

Sveučilište u Rijeci

Fakultet informatike i digitalnih tehnologija

Diplomski studij Informatike

## Detekcija osoba na sportskim snimkama

Seminarski rad

Kolegij: Komunikacija čovjek stroj

Student: Izabela Lesac

Ak. god.: 2022./2023

Rijeka, veljača 2022.

## Sadržaj

1. Uvod .....	3
2. Baza podataka.....	3
3. YOLOv5.....	4
3.2. Učenje YOLOv5s modela s nasumičnim težinama .....	5
3.3. Učenje s unaprijed određenim težinama.....	6
3.4. Podešavanje modela .....	6
4. YOLOv7.....	6
5. Rezultati.....	7
5.2. YOLOv5 s nasumičnim težinama .....	7
5.3. Učenje s predodređenim težinama.....	8
5.4. YOLO v7 .....	9
6. Usporedba.....	10
7. Zaključak .....	11
8. Literatura .....	11
9. Popis slika i tablica.....	12

## 1. Uvod

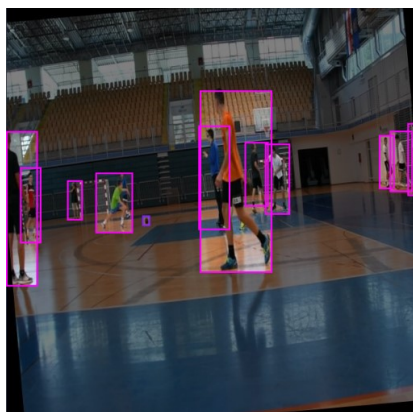
Detekcija objekata u videozapisima posljednjih je godina postala sve važnija tema istraživanja, budući da ima brojne primjene u poljima kao što su nadzor [1], autonomni automobili [2] i medicina [3]. Metode otkrivanja objekata temeljene na dubokom učenju postigle su vrhunske rezultate u otkrivanju i lokaliziranju objekata na slikama, a proširene su i na videozapise pomoću metoda kao što su praćenje i vremensko modeliranje [4]. Detekcija objekata u videozapisima uključuje prepoznavanje i lokaliziranje objekata od interesa u svakom kadru videozapisa. Ove metode obično uključuju obuku neuronske mreže na velikom označenom skupu podataka slika i/ili videozapisa, a zatim korištenje trenirane mreže za otkrivanje objekata u novim videozapisima. Zbog velike količine vizualnih podataka i vremenskih varijacija do kojih može doći dok se objekti pomiču i preklapaju, treniranje duboke mreže za efikasno prepoznavanje objekata je izazovan zadatak. Neke popularne metode detekcije objekata temeljene na dubokom učenju za videozapise uključuju YOLO i Faster R-CNN. Među tim metodama, YOLOv5 se pojavio kao moćan i učinkovit detektor objekata za video zapise, sposoban za postizanje visoke točnosti i performansi u stvarnom vremenu. YOLOv5 temelji se na uspjehu svog prethodnika, YOLOv4, i uključuje nekoliko poboljšanja kao što su pojednostavljena arhitektura i nova strategija povećanja podataka [5]. Novije verzije YOLO algoritama kao YOLOv7 tvrde kako postižu još bolje rezultate uz brže vrijeme izvođenja [6].

Jedno od područja za koje polako raste zanimanje su sportski video zapisi. Oni predstavljaju bogat izvor vizualnih podataka za analizu i razumijevanje pokreta igrača, pozicija momčadi i strategije igre. Mogućnost automatske detekcije igrača i drugih objekata u sportskim videozapisima može pružiti dragocjene uvide trenerima, sportašima i navijačima. Naspram drugih područja, sportska video analiza nije toliko opsežno istražena i postoji nedostatak kvalitetnih označenih setova podataka. Neki od trenutnih projekata koji se fokusiraju na ovo područje su SportsMOT [7] po uzoru na popularne MOT baze podataka, SoccerNet [8] SoccerTrack [9] i projekti sveučilišta u Rijeci s fokusom na rukomet [10].

U ovome radu predstavljam implementaciju YOLOv5s modela za otkrivanje ljudi i lopti u sportskim videozapisima te ju uspoređujem s novijom verzijom modela, YOLOv7. Modeli su trenirani na snimkama rukometnih treninga. Prvi dio rada opisuje označavanje podataka za treniranje i predprocesiranje baze. Zatim slijedi teorijski opis YOLO modela i učenje YOLOv5s i v7 na skupljenim podacima i podešavanje. Izvedba modela se procjenjuje koristeći standardne metrike. Naučeni modeli su demonstrirani na testnim slikama i videozapisima. Naposljetku slijedi analiza i usporedba rezultata.

## 2. Baza podataka

Za ovaj rad kreirala sam malu bazu podataka uz pomoć platforme Roboflow. Video zapise rukometnih treninga podijelila sam u slike i ručno označila. Na slikama sam označila dvije klase „osoba“ i „lopta“. Na ovaj način obradila sam 330 slika izvučenih iz 7 video zapisa. Klasa osoba sadrži 805 instanci a klasa lopta 136. Za YOLOv5 model preporučeno je imati mnogo veću bazu. Preporuka je minimalno 1500 slika po klasi i 10000 instanci klase. Kako bi poboljšala detekciju na manjem setu podataka, prije treniranja sam predprocesirala slike i dodala korak povećanja baze podataka modifikacijom slika. Dimenzija slika je 640x640, na slike je primijenjeno horizontalno zrcaljenje, rotacija i zumiranje.







Slika 1. Primjer preprocesirane slike iz skupa za učenje

Ovime sam uspjela dobiti nešto veću bazu od ukupno 716 slika. Podaci su podijeljeni na 579 slika za treniranje, 77 slika za validaciju i 60 slika za testiranje. Za korištenje s YOLOv5 modelima, baza mora biti u YOLOv5 Pytorch formatu. Roboflow automatski generira datoteke s oznakama u odgovarajućem obliku. Svaka slika ima pripadnu .txt datoteku u kojoj svaki red predstavlja jedan objekt. Ovako pripremljena baza je spremna za treniranje.

### 3. YOLOv5

YOLO je jedan od najpopularnijih modela za duboko učenje danas. Ime „You Only Look Once“ dolazi od glavne značajke ovog pristupa dubokom učenju. Za razliku od drugih modela koji rade na principu procesiranja ulazne slike u segmentima kroz više koraka, YOLO model dubokog učenja obrađuje slike u cijelosti. Ovaj način obrade je brzi i daje preciznije predikcije [11]. Sve verzije karakteriziraju brzina, učinkovitost i visoka preciznost. Popularnost je stekao zbog jednostavnosti korištenja, mogućnosti primjene na mnoge zadatke strojnog učenja, otvorenosti koda i različitim iteracijama koje su pogodne za izvođenje i na slabijim sistemima [12]. YOLOv5 je izdala tvrtka Ultralytics. Izašao je 2020. godine kratko nakon YOLOv4 i prouzročio kontroverzije u znanstvenoj zajednici. Naime, objavljen je bez recenziranog znanstvenog rada koji dokazuje izjave autora o brzini, efikasnosti i općenitim performansama [13]. Dolazi u deset verzija, od kojih su najznačajnije YOLOv5s, YOLOv5m, YOLOv5l i YOLOv5x. Razlikuje ih veličina i preciznost.

			
Small YOLOv5s	Medium YOLOv5m	Large YOLOv5l	XLarge YOLOv5x
14 MB <sub>FP16</sub> 2.2 ms <sub>V100</sub> 36.8 mAP <sub>COCO</sub>	41 MB <sub>FP16</sub> 2.9 ms <sub>V100</sub> 44.5 mAP <sub>COCO</sub>	90 MB <sub>FP16</sub> 3.8 ms <sub>V100</sub> 48.1 mAP <sub>COCO</sub>	168 MB <sub>FP16</sub> 6.0 ms <sub>V100</sub> 50.1 mAP <sub>COCO</sub>

Slika 2. Usporedba YOLOv5 modela. Prva stavka predstavlja 16 bitnu preciznost, druga brzinu zaključivanja (engl. inference) u milisekundama i mAP na COCO bazi podataka

YOLOv5 arhitektura se sastoji od tri glavne komponente: backbone, neck i head. Backbone, okosnica modela je konvolucijska neuronska mreža koja agregira i oblikuje značajke slika. U YOLOv4 i YOLOv5 koristi se CSPNet (Cross Stage Partial Network). CSPNet poboljšavaju učenje mreža, smanjuju opterećenje na memoriju i omogućuju stvaranje manjih modela s većom preciznošću [14]. Slijedeći dio je „vrat“ koji se sastoji od niza slojeva koji kombiniraju značajke slika i proslijeđuju ih „glavi“ gdje se događa predikcija. YOLOv5 za svaku seriju u treniranju na podatke primjeni korak povećanja (engl. augmentation). Slike se skaliraju, naprave se prilagodbe boje i mozaičko povećanje (engl. mosaic augmentation) [15]. Mozaik povećanje je prvi put predstavljeno u YOLOv4. To je proces u kojem se četiri slike iz skupa za treniranje kombiniraju u jedan „mozaik“ [16].



Slika 3. Primjer slike nakon primjene modifikacije mozaik

Novo svojstvo je i 16 bitna preciznost što ubrzava izvedbu detektiranja. YOLOv5 je popularan usprkos svim kontroverzama radi svega navedenog i jednostavne instalacije, malih modela, brze izvedbe treniranja, ugrađenih alata za evaluaciju i mogućnosti prilagodbe za mobilne uređaje [17]. Može se pokrenuti lokalno, koristiti na Google Colab platformi ili Jupyter bilježnici. Potpuni kod, svi modeli i alati su dostupni na Ultralytics Github repozitoriju.

### 3.2. Učenje YOLOv5s modela s nasumičnim težinama

YOLOv5 arhitektura koju sam izabrala za primjenu u detekciji osoba na videozapisima je YOLOv5s. Ovaj model se u prijašnjim eksperimentima pokazao kao najbrži, s dobrom razinom preciznosti [17]. Učenje je provedeno koristeći Google Colab i NVIDIA Tesla T4 GPU. Ovo je standardni GPU koji Google daje na ograničeno korištenje korisnicima Colab-a. Ograničenja vezana uz duljinu korištenja GPU-a su još jedan razlog zašto je izabran najbrži YOLOv5 model. Prije početka rada potrebno je instalirati YOLOv5 i Roboflow. YOLOv5 ima jednostavnu strukturu datoteka i instalacija dolazi sa svim potrebnim alatima i dodacima. Također, potrebno je preuzeti API ključ baze podataka s Roboflow-a kako bi ju mogli koristiti u Colab bilježnici.

Model je moguće naučiti koristeći težine koje su već bile istrenirane na nekom većem setu podataka kao COCO. Ideja je iskoristiti znanje koje je prethodno obučeni model stekao iz velikog skupa podataka, kao što su opće značajke i težine objekata, kako bi se poboljšala izvedba na novom skupu podataka. Ovaj pristup može ubrzati postupak učenja, ali u prvom eksperimentu model treniram „od nule“. Model počinje s nasumično inicijaliziranim težinama i mreža postepeno uči izdvajati značajke i otkrivati objekte. Učenje se pokreće slijedećom naredbom:

```
!python train.py --img 640 --batch 16 --epochs 100 -data
{dataset.location}/data.yaml --weights '' -cfg /content/yolov5/models/yolov5s.yaml
--cache
```

Model se trenira 100 epoha, s veličinom serije od 16 slika i ulaznom veličinom od 640x640 piksela. Lokacija datoteke data.yaml skupa podataka i YOLOv5s konfiguracijske datoteke navedene su pomoću argumenata --data i --cfg. Ne koriste se unaprijed pripremljene težine, i omogućeno je predmemoriranje kako bi se ubrzalo vrijeme vježbanja. Vrijeme učenja s ovim postavkama iznosi nešto više od 22 minute (0.374 sata).

Nakon treniranja pokrenula sam skriptu za detektiranje:

```
!python detect.py --weights runs/train/exp/weights/best.pt --img 640 --conf 0.1 --
source {dataset.location}/test/images
```

Ova naredba pokreće detekciju objekta na skupu testnih slika koristeći datoteku s najboljim težinama generiranu tijekom učenja. Određen je i prag pouzdanosti za detekciju objekta. U ovom slučaju postavljen je na 0.1, što znači da će objekti s niskom ocjenom također biti prijavljeni u izlazu. Ovime

bi se teoretski trebala poboljšati detekcija lopti, pošto su to mali objekti i često su prekriveni u kadru nečijom rukom i sl. Detekciju sam isprobala i na videozapisu iz baze podataka.

### 3.3. Učenje s unaprijed određenim težinama

Nakon učenja modela „od nule“, htjela sam vidjeti ukoliko će se preciznost i ostale metrike poboljšati korištenjem unaprijed definiranih težina. Primjenom prijenosnog učenja (engl. transfer learning) nadala sam se boljim generalnim rezultatima. Broj epoha, veličina slika i baza su isti.

```
!python train.py --img 640 --batch 16 --epochs 100 --data
{dataset.location}/data.yaml --weights yolov5s.pt --cache --freeze 10 --name
'feature_extraction'
```

YOLOv5s model je treniran na COCO bazi podataka za detekciju predmeta koja ima 80 klasa. Ove težine koristim na početku učenja kako bi ubrzali konvergenciju i poboljšali točnost. Također, dio slojeva je zamrznut oznakom „freeze“. Deset početnih slojeva YOLOv5s modela će biti zamrznuto u ovom pokretanju. Ova tehnika ubrzava obuku dopuštajući modelu da se usredotoči na učenje značajki više razine umjesto značajki niske razine koje su već dobro naučene na početnim slojevima [18]. Zajedno, ove postavke bi u teoriji trebale dati preciznije rezultate detekcije i omogućiti modelu efikasnije učenje značajki.

### 3.4. Podešavanje modela

Nakon izvođenja prethodnog koraka učenja, idući faza u procesu je podešavanja modela pomoću hiperparametara i ponovno testiranje. Osnovne postavke i broj epoha su iste kao i u prethodnim koracima.

```
!python train.py --hyp 'hyp.VOC.yaml' --batch 16 --epochs 100 --data
'/content/yolov5/ff-1/ff-1/data.yaml' --weights
'/content/yolov5/runs/train/feature_extraction/weights/best.pt' --name 'fine_tune'
--cache
```

Ova se naredba koristi za fino podešavanje (engl. fine tuning) unaprijed obučenog modela na prilagođenom skupu podataka. Datoteka 'hyp.VOC.yaml' sadrži postavke za brzinu učenja, zamah, opadanje težine itd. Unaprijed pripremljene težine modela koje se koriste za fino podešavanje su one iz prethodnog koraka učenja s već definiranim težinama. Finim podešavanjem unaprijed obučenog modela na prilagođenom skupu podataka, model može naučiti prepoznati specifične objekte koji nisu dio izvornog skupa klasa unaprijed obučenog modela.

## 4. YOLOv7

YOLOv7 je jedna od novijih verzija YOLO arhitekture. Prema radu izdanom u lipnju 2022. predstavlja najbržu i najprecizniju verziju YOLO detektora do sad. Autori su pokazali kako ima bolje performanse od mnogih drugih popularnih detektora. Zanimljivo je model učen isključivo na MS COCO bazi podataka, i to bez korištenja već inicijaliziranih težina. Bazirana je na Scaled YOLOv4 i YOLO-R detektorima [6]. Izvorni kod je dostupan na Github-u <sup>1</sup>. Za učenje YOLOv7 modela koristila sam sve iste postavke kao i u prvom eksperimentu. Model se instalira, zatim se instalira Roboflow i dohvaća se API ključ baze. Koristi se ista baza, ali je format podešen za YOLOv7. Učenje se izvodi 100 epoha, veličina serije je 16, dimenzija slika 640x640. Model se uči od početka, početne težine su inicijalizirane nasumično. Naredba za pokretanje učenja je:

```
!python train.py --batch 16 --img 640 --epochs 100 --data
{dataset.location}/data.yaml --weights '' --cfg
/content/yolov7/cfg/training/yolov7.yaml --device 0 --cache
```

---

<sup>1</sup> <https://github.com/WongKinYiu/yolov7>

## 5. Rezultati

Svi modeli u radu su ocijenjeni na skupu za validaciju koristeći slijedeće mjere procjene: preciznost (P), odziv (R) i srednju prosječnu preciznost (mAP.5) i srednju prosječnu preciznost s višim pragom (mAP.5-.95) za ukupnu izvedbu detekcije i za svaku pojedinačnu klasu. Rezultati su prikazani i grafički, po klasama objekata. Naredbe su navedene u prethodnim poglavljima, koriste se val.py i detect.py skripte iz YOLO paketa.

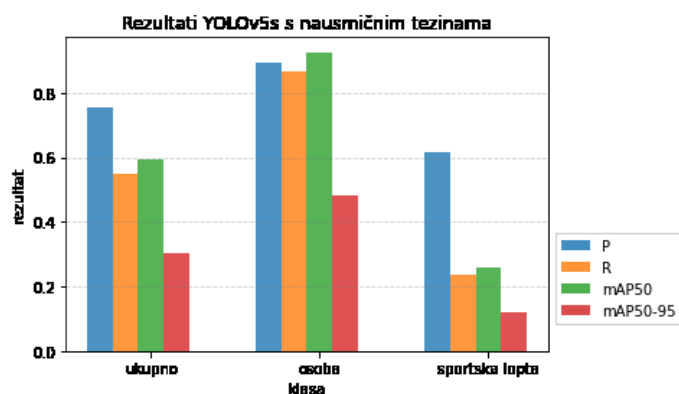
### 5.2. YOLOv5 s nasumičnim težinama

Mjerni podaci pokazuju da model ima P od 0.754, R od 0.551 i mAP.5 od 0.593 za sve klase zajedno. Za klasu "osoba", model ima P od 0.894, R od 0.866 i mAP.5 od 0.926. Međutim, za klasu "sportske lopte", model ima niži P od 0.614, niži R od 0.235 i niži mAP.5 od 0.26.

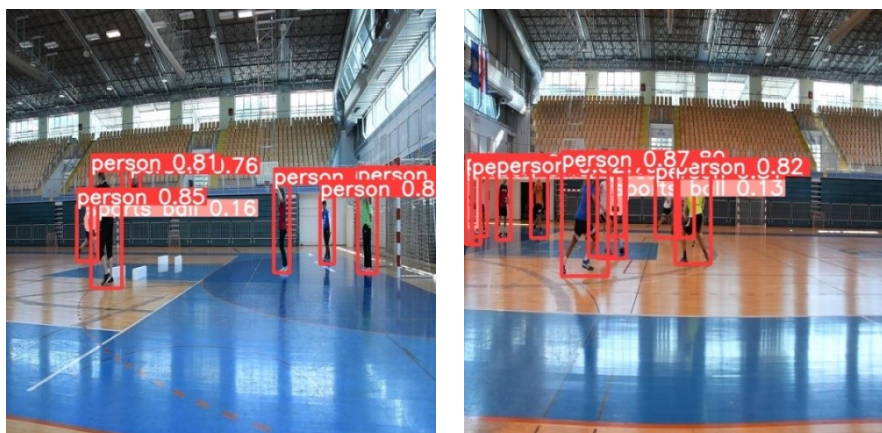
Tabela 1: Rezultati učenja YOLOv5s od nule

Klasa	P	R	mAP.5	mAP.5-.95
Ukupno	0.754	0.551	0.593	0.302
Osoba	0.894	0.866	0.926	0.482
Sportska lopta	0.614	0.235	0.26	0.122

Pri višim IoU (Intersection Over Union) pragovima, model je relativno loš u detekciji objekata. Model je bolji u otkrivanju ljudi od sportskih lopti, što može značiti da se mogu napraviti dodatna poboljšanja u obuci modela kako bi se poboljšala izvedba na klasi sportske lopte. Tablica 1 prikazuje rezultate učenja, slika 4 prikazuje rezultate grafički. Slike 5 i 6 prikazuju detekciju primijenjenu na skup za treniranje i na videozapisu, iz kojeg su naknadno izdvojeni pojedinačni kadrovi.



Slika 4. Rezultati učenja YOLOv5s s nasumičnim težinama po klasi



Slika 5. Rezultati detekcije YOLOv5s na skupu za testiranje





Slika 6. Rezultati detekcije YOLOv5s na videozapisu

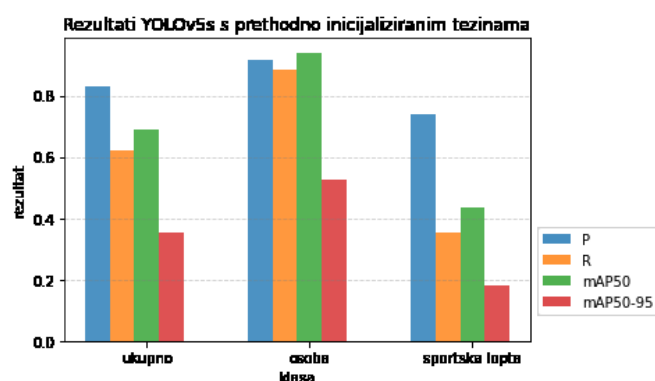
### 5.3. Učenje s predodređenim težinama

Nakon učenja koje je trajalo nekoliko sekundi manje nego prethodni eksperiment (0.369h naspram 0.374h ), dobili smo slijedeće rezultate.

Tabela 2: Rezultati učenja s predodređenim težinama

Klasa	P	R	mAP50	mAP50-95
Ukupno	0.828	0.619	0.688	0.353
Osoba	0.918	0.886	0.94	0.524
Sportska lopta	0.737	0.353	0.436	0.182

Preciznost nad svim klasa iznosi 0.828, odziv 0.619, srednja prosječna preciznost 0.688 i mAP.5 -95 0.353. Za klasu osoba postignuta je P od 0.918, R od 0.886, mAP.5 od 0.94. Vrijednosti su niže za klasu sportska lopta: preciznost iznosi 0.737, odziv 0.35, mAP.5 0.436.



Slika 7. Rezultati treniranja s inicijalnim težinama i zamrzavanjem slojeva

Gledajući rezultate (prikazane na slici 7) čini se da model radi relativno dobro. Ukupni mAP.5 od 0.688 za sve klase zajedno ukazuje na to da model može detektirati objekte na slikama s razumnom točnošću. mAP.5 -95 je niži, što ukazuje da bi se model mogao više boriti s identificiranjem objekata u zahtjevnijim situacijama. Mjerni podaci specifični za klasu pokazuju da je model bolji u otkrivanju



ljudi nego sportskih lopti kao i u prethodnom eksperimentu bez inicijalnih težina. Nakon primjene detaljnog podešavanja modela koristeći hiperparametre rezultati učenja su prikazani u tablici 1:

Tabela 3: Rezultati učenja nakon fine tuning modela

Klasa	P	R	mAP.5	mAP.5-.95
Ukupno	0.856	0.592	0.645	0.351
Osoba	0.923	0.912	0.947	0.546
Sportska lopta	0.788	0.272	0.343	0.157

Preciznost modela je poboljšana nakon podešavanja, što znači da čini manje lažno pozitivnih detekcija. Odziv za klasu osobe se povećao a za klasu sportske lopte smanjio. Srednja prosječna preciznost se generalno smanjila, što može biti znak da dolazi do overfitting. Rezultati sugeriraju da model dobro detektira osobe. Viša preciznost znači da kada model predviđa da je objekt na slici sportska lopta, to je obično točno. Međutim, niski odziv znači da model ne identificira sve sportske lopte na slici. Naredba ispod pokreće skriptu za validaciju. Iako YOLO provodi automatsku validaciju tokom učenja s ovom naredbom dobivamo točniju analizu rezultata. U ovom slučaju rezultati (prikazani u tablici 2) se ne razlikuju puno.

```
!python val.py --data /content/yolov5/ff-1/ff-1/data.yaml --weights
/content/yolov5/runs/train/fine_tune/weights/best.pt --img 640 --name
val_test_finetune
```

Tabela 4: Rezultati validacije nakon učenja

Klasa	P	R	mAP.5	mAP.5-.95
Ukupno	0.862	0.592	0.643	0.349
Osoba	0.928	0.911	0.951	0.547
Sportska lopta	0.796	0.272	0.335	0.15

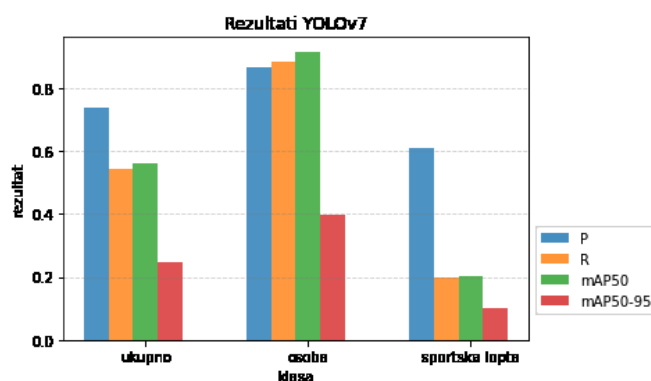
## 5.4. YOLO v7

Početne težine su inicijalizirane nasumično. Proces učenja je sporiji i traje nešto manje od sat vremena. Rezultate treniranja može se vidjeti na slici 11.

Tabela 5: Rezultati treniranja YOLOv7 bez inicijaliziranih težina

Klasa	P	R	mAP.5	mAP.5-.95
Ukupno	0.737	0.541	0.56	0.249
Osoba	0.867	0.883	0.916	0.398
Sportska lopta	0.608	0.199	0.204	0.0999

Gledajući rezultate treniranja, možemo vidjeti da su rezultati preciznosti, odziva i mAP.5 za klasu osoba viši od onih za klasu sportske lopte. Ovo sugerira da je model na danim slikama sposobniji detektirati ljude nego sportske lopte. mAP.5 rezultat od 0.56 ukazuje na to da model ima umjerenu razinu točnosti u otkrivanju objekata. Rezultat mAP.5-.95 je relativno loš i govori nam kako model nije dobar u preciznom lociranju i otkrivanju objekata koji se preklapaju.



Slika 8. Rezultati treniranja YOLOv7 bez inicijalnih težina



Slika 9. Detekcija s YOLOv7 na testnom skupu podataka

Na slici 10 možemo vidjeti primjenu naučenog YOLOv7 modela na videozapis koji nije dio naše baze podataka i koji se dosta razlikuje od slika na kojima je model učen. Rezultati detekcije su loši, što može biti zbog različitih uvjeta na videu, premaloj bazi podataka, činjenice da je model nepodešen i da je učenje krenulo od nule, bez predefiniраниh težina.



Slika 10. YOLOv7 na videozapisu koji nije dio baze podataka

## 6. Usporedba

Analizirajući eksperimente možemo vidjeti kako su performanse modela najgore kod učenja s nasumično inicijaliziranim težinama. Iako su oba modela bila relativno dobra u otkrivanju osoba, mučili se s detekcijom lopti na terenu. Kada koristimo već određene težine na istom skupu podataka i zamrznemo prvih 10 slojeva modela, rezultati učenja su bolji i vidi se veliki napredak u odnosu na prvi eksperiment. Učinkovitost modela u detekciji ljudi ostala je relativno ista, ali došlo je do značajnog poboljšanja u detekciji sportskih lopti. Nakon podešavanja modela koristeći hyp.yaml datoteku s parametrima za optimizaciju, preciznost za obje klase se povećala, ali se odziv smanjio. Srednja prosječna preciznost se također smanjila, ali je još uvijek bolja nego u prvom učenju bez definiranih težina. Općenita točnost modela je smanjena, što može značiti nekoliko stvari. Najvjerojatnije je kako podešavanje nije izvedeno korektno za ovaj specifičan model. Moguće je i da baza podataka nema dovoljno instanci svih klasa ili podataka općenito.

Tabela 6: Usporedba rezultata detekcije svih klasa ukupno

	P	R	mAP.5	mAP.5 -.95
<b>YOLOv5s bez težina</b>	0.754	0.551	0.593	0.302
<b>YOLOv5s s težinama</b>	0.828	0.619	0.688	0.353
<b>YOLOv5s s podešavanjem</b>	0.862	0.592	0.643	0.349
<b>YOLOv7 bez težina</b>	0.737	0.541	0.56	0.249

Uspoređujući rezultate modela YOLOv7 i YOLOv5s obučenih od nule na istom prilagođenom skupu, možemo vidjeti da je YOLOv7 postigao je niže rezultate preciznosti i odziva od YOLOv5s za sve klase, što ukazuje na to da se više borio s točnim otkrivanjem i klasificiranjem objekata na slikama. Razlika nije ogromna i postoje manje varijacije među pojedinačnim klasama.

## 7. Zaključak

Svrha ovog rada je bila usporedba različitih verzija YOLO arhitekture za detekciju sportaša na video snimkama. Htjela sam usporediti razlike u performansama, efikasnosti i točnosti različitih modela i načina primjene (s inicijaliziranim težinama i bez) za zahtjevan zadatak detekcije objekata u pokretu koji su često teško vidljivi i djelomično prekriveni. U radu sam usporedila tri verzije modela na prilagođenoj bazi podataka sačinjenoj od snimki rukometnih treninga. Detekcija ljudi je bila relativno precizna kroz sve eksperimente, ali svi modeli su imali problema s označavanjem sportskih lopti. Moguće je kako bi se detekcija lopti mogla poboljšati kroz daljnja podešavanja modela, i trebalo bi dodatno istražiti najbolji način kako to i primijeniti s ovako malom bazom podataka. Rezultati eksperimenata pokazuju kako je najbolji pristup za detekciju osoba i drugih objekata na video zapisima u ovom slučaju prijenosno učenje, gdje se na bazu podataka primjeni model s postojećim težinama. Važno je napomenuti da postoje i nedostaci koji su vrlo vjerojatno utjecali na izvedbu svih modela: mala baza podataka bez dovoljno slika, instanci klasa, varijacija okoline, svjetlosti. Za daljnja istraživanja na ovu temu, bilo bi potrebno proširiti bazu podataka i temeljito pregledati postojeću za loše označene slike.

## 8. Literatura

- [1] P. Mishra i G. Saroha, »A study on video surveillance system for object detection and tracking,« u *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, 2016.
- [2] Y. Li, H. Wang, L. M. Dang, T. Nguyen, D. Han i A. Lee, »A deep learning-based hybrid framework for object detection and recognition in autonomous driving,« IEEE Access, 2020.
- [3] S. Zou, C. Li, H. Sun, P. Xu, J. Zhang, P. Ma i M. Grzegorzek, »TOD-CNN: An effective convolutional neural network for tiny object detection in sperm videos,« *Computers in Biology and Medicine*, 2022.
- [4] T. Sugirtha i M. Sridevi, »A survey on object detection and tracking in a video sequence,« Springer Singapore, 2020.
- [5] P. Jiang, D. Ergu, F. Liu, Y. Cai i B. Ma, »Review of Yolo Algorithm Developments,« *rocedia Computer Science*, 2022.
- [6] C.-Y. Wang, A. Bochkovskiy i H.-Y. M. Liao, »YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object,« arXiv, 2022.
- [7] J. Wang, C. Meng, D. Li, H. Wang, Y. Yang, L. Jiao i F. Liu, »A Technical Report for SportsMOT Track on Multi-actor Tracking,«.
- [8] S. Giancola, M. Amine, T. Dghaily i B. Ghanem, »Soccernet: A scalable dataset for action spotting in soccer videos,« u *In Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018.
- [9] A. Scott, I. Uchida, M. Onishi, Y. Kameda, K. Fukui i K. Fujii, » Soccertrack: A dataset and tracking algorithm for soccer with fish-eye and drone videos,« IEEE/CVF, 2022.
- [10] M. Buric, M. Ivasic-Kos i M. Pobar, »Player Tracking in Sports Videos,« u *IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, Sydney, 2019.
- [11] R. Joseph i A. Farhadi, »YOLOv3: An Incremental Improvement,« 2018.
- [12] T. Do, » Evolution of YOLO Algorithm and YOLOv5: The State-of-the-art Object Detection,« 2021. [Mrežno].
- [13] V. Meel, »YOLOv5 Is Here! Is It Real or a Fake?,« [Mrežno]. Available: <https://viso.ai/deep-learning/yolov5-controversy/#:~:text=Viso%20Suite%20Platform.-,YOLOv5%3A%20What%20Is%20Different%3F,traction%20soon%20after%20its%20publishing..>

- [14] C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu\*, P.-Y. Chen i J.-W. Hsieh, »CSPNet: A new backbone that can enhance learning capability of CNN,« arxiv preprint, 2019.
- [15] J. Solawetz, »blog.roboflow,« 2020. [Mrežno]. Available: <https://blog.roboflow.com/yolov5-improvements-and-evaluation/>.
- [16] A. Bochkovskiy, C.-Y. Wang i H.-y. M. Liao, »YOLOv4: Optimal Speed and Accuracy of Object Detection,« preprint, 2020.
- [17] H. Marko, J. Ljudevit i G. Gordan, »A comparative study of YOLOv5 models performance for image localization and classification,« u *33rd Central European Conference on Information and Intelligent Systems*, Dubrovnik, 2022.
- [18] G. Jocher, »Github,« 6 Studeni 2020. [Mrežno]. Available: <https://github.com/ultralytics/yolov5/issues/1314>.
- [19] D. Graffox, 2009. [Mrežno]. Available: <http://www.ieee.org/documents/ieeecitationref.pdf>.

## 9. Popis slika i tablica

Slika 1. Primjer preprocesirane slike iz skupa za učenje.....	4
Slika 2. Usporedba YOLOv5 modela. Prva stavka predstavlja 16 bitnu preciznost, druga brzinu zaključivanja (engl. inference) u milisekundama i mAP na COCO bazi podataka.....	4
Slika 3. Primjer slike nakon primjene modifikacije mozaik .....	5
Slika 4. Rezultati učenja YOLOv5s s nasumičnim težinama po klasi .....	7
Slika 5. Rezultati detekcije YOLOv5s na skupu za testiranje .....	7
Slika 6. Rezultati detekcije YOLOv5s na videozapisu.....	8
Slika 7. Rezultati treniranja s inicijalnim težinama i zamrzavanjem slojeva .....	8
Slika 8. Rezultati treniranja YOLOv7 bez inicijalnih težina.....	10
Slika 9. Detekcija s YOLOv7 na testnom skupu podataka.....	10
Slika 10. YOLOv7 na videozapisu koji nije dio baze podataka .....	10