,21	88,88

RF-DETR XLarge ostvaruje viši F_1 u trinaest klasa, YOLO26 Large u klasi Roundabout, dok je u klasi Harbor F_1 vrednost izjednačena. Ova lokalna dijagnostika služi kao referentni kontekst za kasnije original-image rezultate, a ne kao kontrolisana ablacija kompletnog pipeline-a.

4.7.2 Uticaj veličine objekta

Standardne COCO AP vrednosti prema veličini objekta prikazane su u Tabeli 4.25.

Tabela 4.25. AP prema veličini objekta na iSAID validation skupu.

Metod	AP_{small}	AP_{medium}	AP_{large}
Bez konsolidacije	0,1252	0,2673	0,2827
SMM	0,1813	0,4021	0,4785
BC-SMM	**0,1868**	**0,4075**	**0,5031**
BC-SMM + rescue	**0,1869**	**0,4075**	**0,5031**

Najizraženije poboljšanje nastaje prelaskom sa nekonsolidovanih tiled predikcija na globalno konsolidovane instance. SMM povećava AP_{small} sa 0,1252 na 0,1813, AP_{medium} sa 0,2673 na 0,4021 i AP_{large} sa 0,2827 na 0,4785.

Dobitak zato nije ograničen samo na male objekte koji se najčešće povezuju sa problemom tiled inferencije. Konsolidacija posebno snažno utiče i na srednje i velike instance,

kod kojih isti fizički objekat može biti predstavljen većim brojem delimičnih ili međusobno preklopljenih lokalnih predikcija.

BC-SMM dodatno poboljšava SMM u sve tri grupe. Dobitak je naročito izražen za velike objekte, gde AP_{large} raste sa 0,4785 na 0,5031. To je u skladu sa namenom konsolidacije koja uvažava granice pločica: veći objekti češće presecaju granicu pločice ili se pojavljuju kroz više lokalnih posmatranja, pa bolje određivanje prostornih veza direktno utiče na konačnu instancu.

Za male objekte BC-SMM ostvaruje $AP_{\text{small}} = 0,1868$, naspram 0,1813 za SMM i 0,1252 bez konsolidacije. Time se potvrđuje da globalna konsolidacija nije korisna samo za spajanje velikih presečenih objekata, već i za smanjenje redundanse u gustim područjima malih instanci.

Uvođenje rescue grane ne menja rezultate za srednje i velike objekte, dok se AP_{small} neznatno povećava sa 0,1868 na 0,1869. Takav rezultat je očekivan jer je rescue komponenta izrazito selektivna i najveći broj prihvaćenih kandidata pripada klasi Small Vehicle. Njena uloga nije promena globalne AP raspodele, već vraćanje malog broja prethodno odbačenih validnih instanci.

4.7.3 Analiza karakterističnih klasa

Za detaljniju analizu izdvojene su klase Ship, Large Vehicle, Small Vehicle, Tennis Court, Plane i Harbor. One obuhvataju guste skupove malih objekata, veće izdužene instance i prostorno jasno definisane objekte različite geometrije.

Tabela 4.26 prikazuje F1 meru za ove klase.

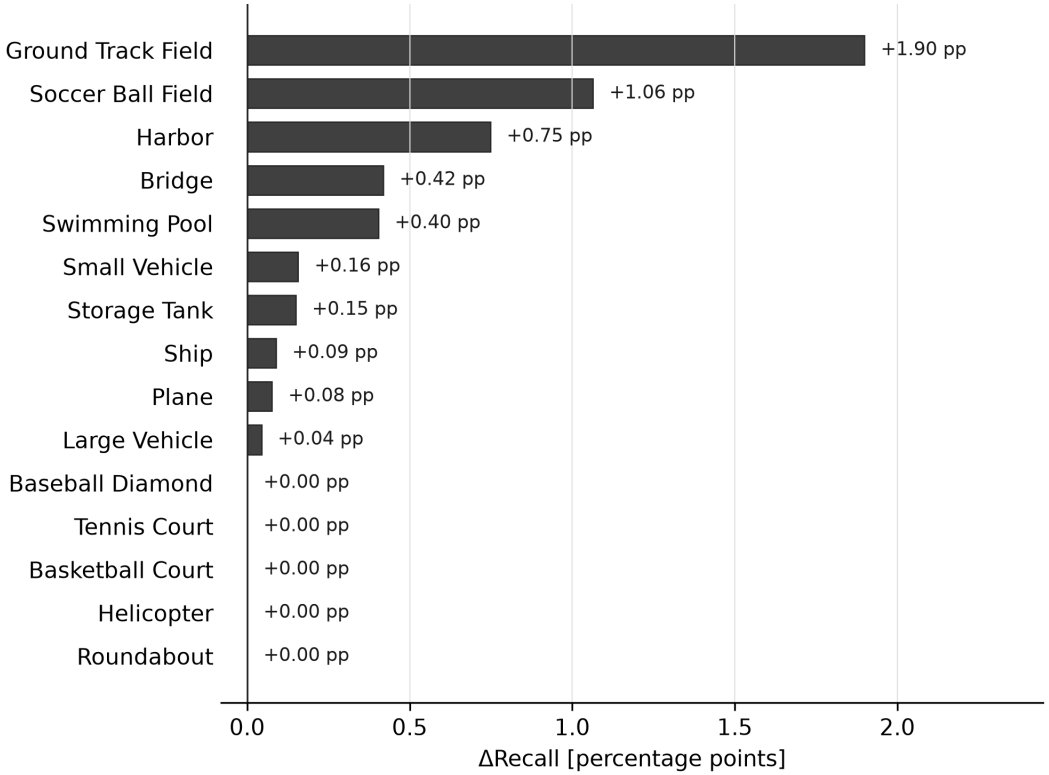
Tabela 4.26. F1 mera za odabrane iSAID klase.

Klasa	Bez kons.	SMM	BC-SMM	BC-SMM + rescue
Ship	70,16%	87,45%	87,96%	**87,98%**
Large Vehicle	52,43%	66,47%	66,85%	**66,87%**
Small Vehicle	48,77%	54,00%	54,03%	**54,13%**
Tennis Court	54,20%	88,44%	**91,36%**	**91,36%**
Plane	67,06%	86,37%	**89,51%**	**89,51%**
Harbor	51,46%	71,22%	**75,37%**	**75,37%**

Dodatno, radi pregleda efekta konačne B3 konfiguracije na odziv svih semantičkih klasa, na Slici 4.4 prikazana je promena recall-a u odnosu na referentni SMM rezultat pri istoj tiled geometriji sa 8% preklapanja.

Najveći porast odziva zabeležen je za klase Ground Track Field ($+1,90$ procentnih poena), Soccer Ball Field ($+1,06$ procentnih poena) i Harbor ($+0,75$ procentnih poena). U pet klasa odziv ostaje nepromenjen, dok negativna promena nije zabeležena ni u jednoj klasi. Prosečno povećanje odziva iznosi približno $0,34$ procentna poena.

Kod svih izdvojenih klasa glavni skok nastaje uvođenjem globalne konsolidacije. Ra-



Slika 4.4. Promena odziva po iSAID klasama za konačnu B3 konfiguraciju u odnosu na referentni SMM rezultat pri istoj tiled geometriji sa 8% preklapanja. Odziv raste u 10 od 15 klasa, dok u preostalih pet ostaje nepromenjen; pad nije zabeležen ni u jednoj klasi.

zlog nije prvenstveno povećanje recall-a, već uklanjanje velikog broja redundantnih lokalnih predikcija.

Klasa Ship predstavlja jasan primer. Bez konsolidacije precision iznosi samo 60,77%, uz recall od 83,00%. Nakon originalnog SMM precision raste na 92,67%, dok recall ostaje praktično nepromenjen. BC-SMM dodatno povećava precision na 93,76%, uz recall od 82,84%. F1 time raste sa 70,16% na 87,96%.

Još izraženiji primer predstavlja Tennis Court. Nekonsolidovani pipeline ostvaruje visok recall od 91,37%, ali precision od svega 38,52%, što je posledica velikog broja redundantnih lokalnih predikcija. Nakon BC-SMM recall ostaje 91,24%, dok precision raste na 91,48%. F1 se zato povećava sa 54,20% na 91,36%.

Sličan obrazac prisutan je kod klase Plane. Recall se kroz sve konfiguracije zadržava oko 93%, dok precision raste sa 52,36% bez konsolidacije na 80,47% nakon originalnog SMM i 86,03% nakon BC-SMM. Ovaj rezultat pokazuje da globalna konsolidacija prvenstveno uklanja višestruka lokalna posmatranja iste fizičke instance, bez značajnog gubitka detekcionog odziva.

Kod klase Harbor efekat BC-SMM je takođe izražen. F1 raste sa 51,46% bez konsolidacije na 71,22% nakon originalnog SMM i 75,37% nakon BC-SMM. Za razliku od prethodnih primera, ovde se uz rast precision-a beleži i veći recall, što pokazuje da kvalitetnija konsolidacija ne mora predstavljati samo uklanjanje redundanse, već može proizvesti i geo-

metrijski pogodniju finalnu instancu za one-to-one evaluaciono uparivanje.

4.7.4 Guste klase vozila i efekat selektivnog oporavka

Klase Small Vehicle i Large Vehicle posebno su značajne zbog velike zastupljenosti prostornih gustih objekata u iSAID skupu i dominantne uloge u eksperimentima selektivnog oporavka.

Za Small Vehicle nekonsolidovani pipeline sadrži 68 461 finalno evaluiranu predikciju, od kojih je 31 258 lažno pozitivno. Nakon BC-SMM broj predikcija smanjuje se na 52 702, a broj FP na 15 749. Precision se time povećava sa 54,34% na 70,12%.

Recall se menja relativno malo, sa 44,24% na 43,94%. To pokazuje da je dominantan problem nekonsolidovanog tiled output-a za ovu klasu redundansa, a ne nedostatak lokalnih detekcija. Zbog velike gustine malih vozila, globalno odlučivanje o pripadnosti istoj fizičkoj instanci postaje ključni korak za povećanje precision-a.

Za Large Vehicle efekat je sličan. Broj FP smanjuje se sa 6 348 bez konsolidacije na 1 788 nakon BC-SMM, dok recall ostaje praktično nepromenjen, sa 60,25% na 60,04%. Precision istovremeno raste sa 46,41% na 75,39%.

Ove dve klase su i najvažnija veza sa rescue granom. Zamrznuti rescue model na zvaničnom validation skupu prihvatio je 159 low-confidence kandidata, od kojih je 139 pripadalo klasi Small Vehicle, a sedam klasi Large Vehicle.

Nakon kompletne reintegracije i BC-SMM konsolidacije broj TP za Small Vehicle povećan je sa 36 953 na 37 061, odnosno za 108 instanci. Broj FN je istovremeno smanjen sa 47 142 na 47 034, dok je broj FP povećan za 31, sa 15 749 na 15 780.

Za Large Vehicle zabeležen je manji, ali isti smer promene: broj TP raste sa 5 478 na 5 482, dok se broj FN smanjuje za četiri.

Ovi rezultati objašnjavaju zašto se efekat rescue grane na agregatni AP_{small} vidi samo kao mala promena. Sistem je konzervativan i prihvata malo kandidata, ali su dodatni TP koncentrisani u najbrojnijoj i jednoj od najzahtevnijih klasa skupa.

Analiza po klasama i veličini pokazuje da se doprinosi komponenti razlikuju: globalna konsolidacija najviše menja precision i redundantnost lokalnih predikcija, dok rescue dodaje mali broj prethodno odbačenih validnih instanci, najviše u klasi Small Vehicle.

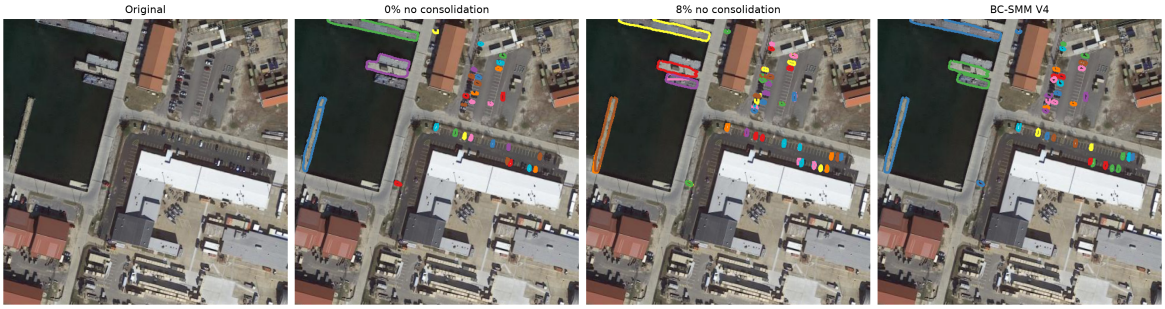
4.8 Kvalitativna analiza i vizuelni primeri

Pored kvantitativnih rezultata, korisno je vizuelno prikazati ponašanje predloženog pipeline-a na reprezentativnim primerima. U ovoj sekciji izdvojeni su slučajevi koji ilustruju uticaj preklapanja pločica, pojavu redundantnih tiled predikcija, razliku između originalnog SMM i predloženog BC-SMM pristupa, kao i uspešan selektivni oporavak predikcija niže pouzdanosti. Na kraju je prikazan i ograničavajući primer u kome globalna konsolidacija ne rešava u potpunosti složenost scene.

Vizuelni primeri organizovani su kao višepanelni prikazi u kojima se isti prostorni region poredi kroz različite faze obrade. Za iSAID primere predikcije su prikazane konturama maski radi preglednijeg poređenja, dok BONAI primeri dodatno uključuju referentne maske krovova, nekonsolidovane lokalne predikcije i konačni globalni rezultat.

4.8.1 Uticaj preklapanja pločica i globalne konsolidacije

Slika 4.5 prikazuje primer u kome preklapanje od 8% daje veći broj lokalnih opažanja istih objekata. Panel bez preklapanja sadrži manje lokalnih predikcija, dok panel sa 8% overlap-a uvodi izraženiju redundansu u istom regionu. Nakon BC-SMM konsolidacije taj skup postaje kompaktniji i konzistentniji.



Slika 4.5. Primer uticaja preklapanja pločica i naknadne konsolidacije. (a) originalni isečak slike, (b) inferencija nad pločicama bez konsolidacije i sa 0% overlap-a, (c) inferencija nad pločicama bez konsolidacije i sa 8% overlap-a, (d) konačni BC-SMM rezultat. Uvođenje overlap-a povećava broj lokalnih opažanja, dok globalna konsolidacija uklanja nastalu redundansu.

Primer pokazuje da preklapanje pločica nije samo mehanizam za ublažavanje sečenja objekata na granicama isečaka, već i izvor dodatne prostorne redundanse koja se koristi u globalnoj konsolidaciji. Time se vizuelno objašnjava zašto overlap sam po sebi nije dovoljan, već mora biti praćen postupkom koji lokalna opažanja vraća u jedinstven globalni skup.

4.8.2 Redundantne predikcije pre i posle konsolidacije

Slika 4.6 prikazuje region sa više izduženih vozila, gde nekonsolidovana inferencija nad pločicama generiše preklopljene i redundantne instance. SMM smanjuje tu redundansu, dok BC-SMM daje stabilniji konačni raspored.

Ovakvi primeri objašnjavaju zašto globalna konsolidacija povećava precision uz minimalan pad recall-a: problem tiled izlaza često nije nedostatak lokalnih detekcija, već njihova višestruka pojava u istom regionu slike. Redukcija tih višestrukih opažanja zato neposredno utiče na broj lažno pozitivnih finalnih instanci.

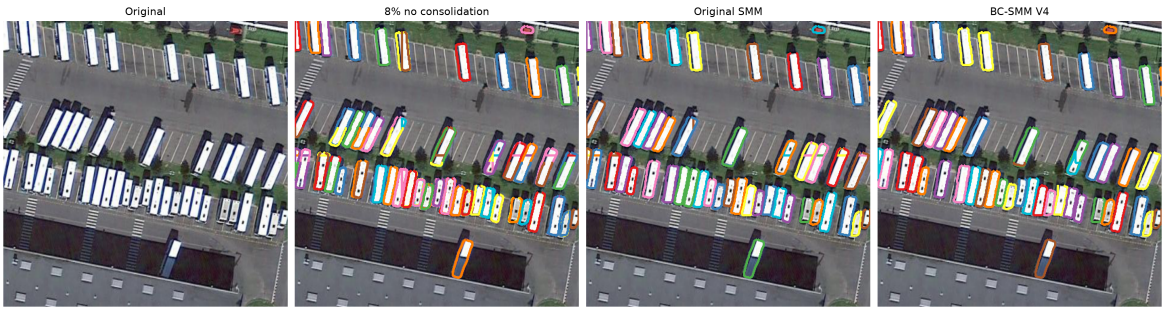
4.8.3 SMM naspram BC-SMM

Slika 4.7 prikazuje primer u kome se razlika između originalnog SMM pristupa i BC-SMM jasno uočava. Oba pristupa smanjuju redundansu u odnosu na nekonsolidovani



Slika 4.6. Primer redukcije redundantnih predikcija. (a) originalni isečak slike, (b) 8% overlap bez konsolidacije, (c) SMM i (d) BC-SMM. Uočava se da nekonsolidovani tiled izlaz sadrži više redundantnih i delimično preklopljenih instanci, dok globalna konsolidacija daje kompaktniji i precizniji konačni skup.

tiled izlaz, ali BC-SMM dodatno ublažava lokalnu fragmentaciju i zadržava bolje usklađene instance.



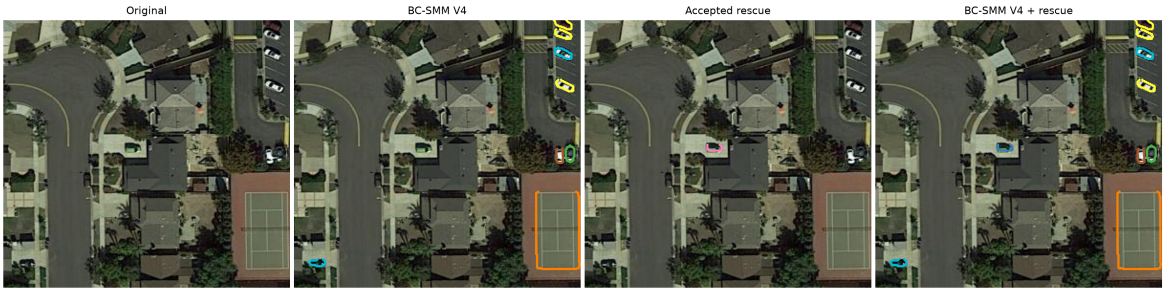
Slika 4.7. Vizuelno poređenje originalnog SMM i BC-SMM pristupa. (a) originalni isečak slike, (b) 8% overlap bez konsolidacije, (c) SMM i (d) BC-SMM. BC-SMM daje dodatno prečišćen i konzistentniji konačni raspored instanci u odnosu na SMM.

Primer pokazuje da doprinos BC-SMM nije samo redukcija broja predikcija, već i kvalitetnije globalno spajanje instanci u složenijim scenama. To je posebno važno u regionima u kojima se parcijalna opažanja nalaze blizu granica pločica i imaju sličnu lokalnu geometriju.

4.8.4 Uspešan primer selektivnog oporavka predikcija

Slika 4.8 prikazuje uspešan selektivni oporavak predikcije. Drugi panel prikazuje BC-SMM bez rescue grane, treći samo jedan prihvaćeni low-confidence kandidat, a četvrti konačni rezultat nakon njegove reintegracije i ponovne konsolidacije.

Primer kvalitativno potvrđuje selektivni oporavak predikcija niže pouzdanosti. BC-SMM konsoliduje postojeće tiled predikcije, dok rescue grana vraća mali broj dodatnih validnih instanci iz prostora niže pouzdanosti; ta dva mehanizma zato rešavaju različite, ali komplementarne probleme.



Slika 4.8. Primer uspešnog selektivnog oporavka predikcije. (a) originalni isečak slike, (b) BC-SMM rezultat, (c) prihvaćeni rescue kandidat izdvojen iz low-confidence prostora i (d) konačni BC-SMM + rescue rezultat. Dodatna instanca se pojavljuje samo nakon primene rescue grane, čime se vizuelno potvrđuje komplementarna uloga selektivnog oporavka u odnosu na globalnu konsolidaciju.

4.8.5 Ograničavajući primer

Slika 4.9 prikazuje i ograničavajući primer scene izuzetno velike gustine, u kojoj globalna konsolidacija ne donosi izraženo lokalno pojednostavljenje rezultata. U takvim regionima brojne prostorne interakcije između susednih instanci ostaju izazov i za tiled inferenciju i za naknadnu konsolidaciju.



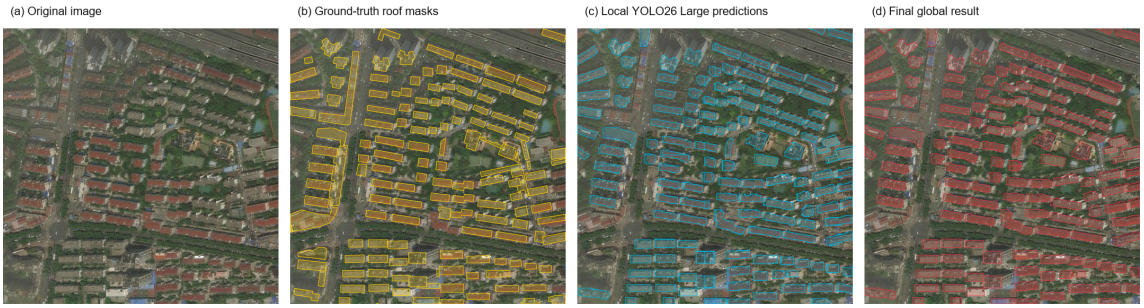
Slika 4.9. Primer ograničenja metode u veoma gustoj sceni. (a) originalni isečak slike, (b) 8% overlap bez konsolidacije, (c) SMM i (d) BC-SMM. Iako konsolidacija i ovde stabilizuje rezultat, ekstremna gustina i složenost scene ograničavaju obim lokalnog poboljšanja.

Ovaj primer pokazuje granice primenljivosti predloženog pipeline-a. Globalna konsolidacija smanjuje redundansu i fragmentaciju, ali u ekstremno gustim scenama ne eliminiše sve izvore neizvesnosti. Time se kvalitativna analiza ne svodi samo na uspešne slučajeve, već prikazuje i uslove u kojima ostaje prostor za dalja poboljšanja.

4.8.6 Kvalitativna analiza prenosivosti na BONAI skup

Kvalitativni primeri na BONAI skupu ilustruju kako globalna konsolidacija izgleda i van malih objekata karakterističnih za iSAID. Posmatrane scene sadrže veće krovne instance, izraženu perspektivnu geometriju i veliki broj objekata koji se pojavljuju u više preklopljenih pločica. Za svaki primer porede se originalna slika, referentne maske krovova, nekonsolidovane lokalne YOLO26 Large predikcije i konačni globalni rezultat.

Na reprezentativnom primeru sa Slike 4.10 globalna konsolidacija ne menja broj tačno pronađenih instanci, ali značajno smanjuje broj redundantnih i fragmentisanih predikcija. Broj TP ostaje 119, dok broj FP opada sa 210 na 19, a broj fragmentisanih GT instanci sa 98 na 2. Recall zato ostaje nepromenjen na 81,51%, dok F_1 raste sa 50,11% na 83,80%. U ovom primeru značajan deo koristi globalne konsolidacije ostvaruje se uklanjanjem redundantnih lokalnih opažanja i fragmentacije.



Slika 4.10. Reprezentativan BONAI primer pre i nakon globalne konsolidacije.

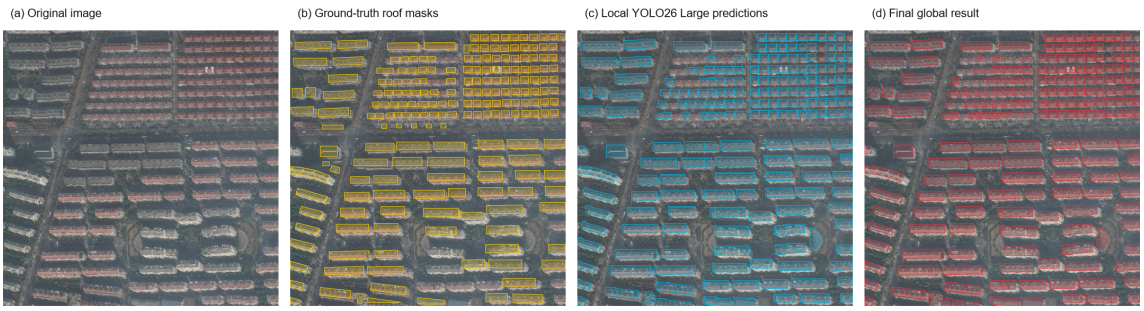
Primer sa Slike 4.11 sadrži veće i geometrijski složenije krovove, uključujući instance koje prelaze granice više lokalnih pločica. Nakon globalne konsolidacije broj TP raste sa 111 na 112, broj FP se smanjuje sa 150 na 23, dok se broj fragmentisanih GT instanci smanjuje sa 76 na 3. Istovremeno F_1 raste sa 55,09% na 80,87%, pa ovaj slučaj ilustruje ponašanje postupka na znatno većim prostornim strukturama.



Slika 4.11. Primer globalne konsolidacije većih krovnih instanci koje se pojavljuju u više pločica.

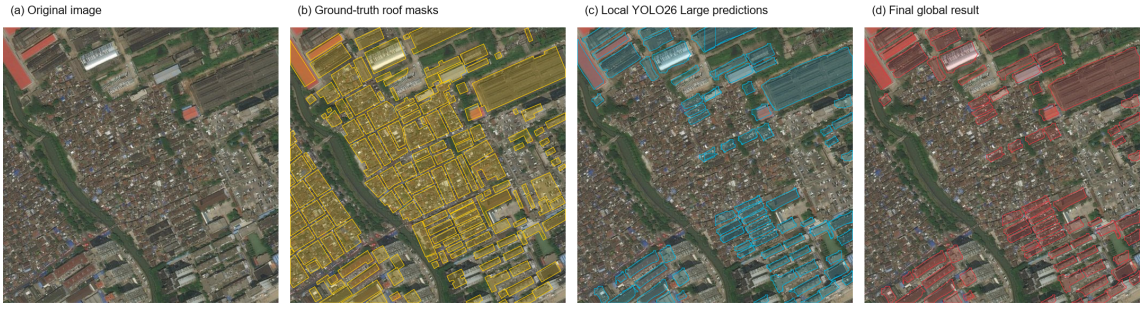
Posebno izražen efekat uočava se u gustom urbanom regionu sa Slike 4.12. U sceni sa 205 referentnih krovnih instanci, od kojih se 163 pojavljuju u više pločica, broj TP raste sa 181 na 183, dok se broj FP smanjuje sa 248 na 28. Broj fragmentisanih GT instanci istovremeno opada sa 142 na 4, a F_1 raste sa 57,10% na 87,98%. Na prikazanoj sceni globalna konsolidacija ostaje korisna i pri veoma velikoj gustini međusobno bliskih objekata.

Granica postupka ilustrovana je na Slici 4.13. Iako globalna konsolidacija smanjuje broj FP sa 109 na 10 i broj fragmentisanih GT instanci sa 46 na 1, broj TP ostaje 63, a broj FN 93, tako da recall ostaje nepromenjen na 40,38%. U ovom primeru konsolidacija efikasno



Slika 4.12. Primer globalne konsolidacije u gustom urbanom regionu.

uklanja redundantne i fragmentisane lokalne predikcije, ali ne nadoknađuje instance koje osnovni segmentacioni model nije pouzdano detektovao ni u jednom lokalnom posmatranju.



Slika 4.13. Primer ograničenja postupka na zahtevnoj BONAI sceni.

Posmatrana četiri BONAI primera ilustruju isti neposredni obrazac kao i iSAID primeri: globalna konsolidacija smanjuje redundantne lokalne predikcije i fragmentaciju, ali ne može nadoknaditi instance koje osnovni model nije pouzdano pronašao ni u jednom lokalnom posmatranju.

4.9 Računske performanse

Pored kvaliteta segmentacije, praktična primenljivost predloženog sistema zavisi i od njegove računske složenosti. Analiza zato razdvaja cenu tiled inferencije, formiranja grupa i naknadne fuzije maski.

SMM i BC-SMM koriste isti prethodno generisan skup RF-DETR predikcija, pa se njihovo poređenje zasniva na vremenu formiranja grupa i odgovarajućem algoritamskom workload-u.

4.9.1 Cena tiled inferencije

Povećanje overlap-a sa 0% na 8% povećava broj generisanih pločica sa 14 222 na 16 617, odnosno za približno 16,8%. Istovremeno broj sirovih predikcija raste sa 102 886 na

115 140.

Tabela 4.27. Računska cena RF-DETR tiled inferencije na iSAID validation skupu.

Konfiguracija	Broj tile-ova	Zbir tile-level vremena [s]	Prosek/tile [s]	P95/tile [s]
0% overlap	14 222	6591,44	0,4635	1,435
8% overlap	16 617	7653,46	0,4606	1,415

Broj pločica povećava se za približno 16,8%, dok zbir tile-level vremena raste za približno 16,1%. Prosečno vreme po pločici ostaje praktično nepromenjeno, pa je dodatni trošak overlap-a gotovo proporcionalan broju dodatnih tile-ova.

4.9.2 Vreme formiranja grupa

Direktno poređenje SMM i BC-SMM sprovedeno je nad istih 458 validacionih slika i istim skupom od 115 140 ulaznih predikcija.

Tabela 4.28. Vreme formiranja grupa po slici za SMM i BC-SMM.

Metod	Medijana [s]	P90 [s]	P95 [s]	Maksimum [s]
SMM	0,635	4,167	8,324	746,59
BC-SMM	**0,483**	**2,607**	**5,130**	**173,72**

BC-SMM smanjuje medijanu vremena sa 0,635 s na 0,483 s, odnosno za oko 24,0%. P90 opada sa 4,167 s na 2,607 s (oko 37,4%), a P95 sa 8,324 s na 5,130 s (oko 38,4%). Najveća razlika javlja se kod računarski najzahtevnijih slika: maksimalno vreme opada sa 746,59 s na 173,72 s, odnosno za približno 76,7%.

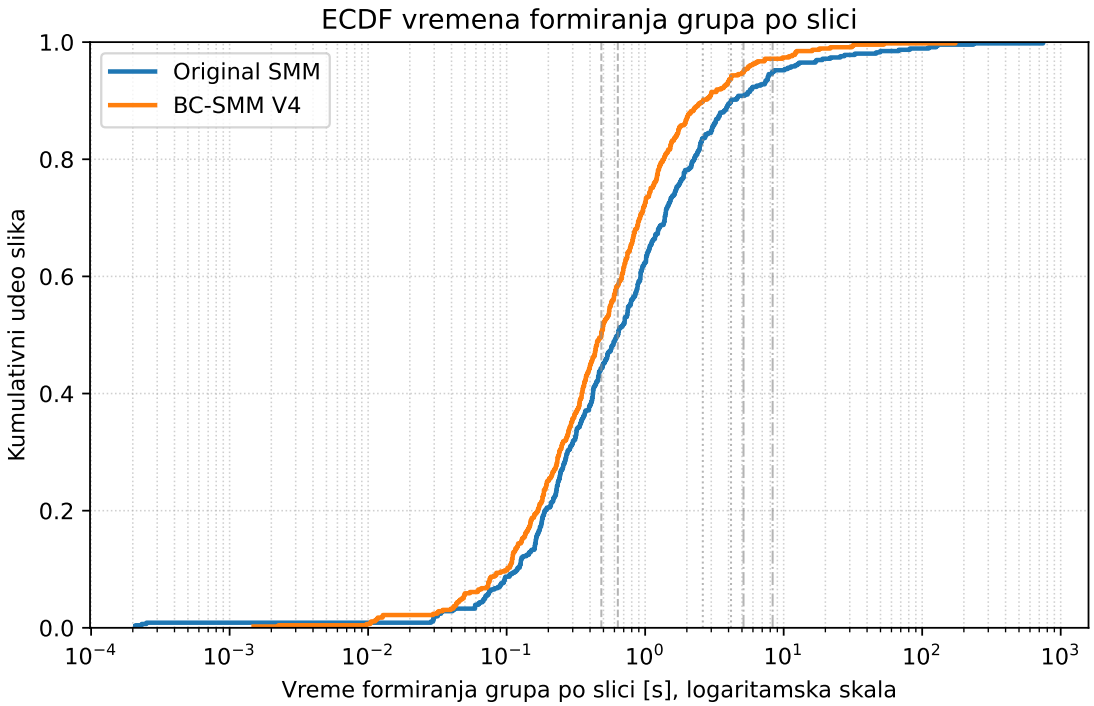
ECDF prikaz pokazuje da ubrzanje nije posledica samo nekoliko ekstremnih slučajeva, već je prisutno kroz veliki deo distribucije.

4.9.3 Redukcija kandidatnog prostora

Tabela 4.29. Algoritamski workload SMM i BC-SMM.

Metod	Ulazne predikcije	Kandidatni parovi	Kompatibilni parovi	Izlazne grupe
SMM	115 140	626 437	43 193	82 919
BC-SMM	115 140	**286 745**	43 029	82 166

BC-SMM smanjuje broj kandidatnih parova sa 626 437 na 286 745, odnosno za približno 54,2%. Broj identifikovanih kompatibilnih parova manji je za svega 164 para, odnosno oko 0,38%. To pokazuje da ubrzanje ne potiče iz agresivnog odbacivanja korisnih veza, već iz uklanjanja prostorno malo verovatnih parova pre skupljih faza evaluacije.



Slika 4.14. ECDF vremena formiranja grupa po slici. BC-SMM kriva je kroz najveći deo distribucije pozicionirana ulevo od SMM krive, što pokazuje kraće vreme obrade za isti kumulativni udeo slika.

4.9.4 Cena fuzije maski

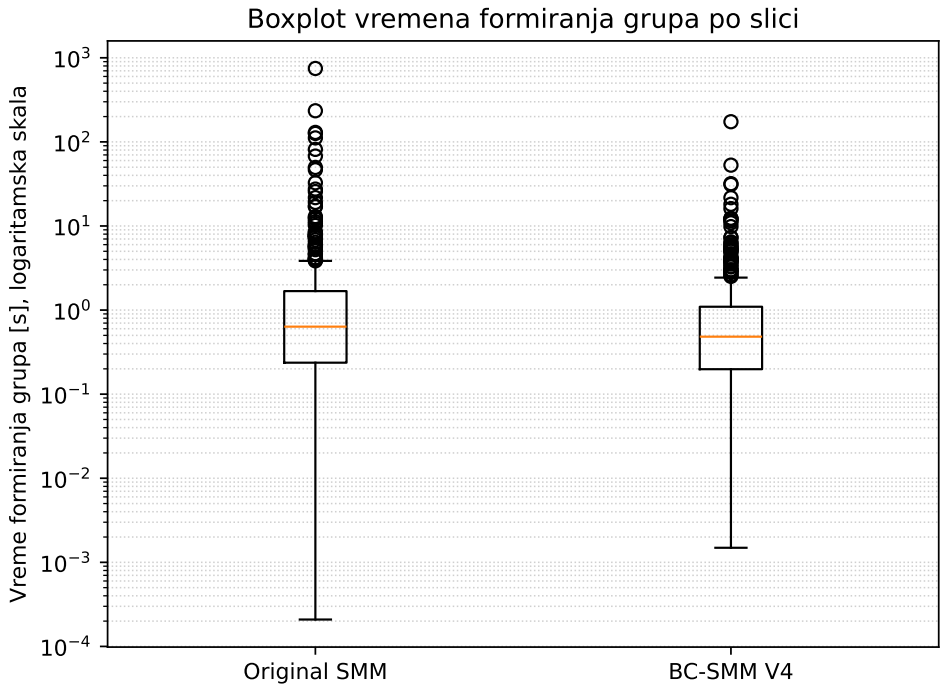
Formiranje grupa i fuzija maski predstavljaju dve odvojene faze. Nakon što su grupe već formirane, funkcija fuzije obrađuje njihove izvorne maske i formira izlazne instance.

Tabela 4.30. Ukupno proteklo vreme kompletne faze fuzije i statistike vremena obrade pojedinačnih slika nakon formiranja grupa.

Metod	Ukupno [s]	Prosek/slika [s]	Medijana/slika [s]	Maksimum [s]
SMM	2273,92	**4,883**	**0,406**	**364,03**
BC-SMM	2388,29	5,133	0,439	380,96

Za razliku od faze formiranja grupa, fuzija maski kod BC-SMM ima neznatno veći računski trošak. Iako BC-SMM razmatra znatno manji broj kandidatnih parova, nakon grupisanja formira nešto više višečlanih grupa (25 564 prema 25 424) i u fazi fuzije obrađuje oko 1,55% više izvornih maski. Zbog toga je ukupno vreme fuzije približno 5% veće.

Ovaj rezultat pokazuje da računska prednost BC-SMM-a ne potiče iz same operacije spajanja maski, već iz efikasnijeg formiranja kandidatnog prostora i određivanja grupa. Boundary-aware ograničenja smanjuju broj razmatranih kandidatnih parova za više od polovine, dok se obim rada u kasnijoj fazi fuzije praktično ne smanjuje.



Slika 4.15. Boxplot vremena formiranja grupa po slici na logaritamskoj vremenskoj osi. BC-SMM ima nižu medijanu, uži centralni opseg i znatno manje ekstremne vrednosti od SMM pristupa.

4.9.5 Računski uticaj rescue grane

Za rescue klasifikator nisu sačuvana zasebna vremena feature ekstrakcije i XGBoost inferencije, pa se ona ne prikazuju kao samostalne referentne vrednosti.

Ipak, moguće je proceniti njegov uticaj na narednu BC-SMM fazu. Zamrznuta rescue konfiguracija dodaje samo 159 predikcija osnovnom skupu, čime broj ulaznih predikcija raste sa 115 140 na 115 299, a broj kandidatnih parova sa 286 745 na 286 803, odnosno za svega 58 parova.

Prema tome, selektivna rescue grana ne uvodi značajno dodatno opterećenje u narednoj BC-SMM fazi: broj kandidatnih parova raste za svega 58. Ovaj rezultat ne obuhvata cenu ekstrakcije rescue karakteristika i XGBoost inferencije, za koje zasebna timing merenja nisu sačuvana.

Računska analiza zato razdvaja tri efekta: overlap povećava cenu inferencije skoro proporcionalno broju dodatnih pločica, BC-SMM smanjuje vreme formiranja grupa i kandidatni prostor, dok je fuzija maski blago skuplja. Za rescue granu nisu sačuvana zasebna timing merenja feature extraction i XGBoost koraka, pa se njen trošak može oceniti samo kroz mali porast narednog BC-SMM kandidatnog prostora.

4.10 Ablacione i dodatne analize

Prethodni odeljci bili su usmereni na konačni pipeline i njegove glavne eksperimentalne rezultate. Tokom razvoja ispitano je i više alternativnih komponenti koje nisu uključene

u završni sistem. Njihova analiza pokazuje koji izvori informacija stvarno doprinose rezultatu, koje intervencije imaju ograničenu praktičnu vrednost i kako su razvojni eksperimenti vodili ka konačnoj konfiguraciji.

U ovom odeljku sažeto su prikazani: razvojna putanja BC-SMM algoritma, morfološko prečišćavanje finalnih maski, post-hoc korekcija semantičke klase, kontrolni klasifikacioni eksperimenti u rescue grani i poređenje različitih reprezentacija low-confidence kandidata.

4.10.1 Razvojna ablacija BC-SMM pristupa

Razvoj BC-SMM algoritma nije predstavljao monotono dodavanje sve većeg broja ograničenja. Naprotiv, eksperimenti su pokazali da previše restriktivan kandidatni domen može značajno umanjiti kvalitet konsolidacije.

Tabela 4.31 prikazuje glavne razvojne konfiguracije na iSAID validation skupu sa 8% overlap-a.

Tabela 4.31. Razvojna putanja algoritama globalne konsolidacije pri 8% overlap-a.

Metod	AP	AP_{50}	F1	TP	FP
Bez konsolidacije	0,1903	0,3373	0,5307	61 508	53 632
SMM	0,3123	0,4950	0,6138	61 249	21 670
BC-SMM V1	0,2381	0,3891	0,5612	61 465	40 918
BC-SMM V2	0,2564	0,4064	0,5608	61 495	41 156
BC-SMM V3	0,2773	0,4550	0,6030	61 307	25 395
EMM V1	0,3162	0,5033	0,6115	60 674	21 118
BC-SMM	**0,3219**	**0,5078**	**0,6167**	61 304	**20 862**

BC-SMM V1 predstavljao je izrazito restriktivnu formulaciju ograničenu na neposredne odnose uz granice između susednih pločica. Broj kandidatnih parova smanjen je sa 626 437 kod SMM-a na 61 156, odnosno za približno 90%, ali je broj kompatibilnih parova istovremeno pao sa 43 193 na 14 137.

To je pokazalo da redukcija kandidatnog prostora nije dovoljna sama po sebi. Analiza propuštenih relacija identifikovala je, između ostalog, 6 821 kompatibilan dijagonalni par i 262 kompatibilna para između fizički preklopljenih, ali indeksno nesusednih pločica, što je motivisalo kasnije proširenje kandidatnog domena.

V2 je uveo razvijeniju kompatibilnost koja uzima u obzir granice između susednih pločica i topološku logiku, ali bez dovoljno širokog kandidatnog domena nije uspeo da dostigne SMM. V3 je zatim proširio pretragu u koridoru uz granice između susednih pločica i približio se SMM-u, povećavajući AP na 0,2773 i F1 na 0,6030.

Finalni BC-SMM dodaje fizički opravdane dijagonalne kontakte, fallback relacije kod stvarnog preklapanja pločica i kontrolisano topološko proširenje grupa. U finalnom run-u kandidatni prostor ima 286 745 parova: 239 764 iz ortogonalnih odnosa na granicama pločica, 45 702 iz dijagonalnih kontakata i 1 279 iz rezervnih relacija zasnovanih na preklapanju.

Time je zadržano znatno manje kandidata nego kod SMM-a, uz gotovo isti broj kompatibilnih veza: 43 029 prema 43 193 kod SMM-a.

Razvojna putanja zato pokazuje da je ključan kompromis između preširoke globalne pretrage i previše restriktivnog lokalnog filtriranja. Finalni BC-SMM koristi fizičku geometriju tiled prostora da smanji kandidatni prostor, ali zadržava relacije potrebne za kvalitetnu konsolidaciju.

4.10.2 Morfološko prečišćavanje maski

Nakon BC-SMM fuzije dodatno je ispitano da li jednostavne morfološke operacije mogu popraviti lokalne nepravilnosti kontura i time povećati konačni kvalitet segmentacije.

Morfološke izmene bile su namerno konzervativne. Korišćeni su uslovi visokog geometrijskog očuvanja originalne instance, uključujući minimalni preservation IoU između originalne i izmenjene maske. Cilj nije bilo agresivno menjanje geometrije, već lokalno zatvaranje manjih prekida i nepravilnosti.

Rezultat na kompletnom validation skupu prikazan je u Tabeli 4.32.

Tabela 4.32. Uticaj morfološkog prečišćavanja nakon BC-SMM konsolidacije.

Konfiguracija	AP	AP_{50}	AP_{75}	F1
BC-SMM	0,321938	0,507839	0,344659	0,616697
BC-SMM + morfologija	0,321821	0,507848	0,344603	0,616707
Promena	-0,000117	+0,000009	-0,000056	+0,000010

Promene svih glavnih metrika praktično su zanemarljive. AP je blago smanjen, AP_{50} gotovo nepromenjen, a F1 povećan za svega 10^{-5}.

Dodatni razvojni eksperimenti doveli su do sličnog zaključka. Pojedine morfološke konfiguracije menjale su stotine ili hiljade maski, posebno u klasi Small Vehicle, ali su proizvele svega nekoliko dodatnih TP instanci ili nikakav merljiv dobitak.

Nalaz pokazuje da nakon tiled inferencije sa overlap-om, globalne konsolidacije i fuzije maski preostale greške dominantno nisu posledica jednostavnih lokalnih nepravilnosti konture. Naknadno morfološko “peglanje” zato nije glavni izvor daljeg poboljšanja i nije uključeno u finalni pipeline.

4.10.3 Naknadna analiza semantičke korekcije klase

Analiza preostalih grešaka pokazala je da je značajan deo izraženih međuklasnih konfuzija prisutan između klasa Small Vehicle i Large Vehicle. Posebno je značajno što su ove greške često povezane sa instancama koje su geometrijski dobro lokalizovane i segmentirane.

Na official validation skupu identifikovano je 1 064 jakih međuklasnih uparivanja, od kojih je 934, odnosno 87,78%, pripadalo upravo paru Small Vehicle–Large Vehicle. Srednji IoU iznosio je približno 0,788 u slučajevima kada je GT Small Vehicle predviđen kao Large

Vehicle, odnosno oko 0,756 u obrnutom smeru. Ovi rezultati pokazuju da kod velikog broja takvih primera osnovni problem više nije lokalizacija ili oblik maske, već dodela odgovarajuće semantičke klase.

Zbog toga je dodatno ispitan post-hoc pristup zasnovan na XGBoost klasifikatoru, čiji je zadatak da proceni da li finalnoj BC-SMM instanci treba promeniti semantičku klasu. Odluka se ne zasniva samo na karakteristikama same konsolidovane instance, već se dopunjuje nezavisnim semantičkim dokazom drugog segmentacionog modela. Za svaku finalnu BC-SMM instancu analizirano je prostorno slaganje sa predikcijama nezavisnog YOLO26 modela, uključujući preklapanje maski, preklapanje ograničavajućih pravougaonika, udaljenost centara, konfidencije i saglasnost ili nesaglasnost predviđene klase.

Na zamrznutom DEV skupu osnovna BC-SMM odluka sadržala je 274 Small/Large semantičke greške. Nakon optimizacije konzervativnog post-hoc korektora pronađen je operativni prag pri kome je predloženo 20 promena klase. Od toga je 17 promena ispravilo prethodno pogrešnu odluku, dok su tri promene uvele novu grešku. Neto dobitak zato iznosi 14 ispravljenih odluka, odnosno broj grešaka se smanjuje sa 274 na 260. Preciznost prihvaćenih korekcija iznosi 85%, dok relativno neto smanjenje posmatranih Small/Large semantičkih grešaka iznosi približno 5,11%.

Tabela 4.33. Rezultat post-hoc semantičke korekcije Small/Large Vehicle na zamrznutom DEV skupu.

Metrika	Vrednost
Broj početnih Small/Large grešaka	274
Predložene korekcije	20
Korisne korekcije	17
Štetne korekcije	3
Neto dobitak	14
Konačan broj grešaka	260
Preciznost korekcija	85,0%
Neto smanjenje semantičkih grešaka	5,11%
PR-AUC	0,2291
ROC-AUC	0,8816

Rezultat potvrđuje da nezavisan model može da obezbedi dodatni signal za prepoznavanje dela rezidualnih semantičkih grešaka. Međutim, dobijeno poboljšanje je relativno malo u odnosu na ukupan broj finalnih instanci, a dominantan doprinos potiče upravo od nezavisnog modelskog dokaza. Informacije povezane sa tiled inferencijom, granicama isečaka i multi-tile konsolidacijom imaju sekundarnu ulogu u ovoj odluci.

Dodatno je uočena izražena asimetrija rezultata. Svih 17 korisnih korekcija odnosilo se na slučajeve u kojima je BC-SMM instanca bila klasifikovana kao Large Vehicle, dok je referentna klasa bila Small Vehicle. Time se potvrđuje da je nezavisni model posebno koristan za jednu dominantnu podvrstu Small/Large konfuzije, ali ne pruža podjednako pouzdan signal u oba smera.

Zbog toga post-hoc semantička korekcija nije uključena u glavni predloženi pipeline. Eksperiment se zadržava kao dodatna analiza prirode preostalih grešaka i pokazuje da, nakon geometrijske konsolidacije maski, deo rezidualne greške prelazi iz domena lokalizacije i segmentacije u domen semantičke klasifikacije. Uvođenje drugog modela može delimično ublažiti ovaj problem, ali uz ograničen praktični dobitak i dodatnu računarsku složenost.

4.10.4 Ablacija reprezentacije rescue kandidata

Najčistiji dodatni nalaz rescue eksperimenata dobija se poređenjem kvaliteta rangiranja kako se reprezentaciji postepeno dodaje novi informacioni sadržaj.

Tabela 4.34. Poređenje informacionog sadržaja različitih rescue reprezentacija.

Reprezentacija	PR-AUC	ROC-AUC	TP	FP
Raw confidence	0,23008	–	–	–
Lokalno-geometrijska	0,39462	0,74457	12	2
Overlap–seam proširena	0,40113	0,75009	12	3
Cross-model proširena	**0,45048**	**0,76894**	**57**	14

Progresija PR-AUC vrednosti od 0,23008 preko 0,39462 i 0,40113 do 0,45048 pokazuje da originalni confidence skor nije dovoljan za procenu kvaliteta low-confidence instance.

Najveći prvi dobitak nastaje uvođenjem geometrijskog i lokalnog konteksta. Dodatne overlap–seam karakteristike daju manje, ali konzistentno poboljšanje, dok nezavisan cross-model signal proizvodi najveći naredni skok.

Ova analiza pokazuje da rangiranje low-confidence kandidata dobija dodatni signal iz lokalnog, overlap–seam i cross-model konteksta.

4.10.5 Zaključak ablacionih i dodatnih analiza

Ablacije pokazuju četiri važne granice finalnog sistema: previše stroga pravila povezivanja u BC-SMM mogu da odbace korisne veze; morfološko prečišćavanje gotovo ne utiče na konačne metrike; korekcija klasa pokazuje nedovoljan pozitivan rezultat na razvojnom skupu i nije uključena u finalni sistem; kod selektivnog oporavka najveći napredak nastaje kada se lokalne i prostorne karakteristike dopune informacijom iz drugog modela.

4.11 Sinteza glavnih eksperimentalnih nalaza

Preklapanje pločica povećava pokrivenost scene, ali u nekonsolidovanom izlazu istovremeno povećava redundansu. U poređenju B0 i B1 broj TP raste za 1 926, ali broj FP raste za 10 328, od čega 6 838 potiče od duplikata. Zato B1 ne predstavlja finalni pipeline, već ulazni režim koji stvara dodatna lokalna opažanja za naknadnu globalnu konsolidaciju.

SMM pokazuje da se ta redundansa može pretvoriti u kvalitetniji skup konačnih instanci: pri 8% overlap-a broj FP opada sa 53 632 na 21 670, dok recall ostaje praktično očuvan. BC-SMM dodatno poboljšava referentni SMM rezultat: broj FP se smanjuje za još 808, AP raste sa 0,3123 na 0,3219, a broj kandidatnih parova opada sa 626 437 na 286 745 uz gotovo isti broj kompatibilnih veza.

Tile-boundary metrike objašnjavaju koji deo poboljšanja standardni AP ne razdvaja. Pri 8% overlap-a BC-SMM u odnosu na SMM smanjuje TBFR sa 0,0866 na 0,0356 i TBEF sa 0,1136 na 0,0529. Bootstrap intervali za ove razlike ne obuhvataju nulu, dok za RIoU i MR nije detektovana pouzdana promena. Time se ukazuje da je glavni detektovani efekat BC-SMM-a smanjenje identitetske fragmentacije, dok za promenu geometrijske pokrivenosti i merge rate nije detektovana pouzdana promena.

Selektivni oporavak predstavlja konzervativan način povećanja odziva sistema. Na razvojnom skupu PR-AUC raste sa 0,23008 za sirovi skor pouzdanosti na 0,45048 kada se uključe informacije iz drugog modela. Na zvaničnom iSAID validation skupu prihvaćeno je 159 od 24 200 kandidata niže pouzdanosti, od čega 124 tačno pozitivna i 35 lažno pozitivnih, uz smanjenje broja FN za 124. Dobitak je mali, ali je ostvaren bez pada ukupne preciznosti.

Lokalna RF-DETR/YOLO evaluacija na pločicama koristi se samo kao deskriptivni referentni kontekst. Detaljna lokalna dijagnostika po klasama prikazana je u Tabeli 4.24. Zbirni kontekst i konačni B3 rezultat dati su u Tabeli 4.35.

Tabela 4.35. Početne performanse lokalnih modela i konačni rezultat kompletnog pipeline-a nad originalnim slikama visoke rezolucije.

Konfiguracija	**Precision**	**Recall**	F_1	$\mathbf{AP}_{50}$	$\mathbf{AP}_{75}$	**AP**
RF-DETR XLarge, lokalna evaluacija	79,08%	50,37%	61,54%	57,45%	38,04%	36,16%
YOLO26 Large, lokalna evaluacija	76,79%	40,86%	53,34%	53,56%	32,07%	31,75%
kompletan predloženi postupak, originalne slike	74,62%	52,66%	61,75%	50,78%	34,47%	32,20%

Vrednosti u Tabeli 4.35 nisu kontrolisano pre/posle poređenje: prva dva reda odnose se na lokalnu evaluaciju pločica veličine 512×512 piksela, dok se poslednji red odnosi na end-to-end evaluaciju B3 sistema nad 458 originalnih slika. Numerička bliskost F_1 vrednosti zato služi samo kao kontekst za prenos lokalnih segmentacionih sposobnosti na nivo originalne slike.

Dodatna BONAI provera koristi drugi skup podataka i YOLO26 Large kao primarni lokalni segmentacioni model. Zbirni rezultati prikazani su u Tabeli 4.36.

Tabela 4.36. Deskriptivni pregled lokalnih segmentacionih modela i kompletnog B3 sistema na skupu BONAI.

Konfiguracija	**Precision**	**Recall**	F_1	**AP**$_{50}$	**AP**$_{75}$	**AP**	**Mean IoU**
RF-DETR XLarge, lokalna evaluacija	89,82%	55,10%	68,30%	53,25%	29,44%	29,13%	78,31%
YOLO26 Large, lokalna evaluacija	85,64%	58,43%	69,47%	54,81%	27,19%	28,45%	76,45%
kompletan predloženi postupak, originalne slike	79,98%	62,23%	69,99%	54,88%	21,94%	26,63%	74,64%

Na BONAI skupu B3 ostvaruje $F_1 = 69{,}99\%$ i $\text{AP}_{50} = 54{,}88\%$ nad originalnim slikama, uz high-density $\text{AP}_{50} = 57{,}39\%$. Ovaj rezultat se tumači kao dodatna evidencija o prenosivosti postupka na drugi skup podataka i drugu osnovnu segmentacionu arhitekturu.

Zajedno, eksperimenti pokazuju sledeći rezultatni tok: overlap obezbeđuje dodatna lokalna opažanja, SMM i BC-SMM ih konsoliduju u globalne instance, tile-boundary metrike izdvajaju smanjenje fragmentacije na granicama pločica, a rescue dodaje ograničen recall-oriented signal uz kontrolisan FP trošak. Šire implikacije, ograničenja i prenosivost ovih nalaza razmatraju se u Poglavlju 5.

5. DISKUSIJA

5.1 Interpretacija glavnih nalaza

Rezultati sprovedenog istraživanja pokazuju da glavni problem tiled inferencije nije samo činjenica da se slika obrađuje u delovima, već način na koji se lokalna opažanja ponovo organizuju u globalni prikaz scene. Segmentacioni model može veoma uspešno obraditi pojedinačnu pločicu, ali prelazak na originalnu sliku visoke rezolucije uvodi dodatni nivo problema: isti fizički objekat može biti viđen više puta, delimično prikazan u različitim pločicama ili predstavljen sa više lokalnih maski. Zbog toga kvalitet lokalnog segmentacionog modela sam po sebi nije dovoljan da garantuje isti kvalitet konačne segmentacije na nivou cele slike.

U tom kontekstu, overlap ne treba posmatrati kao neposredno rešenje problema granica pločica. Njegova važnija uloga je da povećava količinu dostupne informacije. Isti region scene može biti posmatran iz više lokalnih prozora, čime se smanjuje rizik da objekat ostane nepotpuno vidljiv samo zbog položaja granice pločice. Međutim, ta dodatna informacija istovremeno proizvodi redundantne i parcijalne predikcije. Korist overlap-a zato zavisi od toga da li naredni korak sistema može pravilno da protumači odnose između tih predikcija.

Glavni nalaz istraživanja jeste da globalnu konsolidaciju zbog toga nije dovoljno posmatrati kao klasično uklanjanje duplikata. Njena funkcija je bliža rekonstrukciji latentne globalne instance iz više lokalnih opažanja. BC-SMM ovu ideju dodatno razvija korišćenjem poznate geometrije tiled inferencije. Informacija o tome iz kojih pločica predikcije potiču i kakav fizički odnos te pločice imaju postaje deo odluke o tome koje lokalne maske mogu predstavljati isti objekat. Time struktura koja je prvobitno uvedena samo radi podele velike slike postaje aktivan izvor informacije u postprocesiranju.

Razvoj BC-SMM-a dodatno pokazuje da ni potpuno globalna pretraga ni previše strogo lokalno ograničavanje nisu dovoljni za posmatrani problem. Važan rezultat je mogućnost da se kandidatni prostor ograniči na osnovu fizičke strukture tiled mreže, ali bez gubitka veza potrebnih za kvalitetnu globalnu rekonstrukciju.

Uvedene tile-boundary metrike imaju važnu ulogu u tumačenju ovog rezultata, jer omogućavaju da se promena ukupnih referentnih metrika poveže sa konkretnim mehanizmima greške. Standardni AP kvantifikuje agregirani kvalitet konačnih predikcija, ali sam ne izdvaja da li promena potiče od manjeg broja propuštenih objekata, smanjenja fragmentacije, uklanjanja redundantnih predikcija ili boljeg očuvanja identiteta instance. Specijalizovana evaluacija zato omogućava da se efekat globalne konsolidacije posmatra direktno tamo gde problem tiled inferencije nastaje.

Selektivni oporavak dodatno proširuje ovu interpretaciju. Eksperimenti pokazuju da

informacija nastala višestrukim posmatranjem scene nije korisna samo za spajanje maski. Prostorni odnosi, lokalni kontekst i saglasnost ili neslaganje različitih modela mogu pružiti dodatni signal o tome da li predikciju niže pouzdanosti treba zadržati. Dobitak ove grane ostaje namerno ograničen konzervativnim izborom kandidata, ali pokazuje da tiled redundansa može biti iskorišćena i kao izvor informacije o pouzdanosti, a ne samo kao problem koji treba ukloniti.

Početna lokalna evaluacija RF-DETR modela i konačna B3 evaluacija imaju numerički slične F_1 vrednosti, ali se odnose na različite evaluacione nivoe: pločice veličine 512×512 piksela i originalne visokorezolucione slike. Zato ovo poređenje ima samo deskriptivnu ulogu. Efekat pojedinačnih komponenti sistema treba tumačiti kroz kontrolisane B0/B1/SMM/B2/B3 konfiguracije, gde se poređenje izvodi na istom nivou originalne slike.

Ovi eksperimenti zato razdvajaju dva nivoa problema: lokalno prepoznavanje objekta unutar pojedinačnog prozora i očuvanje njegovog identiteta prilikom rekonstrukcije cele scene. Predloženi pristup prvenstveno deluje na ovom drugom nivou, sistematskim korišćenjem prostorne redundanse i poznate strukture tiled inferencije.

5.2 Overlap i globalna konsolidacija

Poređenje SMM i BC-SMM pristupa pokazuje da za kvalitetnu konsolidaciju nije neophodno zadržati potpuno globalni kandidatni prostor. Korišćenjem fizičke topologije tiled mreže BC-SMM zadržava gotovo isti skup relevantnih kompatibilnih odnosa uz približno 54% manje kandidatnih parova. Razvoj ranijih BC-SMM varijanti dodatno pokazuje da ni previše restriktivno ograničavanje samo na neposredne boundary odnose nije dovoljno: dijagonalni kontakti, stvarno prostorno preklapanje i širi topološki kontekst mogu biti neophodni za očuvanje ispravnog identiteta instance.

Ovaj rezultat je posebno važan u gustim scenama i kod malih objekata, gde jedna fizička instanca zauzima mali deo originalne slike i lako postaje zavisna od položaja granice pločice. Upravo takvi slučajevi ostaju najzahtevniji i nakon konsolidacije, što pokazuje da inferencija nad pločicama najviše dobija na značaju kada lokalna rezolucija mora biti očuvana, ali je istovremeno potrebno kontrolisati veliki broj međusobno bliskih opažanja.

Analiza po klasama pokazuje da povećanje odziva nije posledica ponašanja samo nekoliko dominantnih kategorija. U kontrolisanom poređenju B3 konfiguracije sa referentnim SMM rezultatom pri istoj tiled geometriji sa 8% preklapanja, recall raste u 10 od 15 iSAID klasa, u preostalih pet ostaje nepromenjen, a pad nije zabeležen ni u jednoj klasi. Ovakav obrazac podržava tumačenje da višestruka posmatranja nastala preklapanjem pločica povećavaju dostupnost korisnog lokalnog signala, ali da se taj signal mora posmatrati kroz kontrolisano poređenje na nivou originalne slike.

Tile-boundary analiza dodatno pokazuje da poboljšanje BC-SMM-a ne potiče prvenstveno od naknadnog oblikovanja lepših maski. Smanjenje fragmentacije i viška fragmenata,

uz to da nije detektovana pouzdana promena geometrijske pokrivenosti, ukazuje da je glavni detektovani dobitak na nivou identiteta instance, odnosno u pravilnijem povezivanju lokalnih delova istog fizičkog objekta.

Zbog toga se problem tiled inferencije može posmatrati šire od klasičnog uklanjanja duplikata. Suštinski zadatak je rekonstrukcija globalne instance iz skupa parcijalnih i redundantnih lokalnih opažanja, pri čemu poznata geometrija tiled prostora predstavlja koristan strukturni signal za ograničavanje i usmeravanje tog procesa.

5.3 Prostorno indeksiranje i grafovska konsolidacija

Originalni SMM predstavlja važnu polaznu tačku ovog istraživanja jer problem konsolidacije ne svodi na sekvencijalno uklanjanje duplikata, već ga formuliše kao problem odnosa između više lokalnih predikcija. Predikcije se posmatraju kao čvorovi grafa, dok prostorno i geometrijski kompatibilni parovi formiraju njegove veze. Takva reprezentacija omogućava da odluka o pripadnosti istoj fizičkoj instanci zavisi od strukture većeg skupa lokalnih opažanja, a ne samo od jednog izolovanog poređenja dve maske.

Važnu ulogu u ovom pristupu ima R-tree prostorni indeks. Njegova funkcija nije da odlučuje koje maske treba spojiti, već da efikasno pronađe prostorno bliske kandidate nad kojima tek treba izvršiti detaljniju proveru. Time se izbegava potpuno poređenje svake predikcije sa svakom drugom, što je posebno značajno kod visokorezolucionih scena sa velikim brojem lokalnih instanci. R-tree zato ostaje računski mehanizam za efikasno formiranje kandidatnog grafa, dok se semantički važna odluka donosi kroz kompatibilnost lokalnih predikcija.

Ograničenje SMM-a proizlazi iz toga što je prostorna blizina glavni motiv za formiranje kandidatnog prostora. Nakon tiled inferencije, međutim, poznato je i poreklo svake predikcije, odnosno iz koje je pločice nastala i kakav fizički odnos ta pločica ima prema drugim pločicama. BC-SMM zadržava R-tree i grafovski princip, ali dodatno koristi geometriju procesa kojim su lokalne predikcije nastale. Kandidati se zato ne razmatraju samo zato što su bliski u globalnom prostoru, već prvenstveno tamo gde geometrija tiled mreže daje fizičko opravdanje da su dve lokalne predikcije različita opažanja istog objekta.

Time se menja i značenje kandidatnog grafa. Kod SMM-a graf prvenstveno predstavlja mrežu geometrijski kompatibilnih predikcija, dok kod BC-SMM-a njegove veze dodatno nose informaciju o fizičkom kontaktu pločica, regionu uz granicu između susednih pločica i lokalnom kontinuitetu maske. Topologija mreže pločica tako postaje dodatna informacija koja usmerava prostornu pretragu, a ne naknadno pravilo primenjeno posle formiranja globalnog skupa kandidata.

Zato graf u BC-SMM-u više nije samo računska reprezentacija sličnosti, već i način eksplicitnog modelovanja topologije nastanka lokalnih predikcija. Ovo je naročito važno kod tranzitivnog grupisanja: niz pojedinačno prihvatljivih veza ne mora automatski značiti da cela

grupa ima prostorno opravdanu interpretaciju jedne instance.

Ovakav razvoj pokazuje da BC-SMM ne napušta osnovnu ideju SMM-a, već je specijalizuje za tiled inferenciju. R-tree ostaje efikasan mehanizam prostorne pretrage, a graf ostaje prirodna reprezentacija odnosa između predikcija. Ključna promena je u tome što BC-SMM koristi poznatu strukturu dekompozicije slike da ograniči gde se relevantne veze traže i koje njihove kombinacije mogu predstavljati jednu fizičku instancu.

5.4 Granice dodatnog postprocesiranja

Dodatni eksperimenti pokazuju da nakon globalne konsolidacije preostale greške više nisu lako rešive jednostavnim post-hoc intervencijama. Morfološko prečišćavanje praktično nije promenilo konačne rezultate, što ukazuje da dominantni problem nakon BC-SMM-a nije u manjim lokalnim nepravilnostima konture, već u složenijim greškama detekcije, identiteta i semantičke odluke.

Post-hoc korekcija klase daje precizniju granicu. Za usko definisanu klasnu konfuziju dobijen je merljiv DEV signal: većina predloženih korekcija bila je korisna i postignut je pozitivan neto efekat. Ipak, apsolutni broj korigovanih slučajeva ostao je mali, efekat je bio asimetričan i oslanjao se na dodatni modelski signal. Zato ova grana nije uključena u finalni sistem: ne zato što signal ne postoji, već zato što nije bio dovoljno širok i robustan da opravda dodatnu složenost.

Selektivni oporavak predstavlja drugačiji slučaj, jer je pokazao pozitivan efekat kada se primenjuje veoma konzervativno. Overlap, cross-tile i cross-model informacije omogućavaju izdvajanje dela korisnih low-confidence kandidata, ali se sa povećanjem recall-a brzo povećava i broj false-positive predikcija. Zbog toga njegova praktična vrednost zavisi više od kvaliteta selekcije nego od broja vraćenih kandidata.

Razvoj ranijih BC-SMM i alternativnih formulacija vodi ka sličnom zaključku: granice pločica i prostorna struktura tiled inferencije predstavljaju koristan signal, ali njihova primena mora zadržati ravnotežu između ograničavanja kandidatnog prostora i očuvanja relevantnih relacija.

Neutralni ili negativni rezultati dodatnih eksperimenata zato ne ukazuju na odsustvo korisne informacije, već određuju granicu do koje se ona može pouzdano iskoristiti jednostavnim postprocesnim pravilima.

5.5 Generalizacija i prenosivost

Prenosivost predloženog pristupa dodatno je proverena izvan primarne RF-DETR–iSAID konfiguracije. Evaluacija na BONAI skupu, sa YOLO26 Large kao osnovnim segmentacionim modelom, pokazala je da se isti osnovni postupak može primeniti na drugi skup podataka i drugu segmentacionu arhitekturu, bez promene osnovne konfiguracije globalne

konsolidacije.

Rezultati na BONAI skupu istovremeno pokazuju da prenos postupka ne znači nužno poboljšanje svih evaluacionih pokazatelja. Lokalni YOLO26 Large na pločicama ostvaruje recall od 58,43% i $F_1 = 69{,}47\%$, dok B3 na originalnim BONAI slikama ostvaruje recall od 62,23% i $F_1 = 69{,}99\%$. Ovo poređenje je deskriptivno, jer lokalna evaluacija na pločicama i end-to-end evaluacija nad originalnim slikama nisu isti evaluacioni nivo. Istovremeno, B3 ima niže vrednosti AP75, ukupnog AP i srednjeg IoU, što ukazuje na drugačiji kompromis između preciznosti, odziva i geometrijske preciznosti pri strožim kriterijumima poklapanja.

Zbog toga BONAI eksperiment ne treba tumačiti kao dokaz da jedna fiksna postprocesna konfiguracija univerzalno poboljšava sve metrike, već kao dodatnu eksperimentalnu evidenciju da se osnovni princip tiled inferencije, globalne konsolidacije i selektivnog oporavka može primeniti na drugačiju distribuciju podataka i drugačiji osnovni segmentacioni model. Istovremeno, uočeni kompromis pokazuje da optimalna radna tačka sistema može zavisiti od karakteristika skupa podataka, kalibracije osnovnog modela, izabranih pragova i geometrije ciljnih objekata.

Pored iSAID i BONAI skupova, isti osnovni problem dekompozicije i naknadne rekonstrukcije globalnih instanci može se javiti i u drugim domenima visokorezolucione vizuelne inspekcije. Slika 5.1 ilustruje potencijalni scenario UAV inspekcije infrastrukturnih objekata, kao što su mostovi, kod kojih veoma male lokalne nepravilnosti i oštećenja treba analizirati u kontekstu slike rezolucije više hiljada piksela. Taj domen nije eksperimentalno validiran u ovoj disertaciji i zahtevao bi zasebnu evaluaciju.



Slika 5.1. Primer visokorezolucionog UAV snimka mosta sa segmentiranim površinskim oštećenjima i drugim lokalnim nepravilnostima. Primer ilustruje scenario u kome se veliki infrastrukturni objekat obrađuje kroz više lokalnih regiona, dok pojedina oštećenja mogu biti izdužena i prostirati se preko granica susednih pločica, što može zahtevati naknadnu globalnu konsolidaciju lokalnih predikcija.

U takvim scenarijima inferencija nad pločicama mogla bi da sačuva lokalnu rezoluciju bez agresivnog smanjivanja cele slike. Međutim, defekti poput pukotina, površinskih oštećenja ili izloženih delova armature mogu biti izduženi, tanki i prostirati se preko više su-

sednih pločica. Time se ponovo javlja srodan metodološki problem: više lokalnih predikcija treba povezati u jednu fizički konzistentnu globalnu instancu.

Ovakvi objekti mogu predstavljati i zahtevniji slučaj za BC-SMM od većine iSAID instanci. Jedna duga pukotina može prolaziti kroz više pločica i formirati lanac parcijalnih maski, pri čemu susedni delovi mogu imati vrlo mali međusobni IoU iako pripadaju istom fizičkom defektu. To ukazuje na mogućnost daljeg proširenja topološkog modela za instance koje se prostiru kroz veći broj pločica.

Posebna vrednost precizne segmentacije instanci u ovom domenu jeste mogućnost kvantifikacije. Za razliku od graničnog pravougaonika, precizna maska može predstavljati osnovu za procenu površine, dužine ili drugih geometrijskih karakteristika oštećenja. Ovaj rad, međutim, ne sprovodi geometrijsku kalibraciju u fizičke jedinice. Uz poznatu prostornu skalu snimka, takvo preslikavanje iz pikselskog u fizički prostor predstavljalo bi potencijal buduće primene.

5.6 Računska i praktična primenljivost

Računski aspekt tiled inferencije predstavlja važan praktični kompromis. Povećanje overlap-a podrazumeva veći broj pločica i samim tim veći broj modelskih prolaza, pa njegovu korist treba posmatrati zajedno sa dodatnim računskim troškom. Zbog toga nije cilj maksimalno povećavati preklapanje, već pronaći nivo koji donosi dovoljno dodatne informacije bez nepotrebnog uvećanja inferencionog opterećenja.

Kod globalne konsolidacije BC-SMM pokazuje prednost upravo kroz ograničavanje kandidatnog prostora. Korišćenjem fizičke topologije tiled mreže smanjuje se broj parova koji zahtevaju detaljnu proveru, što je posebno važno u gustim scenama sa velikim brojem lokalnih predikcija. Računska korist se ne ogleda samo u izmerenom vremenu group-formation faze, već i u znatno manjem kandidatnom prostoru, što ograničava broj parova koji zahtevaju detaljnu proveru.

Funkcija fuzije maski nije glavni izvor računskog ubrzanja, jer se najveća ušteda ostvaruje pre same fuzije, kroz selektivnije formiranje kandidatnog grafa. Slično tome, finalna rescue grana uvodi veoma mali broj dodatnih predikcija, pa njen uticaj na kasniji BC-SMM workload ostaje ograničen. Ipak, ukupan računski overhead rescue grane nije potpuno kvantifikovan, jer izdvajanje rescue karakteristika i sekundarna XGBoost klasifikacija nisu zasebno vremenski profilisani.

Praktično je značajno i to što su inferencija i postprocesiranje razdvojeni. Lokalne predikcije mogu se sačuvati i zatim više puta koristiti za ispitivanje različitih konfiguracija konsolidacije, fuzije ili rescue grane bez ponovnog pokretanja segmentacionog modela. Takva modularnost olakšava eksperimentalnu analizu, podešavanje sistema i kasniju integraciju novih postprocesnih komponenti.

BC-SMM je zato posebno relevantan u scenarijima sa velikim brojem lokalnih predik-

cija, jer upravo tada nekontrolisana globalna pretraga postaje računski najskuplja. Njegova računska prednost odnosi se prvenstveno na fazu formiranja grupa i poređenja kandidata, gde topološki ograničena konsolidacija zadržava praktičniji odnos između kvaliteta i računskog opterećenja.

5.7 Ograničenja predloženog pristupa

Eksperimentalna validacija predloženog pristupa primarno je sprovedena na iSAID skupu, dok je dodatna evaluacija na BONAI skupu i sa YOLO26 Large modelom pokazala da se kompletan B3 postupak može primeniti i na drugi skup podataka i drugu segmentacionu arhitekturu bez promene osnovne konfiguracije globalne konsolidacije. Dobijeni rezultat pruža dodatnu eksperimentalnu evidenciju o prenosivosti osnovnog metodološkog principa, ali istovremeno pokazuje da odnos preciznosti, odziva i geometrijskog kvaliteta može zavisiti od skupa podataka, osnovnog modela i izabranih pragova.

Globalna konsolidacija zavisi od kvaliteta lokalnih predikcija. Ako objekat nije dovoljno dobro detektovan ni u jednoj pločici, postprocesiranje ga ne može rekonstruisati. Rescue grana može selektivno vratiti deo takvih low-confidence predikcija, ali njen domet ostaje ograničen izraženim kompromisom između preciznosti i odziva i, u cross-model varijanti, dodatnom sistemskom složenošću zbog potrebe za drugim segmentacionim modelom.

Dodatni eksperimenti takođe pokazuju granice naknadne korekcije. Naknadno preciziranje klase pokazalo je pozitivan DEV neto signal, ali ograničen po obimu, asimetričan i zavisan od dodatnog modelskog dokaza. Zbog toga nije uključen u konačni sistem niti predstavlja rešenje za širi semantički problem. Morfološko prečišćavanje takođe nije proizvelo značajan downstream efekat. Ovi rezultati ukazuju da deo preostalih grešaka zahteva bogatiji semantički kontekst ili promenu samog modelskog odlučivanja, a ne jednostavne završne korekcije.

BC-SMM koristi informaciju o položaju i susedstvu pločica kako bi unapred ograničio koje lokalne predikcije imaju smisla razmatrati za spajanje. To predstavlja prednost kada struktura kontakata između pločica dobro opisuje nastanak redundantnih lokalnih opažanja, ali može postati ograničenje kod veoma izduženih objekata koji se prostiru kroz duži niz pločica. Takve instance mogu zahtevati drugačiju topološku formulaciju koja eksplicitnije modeluje kontinuitet kroz više lokalnih segmenata.

Trenutni sistem je zasnovan na nezavisnoj obradi pojedinačnih slika i ne koristi temporalnu koherenciju između uzastopnih UAV snimaka. Takođe, precizne maske ostaju izražene u pikselskom koordinatnom sistemu i nije uvedena eksplicitna geometrijska kalibracija kojom bi se njihove dimenzije i površine direktno preslikale u fizičke mere.

Važno je naglasiti da redundansa analizirana u ovom radu ne potiče iz duplikata jednog globalnog modelskog prolaza ili izostanka cross-tile NMS-a, već iz nezavisnih opažanja iste fizičke instance u više pločica. RF-DETR ONNX graf nema ugrađeni NMS, dok

YOLO26 ONNX graf takođe nema ugrađenu `NonMaxSuppression` operaciju, ali korišćeni Roboflow runtime primenjuje nefuzionisani spoljašnji NMS. U oba slučaja, globalna konsolidacija rešava redundansu koja nastaje zbog obrade slike po pločicama, a ne zbog jedne globalne inferencije.

Ova ograničenja ne menjaju osnovne zaključke istraživanja, ali određuju granice u kojima su oni trenutno eksperimentalno podržani i ukazuju na pravce u kojima je potrebno dalje razvijati metodologiju.

5.8 Pravci budućeg razvoja

Budući razvoj može biti usmeren na dodatno povećanje adaptivnosti globalne konsolidacije. Iako je ista osnovna BC-SMM konfiguracija primenjena na iSAID i BONAI skupovima, adaptivno podešavanje pragova i automatska kalibracija granica i boundary kriterijuma mogli bi dodatno prilagoditi postupak specifičnoj veličini, gustini i geometriji objekata. Dodatni pravac predstavlja naučena verifikacija spajanja, gde bi sekundarni model procenjivao da li predložene lokalne predikcije zaista pripadaju istoj fizičkoj instanci, umesto oslanjanja isključivo na unapred definisana pravila.

Dobijeni rezultati po klasama dodatno motivišu razvoj klasno zavisne i geometrijski adaptivne konsolidacije. Posebno bi trebalo ispitati pravila prilagođena velikim objektima koji prelaze više granica između susednih pločica, korišćenjem položaja u odnosu na granicu pločica, veličine instance, oblika, prostorne kompatibilnosti susednih predikcija i broja nezavisnih posmatranja iz različitih pločica. Cilj takvog pristupa bio bi da se zadrži ostvareno povećanje odziva, uz smanjenje dodatnih FP kandidata kod velikih prostornih struktura.

Selektivni oporavak može se dalje unaprediti boljom procenom nesigurnosti, kalibracijom izlaznih skorova i efikasnijim filtriranjem false-positive kandidata. Posebno je interesantno proširenje postojećeg cross-model pristupa korišćenjem saglasnosti i neslaganja između više nezavisnih segmentacionih modela kao dodatnog signala pouzdanosti. Slično tome, post-hoc semantička korekcija može se ponovo razmotriti samo za unapred identifikovane visokorizične klasne konfuzije i uz reprezentativniji skup takvih primera.

Važan pravac predstavlja i proširenje topologije za veoma izdužene instance koje prolaze kroz više uzastopnih pločica. To je posebno relevantno za infrastrukturne defekte poput pukotina, kod kojih lokalni mask-overlap može biti veoma mali iako parcijalne maske pripadaju istom objektu. Buduće verzije mogle bi zato uključiti kontinuitet linije ili konture, dominantnu orijentaciju i geometrijsku konzistentnost kroz niz susednih tile-ova.

Za UAV primene dodatno je zanimljivo uvođenje temporalne informacije između uzastopnih snimaka, kao i geometrijska rekonstrukcija kojom bi se precizne maske preslikavale iz pikselskog u fizički prostor. Uz poznatu geometriju snimanja, položaj kamere i odgovarajuću prostornu skalu, to bi omogućilo izražavanje dužine, površine i drugih karakteristika segmentisanih objekata ili oštećenja u centimetrima i metrima, čime bi se segmentacija po-

vezala sa direktnom kvantitativnom analizom.

Konačno, nakon početne evidencije o prenosivosti dobijene na iSAID i BONAI skupovima i dve različite segmentacione arhitekture, naredni korak predstavlja proširenje evaluacije na dodatne domene visokorezolucione segmentacije. Posebno su relevantni infrastrukturna inspekcija, medicinsko i mikroskopsko snimanje, kao i drugi satelitski i vazdušni skupovi, gde se isti problem lokalne obrade i naknadne rekonstrukcije globalne instance javlja u drugačijim geometrijskim i semantičkim uslovima.

6. ZAKLJUČAK

Ova disertacija bavila se problemom precizne segmentacije instanci na vazdušnim snimcima visoke rezolucije korišćenjem tiled inferencije i globalne konsolidacije maski. Osnovni izazov proizlazi iz činjenice da obrada slika velikih dimenzija često zahteva njihovu podelu na manje pločice, čime se čuva lokalna rezolucija i olakšava detekcija malih objekata, ali se istovremeno uvode presecanje objekata granicama pločica, višestruka opažanja iste fizičke instance, fragmentacija i redundantne predikcije.

Rezultati pokazuju da preklapanje pločica povećava dostupnost lokalnog signala, ali samo po sebi ne rešava problem konačne organizacije predikcija. Bez globalne konsolidacije dodatna opažanja lako prelaze u redundantne i fragmentisane maske. SMM zato predstavlja važan referentni korak: pokazuje potrebu da se tiled izlaz obrađuje globalno, kao skup međusobno povezanih lokalnih opažanja, a ne samo kao nezavisna lista predikcija.

Centralni algoritamski doprinos rada predstavlja BC-SMM. Za razliku od pristupa koji se oslanja samo na prostornu blizinu maski, BC-SMM koristi eksplicitnu geometriju tiled prostora: poreklo predikcija iz pločica, topološke odnose susednih pločica i fizički opravdane kontakte na granicama. Time se kandidatni prostor smanjuje sa 626 437 na 286 745 parova, dok broj kompatibilnih veza ostaje gotovo isti. U sprovedenim eksperimentima to je dovelo do rasta precision, F1 i AP mera bez relevantnog pada recall-a, uz manji trošak group-formation faze; medijana vremena po slici smanjena je sa 0,635 s na 0,483 s.

Drugi doprinos čine tile-boundary metrike, uvedene da bi se izdvojile greške koje standardni AP ne razdvaja direktno. One odvojeno mere fragmentaciju, višak fragmenata, geometrijsku rekonstrukciju i over-merging na instancama koje dodiruju ili presecaju granice pločica. Najjasniji efekat BC-SMM-a zabeležen je upravo na identitetu instance: TBFR je u poređenju sa SMM-om smanjen sa 0,0866 na 0,0356, dok za promenu geometrijske pokrivenosti i merge rate-a nije detektovana pouzdana promena.

Treći doprinos odnosi se na selektivni oporavak predikcija niže pouzdanosti. Eksperimenti pokazuju da sirovi skor pouzdanosti nije dovoljan za rangiranje takvih kandidata, dok lokalno-geometrijski, overlap/seam i cross-model signali daju dodatnu evidenciju. PR-AUC raste približno sa 0,23 na 0,45, a finalna konfiguracija ostaje izrazito selektivna: vraća 124 tačne predikcije uz 35 dodatnih lažno pozitivnih. Time rescue ne menja osnovni karakter sistema, već dodaje ograničen recall-oriented signal uz kontrolisan FP trošak.

Dodatna evaluacija na BONAI skupu sa YOLO26 Large pruža početnu evidenciju da se isti metodološki princip može preneti na drugu segmentacionu arhitekturu i drugi skup podataka, ali ne predstavlja dokaz univerzalne generalizacije. Naknadno preciziranje klase pokazalo je pozitivan, ali ograničen DEV signal za usko definisanu klasnu konfuziju, nedovoljan da opravda dodatnu složenost finalnog sistema. Glavna ograničenja ostaju zavisnost

od kvaliteta lokalnih predikcija, delimičan domet rescue grane i potreba za zasebnom validacijom u znatno drugačijim domenima, uključujući potencijalne UAV inspekcijske primene.

Praktični značaj ovakvog okvira najizraženiji je u domenima u kojima se veliki snimci moraju obrađivati bez gubitka lokalnih detalja. To uključuje daljinsku detekciju, mapiranje urbanih scena i potencijalno UAV inspekciju infrastrukture, ali primene izvan evaluiranih skupova zahtevaju zasebnu eksperimentalnu proveru. Time se doprinos rada pre svega odnosi na metodološki način organizovanja lokalnih predikcija u globalno konzistentne instance.

Budući rad prirodno proizlazi iz ovih ograničenja: adaptivno podešavanje BC-SMM pragova i naučena verifikacija spajanja, bogatiji modeli nesigurnosti za rescue, kao i topološka pravila za veoma izdužene instance i domene u kojima je potrebna fizička kalibracija maski. Takva proširenja treba da zadrže osnovnu ideju rada, ali da je prilagode različitim geometrijama objekata, gustinama scena i zahtevima primene.

Zaključno, inferencija nad pločicama nije samo tehnička dekompozicija velike slike. Preklapanje pločica stvara dodatnu prostornu informaciju, a kvalitet konačnog sistema zavisi od načina na koji se ta informacija globalno objedinjuje. BC-SMM koristi geometriju prostora pločica za konsolidaciju lokalnih maski, tile-boundary metrike mere karakteristične greške tog procesa, dok selektivni oporavak koristi dodatne prostorne informacije i signal drugog modela za oprezno vraćanje prethodno odbačenih kandidata. Zajedno, ovi elementi čine zaokružen metodološki okvir za preciznu, merljivu i interpretabilnu segmentaciju instanci na slikama visoke rezolucije.

Spisak slika

1.1 Primer karakterističnih izazova na visokorezolucionom iSAID snimku: (a) veoma male instance, (b) velika lokalna gustina objekata, (c) objekat velikih prostornih dimenzija i (d) instanca koja preseca unutrašnju granicu pločice. 4

2.1 Opšta struktura konvolucione neuronske mreže i hijerarhijska ekstrakcija vizuelnih osobina. Sa povećanjem dubine mreže prostorna rezolucija mapa osobina se smanjuje, dok se broj kanala i nivo apstrakcije reprezentacije povećavaju. 8

2.2 Povećavanje teorijskog receptivnog polja kroz dubinu mreže i koncentracija efektivnog uticaja oko centralnog regiona. Slaganjem lokalnih konvolucionih operacija receptivno polje raste, dok efektivni doprinos ulaznih piksela ostaje izraženiji u centralnom delu teorijskog receptivnog polja. 10

2.3 Poređenje lokalne konvolucione obrade i sadržajno adaptivnog povezivanja vizuelnih tokena pomoću mehanizma samopažnje. Kod CNN-a receptivno polje raste kroz dubinu mreže, dok Transformer omogućava neposredno modelovanje odnosa između prostorno udaljenih tokena. 11

2.4 DETR formulacija detekcije kao direktne predikcije skupa objekata uz naučive upite i bipartitno uparivanje jedan-prema-jedan. Transformer dekoder generiše fiksan skup predikcija, dok se tokom obučavanja Hungarian algoritmom svaka anotirana instanca uparuje sa najviše jednom predikcijom, a preostali upiti označavaju kao “bez objekta”. 12

2.5 Koncept selektivnog oporavka predikcija niže pouzdanosti. Predikcije iznad osnovnog praga pouzdanosti zadržavaju se neposredno, dok se kandidati niže pouzdanosti dodatno procenjuju korišćenjem geometrijskih karakteristika, prostornog konteksta između susednih pločica i nezavisne modelske evidencije. Sekundarni postupak prihvata samo podskup kandidata za koje postoji dovoljno dodatnih pokazatelja validnosti, dok se preostali kandidati odbacuju. 22

3.1 Poređenje dve arhitektonske paradigme segmentacije instanci korišćene u eksperimentima. RF-DETR XLarge koristi transformersku obradu sa naučivim upitima objekata i direktnim formiranjem skupa predikcija, dok YOLO26 Large koristi višeskalnu konvolucionu obradu i agregaciju mapa osobina. Oba modela kao završni izlaz generišu klasu, granični pravougaonik i segmentacionu masku instance. 26
3.2 Ilustrativno poređenje domena skupova iSAID i BONAI. iSAID predstavlja višeklasni domen sa velikim brojem malih, gusto raspoređenih i proizvoljno orijentisanih objekata različitih veličina, dok BONAI u ovom istraživanju predstavlja jednoklasni domen segmentacije vidljivih krovova, sa većim instancama, složenijim konturama i izraženom perspektivnom geometrijom. . 29
3.3 Dekompozicija visokorezolucione slike na preklopljene pločice tokom inferencije nad pločicama. Prikazani su fizički obuhvat pločice $T = 512$ piksela, preklapanje $o = 41$ piksel za $r = 0{,}08$ i korak pomeranja $S = 471$ piksel. Uvećani prikaz ilustruje slučaj objekta koji prelazi granicu dve susedne pločice i zato može biti nezavisno posmatran iz oba lokalna regiona. 31
3.4 Rekonstrukcija lokalnih predikcija u koordinatnom sistemu originalne slike. Predikcije dobijene iz više preklopljenih pločica preslikavaju se iz lokalnih koordinata u jedinstven globalni koordinatni sistem korišćenjem poznatih koordinatnih ofseta pločica. Time predikcije koje potiču iz različitih lokalnih pogleda istog fizičkog regiona zauzimaju odgovarajuće položaje u originalnoj slici. 33
3.5 Integrisani pipeline za segmentaciju instanci na snimcima visoke rezolucije. Originalna slika deli se na pločice; u glavnim overlap konfiguracijama susedne pločice se delimično preklapaju. Nakon lokalne inferencije predikcije se rekonstruišu u globalnom koordinatnom sistemu. Predikcije više pouzdanosti zadržavaju se u glavnoj grani, dok kandidati niže pouzdanosti prolaze kroz sekundarnu procenu. Prihvaćeni kandidati se zatim objedinjuju, globalno konsoliduju i njihove maske se spajaju u konačne instance na nivou originalne slike. 35

3.6 Grafički prikaz proširene reprezentacije kandidata za selektivni oporavak predikcija niže pouzdanosti: (a) položaj kandidata u odnosu na unutrašnju granicu pločice i zonu preklapanja, uključujući udaljenost od granice i udeo maske i obuhvatnog pravougaonika u preklopnom regionu; (b) geometrijska i modelska podrška predikcije iz susedne pločice, opisana preklapanjem maski, rastojanjem centroida, odnosom površina i skorovima pouzdanosti; (c) simetrična prostorna podrška sa odgovarajuće strane iste fizičke granice između susednih pločica; (d) višestruka cross-tile podrška, odnosno prisustvo odgovarajućih lokalnih posmatranja u više susednih pločica; (e) karakteristike oblika maske, uključujući površinu, obim, kompaktnost, solidity, ekscentricitet i broj komponenti; (f) lokalni kontekst na nivou pločice, uključujući rang i percentil confidence vrednosti, broj predikcija, njihovu gustinu i lokalnu raspodelu skorova. 38

3.7 Grafički prikaz BC-SMM postupka: (a) globalno rekonstruisane lokalne predikcije sa redundantnim posmatranjima istih fizičkih instanci; (b) ograničavanje kandidatne pretrage korišćenjem topologije pločica i kontaktnih koridora; (c) formiranje kandidatnih parova u prostornom dometu ρ; (d) procena kompatibilnosti parova na osnovu rastojanja granica, preklapanja maski i težine veze, uz lokalnu kontaktnu validaciju marginalnih parova; (e) topološki ograničeno grupisanje predikcija iz susednih pločica; (f) izlazna particija $\mathcal{P}^*$, čije se grupe u narednoj fazi prosleđuju funkciji fuzije maski. 44

3.8 Odnos između BC-SMM grupisanja i naknadne fuzije maski: (a) više lokalnih predikcija iste fizičke instance nakon globalne rekonstrukcije; (b) iste predikcije povezane u jednu BC-SMM grupu, bez promene njihove geometrije; (c) konačna maska formirana fuzijom članova grupe. 50

3.9 Vizuelna interpretacija predloženih tile-boundary metrika: (a) TBMR meri propuštene instance koje presecaju granicu pločica; (b) TBFR registruje pojavu fragmentacije jedne fizičke instance; (c) TBEF meri višak fragmenata nastao takvom fragmentacijom; (d) TBDR meri redundantne predikcije iste fizičke instance; (e) $\mathrm{RIoU}_{\mathrm{cross}}$ opisuje kvalitet prostorne rekonstrukcije instance koja preseca granicu; (f) MR registruje slučajeve prekomernog spajanja različitih fizičkih instanci. 53

4.1 Redukcija kandidatnog prostora primenom BC-SMM pristupa u odnosu na referentni SMM. Broj kandidatnih parova smanjuje se sa 626 437 na 286 745, odnosno za približno 54,2%, dok broj identifikovanih kompatibilnih parova ostaje gotovo nepromenjen (43 193 naspram 43 029). 73

4.2 Primeri karakterističnih događaja na granicama pločica: (a) propuštena instanca koja preseca granicu pločice, bez uparene predikcije; (b) fragmentisana instanca predstavljena sa više lokalnih fragmenata; (c) redundantne predikcije iste fizičke instance; (d) pravilno rekonstruisana instanca koja preseca granicu pločice, sa visokim $\mathrm{RIoU}_{\mathrm{cross}}$. 76
4.3 Precision–recall krive za tri reprezentativna nivoa opisa kandidata za selektivni oporavak. Proširivanje osnovnog confidence signala prostornim karakteristikama i nezavisnim cross-model signalom povećava PR-AUC sa 23,01% na 40,11% i 45,05%. 82
4.4 Promena odziva po iSAID klasama za konačnu B3 konfiguraciju u odnosu na referentni SMM rezultat pri istoj tiled geometriji sa 8% preklapanja. Odziv raste u 10 od 15 klasa, dok u preostalih pet ostaje nepromenjen; pad nije zabeležen ni u jednoj klasi. 87
4.5 Primer uticaja preklapanja pločica i naknadne konsolidacije. (a) originalni isečak slike, (b) inferencija nad pločicama bez konsolidacije i sa 0% overlap-a, (c) inferencija nad pločicama bez konsolidacije i sa 8% overlap-a, (d) konačni BC-SMM rezultat. Uvođenje overlap-a povećava broj lokalnih opažanja, dok globalna konsolidacija uklanja nastalu redundansu. 89
4.6 Primer redukcije redundantnih predikcija. (a) originalni isečak slike, (b) 8% overlap bez konsolidacije, (c) SMM i (d) BC-SMM. Uočava se da nekonsolidovani tiled izlaz sadrži više redundantnih i delimično preklopljenih instanci, dok globalna konsolidacija daje kompaktniji i precizniji konačni skup. . . . 90
4.7 Vizuelno poređenje originalnog SMM i BC-SMM pristupa. (a) originalni isečak slike, (b) 8% overlap bez konsolidacije, (c) SMM i (d) BC-SMM. BC-SMM daje dodatno prečišćen i konzistentniji konačni raspored instanci u odnosu na SMM. 90
4.8 Primer uspešnog selektivnog oporavka predikcije. (a) originalni isečak slike, (b) BC-SMM rezultat, (c) prihvaćeni rescue kandidat izdvojen iz low-confidence prostora i (d) konačni BC-SMM + rescue rezultat. Dodatna instanca se pojavljuje samo nakon primene rescue grane, čime se vizuelno potvrđuje komplementarna uloga selektivnog oporavka u odnosu na globalnu konsolidaciju. 91
4.9 Primer ograničenja metode u veoma gustoj sceni. (a) originalni isečak slike, (b) 8% overlap bez konsolidacije, (c) SMM i (d) BC-SMM. Iako konsolidacija i ovde stabilizuje rezultat, ekstremna gustina i složenost scene ograničavaju obim lokalnog poboljšanja. 91
4.10 Reprezentativan BONAI primer pre i nakon globalne konsolidacije. 92
4.11 Primer globalne konsolidacije većih krovnih instanci koje se pojavljuju u više pločica. 92

4.12 Primer globalne konsolidacije u gustom urbanom regionu. 93
4.13 Primer ograničenja postupka na zahtevnoj BONAI sceni. 93
4.14 ECDF vremena formiranja grupa po slici. BC-SMM kriva je kroz najveći deo distribucije pozicionirana ulevo od SMM krive, što pokazuje kraće vreme obrade za isti kumulativni udeo slika. 95
4.15 Boxplot vremena formiranja grupa po slici na logaritamskoj vremenskoj osi. BC-SMM ima nižu medijanu, uži centralni opseg i znatno manje ekstremne vrednosti od SMM pristupa. 96

5.1 Primer visokorezolucionog UAV snimka mosta sa segmentiranim površinskim oštećenjima i drugim lokalnim nepravilnostima. Primer ilustruje scenario u kome se veliki infrastrukturni objekat obrađuje kroz više lokalnih regiona, dok pojedina oštećenja mogu biti izdužena i prostirati se preko granica susednih pločica, što može zahtevati naknadnu globalnu konsolidaciju lokalnih predikcija. 107

Spisak tabela

3.1 Osnovne karakteristike modela korišćenih u eksperimentima. 26
3.2 ONNX i runtime informacije relevantne za dekodiranje lokalnih predikcija. 27
3.3 Osnovne karakteristike skupa iSAID i njegova zvanična podela. 28
3.4 Osnovne karakteristike skupa BONAI i njegova zvanična podela. 29
3.5 Finalni hiperparametri XGBoost klasifikatora za selektivni oporavak. . . . 40
3.6 Parametri i fiksne konstante finalnog BC-SMM postupka. 48
3.7 Uloge specijalizovanih metrika za evaluaciju tiled inferencije. 58

4.1 Pregled glavnih eksperimentalnih konfiguracija. 62
4.2 Poređenje tiled inferencije bez sistematskog preklapanja i sa 8% preklapanja, pre globalne konsolidacije. 64
4.3 Promena strukture false-positive predikcija nakon uvođenja 8% overlap-a. . 65
4.4 Promena strukture false-negative slučajeva nakon uvođenja 8% overlap-a. . 65
4.5 Promena precision i recall vrednosti za reprezentativne klase. 66
4.6 Poređenje 8% tiled inferencije bez konsolidacije i nakon SMM konsolidacije. 67
4.7 Poređenje strukture SMM particija za 0% i 8% overlap-a. 68
4.8 Raspodela veličina SMM grupa. 68
4.9 Poređenje konačnih SMM rezultata pri 0% i 8% overlap-a. 70
4.10 Poređenje SMM i BC-SMM globalne konsolidacije pri 8% overlap-u. . . . 71
4.11 Promena raspodele veličina grupa između SMM i BC-SMM. 72
4.12 Struktura kandidatnog prostora SMM i BC-SMM. 73
4.13 Klase sa najizraženijom promenom F1 vrednosti između Original SMM i BC-SMM. 74
4.14 Tile-boundary metrike za tiled inferenciju bez preklapanja pločica. Strelice označavaju poželjan smer promene, a podebljane vrednosti predstavljaju numerički najbolji rezultat po pojedinačnoj metrici. Kontrolne metrike TBMR i MR ne tumače se izolovano. 75
4.15 Apsolutni brojevi crossing instanci za režim bez preklapanja. 76
4.16 Tile-boundary metrike za tiled inferenciju sa 8% preklapanja pločica. Strelice označavaju poželjan smer promene, a podebljane vrednosti predstavljaju numerički najbolji rezultat po pojedinačnoj metrici. Kontrolne metrike TBMR i MR ne tumače se izolovano. 77

4.17 Apsolutni brojevi crossing instanci i fragmenata pri 8% preklapanju. 77
4.18 Standardne metrike segmentacije za konfiguracije sa 8% preklapanja. . . . 78
4.19 Uparene bootstrap razlike tile-boundary metrika pri 8% preklapanju pločica. Prikazani su point estimate razlike i 95% percentile bootstrap intervali na osnovu 1000 iteracija. 80
4.20 Udeo pozitivnih rescue kandidata u različitim confidence intervalima. . . . 81
4.21 Progresija kvaliteta reprezentacije low-confidence kandidata na TRAIN izvedenom DEV skupu za interval $[0{,}40, 0{,}50)$. 82
4.22 Razvojna end-to-end evaluacija selektivnog oporavka nakon BC-SMM konsolidacije. 83
4.23 Konačna nezavisna evaluacija selektivnog oporavka na zvaničnom iSAID validation skupu. 83
4.24 Početne lokalne performanse RF-DETR XLarge i YOLO26 Large po klasama na iSAID validation pločicama. 85
4.25 AP prema veličini objekta na iSAID validation skupu. 85
4.26 F1 mera za odabrane iSAID klase. 86
4.27 Računska cena RF-DETR tiled inferencije na iSAID validation skupu. . . . 94
4.28 Vreme formiranja grupa po slici za SMM i BC-SMM. 94
4.29 Algoritamski workload SMM i BC-SMM. 94
4.30 Ukupno proteklo vreme kompletne faze fuzije i statistike vremena obrade pojedinačnih slika nakon formiranja grupa. 95
4.31 Razvojna putanja algoritama globalne konsolidacije pri 8% overlap-a. . . . 97
4.32 Uticaj morfološkog prečišćavanja nakon BC-SMM konsolidacije. 98
4.33 Rezultat post-hoc semantičke korekcije Small/Large Vehicle na zamrznutom DEV skupu. 99
4.34 Poređenje informacionog sadržaja različitih rescue reprezentacija. 100
4.35 Početne performanse lokalnih modela i konačni rezultat kompletnog pipeline-a nad originalnim slikama visoke rezolucije. 101
4.36 Deskriptivni pregled lokalnih segmentacionih modela i kompletnog B3 sistema na skupu BONAI. 102

SPISAK AKRONIMA

AP	Average Precision / prosečna preciznost
AR	Average Recall / prosečni odziv
ASHA	Asynchronous Successive Halving Algorithm / asinhroni algoritam sukcesivnog prepolovljavanja
BC-SMM	Boundary-Constrained Spatial Mask Merging / prostorno spajanje maski ograničeno granicama pločica
BONAI	Buildings in Off-Nadir Aerial Images / skup vazdušnih snimaka zgrada snimljenih pod kosim uglom
CLAHE	Contrast-Limited Adaptive Histogram Equalization / ekvalizacija histograma sa ograničenjem kontrasta
CNN	Convolutional Neural Network / konvoluciona neuronska mreža
COCO	Common Objects in Context / skup podataka i evaluacioni protokol za objekte u kontekstu
DETR	Detection Transformer / transformerski model za detekciju objekata
DEV	Development Set / razvojni skup
F_1	F1 score / harmonijska sredina preciznosti i odziva
FN	False Negative / lažno negativan slučaj
FP	False Positive / lažno pozitivan slučaj
GT	Ground Truth / referentna anotacija
IoU	Intersection over Union / presek nad unijom
iSAID	Instance Segmentation in Aerial Images Dataset / podaci za segmentaciju instanci na vazdušnim snimcima
ML	Machine Learning / mašinsko učenje
MR	Merge Rate / stopa pogrešnog spajanja
NMM	Non-Maximum Merging / spajanje umesto potiskivanja preklopljenih kandidata
NMS	Non-Maximum Suppression / potiskivanje nemaksimalnih kandidata
ONNX	Open Neural Network Exchange / otvoreni format za razmenu neuronskih mreža
PR-AUC	Precision–Recall Area Under the Curve / površina ispod krive preciznost–odziv
RF-DETR	Roboflow Detection Transformer / Roboflow transformerski model za detekciju
RGB	Red–Green–Blue / crvena–zelena–plava reprezentacija boje
RIoU	Reconstruction IoU / rekonstrukcioni presek nad unijom
ROC-AUC	Receiver Operating Characteristic Area Under the Curve / površina ispod ROC krive
SMM	Spatial Mask Merging / prostorno spajanje maski
TBDR	Tile-Boundary Duplicate Rate / stopa duplikata na granicama pločica
TBEF	Tile-Boundary Excess Fragmentation / višak fragmentacije na granicama pločica
TBFR	Tile-Boundary Fragmentation Rate / stopa fragmentacije na granicama pločica
TBMR	Tile-Boundary Miss Rate / stopa propuštanja na granicama pločica
TP	True Positive / tačno pozitivan slučaj
TPE	Tree-structured Parzen Estimator / stablasto strukturisani Parzenov estimator
UAV	Unmanned Aerial Vehicle / bespilotna letelica
XGBoost	Extreme Gradient Boosting / ekstremno gradijentno pojačavanje
YOLO	You Only Look Once / porodica jednostepenih detekcionih modela

SPISAK ENGLESKIH TERMINA

Engleski termin	Značenje u radu
confidence	pouzdanost predikcije, odnosno skor pouzdanosti
cross-model	zasnovano na signalu drugog modela
cross-tile/seam	između susednih pločica i duž njihove zajedničke granice
feature	karakteristika korišćena kao ulaz modela ili analize
feature space	prostor karakteristika
ground truth	referentna anotacija
hard-negative mining	izdvajanje teških negativnih primera
high-confidence	visoka pouzdanost predikcije
low-confidence	niska pouzdanost predikcije
NMS-free	bez NMS postprocesiranja u užem lokalnom izlazu modela
one-to-many	dodeljivanje ili signal jedan-prema-više tokom obučavanja
one-to-one	dodeljivanje ili uparivanje jedan-prema-jedan
overlap	preklapanje pločica
pipeline	cevovod ili tok obrade
query-based	zasnovan na upitima
rescue	selektivni oporavak predikcija niže pouzdanosti
self-attention / cross-attention	samopažnja / unakrsna pažnja
tile	pločica, lokalni isečak originalne slike
tile-boundary	vezan za granice između pločica
tiled inference	inferencija nad pločicama

LITERATURA

[1] Nicolas Audebert i dr. „Deep learning for urban remote sensing”. U: *2017 Joint Urban Remote Sensing Event (JURSE)*. IEEE. 2017., str. 1–4.

[2] Aisyah Marliza Muhmad Kamarulzaman i dr. „UAV implementations in urban planning and related sectors of rapidly developing nations: A review and future perspectives for Malaysia”. U: *Remote Sensing* 15.11 (2023.), str. 2845.

[3] Syed Waqas Zamir i dr. „iSAID: A Large-Scale Dataset for Instance Segmentation in Aerial Images”. U: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2019., str. 28–37.

[4] Qi Li i dr. „Memory-constrained semantic segmentation for ultra-high resolution UAV imagery”. U: *IEEE Robotics and Automation Letters* 9.2 (2024.), str. 1708–1715.

[5] Arian Bakhtiarnia, Qi Zhang i Alexandros Iosifidis. „Efficient high-resolution deep learning: A survey”. U: *ACM Computing Surveys* 56.7 (2024.), str. 1–35.

[6] Fatih Cagatay Akyon, Sinan Onur Altinuc i Alptekin Temizel. „Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection”. U: *Proceedings of the IEEE International Conference on Image Processing*. IEEE, 2022., str. 966–970. DOI: 10.1109/ICIP46576.2022.9897990.

[7] G. Anthony Reina i dr. „Systematic Evaluation of Image Tiling Adverse Effects on Deep Learning Semantic Segmentation”. U: *Frontiers in Neuroscience* 14 (2020.), str. 65. DOI: 10.3389/fnins.2020.00065.

[8] Marko Mihajlovic i Marina Marjanovic. „Enhancing Instance Segmentation in High-Resolution Images Using Slicing-Aided Hyper Inference and Spatial Mask Merging Optimized via R-Tree Indexing”. U: *Mathematics* 13.19 (2025.). ISSN: 2227-7390. DOI: 10.3390/math13193079. URL: https://www.mdpi.com/2227-7390/13/19/3079.

[9] Tsung-Yi Lin i dr. „Feature pyramid networks for object detection”. U: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017., str. 2117–2125.

[10] Jingdong Wang i dr. „Deep high-resolution representation learning for visual recognition”. U: *IEEE transactions on pattern analysis and machine intelligence* 43.10 (2020.), str. 3349–3364.

[11] Yann LeCun i dr. „Gradient-Based Learning Applied to Document Recognition”. U: *Proceedings of the IEEE* 86.11 (1998.), str. 2278–2324. DOI: 10.1109/5.726791.

[12] Alex Krizhevsky, Ilya Sutskever i Geoffrey E. Hinton. „ImageNet Classification with Deep Convolutional Neural Networks". U: *Advances in Neural Information Processing Systems*. Sv. 25. Curran Associates, 2012., str. 1097–1105.

[13] Yann LeCun, Yoshua Bengio i Geoffrey Hinton. „Deep Learning". U: *Nature* 521.7553 (2015.), str. 436–444. DOI: 10.1038/nature14539.

[14] Ian Goodfellow, Yoshua Bengio i Aaron Courville. *Deep Learning*. Cambridge, Massachusetts: MIT Press, 2016. ISBN: 978-0-262-03561-3.

[15] Zewen Li i dr. „A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects". U: *IEEE Transactions on Neural Networks and Learning Systems* 33.12 (2022.), str. 6999–7019. DOI: 10.1109/TNNLS.2021.3084827.

[16] Wenjie Luo i dr. „Understanding the Effective Receptive Field in Deep Convolutional Neural Networks". U: *Advances in Neural Information Processing Systems*. Sv. 29. arXiv version published in 2017 as arXiv:1701.04128. Curran Associates, 2016., str. 4898–4906.

[17] Taco S. Cohen i Max Welling. „Group Equivariant Convolutional Networks". U: *Proceedings of the 33rd International Conference on Machine Learning*. Sv. 48. Proceedings of Machine Learning Research. PMLR, 2016., str. 2990–2999.

[18] Ashish Vaswani i dr. „Attention Is All You Need". U: *Advances in Neural Information Processing Systems*. Sv. 30. Curran Associates, 2017., str. 5998–6008.

[19] Alexey Dosovitskiy i dr. „An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale". U: *International Conference on Learning Representations*. 2021.

[20] Nicolas Carion i dr. „End-to-End Object Detection with Transformers". U: *Computer Vision – ECCV 2020*. Sv. 12346. Lecture Notes in Computer Science. Springer, 2020., str. 213–229. DOI: 10.1007/978-3-030-58452-8_13.

[21] Xizhou Zhu i dr. „Deformable DETR: Deformable Transformers for End-to-End Object Detection". U: *International Conference on Learning Representations*. 2021.

[22] Bowen Cheng i dr. „Masked-Attention Mask Transformer for Universal Image Segmentation". U: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022., str. 1290–1299. DOI: 10.1109/CVPR52688.2022.00135.

[23] Feng Li i dr. „Mask DINO: Towards a Unified Transformer-Based Framework for Object Detection and Segmentation". U: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023., str. 3041–3050. DOI: 10.1109/CVPR52729.2023.00297.

[24] Ze Liu i dr. „Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows”. U: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021., str. 10012–10022. DOI: 10.1109/ICCV48922.2021.00986.

[25] Jan Hosang, Rodrigo Benenson i Bernt Schiele. „Learning Non-Maximum Suppression”. U: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017., str. 4507–4515.

[26] Navaneeth Bodla i dr. „Soft-NMS—Improving Object Detection with One Line of Code”. U: *Proceedings of the IEEE International Conference on Computer Vision*. 2017., str. 5561–5569.

[27] Jinglu Chu i dr. „Syncretic-NMS: A Merging Non-Maximum Suppression Algorithm for Instance Segmentation”. U: *IEEE Access* 8 (2020.), str. 114705–114714. DOI: 10.1109/ACCESS.2020.3003917.

[28] Marko Mihajlovic i Marina Marjanovic. „Medical Segmentation of Kidney Whole Slide Images Using Slicing Aided Hyper Inference and Enhanced Syncretic Mask Merging Optimized by Particle Swarm Metaheuristics”. U: *BioMedInformatics* 5.3 (2025.), str. 44. DOI: 10.3390/biomedinformatics5030044.

[29] Peize Sun i dr. „What Makes for End-to-End Object Detection?” U: *Proceedings of the 38th International Conference on Machine Learning*. Sv. 139. Proceedings of Machine Learning Research. PMLR, 2021., str. 9934–9944.

[30] Jianfeng Wang i dr. „End-to-End Object Detection with Fully Convolutional Network”. U: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021., str. 15844–15853. DOI: 10.1109/CVPR46437.2021.01559.

[31] Zhuofan Zong, Guanglu Song i Yu Liu. „DETRs with Collaborative Hybrid Assignments Training”. U: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2023., str. 6725–6735. DOI: 10.1109/ICCV51070.2023.00621.

[32] Ao Wang i dr. „YOLOv10: Real-Time End-to-End Object Detection”. U: *Advances in Neural Information Processing Systems*. Sv. 37. 2024.

[33] Kaiming He i dr. „Mask R-CNN”. U: *Proceedings of the IEEE International Conference on Computer Vision*. 2017., str. 2961–2969.

[34] Fatih Cagatay Akyon, Sinan Onur Altinuc i Alptekin Temizel. „Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection”. U: *2022 IEEE International Conference on Image Processing*. IEEE, 2022., str. 966–970. DOI: 10.1109/ICIP46576.2022.9897990.

[35] Nikhil Bansal, Avrim Blum i Shuchi Chawla. „Correlation Clustering”. U: *Machine Learning* 56 (2004.), str. 89–113. DOI: 10.1023/B:MACH.0000033116.57574.95.

[36] Antonin Guttman. „R-Trees: A Dynamic Index Structure for Spatial Searching”. U: *Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data*. ACM, 1984., str. 47–57. DOI: 10.1145/602259.602266.

[37] Marko Mihajlovic i Marina Marjanovic. „Enhancing Instance Segmentation in High-Resolution Images Using Slicing-Aided Hyper Inference and Spatial Mask Merging Optimized via R-Tree Indexing”. U: *Mathematics* 13.19 (2025.), str. 3079. DOI: 10.3390/math13193079.

[38] Chuan Guo i dr. „On Calibration of Modern Neural Networks”. U: *Proceedings of the 34th International Conference on Machine Learning*. Sv. 70. Proceedings of Machine Learning Research. PMLR, 2017., str. 1321–1330.

[39] Fabian Küppers i dr. „Multivariate Confidence Calibration for Object Detection”. U: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020., str. 326–327.

[40] Buddhika Pathiraja i dr. „Multiclass Confidence and Localization Calibration for Object Detection”. U: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

[41] Takaya Saito i Marc Rehmsmeier. „The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets”. U: *PLOS ONE* 10.3 (2015.), e0118432. DOI: 10.1371/journal.pone.0118432.

[42] Kemal Oksuz i dr. „Localization Recall Precision (LRP): A New Performance Metric for Object Detection”. U: *Proceedings of the European Conference on Computer Vision*. 2018., str. 504–519.

[43] Yonatan Geifman i Ran El-Yaniv. „Selective Classification for Deep Neural Networks”. U: *Advances in Neural Information Processing Systems*. Sv. 30. 2017.

[44] Yonatan Geifman i Ran El-Yaniv. „SelectiveNet: A Deep Neural Network with an Integrated Reject Option”. U: *Proceedings of the 36th International Conference on Machine Learning*. 2019., str. 2151–2159.

[45] Matthias Rottmann i dr. „Prediction Error Meta Classification in Semantic Segmentation: Detection via Aggregated Dispersion Measures of Softmax Probabilities”. U: *2020 International Joint Conference on Neural Networks*. IEEE, 2020., str. 1–9. DOI: 10.1109/IJCNN48605.2020.9206659.

[46] Marius Schubert, Karsten Kahl i Matthias Rottmann. „MetaDetect: Uncertainty Quantification and Prediction Quality Estimates for Object Detection”. U: *2021 International Joint Conference on Neural Networks*. IEEE, 2021., str. 1–10. DOI: 10.1109/IJCNN52387.2021.9534289.

[47] Abhinav Shrivastava, Abhinav Gupta i Ross Girshick. „Training Region-Based Object Detectors with Online Hard Example Mining". U: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016., str. 761–769.

[48] Edgar Heinert i dr. „Uncertainty and Prediction Quality Estimation for Semantic Segmentation via Graph Neural Networks". U: *Proceedings of the BMVC 2024 Workshop on Robust Recognition in the Open World*. 2024.

[49] Balaji Lakshminarayanan, Alexander Pritzel i Charles Blundell. „Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles". U: *Advances in Neural Information Processing Systems*. Sv. 30. 2017.

[50] Dimity Miller i dr. „Dropout Sampling for Robust Object Detection in Open-Set Conditions". U: *2018 IEEE International Conference on Robotics and Automation*. 2018.

[51] Dimity Miller i dr. „Evaluating merging strategies for sampling-based uncertainty techniques in object detection". U: *2019 international conference on robotics and automation (icra)*. IEEE. 2019., str. 2348–2354.

[52] Bowen Cheng i dr. „Boundary IoU: Improving Object-Centric Image Segmentation Evaluation". U: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021., str. 15334–15342.

[53] Gui-Song Xia i dr. „DOTA: A Large-Scale Dataset for Object Detection in Aerial Images". U: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018., str. 3974–3983. DOI: 10.1109/CVPR.2018.00418.

[54] Jinwang Wang i dr. „Learning to Extract Building Footprints from Off-Nadir Aerial Images". U: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.4 (2023.), str. 4434–4449. DOI: 10.1109/TPAMI.2022.3186860.

[55] Tsung-Yi Lin i dr. „Microsoft COCO: Common Objects in Context". U: *Computer Vision – ECCV 2014*. Springer International Publishing, 2014., str. 740–755.