Stručni rad
https://doi.org/10.31784/zvr.11.1.15
Datum primitka rada: 1. 12. 2021.
Datum prihvaćanja rada: 9. 12. 2021.

BAZA SLIKA ZA STROJNO UČENJE MODELA ZA DETEKCIJU PLIVAČA

Ivan Šimac
Asistent, Veleučilište u Rijeci, Vukovarska 58, 51000 Rijeka, Hrvatska; *e-mail:* isimac@veleri.hr

Miran Pobar
Dr. sc., docent, Sveučilište u Rijeci, Odjel za informatiku, Radmile Matejčić 2,
51 000 Rijeka, Hrvatska; *e-mail:* mpobar@uniri.hr

Marina Ivašić-Kos
Dr. sc., izvanredna profesorica, Sveučilište u Rijeci, Odjel za informatiku, Radmile Matejčić 2,
51 000 Rijeka, Hrvatska; *e-mail:* marinai@uniri.hr

SAŽETAK

Velika količina podataka koja se svaki dan kreira može se upotrijebiti za razvoj algoritama umjetne inteligencije u domeni računalnog vida koji rješavaju zadatke poput klasifikacije slika, detekcije osoba i raspoznavanja akcija. Ti skupovi podataka su najčešće izrađeni od videozapisa i slika preuzetih s televizijskih kanala ili s društvene mreže YouTube i prikupljeni su i pripremljeni za odgovarajući zadatak. Nas je zanimao zadatak detekcije plivača, kako bi se model mogao koristiti za raspoznavanje i unaprjeđenje plivačkih tehnika. Iako danas postoje ogromne otvorene baze slika poput COCO i ImageNet, pripremljene za nadzirano strojno učenje te baze sportskih scena poput Olympic Sports Dataset, UCF Action Sport dataset ili Sport-1M koje uključuju slike popularnijih (gledanijih) sportova, nijedna od njih ne uključuje slike koje bi se mogle koristiti za izradu našeg modela za detekciju plivača. Stoga je u ovom radu opisan postupak snimanja i prikupljanja video materijala te priprema skupa slika UNIRI-SWM za detekciju plivača. Skup uključuje snimke plivača u realnim, situacijskim uvjetima treninga i natjecanja snimljenih akcijskim kamerama iz različitih kutova snimanja. U radu su dani rezultati detekcije plivača korištenjem dubokih konvolucijskih neuronskih mreža Mask R-CNN i Yolov3, naučenim na skupu općih slika prije i nakon učenja na skupu UNIRI-SWM. Rezultati pokazuju da se nakon prilagodbe modela na odgovarajućem skupu slika iz domene plivanja mogu postići jako dobri rezultati detekcije plivača.

Ključne riječi: *detekcija osoba, konvolucijska neuronska mreža, skup podataka, plivanje*

1. UVOD

Detekcija osoba na slikama i video zapisima je postupak kojim se automatski na slici ili okviru video zapisa označavaju sva područja na kojima se nalaze osobe, najčešće tako da je svaka osoba označena pravokutnim okvirom unutar kojeg se nalazi (Paul *et al.*, 2013.).

Automatska detekcija osoba je važan zadatak u području umjetne inteligencije i računalnog vida koji se aktivno istražuje zbog mogućnosti široke primjene. Važna je npr. za sigurnost i nadzor javnih prostora poput aerodroma i željezničkih postaja, za razvoj autonomnih vozila koja moraju biti svjesna pješaka u okolini (Paul *et al.*, 2013), ali i za analizu sportskih scena u svrhu izrade atraktivnih i informativnih TV prijenosa, analize napretka sportaša na treningu ili analize taktike određenih timova.

Trenutno se za detekciju osoba koriste metode dubokog učenja (engl. *deep learning*) kao što su YOLOv3 (Redmon i Farhadi, 2018), SSD (Liu *et al.*, 2016), Mask R-CNN (He *et al.*, 2017), R-FCN (Dai *et al.*, 2016) koje se temelje se na konvolucijskim neuronskim mrežama (engl. *convolutional neural network*, CNN). Proces učenja konvolucijskih neuronskih mreža pripada nadziranom strojnom učenju, što znači da zahtijeva prethodno pripremljenu bazu označenih pozitivnih i negativnih primjera temeljem kojih se definira decizijska funkcija i model koji će, po uzoru na primjere u skupu za učenje, moći raspoznati i nove još neviđene primjere.

U slučaju učenja modela za detekciju osobe, potrebna je baza slika na kojima su označene osobe. Prikupljanje slika je olakšano s obzirom na veliki broj podataka koji se danas snima ili postavlja na javne servise (video nadzor, snimke postavljene na YouTube, fotografije na društvenim mrežama), međutim te podatke se ne može direktno koristiti već je potrebna ručna obrada. Ručna obrada i priprema slike za strojno učenje uključuje označavanje područja slike i pridruživanje odgovarajuće oznake ili klase kojoj pripada označeno područje.

Naučeni modeli strojnog učenja su dobro prilagođeni domeni na kojoj su učeni i daju dobre rezultate na podacima, u ovom slučaju slikama, koje nalikuju onima iz skupa za učenje. Cilj je da model bude generalan, što se uglavnom odnosi na donošenje zaključaka o podacima iz iste domene, sličnim onima na kojima je bio učen, ali koje model još nije „vidio". Problem generalizacije na način kako ga razumiju i koriste ljudi još je uvijek daleko od realizacije. Npr. već u najranijoj dobi, dijete kojem se pokaže npr. mačka na slici, moći će u realnom životu u najrazličitijim scenarijima raspoznati mačku iako je nikada ranije nije vidjelo u toj situaciji, iako se razlikuju po veličini, boji, položaju tijela, i slično (Ivašić-Kos *et al.*, 2009). Kod modela dubokog učenja to nije slučaj i potreban je veliki broj primjera, u različitim situacijama, s promjenom osvjetljenja, položaja kamere, boje, pozadine, promjene položaja i razmještaja objekata na slici i slično, da bi se model naučio raspoznavati objekte neke klase.

S obzirom na to da je detekcija osoba i općenito objekata na slikama značajna za cijeli niz područja, postoji veći broj javno dostupnih baza za učenje detektora, no zbog najšire primjene naglasak je najčešće na detekciji pješaka. Mi smo pak bili zainteresirani za detekciju sportaša, poglavito plivača. S obzirom na razlike u kutovima snimanja javnih prostora kojima se kreću pješaci i onih koji se najčešće koriste za snimanje sportskih događaja, te velike raznolikosti pokreta koju sportaši

izvode i opreme koju nose, baze koje su namijene detekciji pješaka nisu dovoljne niti adekvatne za definiranje uspješnih metoda detekcije sportaša. Zbog toga su formirane baze iz domene sporta, pripremljene za strojno učenje i prilagođene zadacima računalnog vida kao što je klasifikacija slika iz domene sporta ili detekcija sportaša.

U sljedećem poglavlju prikazane su postojeće i javno dostupne označene baze slika iz domene sporta. Ustanovili smo da one uglavnom ne sadrže sportove u vodi gdje je vizualna razlika između pojavljivanja osobe još veća zbog samog položaja osobe u odnosu na podlogu i medija vode u kojem se sport odvija. Iz tog razloga, u trećem je poglavlju opisana baza i proces izrade baze za učenje modela detekcije plivača i raspoznavanje plivačkih tehnika na slikama, nazvane UNIRI-SWM.

U četvrtom poglavlju ispitali smo mogućnost transfera znanja i korištenja modela Mask R-CNN i YOLOv3 koji su naučeni za detekciju osoba na skupu slika opće namjene za detekciju plivača. Analizirani su i objašnjeni kvalitativni rezultati detekcije i naglašena je potreba učenja modela za detekciju plivača na odgovarajućem skupu slika iz domene plivanja. Eksperimentalno je pokazano značajno poboljšanje rezultata detekcije plivača nakon dodatnog učenja modela YOLOv3 na pripremljenom skupu slika UNIRI-SWM iz domene plivanja te da se izrađena baza može uspješno iskoristiti za učenje modela za detekciju plivača. Rad završava zaključkom i planom budućih istraživanja.

2. POSTOJEĆE BAZE SLIKA ILI VIDEA IZ DOMENE SPORTA

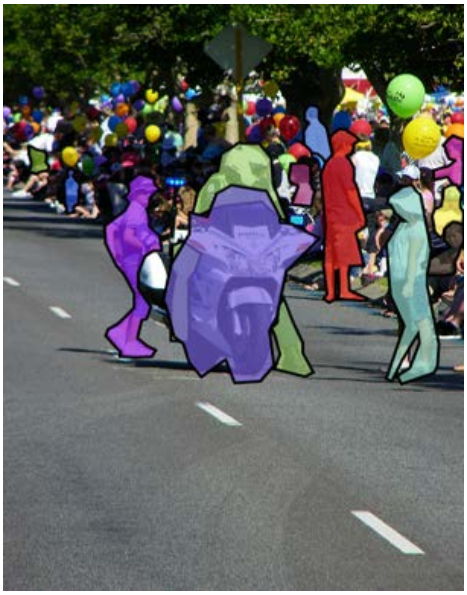
Postoji veći broj javno dostupnih skupova podataka za učenje modela za različite zadatke računalnog vida, poput detekcije objekata ili semantičke segmentacije i za različite domene kao što su medicina, svakodnevni život ili autonomna vožnja. Među najpoznatijim skupovima podataka su COCO (Lin *et al.*, 2014) i ImageNet (Deng *et al.*, 2009), koje sadrže veliki broj označenih fotografija različitih objekata u prirodnom okruženju (kontekstu). Primjerice, COCO baza fotografija trenutno sadrži oko 330.000 fotografija od kojih su preko 200.000 označene s barem jednom oznakom objekta iz 80 kategorija (Lin *et al.*, 2014).

Fotografije mogu biti označene na razini klasa koje postoje na sceni, ili na razini objekta tako da su objekti označeni graničnim okvirima ili isticanjem područja slike (piksela) koje odgovara nekom objektu. Na slici 1. dan je primjer slike na kojem su objekti označeni s graničnim okvirima koji su prikladni za zadatak detekcije osoba, a na slici 2. je dan primjer na kojem je svaki piksel, odnosno segment pridružen odgovarajućem objektu tako da je prikladna za zadatak detekcije i segmentacije objekata.

Slika 1. Detekcija osoba

Izvor: Sambolek i Ivasic-Kos (2021)

Slika 2. Primjer detekcije i segmentacije objekata iz COCO dataseta



Izvor: Lin *et al.* (2014)

Pored baza opće namjene, postoje baze slika specijalizirane za pojedinu domenu. Primjerice baza slika za detekciju i klasifikaciju životinja - KTH, (Schüldt *et al.*, 2004) ili živog svijeta - iNaturalist, (Van Horn *et al.*, 2018), baza za detekciju i klasifikaciju bolesti na rendgenskim snimkama pluća (NIH) (Wang *et al.*, 2017), ili baza za detekciju osoba u termalnim slikama - UNIRI-ITD, (Kristo *et al.*, 2020).

U domeni sporta češći su skupovi podataka koji uključuju označene video sekvence. Primjerice za zadatak klasifikacije sportskih scena popularna je baza Olympic Sports Dataset (Niebles, 2010) i SVW (Safdarnejad *et al.*, 2015) uključuju kratke sekvence akcija koje se odnose na 16, odnosno 30 sportova. Postoje i specijalizirane baze slika i videa koje se odnose na pojedini sport, npr. UNIRI-

HBD (Ivašić-Kos i Pobar, 2018) za detekciju sportaša i akcija u rukometu, u košarci (Ramanathan *et al.*, 2016), te u odbojci (Ibrahim *et al.*, 2016).

Također, sa domenom sporta se mogu povezati i neke od javno dostupnih baza slika poput KTH (Schüldt *et al.*, 2004) i Weizmann (Blank *et al.*, 2005) čiji primarni fokus nije sport, ali sadrže razne scene za detekciju aktivnosti osoba. Npr. skup podataka KTH uključuje šest klasa aktivnosti osoba (hodanje, džoging, trčanje, boks, mahanje i pljeskanje) koje izvodi 25 ljudi u četiri različita scenarija. Skup Weizmann ima 10 akcija od kojih se neke poput trčanja, skoka, skoka u mjestu, i preskakanja mogu povezati sa sportom. Svaku od ovih akcija izvodi 9 glumaca pa u bazi ukupno ima 90 videozapisa. Obje ove baze predstavljene su početkom 21. stoljeća, i u usporedbi s današnjim skupovima podataka sadrže malen broj klasa i uzoraka snimljenih u laboratorijskim uvjetima te u niskoj rezoluciji.

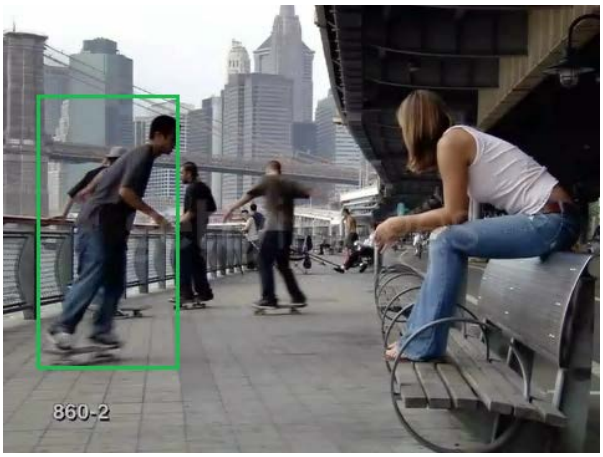
Suprotno tome, skupovi poput HACS (Zhao *et al.*, 2019) i Kinectics 700-2020 (Smaira *et al.*, 2020) su snimljeni u realnim uvjetima i imaju znatno više klasa i podataka. Primjerice, Kinectics je velik skup podataka (s 400 do 700 klasa koje odgovaraju različitim aktivnostima ljudi, ovisno o verziji) koji sadrži ručno označene videozapise preuzete s YouTube-a.

U nastavku će biti detaljnije opisani popularna skupa podataka u sportskoj domeni: UCF Sports Action Data Set (Soomro i Zamir, 2014), Olympic Sports Dataset (Niebles, 2010) i Sports-1M (Karpathy *et al.*, 2014).

2. 1 UCF Sports Action Data Set

Skup podataka UCF Sports (Rodrigez *et al.*, 2008) sastoji se od 150 sekvenci različitih akcija prikupljenih iz 10 različitih sportova koji se obično prikazuju na televizijskim kanalima. Akcije uključuju skokove u vodu, golf zamah, udaranje nogom, podizanje utega, jahanje konja, trčanje, vožnju *skateboarda*, vježbe na gimnastičkom konju, vježbe na vratilu i hodanje, a odvijaju se u različitim okruženjima (dvorana, igralište, priroda). Broj sekvenci nije jednak za svaku klasu te su neke akcije kratke (udarac nogom) a neke duže (trčanje). Sekvence su snimljene pri 10 sličica u sekundi, trajanja od 2 do 14s. Sve slike snimljene su u realnom okruženju, samo neke imaju samo objekt od interesa a neke imaju složeniju scenu koja uključuje i druge osobe koje ne obavljaju promatranu aktivnost. Na slici 3. dan je primjer scene sa više ljudi, ali je označena samo jedna osoba koja izvodi danu akciju (slika 3). Primjer različitih akcija iz skupa UCF Sports prikazan je na slici 4.

Slika 3. Primjer označene akcije skateboarding u skupu UCF Sports.



Izvor: Soomro i Zamir (2014)

Slika 4. Primjer označenih slika iz UFC Sport skupa podataka

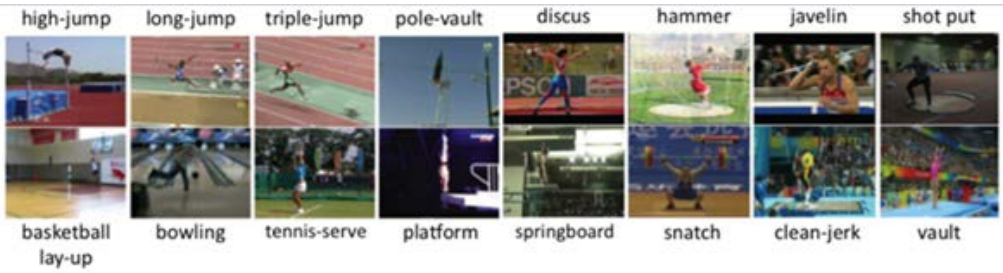
Izvor: Soomro i Zamir (2014)

2. 2 Olympic Sports Dataset

Olympic Sports Dataset (Niebles, 2010) sadrži video sekvence sportaša koji se bave sa 16 različitih sportova. Sadrži po 50 videozapisa iz svake od 16 klasa: skok u vis, skok u dalj, troskok, skok s motkom, bacanje diska, bacanje kladiva, bacanje koplja, bacanje kugle, košarkaški dvokorak, kuglanje, teniski servis, skokovi u vodu-platforma, skokovi u vodu-odskočna daska, dizanje utega-trzaj, dizanje utega-izbačaj i skok u gimnastici. Videozapisi su preuzeti s YouTubea i sadrže realne scene. Označeni su uz pomoć *Amazon Mechanical Turka*. Scene su snimljene u realnom okruženju za odgovarajući sport, sportska dvorana ili sportski teren a pri tom je sportaš snimljen iz različitih kutova i na različitoj udaljenosti od kamere. Na nekim snimkama su prisutni i osobe koje ne obavljaju akciju od interesa što dodatno komplicira zadatak detekcije ili raspoznavanja akcija. Oznaka klase je pridružena čitavoj sekvenci, tako da objekti nisu označeni graničnim okvirima. Slika 5 prikazuje po jedan primjer iz svake klase u skupu podataka Olympic sport. Iako svaka sekvenca predstavlja pojedinačnu akciju, većina akcija poput skoka u vis, skoka u dalj i troskoka je složena pa tako npr. sekvence košarkaškog dvokoraka uključuju vođenje lopte, skok i ubacivanje lopte u koš, a

sekvence iz klase skoka u dalj pokazuju sportaša koji prvo stoji na mjestu u pripremi za skok, nakon čega slijedi trčanje, skakanje, slijetanje i konačno ustajanje.

Slika 5. Olympic Sport skup podataka



Izvor: Niebles (2010)

2. 3 Sports-1M

Skup podataka Sports-1M (Karpathy *et al.*, 2014) sadrži preko milijun URL-a YouTube videozapisa, na koje je automatski nadodano 487 oznaka sportova pomoću *YouTube Topics API-ja*.

Akcije su razne, snimane u teretanama, bazenima, sportskim dvoranama, cestama, šumama, skijalištima i drugim mjesta s kojima se ljudi svakodnevno susreću. Neki primjeri sportskih scena i pridodane oznake prikazani su na slici 6. S obzirom da je u ovom skupu naglasak na razlikovanju vrsta sporta kao što su tenis, hokej, plivanje, vaterpolo i skijanje, čitave sekvence su označene samo oznakom sporta, dok pojedine radnje koje se pojavljuju unutar sporta ili u više sportova, poput trčanja, skoka, dodavanja, zamaha i ostalog nisu označene. Također nisu označeni niti pojedini sportaši ili druge osobe na sceni, pa se skup bez dodatne pripreme podataka i označavanja ne može koristiti za učenje detektora sportaša, već samo za klasifikaciju sportova.

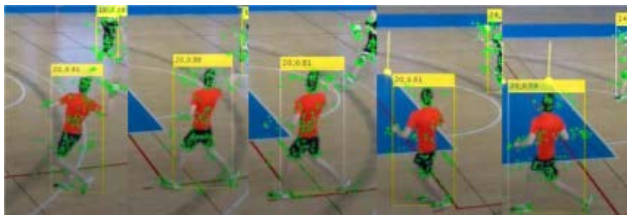
Slika 6. Sports-1M skup podataka

Izvor: Karpathy *et al.* (2014)

Pored navedenih primjera, postoje i skupovi podataka iz sportske domene koji su fokusirani na analizu sportskih rezultata i ne uključuju video zapise, poput rezultata *baseballa* (Lahman, 2017) ili na analizu pojedinih sportova ili zadataka unutar sporta poput detekcije akcija ili praćenja igrača. Primjerice za nogomet, skup SoccerNet (Giancola *et al.*, 2018) sadrži anotacije za 500 nogometnih utakmica u kojem su označeni vremenski trenuci ključnih događaja (gol, karton i zamjena) u svrhu učenja modela za njihovu detekciju, dok skup (Pettersen *et al.*, 2014) uz video zapise utakmica sadrži i senzorske podatke očitanja pozicije pojedinih igrača, s glavnom svrhom razvoja i testiranja modela za praćenje osoba (nogometaša) u videu.

Primjer prilagođenog skupa podataka može se naći i za rukomet (Pobar i Ivašić-Kos, 2020), gdje su istraživači snimali rukometne treninge organizirane u zatvorenom prostoru u sportskoj dvorani ili na otvorenom terenu. Skup podataka sastoji se od 751 videozapisa s promjenjivim brojem igrača, u prosjeku njih 12, koji istovremeno izvode razne akcije. Na svakoj sekvenci posebno je označen jedan igrač koji izvodi rukometnu akciju od interesa poput dodavanja, šutiranja, skoka ili vođenja lopte (slika 7). Prizori su snimljeni u dvorani i na otvorenom igralištu pomoću nepokretnih GoPro kamera postavljenih na lijevoj ili desnoj strani igrališta.

Slika 7. Primjer slika iz baze za rukomet



Izvor: Ivašić-Kos i Pobar (2018)

U današnje vrijeme, servisi poput YouTubea olakšavaju prikupljanje video materijala različitih sportova, međutim da bi takvi videozapisi bili korisni za strojno učenje nužno je da su označeni, a to je dugotrajan i zamoran posao pa za veliki broj sportova još ne postoji reprezentativni skup podataka koji bi se mogao koristiti za strojno učenje modela za detekciju sportaša ili raspoznavanje njihovih akcija.

Za uspješno učenje modela za raspoznavanje akcija u nekom sportu potreban je skup podataka koji realno predstavlja različite situacije i akcije u tom sportu i koji je dovoljno velik tako da za svaku akciju postoji dovoljno primjera iz različitih kutova snimanja, različitih sportaša itd. tako da su istraživači često prisiljeni stvoriti vlastite skupove za sport koji istražuju. U idućem ćemo poglavlju opisati stvaranje vlastite baze označenih snimaka plivanja koja će se koristiti za detekciju plivača i raspoznavanje plivačkih tehnika.

3. PRIPREMA BAZE ZA DETEKCIJU PLIVAČA I PLIVAČKE TEHNIKE

Istraživanjem smo utvrdili kako ne postoji baza plivača koja je snimljena s ciljem učenja modela za raspoznavanje akcija ili detekciju plivača te smo pristupili izradi vlastite baze.

U ovom radu koristimo četiri akcijske kamere koje su postavljene na točno određene, predefinirane lokacije, opisane niže. Kamere su postavljene na visinu od 30 cm, 2 m i 13 m od razine vode te 2,20 m ispod razine vode.

Snimanje je izvršeno u uvjetima treninga i u uvjetima natjecanja.

U uvjetima treninga, istu plivačku stazu unutar bazena najčešće dijeli veći broj plivača, zbog ograničenog kapaciteta bazena, a velikog broja zainteresiranih plivača. Snimani su uglavnom plivači u stanju razvoja tehnike, te se iz tog razloga plivačka tehnika znatno razlikuje od plivača do plivača i generalno je na nižoj razini. Profesionalnih plivača s bolje razvijenom tehnikom je na treninzima općenito bio manji broj.

Kod snimanja koja su izvršena u uvjetima natjecanja, u jednoj plivačkoj stazi se istovremeno nalazi samo jedan plivač. Snimani su plivači starije dobi i s bolje razvijenom plivačkom tehnikom od plivača snimanih na treninzima. Kod snimanja natjecanja znatno je izraženiji problem smetnji u vidu kapljica koje i pjene koje pojačano stvaraju plivači za vrijeme natjecanja u odnosu na treninge.

Snimljeno je ukupno četiri sata materijala, koji zauzimaju 170 GB.

3. 1 Položaj i postavke kamera

Snimanje se odvijalo na bazenu duljine 25 m, širine 25 m, dubine 2,20 m s ukupno deset pruga širine 2,5 m. Plivači su snimani s više pozicija te s kamerama iznad i ispod vode, jer se potpuno različiti dijelovi neke plivačke tehnike odvijaju pod i nad vodom. U snimanju su korištene četiri akcijske kamere GoPro Hero 4, raspoređene na bazenu prema shemi prikazanoj na Slici 8.

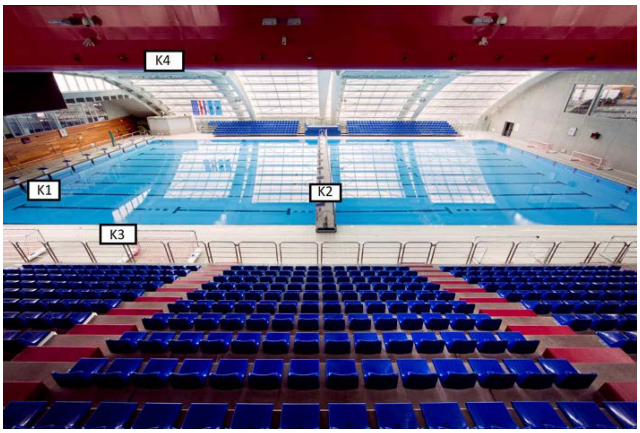
Kamera K1 je fiksirana na zidu bazena na kraju plivačke staze i uronjena u vodu na dubini od 2,20 m. Zbog gustoće i neprozirnosti vode, plivači su bili vidljivi do maksimalne udaljenosti 10 m od kamere.

Kamera K2 bila je postavljena pored startnog bloka na ponton, 30 cm iznad razine vode. Kamera K2 se nalazi u istoj plivačkoj stazi kao kamera K1, ali na suprotnom kraju staze.

Kamera K3 nalazila se na bočnoj strani bazena, 15 m od početka plivačkih staza, na visini od 2 m iznad vode.

Kamera K4 nalazila se na stropu bazena na visini od 13 m približno u sredini bazena.

Slika 8. Bazen i položaj kamera na kojem je vršeno snimanje



Izvor: autori

Broj sličica u sekundi (FPS) svih kamera je bio postavljen na 60, što je smanjilo zamućenje zbog pokreta s obzirom da se radi o akcijskim snimkama.

Za svaku kameru posebno je, s obzirom na njen položaj, odabrana odgovarajuća širina kuta snimanja, kako bi se obuhvatilo što veće područje samog bazena gdje se nalaze plivači, a smanjilo područje oko bazena.

S obzirom na ograničeno trajanje baterije korištenih kamera, koja omogućuje neprekinuto snimanje otprilike 30ak minuta videa pri 4K rezoluciji, te nepristupačnu poziciju kamere K4 koja ne omogućuje laku izmjenu baterija u tijeku sportskog događanja, broj i trajanje snimki tom kamerom je manji nego kod ostalih kamera.

Dodatna teškoća sa svim kamerama, a posebno s kamerom K2 koja je bila uronjena u vodu, je bila kontrola tijekom snimanja, zbog nemogućnosti ostvarivanja kvalitetne veze između kamere i mobilnog telefona ili drugog ekrana. Zbog uvjeta na bazenu, veza se učestalo prekidala te nije bilo moguće pratiti je li se kut snimanja promijenio uslijed vodene struje ili plivača koji bi pomaknuli kameru, je li leća objektiva čista, je li se baterija potrošila i slično. Iz tog razloga, jedan dio snimki je nakon pregleda morao biti odbačen kao neupotrebljiv.

3. 2 Problemi kod snimanja

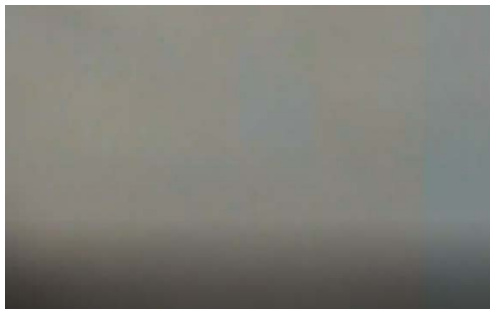
Snimke snimane pod vodom su vrlo često mutne zbog gustoće vode i lošijih svjetlosnih uvjeta te zbog nečistoće same vode i raznih spojeva klora koji vodu zamućuju. S druge strane, objektivi kamera K2 i K3 koje su bile iznad vode, ali blizu bazena, često bi bili mokri ili bi na njima bile kapi vode zbog prskanja vode dolaskom plivača (slika 9). U tim slučajevima daljnja snimka ne bi bila čista kao što bismo očekivali, a često je snimka bila i posve neupotrebljiva nakon prolaska samo jednog plivača. Primjer takve situacije prikazan je na slici 10. Dodatno, prirodne pojave, poput odsjaja Sunca, nisu se mogle eliminirati, što je sve utjecalo na broj i kvalitetu snimki koje su uključene u izrađenu bazu.

Slika 9. Primjer snimke kada je objektiv djelomično prekriven kapljicom vode

Izvor: autori

Slika 10. a) Prilaz kameri b) Izgled leće kamere nakon dolaska plivača

(a)



(b)

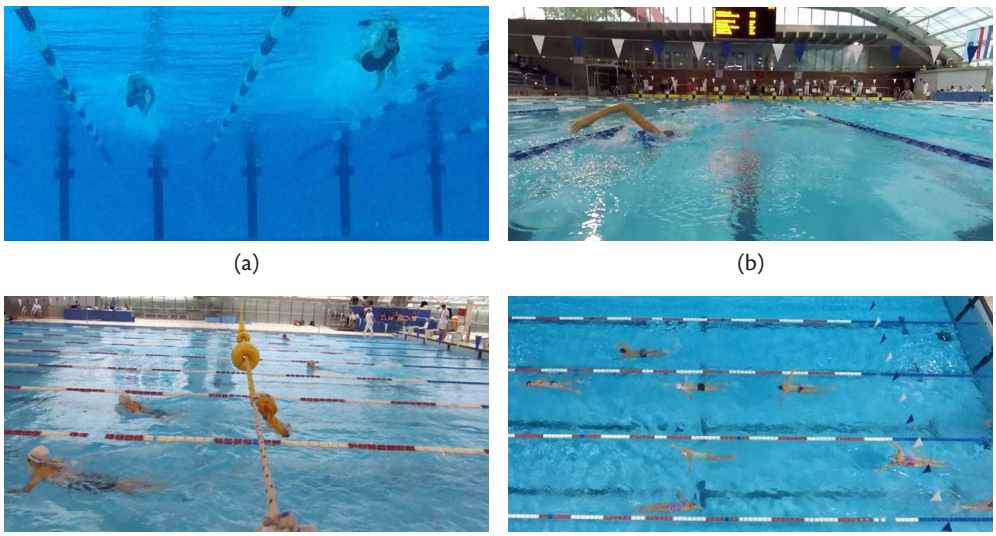
Izvor: autori

3. 3 Dobiveni *dataset*

Od ukupno 4 sata materijala koji je snimljen, ručnim pregledom su odabrane snimke koje su ocijenjene kao dovoljno kvalitetne za uključivanje u bazu. Snimke uključuju četiri plivačke tehnike: prsno, leđno, kraul i delfin. Za svaku tehniku je snimljeno i uključeno u bazu između 150 i 230 snimki, prosječnog trajanja 1,41 s. Baza je nazvana UNIRI-SWM.

Primjer kadrova dobivenih kamerama K1-K4 prikazan je na slici 11. a)-d).

Slika 11. Primjeri snimki kamerama K1-K4: a) kamera K1, 2,20m pod vodom, b) kamera K2, 30 cm iznad vode, c) kamera K3, 2 m iznad vode, d) kamera K4, 13 m iznad vode.



(a) (b)

(c) (d)

Izvor: autori

3. 4 Predobrada podataka

Nakon završetka snimanja i pregleda materijala, uslijedila je predobrada materijala koji će se označiti i pripremiti za strojno učenje modela za raspoznavanje plivačkih tehnika. Kako je bilo nemoguće precizno daljinski upravljati kamerama tijekom snimanja odabranog događaja (treninga ili natjecanja), kamere su nakon uključivanja snimale bez prekida i za vrijeme kratkih odmora ili izmjena plivača. Snimke su stoga rezane na kraće dijelove koji sadrže relevantne plivačke tehnike. Jedan je kadar (od početnog dijela do prekida snimke) tada bio čitavo plivanje jednakom tehnikom. Kadrovi su trajali po nekoliko minuta. Nakon analize snimke, nismo dobili zadovoljavajuće rezultate detekcije.

S obzirom na to da je plivanje ciklički sport, što znači da se čitavo vrijeme ponavlja jednaka radnja (svaki je zaveslaj jednak, ruke se okreću u istom smjeru, na isti način, a noge udaraju bez prekida), snimke su rezane upravo nakon svakog zaveslaja kako bi se dobila čim manja jedinica pokreta. Primjenom ove metode, dobili smo veću količinu podataka koja je dala bolje rezultate detekcije. Iako bolji, nedovoljno dobri jer je uspješnost bila manja od četrdeset posto.

Za pregledavanje i rezanje videa korišten je besplatan alat VLC media player[1] (slika 12. a) koji nudi mogućnost izrezivanja i spremanja odabranog dijela video zapisa pritiskom na jednu tipku, što olakšava mukotrpan ručni posao rezanja videa.

[1] www.videolan.org

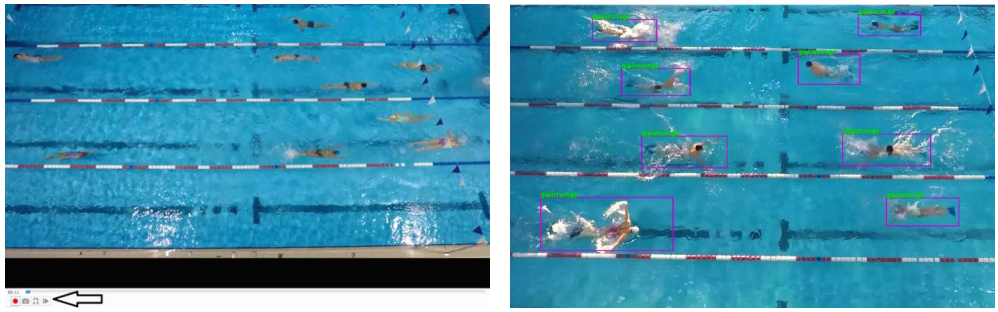
Kako bez dodatnog označavanja plivača razina detekcije nije bila zadovoljavajuća, svakom je od novonastalih videozapisa dodijeljena odgovarajuća oznaka, to jest jedna od četiri klase koje odgovaraju plivačkim tehnikama (delfin, leđno, prsno i kraul). Snimke za svaku od tehnika bile su imenovane ime_tehnikeRedniBroj, npr. *delfin1*, *delfin2*, …

Ovim postupkom, stvorena je velika količina podataka, između 15 i 25 snimki za isplivanih 25 metara.

Za zadatak detekcije plivača, na pojedinim okvirima dobivenih snimki označena su područja slike koje pripadaju plivačima (slika 12. b). Nakon označavanja, uz svaku označenu sliku postoji lista koordinata, u tekstualnom obliku, za svaki označeni okvir na slici.

S obzirom na to da je prilikom snimanja korišten velik broj sličica u sekundi (60 fps) te su uzastopne sličice vrlo slične, za označavanje se koristio svaki 20. frame izvornih snimki.

Slika 12. Predobrada podataka. a) izrezivanje video segmenata u alatu VLC Media player b) označavanje područja slike na kojima su plivači pravokutnim okvirima



(a) (b)

Izvor: autori

4. EKSPERIMENT

Na formiranom skupu podataka UNIRI-SWM izvršen je preliminarni eksperiment s ciljem detekcije plivača. Za detekciju su korištene konvolucijske neuronske mreže YOLOv3 i Mask R-CNN koje su naučene za detekciju osoba na bazi slika iz svakodnevnog života (MS COCO). Ideja je koristiti transfer znanja i provjeriti koliko modeli za detekciju osobe naučeni na generalnim slikama mogu biti uspješni na slikama iz domene sporta, točnije plivanja i detektirati osobu koja pliva, dakle sportaša-plivača.

Modeli strojnog učenja su u pravilu dobro adaptirani na domenu na kojoj su učeni i daju dobre rezultate na slikama koje sliče slikama iz skupa za učenje, međutim ne pokazuju istu efikasnost kada se primjenjuju na slikama izvan domene. Zbog toga je neuronska mreža YOLOv3 dodatno naučena na dijelu skupa UNIRI-SWM kako bi se naučio model YOLOv3 (plivači), prilagođen za detekciju plivača. Učenje je izvedeno na 195 označenih slika iz skupa za učenje, na kojima je bilo ukupno 528 pojavljivanja plivača. Izvršeno je u 15000 iteracija, uz parametre stope učenja (engl.

learning rate) 0.001 te veličinom ulazne slike 608x608 piksela. Za testiranje modela korišten je dio skupa UNIRI-SWM za testiranje od 84 slike, s ukupno 239 označenih plivača.

S obzirom da su za učenje mreže Mask R-CNN uz označene pravokutne okvire koji omeđuju objekte potrebni i precizni obrisi objekata, u ovom eksperimentu Mask R-CNN nije dodatno učen na skupu slika UNIRI-SWM.

Rezultati detekcije su kvalitativno uspoređeni i objašnjeni na većem broju primjera kod testiranja performansi modela YOLOv3 i Mask R-CNN bez dodatnog učenja na skupu UNIRI-SWM. Usporedba performansi modela YOLOv3 bez učenja na dijelu skupa UNIRI-SWM i modela YOLOv3 (plivači) nakon učenja na tom skupu dana je kvantitativno s obzirom na standardne metrike za ocjenu rezultata klasifikacije.

Standardne metrike koje se koriste za ocjenu uspješnosti klasifikacije i detekcije su:

- Preciznost – broj ispravnih detekcija s obzirom na sve detekcije

$$Prec = \frac{TP}{TP + FP}$$

- Odziv – omjer broja ispravnih detekcija s brojem detekcija koje su morale biti prijavljene

$$Rec = \frac{TP}{TP + FN}$$

- F1- mjera - harmonijska sredina preciznosti i odziva, često se koristi kao mjera koja pokazuje pravu točnost modela

gdje je

- TP (*True positive*) – broj slika na kojima su objekti točno detektirani (u radu, klasifikacija plivača pod klasu osobe)
- FP (*False positive*) – broj slika netočno detektiranih (u radu, nije plivač, a označen kao osoba)
- FN (*False negative*) – broj slika pogrešno ne detektiranih (u radu, plivač koji nije označen kao plivač)

Kako je provođenje ovog eksperimenta i testiranje već naučenih modela detekcije na našim slikama za računalo kućnih performansi zahtjevan posao, taj je dio odrađen na računalima s GPU jedinicom u laboratoriju Odjela za informatiku Sveučilišta u Rijeci.

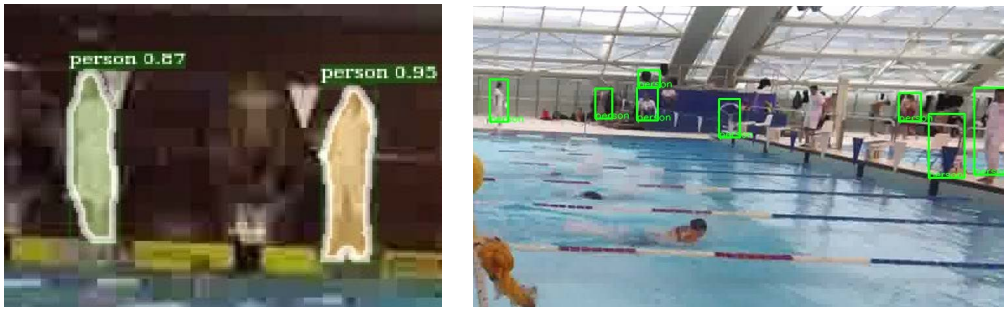
4. 1 Analiza rezultata

U prvom dijelu prikazani su rezultati testiranja modela Mask R-CNN i YOLOv3 naučenih na bazi opće namjene, koja sadrži scene i objekte iz svakodnevnog života, ali koja nije prilagođena za domenu plivanja ili sporta. Primjer detekcije korištenjem Mask R-CNN i YOLOv3 prikazan je na slici 13.

Iz fotografije koje je detektor Mask R-CNN obradio vidljivo je da su modeli za objekte koji su postojali u skupu za učenje dobro naučeni jer se osobe koje su na suhom detektirane i segmentirane vrlo točno, iako su prilično udaljene od same kamere i pored nejednolične pozadine. Riječ je o sucima, osobama koje šeću po rubu bazena i koje nalikuju pješacima na kojima je uglavnom model i učen, slika 13. a). Za razliku od sudaca, plivače koji prilaze kameri, dakle nalaze se bliže kameri, isti Mask R-CNN detektor uopće ne detektira.

Slično uočavamo i u slučaju YOLOv3 detektora koji detektira veći broj osoba izvan bazena, ali ne i plivače blizu kamere, slika 13. b).

Slika 13. Primjer detekcije osoba na rubu bazena korištenjem a) Mask R-CNN b) YOLOv3

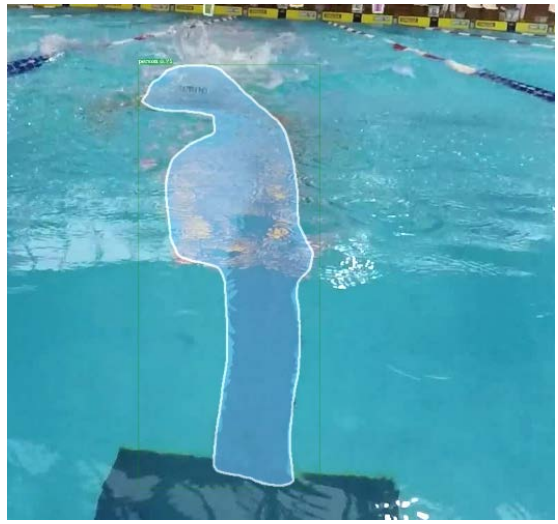


(a) (b)

Izvor: autori

Također, Mask R-CNN detektor je u velikom broju slučajeva imao lažne detekcije osoba, tj. detektirao je da postoji osoba tamo gdje ona nije bila (engl. *false positive* – FP detekcija). Npr. položaj crne oznake na bazenu i vrtlog vode stvorio je siluetu koja je lažno prepoznata kao osoba (slika 14).

Slika 14. Lažna detekcija osobe



Izvor: autori

Na slici 15., detektor nije prepoznao glavu plivača kada je plivač bio blizu kameri, dok na nekim slikama prepoznaje osobu samo na osnovu dijela ruke koja izviruje iz vode dok plivač pliva kraul. Vjerojatno je u ovom slučaju problem pjena koju stvara plivač prilikom plivanja i koja značajno mijenja izgled lica osobe. S druge strane, osoba u daljini, sudac na rubu bazena je uspješno detektiran (slika 15).

Slika 15. Primjer uspješne detekcije udaljene osobe i neuspješne detekcije plivača

Izvor: autori

Mreškanje vode, odsjaj i zrcaljenje zbog utjecaja svjetla te kapljice vode na objektivu također su ponekad pogrešno detektirane kao neki objekt (slika 16).

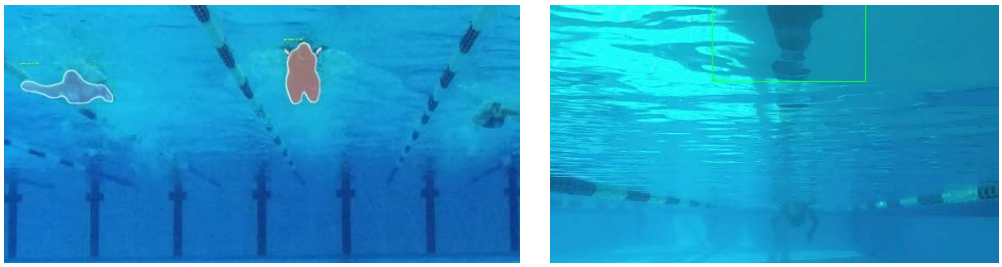
Slika 16. a) lažna detekcija čamca (Mask R-CNN) b) lažna detekcija osobe (YOLOv3)

(a) (b)

Izvor: autori

Pomalo iznenađujuće, detekcija s kamerom postavljenom ispod razine vode kod Mask R-CNN češće je dala bolje rezultate, kao na primjeru na slici 17. a). Kod YOLO detektora to nije bio slučaj i u slučaju podvodnih snimki najčešće ili nije bilo detekcija ili su lažno detektirane osobe. Na slici 17. b) prikazan je primjer detekcije s YOLOv3 gdje je vidljiva lažna detekcija osobe na vrhu slike te propuštena detekcija plivača pri dnu slike.

Slika 17. Detekcija osoba s kamerom postavljenom na dno bazena, a) Mask R-CNN, b) YOLOv3

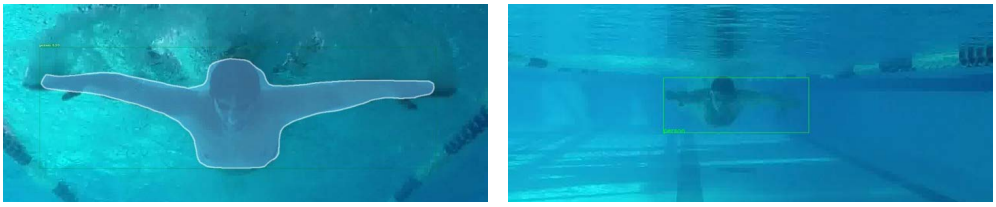


(a) (b)

Izvor: autori

Detekcija plivača bila je najbolja kada nije bilo smetnji u vidu dodatnih osoba ili objekata sa strane, s kamerom ispod ili iznad vode (slika 18, slika 19).

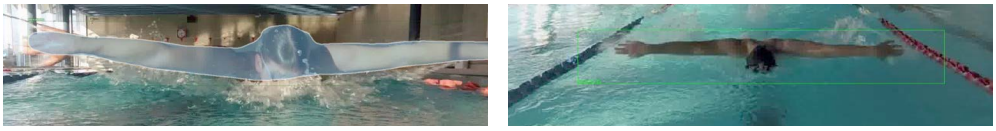
Slika 18. Potpuna, pravila detekcija osobe pod vodom s detektorom a) Mask R-CNN b) YOLOv3



(a) (b)

Izvor: autori

Slika 19. Potpuna, pravilna detekcija plivača a) Mask R-CNN b) YOLOv3



(a) (b)

Izvor: autori

Međutim, detekcija plivača nije bila konzistentna čak ni u slučaju vrlo slične poze (slika 20. a, b i c), gdje su plivači nekad detektirani u potpunosti (a), samo djelomično (b), ili nikako (c).

Slika 20. Rezultati detekcije plivača u sličnoj pozi pomoću YOLOv3

(a) (b) (c)

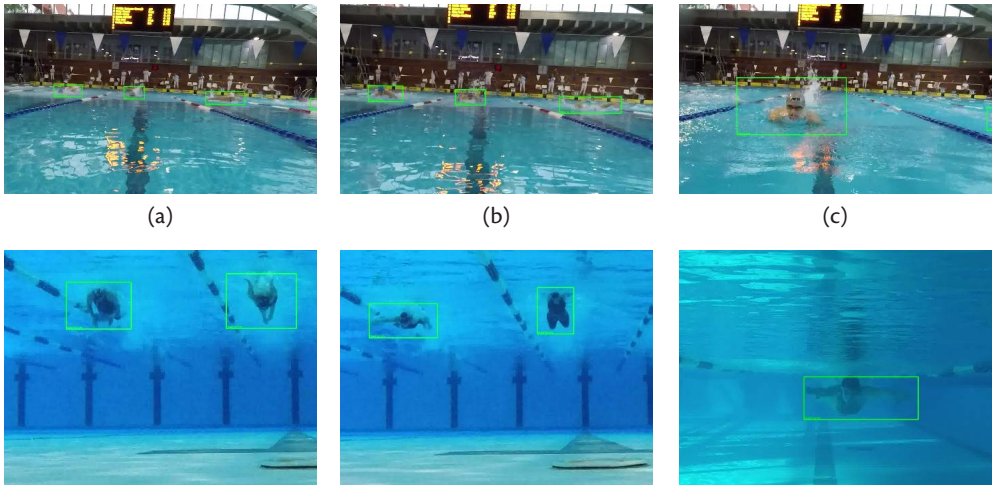
Izvor: autori

Dobiveni rezultati modelima koji su naučeni na bazi slika iz svakodnevnog života nisu dovoljno dobri za zadatak detekcije plivača, a niti za zadatak klasifikacije plivačkih tehnika.

4. 2 Analiza rezultata modela naučenog na bazi UNIRI-SWM

Nakon učenja mreže YOLOv3 na 195 slika iz baze UNIRI-SWM, detekcija plivača je postala konzistentna i pouzdana, s malim brojem i lažnih detekcija i propuštenih detekcija (slika 21).

Slika 21. Primjeri detekcije plivača korištenjem mreže YOLOv3 nakon učenja na skupu UNIRI-SWM



(a) (b) (c)

(a) (b) (c)

Izvor: autori

U tablici 1 prikazani su kvantitativni rezultati testiranja neuronske mreže YOLOv3 naučene na skupu slika COCO iz svakodnevnog života (označeno YOLOv3) i iste mreže naučene na skupu od 193 slike za učenje iz baze UNIRI-SWD, (označeno YOLOv3 (plivači)). Za testiranje je korišten skup od 93 slike iz baze UNIRI-SWM koje nisu korištene za učenje modela.

Tablica 1. Rezultati detekcije plivača korištenjem mreže YOLOv3 naučene na skupu COCO i YOLOv3(plivači) naučene na skupu UNIRI-SWM

Metrika	YOLOv3 / COCO	YOLOv3 (plivači) / UNIRI-SWM
TP	3	213
FP	105	33
FN	236	26
Preciznost	2,78 %	86,59 %
Odziv	1,26 %	89,12 %
F1	1,73 %	87,84 %

Izvor: autori

Rezultati pokazuju značajno poboljšanje performansi u detekciji plivača nakon učenja YOLOv3 modela na skupu za učenje UNIRI-SWD. Bez dodatnog učenja na skupu UNIRI-SWD, YOLOv3 uspješno detektira osobe koje podsjećaju na pješake, ali ne uspijeva detektirati plivače u bazenu. To je vidljivo iz malog broja TP, tek 3 i velikog broja FP koji se uglavnom odnosi na detektirane osobe uz rub bazena kao što su suci, dok je zadatak bio detektirati plivače koje u najvećem broju slučajeva nisu detektirani (veliki broj FN). S druge pak strane, nakon dodatnog učenja na relativno malom skupu za učenje, YOLOv3 (plivači) u najvećem broju slučajeva uspješno detektira plivače u bazenu (veliki TP, i relativno mali FP i FN). Sve to se ogleda i u metrikama preciznosti, odziva i F1 koje su nakon učenja na skupu za učenje UNIRI-SWD značajno popravljene. Napomenimo kako je u ovom eksperimentu cilj bio detektirati plivače, a ne i tehniku kojom plivaju.

5. ZAKLJUČAK

U radu je opisan postupak izrade vlastite baze slika za strojno učenje modela za detekciju plivača. Opisali smo postupak snimanja plivača koristeći različite pozicije kamere i poglede na plivače koji su karakteristični za bazen. Naveli smo probleme koji su se pojavljivali tijekom snimanja zbog vodenog medija kao što je šum, prskanje vode i nepovezanost s kamerom kada je postavljena pod vodom te zatvorenog grijanog prostora (magljenje kamere) i umjetne rasvjete ili intenzivnih prodora svjetlosti kroz velike ostakljene površine. Po završetku snimanja izvršen je probir materijala kako bi se odabrale odgovarajuće slike za strojno učenje, te predobrada slika na način da se na njima označe plivači.

Na formiranom skupu podataka UNIRI-SWM testirana je uspješnost modela neuronskih mreža YOLOv3 i Mask R-CNN učenih na bazi slika iz svakodnevnog života (MS COCO) za detekciju plivača. Pokazalo se da ti modeli koji uspješno detektiraju osobe na generalnim slikama loše detektiraju osobe dok su u vodi i plivaju, odnosno plivače.

Zbog toga je neuronska mreža YOLOv3 dodatno naučena na dijelu skupa UNIRI-SWM te je napravljen model YOLOv3 (plivači) za detekciju plivača koji značajno premašuje rezultate polaznog modela.

Rezultati eksperimenta pokazuju da se detektori naučeni na generalnom skupu slika ne mogu direktno koristiti u domeni sporta, točnije plivanju jer postižu loše rezultate te da je bilo nužno formirati bazu za učenje modela. Također pokazalo da se s učenjem modela na slikama iz domene plivanja mogu značajno popraviti performanse modela i poboljšati rezultati detekcije plivača do razine da su upotrebljivi za daljnje analize tehnike i stilova plivanja.

Daljnja istraživanja obuhvaćat će novo snimanje iz jednakih (uklanjanje do sada uočenih šumova), ali i različitih pozicija kamere (moguće bolje detekcije iz novih pozicija). Također, pokušat će se izvršiti detekcija osobe koja pliva. Za potrebe takve detekcije, već se provodi snimanje akcijskim kamerama i senzorima postavljenim na plivačevo tijelo.

LITERATURA

Blank, M. et al. (2005, October). Actions as space-time shapes. In Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 (Vol. 2, pp. 1395-1402).

Dai, J. et al. (2016). R-fcn: Object detection via region-based fully convolutional networks. In Advances in neural information processing systems (pp. 379-387).

Deng, J. et al. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255).

Giancola, S. et al. (2018). Soccernet: A scalable dataset for action spotting in soccer videos. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 1711-1721).

He, K. et al. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

Ibrahim, M. S. et al. (2016). A hierarchical deep temporal model for group activity recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1971-1980).

Ivašić-Kos, M., Pavlić, M. and Pobar, M. (2009) Analyzing the semantic level of outdoor image annotation. Proceedings of MIPRO 2009 - 32nd International convention on information and communication technology, electronics and microelectronics, Opatija

Ivašić-Kos, M. and Pobar, M. (2018, November). Building a labeled dataset for recognition of handball actions using mask R-CNN and STIPS. In 2018 7th European Workshop on Visual Information Processing (EUVIP) (pp. 1-6).

Karpathy, A. et al. (2014). Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 1725-1732).

Krišto, M., Ivasic-Kos, M., & Pobar, M. (2020). Thermal object detection in difficult weather conditions using YOLO. IEEE Access, 8, 125459-125476.

Lahman, S. (2017), Lahman's Baseball Database, 1871-2017, v.2017, Comma-delimited version, http://seanlahman.com/files/database/baseballdatabank-master_2018-03-28.zip

Lin, T.Y. et al. (2014). September. Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.

Liu, W. et al. (2016, October). Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.

Niebles J.C., Olympic Sports Dataset, (2010). http://vision.stanford.edu/Datasets/OlympicSports/ (last access 6/12/2021).

Paul, M., Haque, S. M. and Chakraborty, S. (2013). Human detection in surveillance videos and its applications-a review. EURASIP Journal on Advances in Signal Processing, 2013(1), 1-16.

Pettersen, S.A. et al. (2014). Soccer video and player position dataset. In Proceedings of the 5th ACM Multimedia Systems Conference MMSys (pp. 18-23).

Pobar M. and Ivašić-Kos M. (2020). Active Player Detection in Handball Scenes Based on Activity Measures. Sensors, 20(5), 1475. https://doi.org/10.3390/s20051475.

Ramanathan, V. et al. (2016). Detecting events and key actors in multi-person videos. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3043-3053).

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Rodriguez, M. D., Ahmed, J. and Shah, M. (2008, June). Action mach a spatio-temporal maximum average correlation height filter for action recognition. In 2008 IEEE conference on computer vision and pattern recognition (pp. 1-8).

Safdarnejad, S. M. et al. (2015, May). Sports videos in the wild (SVW): A video dataset for sports analysis. In 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) (Vol. 1, pp. 1-7).

Sambolek, S. and Ivasic-Kos, M. (2021). Automatic Person Detection in Search and Rescue Operations Using Deep CNN Detectors. IEEE Access, 9, 37905-37922.

Schuldt, C., Laptev, I. and Caputo, B. (2004, August). Recognizing human actions: a local SVM approach. In Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. (Vol. 3, pp. 32-36).

Smaira, L. et al. (2020). A short note on the kinetics-700-2020 human action dataset. arXiv preprint arXiv:2010.10864.

Soomro, K. and Zamir, A. R. (2014). Action recognition in realistic sports videos. In Computer vision in sports (pp. 181-208). Springer, Cham.

Van Horn, G. et al. (2018). The inaturalist species classification and detection dataset. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8769-8778).

Wang, X. et al. (2017, July). Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In IEEE CVPR (Vol. 7).

Zhao H. et al. HACS (2019) Human action clips and segments dataset for recognition and temporal localization, in: Proceedings of the IEEE International Conference on Computer Vision, Institute of Electrical and Electronics Engineers Inc.: pp. 8667–8677. https://doi.org/10.1109/ICCV.2019.00876.

Professional paper
https://doi.org/10.31784/zvr.11.1.15
Received: 1. 12. 2021.
Accepted: 9. 12. 2021.

DATASET PREPARATION FOR SWIMMER DETECTION

Ivan Šimac
Teaching Assistant, Polytechnic of Rijeka, Vukovarska 58, 51000 Rijeka, Croatia; e-mail: isimac@veleri.hr

Miran Pobar
PhD, Assistant Professor, University of Rijeka, Department of Informatics, Radmile Matejčić 2, 51000 Rijeka, Croatia; e-mail: mpobar@uniri.hr

Marina Ivašić-Kos
PhD, Associate Professor, University of Rijeka, Department of Informatics, Radmile Matejčić 2, 51000 Rijeka, Croatia; e-mail: marinai@uniri.hr

SUMMARY

The large amount of data that is created every day can be used to develop artificial intelligence algorithms in the domain of computer vision that solve tasks such as image classification, face detection and action recognition. These datasets are most often created from videos and images downloaded from television channels or the YouTube social network and are collected and prepared for the appropriate task. We were interested in the task of detecting swimmers, so that the model could be used to recognize and improve swimming techniques. Although today there are huge open image databases like COCO and ImageNet, prepared for supervised machine learning and sports scene databases like Olympic Sports Dataset, UCF Action Sport dataset or Sport-1M that include images of more popular (watched) sports, none of them include images that could be used to make our swimmer detection model. Therefore, this paper describes the process of recording and collecting video material and preparing a set of UNIRI-SWM images for swimmer detection. The set includes shots of swimmers in real, situational training and competition conditions filmed by action cameras from different shooting angles. The paper presents the results of swimmer detection using deep convolutional neural networks Mask R-CNN and Yolo v3, learned in the set of general images before and after learning in the set UNIRI-SWM. The results show that after adjusting the model on the appropriate set of images from the swimming domain, very good results of swimmer detection can be achieved.

Key words: *person detection, convolutional neural network, data set, swimming*