

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

Praćenje nogometaša primjenom YOLOv5 modela i DeepSORT algoritma

Marko Tutić

Voditelj: Izv. prof. dr. sc. Zoran Kalafatić

Zagreb, svibanj 2022.

SADRŽAJ

1. Uvod	1
2. Skup podataka	3
3. Arhitektura sustava	6
3.1. Detekcija nogometaša	6
3.2. Praćenje objekata	6
3.2.1. Kalmanov filter	7
3.2.2. Dodjela trajektorija detekcijama	8
3.3. Izrada mape aktivnosti	9
4. Implementacija sustava	12
4.1. Podsustav za detekciju nogometaša	12
4.2. Podsustav za praćenje nogometaša	12
4.3. Podsustav za izradu mape aktivnosti nogometaša	13
4.4. Povezivanje podsustava u cjelinu	14
4.5. Podsustav za vrednovanje praćenja i detekcije objekata	14
5. Analiza rezultata	15
5.1. Vrednovanje modela detekcije nogometaša	15
5.2. Vrednovanje algoritma praćenja objekata	17
6. Zaključak	19
7. Literatura	20

1. Uvod

Praćenje objekata je grana računalnog vida u kojoj se detekcije objekata u određenom okviru nastoje pratiti u nadolazećim okvirima. Metode praćenja objekata se najčešće razvrstavaju u metode praćenja jednog objekta i metode praćenja više objekata. U metode praćenja jednog objekta ubrajaju se metode koje ne zahtijevaju detekciju objekta u svakom okviru videozapisa (engl. *detection-free tracking*). U ovim metodama praćenja dovoljno je u početnom okviru zadati poziciju objekta koji se želi pratiti, a algoritam praćenja objekta zatim nastoji pratiti objekt u daljnjim okvirima na temelju karakteristika objekta i/ili promjena koje se događaju u slici.

Praćenje više objekata se realizira kao detektiranje objekata u nekom okviru i dodjeljivanje identifikatora dobivenim detekcijama. Cilj praćenja više objekata je očuvati identifikatore pojedinih objekata kroz više okvira videozapisa. Ovakva metoda praćenja objekata zahtjeva detekcije objekata u svakom okviru videozapisa (engl. *tracking-by-detection*). Praćenje više objekata često se vrši pod pretpostavkom da objekti neće naglo mijenjati brzinu i smjer kretanja.

Praćenje više objekata predstavlja veliki izazov u području računalnog vida jer se redovito javljaju situacije poput zaklanjanja objekta ili mimoilaženja objekata u kojima može doći do pogrešnog ili propuštenog dodjeljivanja identifikatora objektima. Zaklanjanje objekta je pojava kada objekt nije vidljiv na slici, ali je pozicija objekta i dalje sadržana unutar slike. Mimoilaženje objekata može izazvati zamjenu identiteta te se ta pojava manifestira tako da se identifikator nekog praćenog objekta dodjeli drugom praćenom objektu, a identifikator drugog objekta dodjeljuje se početnom objektu.

Praćenje nogometaša predstavlja situaciju u kojoj je potrebno pratiti više objekata pri čemu su moguće česte zamjene identiteta s obzirom na to da se igrači kroz trajanje cijele utakmice mimoilaze. Dodatan izazov pri praćenju nogometaša predstavlja činjenica da ako je kamera udaljena, igrači izgledaju vrlo slično zbog identičnih boja dresova te ih je stoga teže identificirati na temelju izgleda.

Za praćenje nogometaša na videozapisima izrađen je sustav za praćenje nogometaša koji se sastoji od djela za detekciju nogometaša i djela za praćenje na temelju de-

tekcija. Dio sustava koji je zadužen za detekciju nogometaša koristi YOLOv5 model koji se uči na prikladnom skupu podataka, dok za se praćenje nogometaša koristi DeepSORT algoritam. DeepSORT algoritam koristi značajke koje opisuju izgled objekta te udaljenost detekcija i trajektorija kao kriterij dodjele detekcija trajektorijama.

Pozicije nogometaša u okvirima bilježe se za svakog nogometaša zasebno te se na temelju tih pozicija izrađuje mapa aktivnosti nogometaša. Pozicija nogometaša na terenu tretira se kao slučajna vektorska varijabla koja ima nepoznatu razdiobu pri čemu je mapa aktivnosti procjena funkcije gustoće te varijable. Funkcija gustoće se procjenjuje metodom procjene gustoće jezgre (engl. *kernel density estimation*) pri čemu pozicije nogometaša dobivene analizom videozapisa predstavljaju uzorke pomoću kojih se vrši procjena.

2. Skup podataka

Za potrebe učenja i vrednovanja sustava za praćenje nogometaša korišten je videozapis duljine 5 minuta koji sadrži snimku jedne nogometne utakmice te tekstualna datoteka koja sadrži oznake pozicija i identifikatora nogometaša u svakom okviru tog videozapisa. Kamera kojom je utakmica snimana je statična. Svi okviri u videozapisu sadrže pregled cijelog nogometnog terena. Jedna sekunda navedenog videozapisa sadrži 25 slikovnih okvira i širina okvira videozapisa iznosi 3260 piksela dok visina okvira iznosi 570 piksela. Slika 2.1 prikazuje primjer jednog okvira videozapisa dok slika 2.2 prikazuje primjer oznaka pozicija i identifikatora igrača u tekstualnoj datoteci za prvi okvir videozapisa.

Oznake pozicija igrača u okvirima videozapisa zapisane su u tekstualnoj datoteci u obliku propisanom MOT16 specifikacijom. Jedan redak tekstualne datoteke ima sljedeći format:

<redni broj okvira>, <identifikator igrača>, <x koordinata>, <y koordinata>, <širina>, <visina>, -1, -1, -1, -1.

Koordinate x i y te širina i visina odnose se na pravokutnik koji omeđuje igrača na slici. Koordinate x i y označuju poziciju gornje lijeve točke tog pravokutnika. Na temelju gornje lijeve točke te širine i visine pravokutnika mogu se izvesti i ostale točke omeđujućeg pravokutnika čime je omeđujući pravokutnik u potpunosti definiran. Niz brojeva -1 se ne koristi, no služe kao nadopuna za zadovoljavanje MOT16 formata.

Skup podataka dijeli se na skup za učenje, skup za provjeru i skup za testiranje. Okviri videozapisa se dijele u navedene skupove slijedno kako bi se očuvala mogućnost



Slika 2.1: Primjer okvira videozapisa

```

1,0,841,175,13,35,-1,-1,-1,-1
1,1,1551,405,24,54,-1,-1,-1,-1
1,2,1341,217,15,40,-1,-1,-1,-1
1,3,1374,154,14,34,-1,-1,-1,-1
1,4,1457,103,12,27,-1,-1,-1,-1
1,5,1448,186,13,36,-1,-1,-1,-1
1,6,1466,168,12,33,-1,-1,-1,-1
1,7,1621,269,15,44,-1,-1,-1,-1
1,8,1624,175,16,37,-1,-1,-1,-1
1,9,1613,142,13,28,-1,-1,-1,-1
1,10,1627,105,11,26,-1,-1,-1,-1
1,11,1528,133,16,27,-1,-1,-1,-1
1,12,2429,176,15,34,-1,-1,-1,-1
1,13,2056,195,18,39,-1,-1,-1,-1
1,14,1997,263,16,42,-1,-1,-1,-1
1,15,1964,151,14,33,-1,-1,-1,-1
1,16,1925,137,13,29,-1,-1,-1,-1
1,17,1812,283,16,47,-1,-1,-1,-1
1,18,1846,197,15,37,-1,-1,-1,-1
1,19,1758,200,14,35,-1,-1,-1,-1
1,20,1758,155,12,33,-1,-1,-1,-1
1,21,1756,119,12,29,-1,-1,-1,-1
1,22,1618,223,19,37,-1,-1,-1,-1

```

Slika 2.2: Primjer oznaka skupa podataka

praćenja objekata u navedenim skupovima podataka.

Skup podataka dijeli se tako da skup za učenje sadrži prvih 4500 okvira videozapisa, skup za provjeru sadrži idućih 1500 okvira, dok skup za testiranje sadrži preostalih 1515 okvira.

Skup za učenje koristi se za učenje modela detekcije objekata. Skup za provjeru koristi se za provjeru svojstva generalizacije modela detekcije objekata tijekom različitih epoha učenja. Konačno, skup za testiranje koristi se za vrednovanje modela detektora objekata nakon što je završen proces učenja.

Za potrebe učenja detektora objekata oznake koje su zadane u MOT16 formatu potrebno je pretvoriti u oznake koje zadovoljavaju YOLOv5 format. YOLOv5 format specificira izradu tekstualne datoteke s oznakama za svaki okvir zasebno te se omeđujući pravokutnici u tim datotekama opisuju središtem omeđujućeg pravokutnika i njegovom širinom i visinom. Identifikatori igrača se ne uzimaju u obzir prilikom pretvorbe u YOLOv5 format jer je cilj detekcije objekata lokalizirati objekte i klasificirati ih.

Dodatno je potrebno x koordinatu i širinu omeđujućeg pravokutnika podijeliti sa širinom slike, odnosno y koordinatu i visinu omeđujućeg pravokutnika potrebno je podijeliti s visinom slike. Takva pretvorba je potrebna jer model YOLOv5 na ulazu može primiti slike različitih veličina te se pretvorbom postiže invarijantnost pozicije objekta na visinu i širinu slike.

Iako se u DeepSORT algoritmu za praćenje objekata može naučiti konvolucijska neuronska mreža za izlučivanje značajki izgleda objekta pomoću istog skupa podataka, taj postupak je preskočen s obzirom na to da izvorni algoritam već sadrži naučenu

mrežu za izlučivanje značajki osoba. Navedena mreža je prethodno naučena na MARS (Motion Analysis and Re-Identification Set) skupu podataka koji sadrži slike različitih osoba. Sustav za praćenje objekata vrednovat će se na skupovima za učenje i testiranje.

3. Arhitektura sustava

Sustav za praćenje nogometaša i generiranje mape aktivnosti sastoji se od 3 komponente:

- Podsustav za detekciju pozicija nogometaša u okviru videozapisa
- Podsustav za praćenje nogometaša između okvira videozapisa
- Podsustav za generiranje mape aktivnosti na temelju rezultata dobivenih praćenjem nogometaša

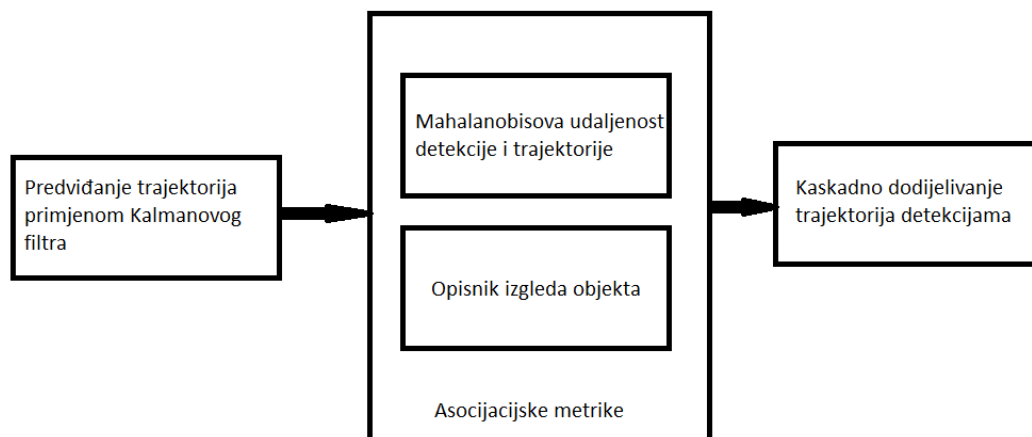
3.1. Detekcija nogometaša

Za detekciju nogometaša na slici koristi se YOLOv5s model iz skupa modela YOLOv5. Karakteristike YOLOv5 modela je da sliku podijeli na ćelije te je svaka ćelija zadužena za detektiranje objekata kojima se središte nalazi u toj ćeliji [9]. Sustav na ulazu prima slike koje predstavljaju okvire videozapisa te na izlazu vraća skup detekcija objekata na slici. Model vraća uvijek isti broj detekcija te je potrebno ukloniti redundantne detekcije primjenom algoritma potiskivanja nemaksimalnih odziva (engl. non-maximum suppression). S obzirom na to da se na terenu većinu vremena nalazi 22 nogometaša i jedan sudac, očekivani broj detekcija u svakom okviru iznosi 23. Model za detekciju objekata služi isključivo za određivanje pozicija nogometaša u trenutnom okviru, odnosno ne može identificirati pojedine nogometaše jer ne uzima u obzir detekcije iz prethodnih okvira.

3.2. Praćenje objekata

Za praćenje objekata koristi se algoritam DeepSORT.

Algoritam DeepSORT [11] nastao je kao proširenje SORT algoritma koji se bazira na korištenju Kalmanovog filtra za predviđanje trajektorija te mađarskog algoritma za dodjelu detekcija trajektorijama na temelju njihovog omjera presjeka i unije[2]. Iako



Slika 3.1: Arhitektura DeepSORT algoritma

takav algoritam ima veliku brzinu izvođenja, navedeni algoritam karakterizira česta pojava zamjene identiteta. Za praćenje nogometaša, zamjena identiteta predstavlja veliki problem zbog česte interakcije među nogometašima te može rezultirati niskom preciznošću. Algoritam DeepSORT nastoji riješiti problem zamjene identiteta uvođenjem opisnika izgleda objekta koji se također uzima u obzir prilikom dodjele detekcija trajektorijama. Arhitektura neuronske mreže za dobivanje opisnika izgleda objekta prikazana je tablicom na Slici 3.2.

Algoritam također koristi Kalmanov filtar u svakoj trajektoriji za praćenje trenutnog stanja i predviđanje idućeg stanja, odnosno pozicije trajektorije u sljedećem okviru videozapisa. Predviđene trajektorije uspoređuju se s detekcijama na izlazu detektora objekata te se na temelju međusobne udaljenosti uparuju. Udaljenost trajektorije i detekcije se definira kao težinska suma Mahalanobisove udaljenosti trajektorije i detekcije te kosinusne udaljenosti opisnika izgleda detekcije i trajektorije [11].

Uparivanje se vrši primjenom kaskadnog uparivanja koje se bazira na iterativnoj primjeni mađarskog algoritma pri čemu se prilikom uparivanja daje prioritet češće ažuriranim trajektorijama. Rezultat algoritma su uparene detekcije i trajektorije. Arhitektura sustava prikazana je na Slici 3.1.

3.2.1. Kalmanov filtar

Kalmanov filtar se u algoritmu DeepSORT koristi za praćenje internog stanja trajektorije te predviđanje pozicije objekta u nadolazećem okviru. Interno stanje trajektorije

Naziv	Vel. jezgre/Korak	Vel. izlaza
Conv 1	$3 \times 3/1$	$32 \times 128 \times 64$
Conv 2	$3 \times 3/1$	$32 \times 128 \times 64$
Max Pool 3	$3 \times 3/2$	$32 \times 64 \times 32$
Residual 4	$3 \times 3/1$	$32 \times 64 \times 32$
Residual 5	$3 \times 3/1$	$32 \times 64 \times 32$
Residual 6	$3 \times 3/2$	$64 \times 32 \times 16$
Residual 7	$3 \times 3/1$	$64 \times 32 \times 16$
Residual 8	$3 \times 3/2$	$128 \times 16 \times 8$
Residual 9	$3 \times 3/1$	$128 \times 16 \times 8$
Dense 10		128
Batch and ℓ_2 normalization		128

Slika 3.2: Arhitektura opisnika izgleda

ažurira se kada je dostupan vektor mjerenja. Vektor mjerenja u kontekstu algoritma jednak je jednoj detekciji objekta.

Algoritam DeepSORT stanje trajektorije definira kao 8-dimenzionalni vektor koji se sastoji od: x koordinate objekta, y koordinate središta objekta, omjera visine i širine, visine te brzina promjena za svaku od navedenih komponenti. Mjerenje se definira kao 4-dimenzionalni vektor koji se sastoji od: x koordinate pozicije objekta, y koordinate pozicije objekta, omjera visine i širine omeđujućeg pravokutnika objekta te visine tog pravokutnika.

Za detekcije koje nisu uparene s postojećim trajektorijama prilikom procesa uparivanja, stvaraju se nove trajektorije koje su na početku u probnom stanju, odnosno još se ne tretiraju kao rezultat praćenja objekata. Tek ako se takvoj novostvorenoj trajektoriji dodjeli mjerenje kroz određen broj uzastopnih okvira, trajektorija prelazi u aktivno stanje te se tek tada tretira kao ispravna trajektorija. Tako se sprječava praćenje nepostojećih objekata koji bi nastajali kao posljedica lažno pozitivnih detekcija [11].

Kada je trajektorija u aktivnom stanju te vektor mjerenja nije dostupan, odnosno detekcija nije dodijeljena, sljedeće stanje se predviđa pomoću internog stanja trajektorije [2]. Ako se nekoj aktivnoj trajektoriji ne dodjeli mjerenje unutar određenog broja uzastopnih okvira, trajektorija i njeno stanje se briše. [11]. Trajektorija kojoj nije dodijeljeno mjerenje se ne prikazuje na izlazu sustava sve dok se trajektoriji ponovo ne dodijeli neka detekcija.

3.2.2. Dodjela trajektorija detekcijama

U svakom okviru potrebno je dodijeliti detekcije trajektorijama tako da parovi detekcija i trajektorija budu minimalno udaljeni. Algoritam DeepSORT ne pokušava upariti sve detekcije i trajektorije odjednom već daje prednost trajektorijama kojima se interno stanje češće ažurira pomoću novih mjerenja, odnosno detekcija. Algoritam kreće od

dodjeljivanja nedodijeljenih detekcija najčešće ažuriranim trajektorijama.

Kriterij koji se razmatra prilikom uparivanja trajektorija i detekcija sastoji se od dvije komponente: kvadrirane Mahalanobisove udaljenost između predviđene pozicije objekta (dobiva se primjenom Kalmanovog filtra) i detekcije te minimalne kosinusne udaljenosti između opisnika izgleda detektiranog objekta i prethodnih opisnika objekta za određenu trajektoriju. Za jednu trajektoriju se pamti određen broj opisnika izgleda detekcija koje su bile prethodno uparene s tom trajektorijom. Navedene udaljenosti se linearno kombiniraju u konačnu udaljenost koja se razmatra prilikom dodjele detekcija trajektorijama.

Trajektorije i detekcije s minimalnom udaljenošću uparuju se uzimajući u obzir ažurnost trajektorija i prag udaljenosti koji ni jedna dodjela ne smije prekoračiti. Rezultat uparivanja s ovim ograničenjima je da će neke trajektorije biti uparene s nekim detekcijama, ali i dalje mogu ostati neuparene detekcije i neuparene trajektorije. U idućem koraku ponovo se uzimaju neuparene trajektorije koje su rjeđe ažurirane od prethodno uparenih te se pokušavaju upariti s preostalim neuparenim detekcijama.

Kao parametar algoritma se postavlja maksimalni broj okvira, odnosno maksimalno vrijeme koje je prošlo od ažuriranja trajektorije te se ono koristi za određivanje koliko puta će pokušati dodijeliti neuparene detekcije neuparenim trajektorijama.

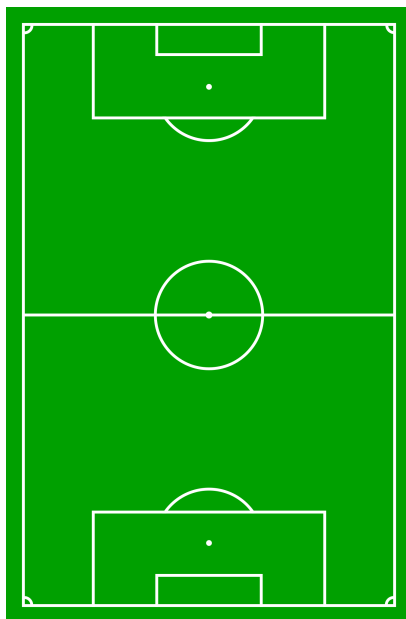
Ako nakon ovakvog iterativnog dodjeljivanja i dalje postoje neuparene detekcije i trajektorije, one se uparuju primjenom mađarskog algoritma gdje se kao kriterij optimizacije prilikom dodjeljivanja koristi omjer presjeka i unije između područja detekcije i predviđenog područja trajektorije. Za ovo dodjeljivanje se također uvodi prag koji mora biti zadovoljen kako bi se uparivanje smatralo uspješnim[11].

Ako i dalje postoje neuparene detekcije, za njih se stvaraju nove trajektorije koje zatim moraju proći probnu fazu. Neuparene trajektorije ne dobivaju mjerenje u tom okviru te ako su u aktivnom stanju, same predviđaju svoju poziciju primjenom Kalmanovog filtra.

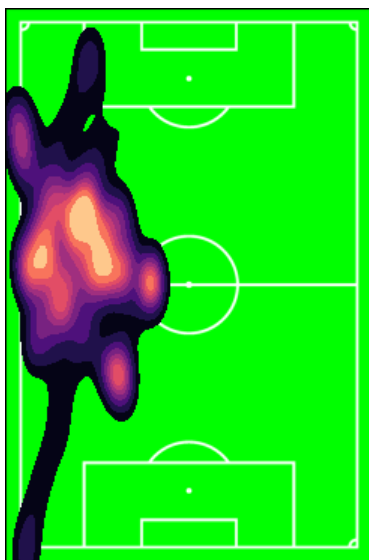
3.3. Izrada mape aktivnosti

Mape aktivnosti izrađuju se nakon analize cijelog videozapisa. Mape aktivnosti se izrađuju za svakog nogometaša zasebno te se pohranjuju kao slike. Za izradu mape aktivnosti, koristi se slika igrališta iz ptičje perspektive kao pozadina. Navedena slika prikazana je na Slici 3.3.

Na početku se za svakog igrača izrađuju liste x i y koordinata njihovih pozicija na okvirima videozapisa. Za poziciju igrača odabrana je točka koja se nalazi na polovištu



Slika 3.3: Nogometni teren iz ptičje perspektive



Slika 3.4: Mapa aktivnosti za igrača

donje lijeve i donje desne točke omeđujućeg pravokutnika detekcije igrača.

Pozicije igrača na okvirima videozapisa ne moraju odgovarati pozicijama na slici terena iz ptičje perspektive te je stoga potrebno provesti perspektivnu transformaciju. Za dobivanje matrice perspektivne transformacije potrebno je zadati četiri točke koje predstavljaju pozicije rubova igrališta u videozapisu te četiri točke koje predstavljaju rubove igrališta na slici terena iz ptičje perspektive.

Na temelju tih točaka izračunava se transformacijska matrica pomoću koje se sve točke koje predstavljaju pozicije igrača u okviru videozapisa preslikavaju u poziciju igrača na terenu snimljenom iz ptičje perspektive. S obzirom na to da se pozicija kamere može razlikovati u pojedinim videozapisima, potrebno je za svaki videozapis precizno zadati točke koje predstavljaju rubove terena, jer će u protivnom rezultirajuće mape aktivnosti biti netočne.

Jednom kada su sve točke pozicija u okvirima preslikane u odgovarajuće točke pozicija na terenu iz ptičje perspektive, izračunava se mapa aktivnosti. Pozicije nogometaša tretiraju se kao slučajne varijable koje imaju nepoznatu funkciju gustoće. Za procjenu te razdiobe koristi se metoda procjena gustoće jezgre. Konačan rezultat je procijenjena funkcija gustoće pozicija koja ujedno predstavlja i mapu aktivnosti nekog igrača. Primjer jedne mape aktivnosti prikazan je na Slici 3.4

4. Implementacija sustava

4.1. Podsustav za detekciju nogometaša

Podsustav za detekciju nogometaša sastoji se YOLOv5 detektora objekata koji na ulazu prima slikovne okvire videozapisa, a na izlazu vraća detekcije s podacima o detekciji. Implementacija YOLOv5 detektora preuzeta je s izvornog repozitorija [5]. Iz repozitorija je iskorištena skripta za učenje modela detekcije i skripta za vrednovanje naučenog modela.

Prilikom implementacije sustava praćenja nogometaša iz repozitorija se dodatno iskorištavaju sljedeće funkcije:

- funkcija za izvođenje unaprijednog prolaza za određenu sliku
- funkcija za primjenu potiskivanja nemaksimalnih odziva s ciljem uklanjanja redundantnih i višestrukih detekcija
- funkcija za dohvaćanje naučenog modela
- funkcije za iscertavanje izlaznih detekcija na okvirima izlaznog videozapisa

Za potrebe prilagodbe skupa podataka YOLOv5 formatu izrađena je skripta *dataset_builder.py*. Navedena skripta slijedno čita okvire videozapisa i za svaki okvir stvara sliku i prikladnu tekstualnu datoteku s oznakama u YOLOv5 formatu. Dodatno je izrađena skripta *dataset_splitter.py* koja slike i oznake dobivene skriptom *dataset_builder.py* dijeli u skup za učenje, skup za provjeru i skup za testiranje.

4.2. Podsustav za praćenje nogometaša

Za praćenje nogometaša koristi se algoritam DeepSORT. Implementacija navedenog algoritma također je preuzeta s izvornog repozitorija [10][11]. Algoritam DeepSORT koristi detekcije objekata i značajke izgleda prilikom izračuna udaljenosti detekcija i trajektorija. Detekcije objekata se dobivaju iz modela YOLOv5s, dok se značajke izgleda objekata dobivaju iz neuronske mreže za izlučivanje značajki.

Za izlučivanje značajki može se koristiti bilo koji model neuronske mreže koji na izlazu daje vektor značajki, odnosno model neuronske mreže koji ima ulogu enkodera. Učenje neuronske mreže za izlučivanje značajki u implementaciji sustava nije napravljeno, već je iskorišten izvorni model koji je naučen izlučivati značajke osoba.

U implementaciji sustava korišteni su sljedeći parametri algoritma DeepSORT:

- potrebne su 3 uzastopne detekcije dodijeljene trajektoriji prije nego što trajektorija postane aktivna
- najveća dozvoljena kosinusna udaljenost prilikom dodjele detekcija trajektorijama iznosi 0.3
- trajektorija se briše ako se ne upari ni s jednom detekcijom u uzastopnih 100 okvira
- jedna trajektorija sprema posljednjih 200 vektora značajki izgleda objekata s kojima uspoređuje vektore značajki detekcija prilikom dodjeljivanja detekcija trajektorijama
- najmanji dozvoljeni omjer presjeka i unije između detekcije i trajektorije prilikom dodjele iznosi 0.1

Navedene vrijednosti parametara određene su eksperimentalno na skupu podataka za učenje.

4.3. Podsustav za izradu mape aktivnosti nogometaša

Dio sustava zadužen za generiranje mapi aktivnosti implementiran je pomoću skripte *heatmap_builder.py*. Skripta za izradu mape aktivnosti pojedinih nogometaša koristi tekstualnu datoteku s pozicijama nogometaša u MOT16 formatu iz koje se za svakog nogometaša stvara lista njegovih pozicija u videozapisu. Pozicija nogometaša se tretira kao slučajna varijabla, odnosno mapa aktivnosti predstavlja funkciju gustoće distribucije pozicije igrača. Navedena funkcija gustoće aproksimira se pomoću metode procjene gustoće jezgre pri čemu pozicije nogometaša predstavljaju uzorke funkcije gustoće koja se procjenjuje. Procjena funkcije gustoće definirana je formulom:

$$\hat{f}_h(\vec{x}) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{\vec{x} - \vec{x}_i}{h}\right)$$

Za aproksimaciju funkcije i njen slikovni prikaz koristi se funkcija *kdeplot* iz knjižnice *Seaborn*. Slika 3.3 koja prikazuje nogometni teren iskorištena je kao pozadina

mape aktivnosti. Rezultat izvođenja skripte *heatmap_builder.py* su slike mapa aktivnosti svih detektiranih igrača.

4.4. Povezivanje podsustava u cjelinu

Glavna skripta *system.py* na početku inicijalizira podsustav za praćenje objekata. Za inicijalizaciju podsustava potrebno je navesti putanju do datoteke s informacijama o modelu za izlučivanje značajki te je potrebno navesti parametre algoritma DeepSort navedene u potpoglavlju 4.2.

Nakon inicijalizacije sustava za praćenje, inicijalizira se sustav za detekciju objekata. Za inicijalizaciju navedenog sustava potrebno je navesti putanju do datoteke koja sadrži informacije o naučenom modelu te sklopovlje na kojem će se izvršavati detekcija nogometaša.

Sustav na izlazu daje videozapis s označenim nogometašima te je stoga potrebno stvoriti razrede za izradu videozapisa te se oni stvaraju nakon inicijalizacije sustava za detekciju objekata.

Nakon inicijalizacije podsustava za detekciju i praćenje, iterativno se obrađuju okviri videozapisa tako da se za svaki okvir provodi detekcija objekata, potiskivanje ne-maksimalnih odziva, dodjeljivanje detekcija trajektorijama te na kraju izrada i dodavanje okvira s rezultatima na kraj izlaznog videozapisa. Implementacija navedene skripte inspirirana je postojećim implementacijama sličnih algoritama dostupnim na javnim izvorima[3][1].

4.5. Podsustav za vrednovanje praćenja i detekcije objekata

Za evaluaciju detekcije objekata korištena je skripta *val.py* iz implementacije YOLOv5 detektora. Navedena skripta na ulazu prima model i skup podataka nad kojima vrši vrednovanje. Skripta na temelju ulaza generira podatke koji opisuju kvalitetu modela detekcije poput preciznosti, odziva, F1-mjere i mnogih drugih.

Za evaluaciju praćenja objekata iskorišten je program TrackEval [8][6][7][4] koji kao ulaz zahtjeva datoteku s pozicijama pojedinih objekata koje je dao sustav i datoteku sa stvarnim pozicijama objekata. Program zahtjeva da obje datoteke sadrže pozicije i identitete objekata u MOT16 formatu. Rezultat programa je ispis raznih mjera koje ocjenjuju performanse algoritma za praćenje objekata.

5. Analiza rezultata

5.1. Vrednovanje modela detekcije nogometaša

Model detekcije nogometaša vrednovat će se na skupu za učenje te na skupu za testiranje. Za svaki skup mjerit će se preciznost, odziv, F1-mjera, sredina prosječnih preciznosti kada je prag omjera presjeka i unije 0.5 te sredina prosječnih preciznosti za pragove omjera presjeka i unije u rasponu od 0.5 do 0.95.

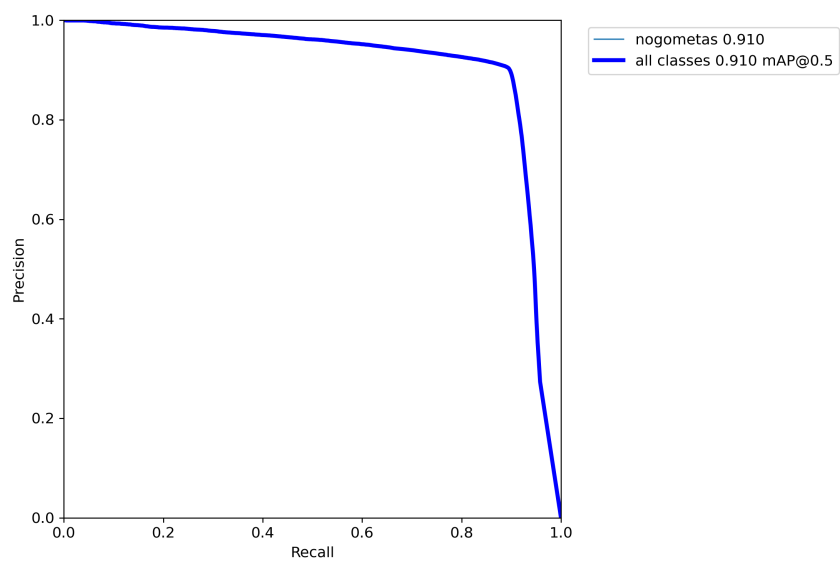
Preciznost za binarnu klasifikaciju se definira kao omjer stvarno pozitivnih detekcija i zbroja stvarno pozitivnih detekcija i lažno pozitivnih detekcija. Preciznost modela detekcije nogometaša iznosi na skupu za učenje iznosi 0.904, dok na skupu za testiranje preciznost iznosi 0.868.

Odziv se definira kao omjer stvarno pozitivnih detekcija i zbroja stvarno pozitivnih detekcija i lažno negativnih detekcija te ona za naučeni model na skupu za učenje iznosi 0.899, dok na skupu za testiranje iznosi 0.852.

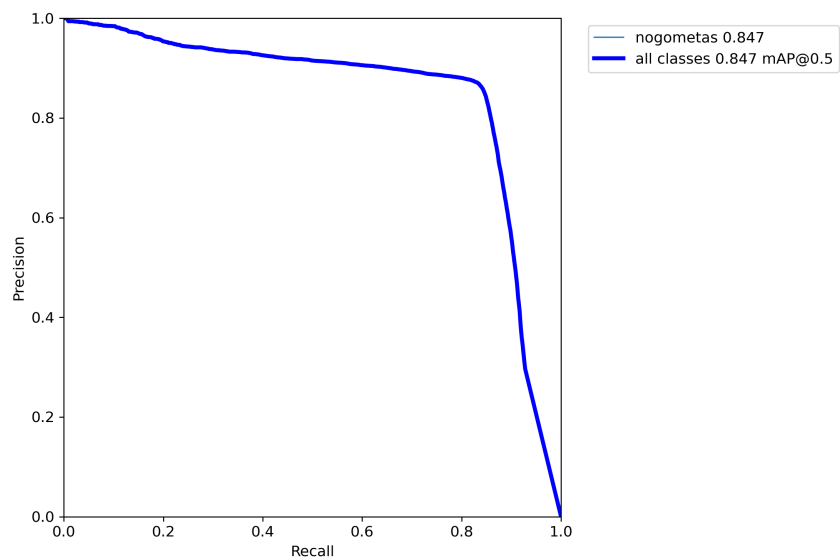
F1-mjera definira se kao harmonijska sredina preciznosti i odziva te ona iznosi 0.892 na skupu za učenje, odnosno 0.846 na skupu za testiranje. Na slikama 5.1 i 5.2 prikazana je krivulja preciznosti u ovisnosti o odzivu (engl. *precision-recall curve*).

Sredina prosječnih preciznosti za prag omjera presjeka i unije 0.5 ($mAP@0.5$) definira se kao aritmetička sredina prosječnih preciznosti ($AP_i@0.5$) za svaki razred i pri čemu prosječne preciznosti također koriste prag omjera presjeka i unije iznosa 0.5 prilikom određivanja točnih detekcija. Prag omjera presjeka i unije definira kada se neka detekcija smatra ispravnom. Za konkretni slučaj kada je omjer presjeka i unije detekcije objekta i stvarne pozicije objekta veći od 0.5, detekcija se smatra ispravnom. Prosječna preciznost definira se kao površina ispod krivulje preciznosti u ovisnosti o odzivu. Sredina prosječnih preciznosti na skupu za učenje iznosi 0.91, dok na skupu za testiranje iznosi 0.847.

Sredina prosječnih preciznosti za pragove omjera presjeka i unije u rasponu od 0.5 do 0.95 definira se kao prosjek sredina prosječnih preciznosti za sljedeće pragove omjera presjeka i unije: 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95. Navedena



Slika 5.1: Krivulja preciznosti u ovisnosti o odzivu na skupu za učenje



Slika 5.2: Krivulja preciznosti u ovisnosti o odzivu na testnom skupu

mjera na skupu za učenje iznosi 0.424, dok na skupu za testiranje iznosi 0.333.

Visoka preciznost, odziv i F1-mjera ukazuju na dobra svojstva generalizacije naučenog modela. Mala vrijednost sredine prosječnih preciznosti za pragove omjera presjeka i unije u rasponu od 0.5 do 0.95 ukazuje na manje preciznu lokalizaciju pri visokim pragovima omjera presjeka i unije, odnosno kada je prag omjera presjeka i unije vrlo visok, detektor daje manje ispravnih detekcija. Vrijedi napomenuti kako je sustav naučen detektirati nogometaše samo na jednom videozapisu, odnosno snimci jedne utakmice. Vrednovanje ovakvog modela detekcije na nekoj drugoj snimci mogao bi rezultirati manjom preciznošću. Za ostvarenje mogućnosti šireg korištenja modela, potrebno je naučiti model na više snimki nogometnih utakmica te bi bilo poželjno da su uvjeti u kojima se igraju utakmice različiti.

5.2. Vrednovanje algoritma praćenja objekata

Za vrednovanje praćenja objekata koristit će se sljedeće metrike [8]:

- točnost praćenja više objekata (engl. *multiple object tracking accuracy*, MOTA)
- preciznost praćenja više objekata (engl. *multiple object tracking precision*, MOTP)
- broj većinom praćenih objekata (engl. *mostly tracked*, MT)
- broj djelomično praćenih objekata (engl. *partially tracked*, PT)
- broj većinom izgubljenih objekata (engl. *mostly lost*, ML)
- broj zamjena identiteta (IDSW)
- broj fragmentacija (FM)

Točnost praćenja više objekata definira se formulom:

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + IDSW_t)}{\sum_t GT_t}$$

pri čemu se u brojniku za svaki okvir zbrajaju broj lažno pozitivnih i lažno negativnih objekata te broj zamjena identiteta. U nazivniku se u svakom okviru zbraja broj objekata koji je u tom okviru sadržan [8]. Točnost praćenja na skupu za učenje iznosi 57.956%, a na testnom skupu iznosi 38.964%.

Preciznost praćenja više objekata definira se formulom:

$$MOTP = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t}$$

gdje $d_{t,i}$ predstavlja preklapanje omeđujućeg pravokutnika i -te trajektorije i stvarne vrijednosti, dok c_t predstavlja broj uspješnih dodjeljivanja u okviru t [8]. Preciznost praćenja nogometaša na skupu za učenje iznosi 70.872%, dok na testnom skupu iznosi 67.96%.

Većinom praćen objekt je onaj objekt kojeg algoritam za praćenje objekata uspješno prati u više od 80% okvira[8]. Broj većinom praćenih objekata u skupu za učenje iznosi 11, dok u skupu za testiranje iznosi 1.

Djelomično praćen objekt je objekt kojeg algoritam za praćenje prati u rasponu od 20% do 80% okvira[8] te broj takvih objekata u skupu za učenje iznosi 12, a u skupu za testiranje iznosi 22.

Većinom izgubljeni objekti su oni objekti koje algoritam za praćenje uspijeva pratiti u manje od 20% okvira[8]. Sustav za praćenje nogometaša na skupu za testiranje kao ni na skupu za učenje nema takvih objekata.

Broj zamjena identiteta je značajan unatoč korištenju opisnika izgleda objekata te iznosi 135 u skupu za učenje i 92 u skupu za testiranje. Vrijedi napomenuti kako je veći broj zamjena identiteta u skupu za učenje očekivan jer se u skupu za učenje analizira videosekvencu koja je trostruko duža od videosekvence za testiranje.

Prilikom praćenja objekata zbog mimoilaženja nogometaša na slici, često dolazi do zaklanjanja koje rezultira fragmentacijom. Fragmentacija mjeri broj situacija u kojima se objekt prestaje pratiti te se nakon određenog broja okvira ponovo nastavi pratiti[8]. Broj fragmentacija na skupu za učenje iznosi 5931, a na testnom skupu iznosi 2140. Iz istih razloga kao i za veći broj zamjena identiteta, broj fragmentacija na skupu za učenje je očekivan.

Velike razlike u preciznostima algoritma za praćenje ukazuju na važnost korištenja preciznog detektora objekata. Praćenje objekata davalo je bolje rezultate na skupu za treniranje jer detektor nogometaša YOLOv5s daje preciznije detekcije na skupu za učenje.

6. Zaključak

Praćenje nogometaša zahtjevan je zadatak zbog čestih promjena u kretanjima nogometaša. Veliki broj algoritama praćenja objekata pretpostavlja linearan model kretanja objekata u kojem se između okvira ne događaju velike promjene u brzini kretanja objekata što u slučaju kretanja nogometaša ne mora vrijediti. Praćenje dodatno otežavaju česte interakcije nogometaša koje vrlo lako mogu izazvati zamjenu identiteta i zaklanjanje pojedinih nogometaša.

Za praćenje nogometaša izrađen je sustav koji se bazira na detekciji svih nogometaša primjenom YOLOv5 modela te na dodjeli identiteta dobivenim detekcijama primjenom DeepSORT algoritma.

Model za detekciju objekata YOLOv5s pokazuje dobra generalizacijska svojstva na korištenom skupu podataka, unatoč tome što je za detekciju bio korišten najmanje složen model iz skupa modela koje sadrži YOLOv5. S obzirom na to da igrači na snimci zauzimaju relativno mali udio okvira u odnosu na pozadinu, potreban je detektor koji radi sa slikama visoke rezolucije te je stoga naučeni model za detekciju koristio ulazne slike širine 1504. U protivnom se događa veliki gubitak informacije i dolazi do značajnog pada preciznosti detektora.

Algoritam za praćenje DeepSORT unatoč korištenju opisnika izgleda objekta i dalje rezultira relativno čestim zamjenama identiteta objekata prilikom interakcija između nogometaša. Kada nema navedenih interakcija, algoritam daje zadovoljavajuće rezultate, odnosno prati objekte s visokom preciznošću. Preciznost algoritma praćenja ovisi i o korištenom modelu detekcije objekata jer algoritam DeepSORT ažurira trajektorije pomoću detekcija objekata u svakom okviru videozapisa te je stoga poželjno imati detektor objekata vrlo visoke preciznosti.

Generiranje mape aktivnosti uvelike se oslanja na pozicije objekata dobivene praćenjem objekata te su poželjna što duža praćenja nogometaša u snimci, u protivnom se pojedina područja aktivnosti nogometaša ne bilježe i mapa aktivnosti poprima drugačiji oblik od očekivanog. Također na izgled mape aktivnosti utječu i zamjena identiteta te privremena zaklanjanja nogometaša.

7. Literatura

- [1] Object tracking implemented with yolov4, deepsort, and tensorflow. <https://github.com/theAIGuysCode/yolov4-deepsort>, 2022.
- [2] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, i Ben Upcroft. Simple online and realtime tracking. *2016 IEEE International Conference on Image Processing (ICIP)*, Sep 2016. doi: 10.1109/icip.2016.7533003. URL <http://dx.doi.org/10.1109/ICIP.2016.7533003>.
- [3] Mikel Broström. Real-time multi-camera multi-object tracker using yolov5 and deepsort with osnet. https://github.com/mikel-brostrom/Yolov5_DeepSort_OSNet, 2022.
- [4] Patrick Dendorfer, Aljosa Osep, Anton Milan, Konrad Schindler, Daniel Cremers, Ian Reid, Stefan Roth, i Laura Leal-Taixé. Motchallenge: A benchmark for single-camera multiple target tracking. *International Journal of Computer Vision*, stranice 1–37, 2020.
- [5] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, NanoCode012, Yonghye Kwon, TaoXie, Jiacong Fang, imyhxy, Kalen Michael, Lorna, Abhiram V, Diego Montes, Jebastin Nadar, Laughing, tkianai, yxNONG, Piotr Skalski, Zhiqiang Wang, Adam Hogan, Cristi Fati, Lorenzo Mammanna, AlexWang1900, Deep Patel, Ding Yiwei, Felix You, Jan Hajek, Laurentiu Diaconu, i Mai Thanh Minh. ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference, Veljača 2022. URL <https://doi.org/10.5281/zenodo.6222936>.
- [6] Arne Hoffhues Jonathon Luiten. Trackeval. <https://github.com/JonathonLuiten/TrackEval>, 2020.
- [7] Jonathon Luiten, Aljosa Osep, Patrick Dendorfer, Philip Torr, Andreas Geiger, Laura Leal-Taixé, i Bastian Leibe. Hota: A higher order metric for evaluating

- multi-object tracking. *International Journal of Computer Vision*, stranice 1–31, 2020.
- [8] A. Milan, L. Leal-Taixé, I. Reid, S. Roth, i K. Schindler. MOT16: A benchmark for multi-object tracking. *arXiv:1603.00831 [cs]*, Ožujak 2016. URL <http://arxiv.org/abs/1603.00831>. arXiv: 1603.00831.
- [9] Joseph Redmon, Santosh Divvala, Ross Girshick, i Ali Farhadi. You only look once: Unified, real-time object detection, 2015. URL <https://arxiv.org/abs/1506.02640>.
- [10] Nicolai Wojke i Alex Bewley. Deep cosine metric learning for person re-identification. U *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, stranice 748–756. IEEE, 2018. doi: 10.1109/WACV.2018.00087.
- [11] Nicolai Wojke, Alex Bewley, i Dietrich Paulus. Simple online and realtime tracking with a deep association metric, 2017.

Praćenje nogometaša primjenom YOLOv5 modela i DeepSORT algoritma

Sažetak

Cilj seminarskog rada je izgraditi i opisati sustav za praćenje nogometaša i generiranje mape aktivnosti pojedinih igrača. Sustav se sastoji od detekcije svih nogometaša u pojedinom okviru videozapisa i od povezivanja dobivenih detekcija sa detekcijama iz prethodnih okvira. Za detekciju nogometaša koristi se YOLOv5 model konvolucijske neuronske mreže dok se za praćenje detektiranih nogometaša koristi DeepSORT algoritam. Model detekcije i praćenja objekata se evaluiraju prikladnim mjerama te se dobiveni rezultati analiziraju. Pozicije nogometaša dobivene praćenjem objekata koriste se za izradu mapi aktivnosti za svakog pojedinog nogometaša. Pozicije nogometaša na okvirima videozapisa ne moraju u potpunosti odgovarati pozicijama promatranim iz ptičje perspektive te je potrebno odraditi određene transformacije nad tim pozicijama s ciljem dobivanja pravilnih mapi aktivnosti.

Ključne riječi: praćenje objekata, praćenje više objekata, YOLOv5, DeepSORT, računalni vid, duboko učenje

Tracking football players using YOLOv5 and DeepSORT algorithm

Abstract

The goal of the research is to build and analyze a system for tracking football players and generating heat maps of the individual players. The system is consisted of object detection in each frame of the video and assigning resulting detections with previous detections. For detecting football players, a YOLOv5 convolutional neural network model is trained. DeepSORT algorithm is used for tracking football players based on detections in the current frame and previous frames. Object detection and tracking models are evaluated with appropriate metrics and the results are analyzed. Players' positions in frames don't necessarily match their positions on the football pitch viewed from the top. It is necessary to make specific geometric transformations on the player's positions for getting the appropriate heatmaps.

Keywords: object tracking, multiple object tracking, YOLOv5, DeepSORT, computer vision, deep learning