

Detekcija i raspoznavanje osoba u otežanim vremenskim uvjetima primjenom infracrvene termovizije

Krišto, Mate

Doctoral thesis / Disertacija

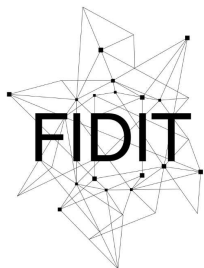
2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka / Sveučilište u Rijeci**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:195:956338>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-04-19**



Sveučilište u Rijeci
**Fakultet informatike
i digitalnih tehnologija**

Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Informatics and Digital Technologies - INFORI Repository](#)



SVEUČILIŠTE U RIJECI
ODJEL ZA INFORMATIKU

Mate Krišto

**DETEKCIJA I RASPOZNAVANJE OSOBA U
OTEŽANIM VREMENSKIM UVJETIMA
PRIMJENOM INFRACRVENE TERMOVIZIJE**

DOKTORSKI RAD

Rijeka, 2020.

SVEUČILIŠTE U RIJECI
ODJEL ZA INFORMATIKU

Mate Krišto

**DETEKCIJA I RASPOZNAVANJE OSOBA U
OTEŽANIM VREMENSKIM UVJETIMA
PRIMJENOM INFRACRvene TERMOVIZIJE**

DOKTORSKI RAD

Mentor: izv.prof. dr. sc. Marina Ivašić-Kos

Rijeka, 2020.

UNIVERSITY OF RIJEKA
DEPARTMENT OF INFORMATICS

Mate Krišto

**HUMAN DETECTION AND RECOGNITION
IN HARSH WEATHER CONDITIONS USING
INFRARED THERMAL IMAGING**

DOCTORAL THESIS

Rijeka, 2020.

Mentorica rada: izv.prof. dr. sc. Marina Ivašić-Kos

Doktorski rad obranjen je dana _____ na Odjelu za informatiku Sveučilišta u Rijeci, pred povjerenstvom u sastavu:

1. _____

2. _____

3. _____

4. _____

5. _____

Zahvale

"*I najduže putovanje počinje prvim korakom*", izreka je koja se pripisuje Lao Tseu, jednom od najznačajnijih kineskih filozofa koja u kontekstu mog doktorskog studija, posebice pripreme potrebnih materijala ima više značenja, od kojih su neka doista doslovna jer je bilo potrebno napraviti prvi korak kako bi bilo moguće termalnom kamerom snimiti osobe, dok su druga više u prenesnom značenju, ali ništa manje značajna.

Kao i kod svakog drugog projekta, uspjeh pojedinca, odnosno moj uspješan završetak doktorskog studija, ovisio je o ljudima kojima sam okružen, a na moju veliku sreću, okružen sam ljudima koji su vjerovali u moje ideje, čak toliko da su se dobrovoljno prijavili kao "testni subjekti" i pristali na snimanje scenarija po otežanim vremenskim uvjetima.

Zbog svega navedenog, najnezahvalnije je napisati zahvalu svima koji su pripomogli na ovom putovanju zvanom doktorski studij, stoga svima zahvaljujem na pomoći koju su mi pružili.

Unatoč tomu, nužno je zahvale uputiti onima koji su najviše doprinijeli, a to je u prvom redu moja supruga koja je iskazala beskrajno strpljenje i razumijevanje zbog moje mentalne, a povremeno i fizičke odsutnosti zbog posvećenosti istraživanju za potrebe doktorskog studija te u konačnici izrade ovog doktorskog rada. Osim supruge, strpljenje i razumijevanje iskazali su i moji sinovi, kojima također zahvaljujem.

Osim obitelji, zahvalu sam dužan uputiti kolegama i nadređenima koji su mi hrabro ustupili termalnu kameru i popratnu opremu kako bih mogao provesti nužna snimanja za kreiranje potrebnog skupa slika, a samim tim i stvoriti uvjete za izvođenje eksperimenata. No, što je oprema bez ljudi? Kao i u svemu, bez prijatelja koji su bili spremni na ozeblinje zbog snimanja u zimskim uvjetima, kao i onih spremnih na ozbiljno tuširanje jakom zimskom kišom na svega nekoliko stupnjeva iznad nule, ne bi bilo slika potrebnih za kreiranje spomenutog skupa za učenje i testiranje modela za neuronsku mrežu. Stoga, zahvala ide i njima.

Zasebnu zahvalu upućujem i mentorici prof. Ivašić-Kos koja mi je svojim savjetima, mentorskim vodstvom i vlastitim iskustvom neizmjereno pomogla da doktorski studij uspješno privedem kraju, a krajnji rezultat je ovaj doktorski rad koji je *summa summarum* svih istraživanja i eksperimentiranja tijekom doktorskog studija.

Sažetak

Naslov doktorskog rada je ujedno i najbolji sažetak rada i u jednoj rečenici opisuje ono što se htjelo istražiti te eksperimentalno potvrditi ili odbaciti. Uvodno treba istaknuti da su ciljevi istraživanja uspješno postignuti te je kroz provedene eksperimente dokazano da je korištenjem termalne kamere kao senzora računalnog vida moguće povećati razine sigurnosti štićenih prostora i područja. Uspješnost se dodatno očituje u kombinaciji s korištenjem detektora objekata temeljenog na konvolucijskoj neuronskoj mreži zbog čega je moguća detekcija i raspoznavanje osoba u gotovo svim vremenskim uvjetima (vedro, magla, kiša) u noćnim uvjetima na različitim udaljenostima, pri promjeni brzine i načina kretanja, te položaja tijela osoba koje namjerno, ali ilegalno pokušavaju proći kroz nadzirano područje. Pored toga, u radu je pokazano da je moguće uspješno raspoznavanje osoba od drugih pokretnih objekata na slikama, što je u ovom slučaju bio pas. S tim u svezi treba napomenuti da su testirana dva naučena modela. Prvi model za detekciju osoba učen je na slikama na kojima je bila označena samo osoba (Human, Person), a isti je naknadno testiran i na slikama te video zapisima koji su sadržavali i druge pokretne objekte (psa). Također je značajno istaknuti da naučeni model niti u jednom slučaju nije druge objekte iz okoline, kao što su stabla ili lišće detektirao kao osobe. Drugi model je naučen kao model s dvije klase (Human – Non-Human) s ciljem poboljšanja rezultata detekcije i raspoznavanja, a postignuti rezultati su to i potvrdili. Nadalje, kreirani skup slika, prigodno nazvan **UNIRI Thermal**, odnosno u eksperimentalnoj fazi označen kao **UNIRI-TID**, iskorišten je i za provjeru mogućnosti detektora temeljenog na neuronskoj mreži u kontekstu raspoznavanja aktivnosti osoba, odnosno da li osoba hoda uspravno, hoda pogureno, trči, hoda četveronoške ili leži na tlu. I u ovom slučaju su postignuti odlični rezultati koji su pokazali da je moguće s velikom pouzdanošću prepoznati o kojem načinu kretanja je riječ. Navedeno u kontekstu nadzora štićenih područja može ukazivati na eventualne neprijateljske ili druge nezakonite namjere osoba koje se kreću u tom području. Osim navedenog, u eksperimentalnoj fazi je pokazano i da je moguće pouzdan model naučiti na relativnom malom skupu termalnih slika. Također su uspoređeni aktualni vrhunski detektori objekata kao što su Faster R-CNN, SSD, Cascade R-CNN i YOLOv3, koji su učeni na skupu termalnih slika ekstrahiranih iz videozapisa koji simuliraju ilegalna kretanja u području granica te drugim štićenim područjima.

Zaključno, postignuti rezultati daju dobar temelj za daljnja istraživanja potencijalno opasnih situacija za sigurnost ljudi ištićenih područja u kontekstu raspoznavanja i drugih objekata na sceni kao što su civilna i vojna vozila, različite letjelice, padobrani, ali i prepoznavanje lažnih detekcija kao što su prolazi životinja i identifikacije osoba koje su zadužene za nadzor tih područja.

Ključne riječi:

Konvolucijske neuronske mreže, detektor objekata, nadzor, termovizijska slika, YOLO, otkrivanje i prepoznavanje ljudi, nacionalna sigurnost, nadzor zaštićenih područja i objekata

Abstract

Global terrorist threats and illegal migration have intensified concerns for the security of citizens, and every effort is made to exploit all available technological advances to prevent adverse events and protect people and their property. Due to the ability to use at night and in weather conditions where RGB cameras do not perform well, thermal cameras have become an important component of sophisticated video surveillance systems. In this paper, we investigate the task of automatic person detection in thermal images using convolutional neural network models originally intended for detection in RGB images. We compare the performance of the standard state-of-the-art object detectors such as Faster R-CNN, SSD, Cascade R-CNN, and YOLOv3, that were retrained on a dataset of thermal images extracted from videos that simulate illegal movements around the border and in protected areas. Videos are recorded at night in clear weather, rain, and in the fog, at different ranges, and with different movement types. YOLOv3 was significantly faster than other detectors while achieving performance comparable with the best, so it was used in further experiments. We experimented with different training dataset settings in order to determine the minimum number of images needed to achieve good detection results on test datasets. We achieved excellent detection results with respect to average accuracy for all test scenarios although a modest set of thermal images was used for training. We test our trained model on different well known and widely used thermal imaging datasets as well. In addition, we present the results of the recognition of humans and animals in thermal images, which is particularly important in the case of sneaking around objects and illegal border crossings. Also, we present our original thermal dataset used for experimentation that contains surveillance videos recorded at different weather and shooting conditions.

In conclusion, the achieved results provide a good basis for further research of potentially dangerous situations with a goal to increase personal security and protected areas in the context of recognition human and other objects on the scene such as civilian and military vehicles, various aircrafts, parachutes, but also determination of false detections such as animal passages, and the identification of security personnel in charge of monitoring these areas.

Keywords:

Convolutional Neural Networks, Object Detector, surveillance, Thermal imaging, YOLO, Human detection and recognition, national security, surveillance of protected areas and facilities

Sadržaj

1	Uvod.....	1
1.1	Problem i objašnjenje osnovnih pojmova.....	4
1.2	Motivacija.....	5
1.3	Cilj, hipoteze i znanstveni doprinosi istraživanja.....	6
1.4	Struktura rada.....	7
2	Osnove infracrvene termovizije.....	8
2.1	Elektromagnetski spektar i infracrveno zračenje.....	8
2.2	Detekcija IR zračenja.....	12
2.3	Snimanje IR zračenja – Termalne kamere.....	14
2.4	Termalna kamera i snimanje slike/videoa.....	15
2.5	Termalna kamera i detektori IR zračenja.....	16
2.6	Termalno snimanje i negativni utjecaji na termografiju.....	22
3	Pregled dosadašnjih istraživanja.....	31
3.1	Klasični pristupi detekcije i raspoznavanja osoba.....	32
3.1.1	Termovizijski sustavi za detekciju osoba temeljeni na oduzimanju pozadine.....	33
3.1.2	Sustavi termalnog nadzora temeljeni na praćenju.....	36
3.1.3	Sustavi termalnog nadzora temeljeni na prostorno-vremenskim značajkama.....	36
3.1.4	Sustavi termalnog nadzora s pokretnim kamerama.....	38
3.2	Metode detekcije i raspoznavanja osoba temeljene na CNN.....	41
3.3	Baze slika.....	44
4	Umjetne neuronske mreže (engl. <i>ANN – Artificial Neural Networks</i>).....	47
4.1	Povijesni pregled razvoja neuronskih mrež, tipovi NN mreža i konvolucijske neuronske mreže.....	50
4.2	Elementi neuronskih mreža.....	60
4.3	Aktivacijske funkcije.....	64
4.3.1	Funkcija s pragom osjetljivosti (engl. <i>Step funkcija</i>).....	67
4.3.2	Linearna i nelinearna funkcija.....	68
4.3.3	Sigmoidna funkcija.....	69
4.3.4	Funkcija hiperbolnog tangensa ili Tanh funkcija.....	70
4.3.5	ReLu funkcija.....	71

4.3.6 Softmax funkcija.....	72
4.3.7 SWISH aktivacijska funkcija.....	73
4.4 Učenje umjetnih neuronskih mreža.....	75
4.4.1 Tipovi učenja neuronskih mreža.....	78
4.4.1.1 Nadzirano učenje.....	78
4.4.1.2 Polunadzirano učenje.....	79
4.4.1.3 Nenadzirano učenje.....	79
4.4.1.4 Duboko podržano učenje.....	80
4.4.2 Algoritmi učenja neuronskih mreža.....	81
4.4.2.1 Učenje korekcijom pogreške.....	81
4.4.2.2 Učenje zasnovano na memoriranju.....	82
4.4.2.3 Hebbovo učenje.....	83
4.4.2.4 Kompetitivno učenje.....	84
4.4.2.5 Boltzmanovo učenje.....	84
4.4.2.6 Algoritam "propagacije unazad".....	85
4.5 Umjetne neuronske mreže i duboko učenje	86
4.6 Aktualno naraširenije platforme.....	89
4.6.1. TensorFlow (TF).....	89
4.6.2. Keras.....	90
4.6.3. Caffé.....	90
4.6.4 PyTorch.....	90
4.6.5 Microsoft Cognitive Toolkit.....	91
5 Konvolucijske neuronske mreže.....	91
5.1 Metode evaluacije točnosti i pouzdanosti rezultata detekcije i klasifikacije.....	96
5.2 YOLO.....	99
5.2.1. YOLOv1.....	99
5.2.2. YOLOv2.....	101
5.2.3. YOLOv3.....	102
5.2.4. YOLOv4.....	104
6 Detekcija i raspoznavanje osoba u otežanim vremenskim uvjetima.....	108
6.1 Snimanje i obrada slika.....	110
6.1.1 Kriteriji za kreiranje baze slika.....	110

6.1.1.1. Vedro vrijeme.....	110
6.1.1.2. Gusta magla.....	110
6.1.1.3. Jaka kiša.....	111
6.1.2 Prikupljanje podataka	111
6.1.2.1. Snimanje po vedrom vremenu.....	113
6.1.2.2. Snimanje po gustoj magli.....	115
6.1.2.3. Snimanje po jakoj kiši.....	116
6.2 Pretporcesiranje snimljenog materijala.....	117
6.3 Opis modela, korišteni hiperparametri.....	126
6.4 Rezultati detekcije i raspoznavanja osoba primjenom YOLOv3 detektora.....	127
6.4.1 Rezultati evaluacije detekcija osoba na termovizijskim slikama i video snimkama i usporedba rezultata.....	129
6.4.1.1 Rezultati evaluacije detekcija osoba primjenom bY i tY modela na skupu slika UNIRI-Thermal.....	129
6.4.1.2 Rezultati evaluacije detekcija osoba primjenom bY i tY modela na podskupu slika <i>clear_ALL</i>	140
6.5 Detekcija i raspoznavanje osoba u odnosu na druge objekte na slici (Human/Non-Human).....	141
6.6. Raspoznavanje aktivnosti na termovizijskim slikama i video snimkama.....	144
6.7. Treniranje YOLO modela na malom skupu termovizijskih slika.....	150
6.8 Testiranje naučenog YOLO modela na drugim skupovima termalnih slika i video zapisa.....	153
7 Sustav za detekciju i raspoznavanje osoba u otežanim uvjetima primjenom IR termovizije.....	161
7.1 Arhitektura sustava za detekciju i raspoznavanje osoba.....	165
7.2 Modul za snimanje.....	166
7.3 Modul za detekciju i klasifikaciju temeljen na dubokom učenju.....	166
7.4 Izvještajni modul.....	167
7.5. Mogućnosti skrivanja od detekcije termalnom kamerom.....	169
8 Zaključak.....	171
9 Znanstveni doprinos i perspektiva budućih istraživanja.....	173

Literatura.....	177
Popis slika.....	195
Popis tablica.....	200
Popis kratica.....	201
Dodatak: Programska implementacija.....	206
Životopis.....	209
Sudjelovanja u projektima.....	209
Popis priloga i dodataka.....	209

1 Uvod

Gotovo nikada u povijesti nije tolika pažnja i količina resursa usmjeravana u pitanja sigurnosti, kako izravne sigurnosti osoba tako i sigurnosti imovine. Naime, prema procjenama ukupna vrijednost tržišta video nadzora na globalnoj razini će do 2025. godine narasti na vrijednost od oko 89,7 milijardi američkih dolara. Za usporedbu, isto tržište je 2018. godine vrijedilo oko 47,4 milijarde američkih dolara. [1]. Nadalje, suvremeno društvo i svijet u globalu, također, kao nikada ranije nisu bili izloženi tolikom broju izazova i prijetnji. Primjerice, tijekom posljednjih 10 godina zbog terorističkih aktivnosti, godišnje je prosječno ubijeno oko 21.000 ljudi. U svezi toga treba napomenuti da podaci znatno osciliraju od godine do godine, kao i ovisno o pojedinom dijelu svijeta, odnosno ovisno o pojedinoj državi. [2]. U tom kontekstu, primarno se ističu pitanja javne sigurnosti zbog globalne prijetnje terorizmom, a prisutni su i drugi oblici ilegalnih i nezakonitih aktivnosti koji također značajno ugrožavaju kako sigurnost na osobnoj razini, tako i sigurnost na razini društva (ilegalne migracije, povećana stopa razbojstava, razbojničkih krađa, uništavanja osobne i javne imovine i sl.). Pored navedenog, suvremeno društvo se aktualno suočava s novim oblikom prijetnje, a to je prijetnja javnom zdravlju zbog aktualne pandemije korona virusa, odnosno bolesti uzrokovane tim virusom COVID-19. Zbog svega navedenog, povećana je zabrinutost za sigurnost građana i imovine, a kako bi se spriječili neželjeni događaji i zaštitili ljudi i njihova imovina, ulaganje u sigurnosne sustave doseglo je rekordne razine pokušavajući iskoristiti sva dostupna tehnološka dostignuća za razvoj sofisticiranih sustava [1].

Toplinske, termovizijske ili termalne kamere (svi nazivi se podjednako koriste u hrvatskom jeziku, nap.a.), a u ovom radu bit će korišten pojam "termalna kamera". Nadalje, termalne kamere danas su sveprisutne u sustavima video nadzora namijenjenim povećanju zaštite i sigurnosti ljudi i objekata kako u urbanim područjima, tako i na državnim granicama i ostalim nadziranim i štićenim područjima, a zbog aktualne pandemije korona virusa, učestalo se koriste i u trgovačkim centrima, zračnim lukama, graničnim prijelazima i sl., za potrebe mjernja temperature kupaca i putnika. Termalne kamere svoju ulogu u sustavima za nadzor štićenih prostora i područja dugo su prisutne, uglavnom u domeni vojne i policijske primjene, ali ništa manje nisu važne za primjenu u domeni osobne sigurnosti, odnosno zaštiti osobne imovine, a primarno iz razloga jer se mogu koristiti u vremenskim uvjetima kada se ne mogu koristiti

standardne RGB kamere ili kada standardne kamere daju loše rezultate, poput noćnih uvjeta ili otežanih vremenskih uvjeta.

Kada se govori o zaštiti štićenih područja, najvažniji korak je pravodobna detekcija neovlaštenih osoba, sumnjivih kretanja u zaštićenim područjima, te posljedično spriječavanje ilegalnih radnji kao što su upadi u štićena područja vojnih ili drugih značajnih infrastrukturnih objekata ili nezakoniti prelasci državne granice. Sukladno tomu, mogućnost nadzornog sustava da automatski i pravovremeno detektira kretanje te raspozna osobe od drugih objekata na snimci štićenog područja te posljedičnog pravovremenog upozoravanja na sumnjivu situaciju jedan je od ključnih parametara za pouzdanost takvog sustava.

Do sada su razvijeni mnogi uspješni algoritmi strojnog učenja za otkrivanje i praćenje objekata kao što su oni usmjereni na ljudsko lice [3, 4, 5] ili ljudsku figuru [6] na RGB optičkim slikama. Svrha otkrivanja objekata je njihova klasifikacija i označavanje njihovog točnog položaja na slici ili videozapisu.

Neki od najboljih rezultata detekcije objekata postižu primjenom metoda dubokog učenja temeljenog na konvolucijskim neuronskim mrežama (engl. *Convolutional Neural Networks*) [7]. Popularnost konvolucijskih neuronskih mreža i duboko učenje započeli su velikim uspjehom *AlexNeta* za zadatak prepoznavanja objekata na međunarodnom natjecanju *ImageNet Challenge 2012*. [8]. Od tada su razvijene mnoge uspješne CNN arhitekture za otkrivanje objekata, poput Fast R-CNN [9], SSD [10], Mask R-CNN [11], R-FCN [12] i YOLO [13 – 16] i prilagođeni za različite zadatke [17 – 24].

Zbog razlika u značajkama vizualne i termalne slike, cilj ovog doktorskog rada je istražiti kako se metode dubokog učenja, koje su uspješne za detekciju objekata na optičkim (RGB) slikama, ponašaju na termalnim slikama i videozapisima.

U tu svrhu, kreiran je novi skup termalnih slika i videozapisa koji su snimljeni u različitim vremenskim uvjetima, na različitim udaljenostima, korištenjem različitih tipova opreme, a snimane osobe su mijenjale svoje načine kretanja, a sve s ciljem da maksimalno simuliraju ilegalno kretanje unutar nadziranih i štićenih prostora i područja [25].

U eksperimentalnom dijelu, primarno je korišten spomenuti algoritam YOLOv3 [15], koji postiže odlične rezultate detekcije objekata na RGB slikama. U kontekstu

eksperimenata u cilju detekcije osoba na termalnim videozapisima YOLOv3 je u prvom eksperimentu treniran na skupu slika snimljenih po vedrom vremenu, a model je potom treniran na skupovima slika snimljenih po jakoj kiši i gustom magli. Osim primarnog cilja, a to je detekcija osoba u otežanim vremenskim uvjetima primjenom termovizije, provedeni su i drugi eksperimenti, kao što je uspješnost raspoznavanja osoba od životinja, a ispitana je i mogućnost raspoznavanja aktivnosti osoba u smislu utvrđivanja da li osoba trči, hoda normalnim hodom, hoda pogureno ili četveronoške. Cilj ovog eksperimenta je, osim ispitivanja mogućnosti razlikovanja ovih aktivnosti, ujedno i baza za buduća ispitivanja mogućnosti identifikacije osobe prema načinu hoda, te eventualnog otkrivanja budućeg ponašanja osoba, s obzirom da je opće poznato da određen položaj tijela kao i način kretanja može upućivati na određene buduće aktivnosti osobe, primjerice u smislu da li osoba nešto skriva, da li se pokušava neopaženo kretati u određenom području i sl. Osim YOLOv3, na istim skupovima podataka testirani su i drugi aktualno najpouzdaniji detektori objekata temeljeni na neuronskim mrežama kao što su Faster R-CNN [26], SSD [10], kaskadni R-CNN [27], YOLOv4 te FCOS [28].

Osim toga, u sklopu istraživanja, ispitane su generalizacijske mogućnosti treniranog YOLOv3 modela, a isti je ispitan i na vanjskim skupovima podataka koji se naširoko koriste u domeni detekcije osoba na termalnim slikama, kao i na videozapisima iz realnih okolnosti kao što je vojna vježba. Ovdje treba napomenuti da je ovaj YOLOv3 model treniran na baznom skupu slika, odnosno skupu koji obuhvaća slike u svim vremenskim uvjetima, svim udaljenostima, brzinama kretanja, položajima tijela, slike snimljene standardnim i teleobjektivom, a slike iz skupa dodatno su transformirane i RGB u dvije verzije crno-bijelih slika, s tim da jedna verzija topla područja prikazuje svijetlim, a druga verzija tamnim nijansama.

1.1 Problem i objašnjenje osnovnih pojmova

Problem i cilj istraživanja u sklopu doktorskog rada je ispitivanje mogućnosti metoda dubokog učenja za detekciju i raspoznavanje osoba u otežanim vremenskim uvjetima primjenom infracrvene termovizije. Prethodno proučavanje literature pokazalo je da ovo područje ima jako puno prostora za nova istraživanja, posebice kada je riječ o razvoju autonomnog sustava temeljenog na dubokom učenju i neuronskim mrežama s krajnjim ciljem pouzdane detekcije i raspoznavanja osoba. Iako je područje detekcije osoba primjenom infracrvene termovizije ranije već istraživano, iz pregleda dostupne literature za ovo područje razvidno je da su provedena istraživanja uglavnom načelnog karaktera u smislu detekcije osoba kao objekata na slikama ili videozapisima, međutim nisu uočena istraživanja detekcije i raspoznavanja osoba u otežanim vremenskim uvjetima, posebice ne u kontekstu eventualnog razvoja sustava i njegove implementacije u realni sektor. Moguće je da su istraživanja ovog tipa provedena u okviru vojno-istraživačkih aktivnosti neke od svjetskih vojski ili vojnih saveza, međutim izgledno je da su takva istraživanja klasificirana određenim stupnjem tajnosti, te kao takva nisu dostupna široj javnosti i znanstvenoj zajednici.

Nadalje, pregled dostupne literature također je pokazao nedostatak odgovarajućih skupova slika za treniranje modela dubokog učenja za detekciju i raspoznavanje osoba na termovizijskim slikama i video zapisima, posebice kada je riječ o teškim vremenskim uvjetima kao što su kiša ili magla, stoga je uočeno da postoji i potreba za kreiranjem novog skupa slika. S tim u svezi, u okviru doktorskog rada, kreiran je novi skup slika, koji će biti šire elaboriran i opisan u nastavku rada.

1.2 Motivacija

Motivacija i znanstvena znatiželja za istraživanje mogućnosti metoda dubokog učenja na termovizijskim slikama i videozapisima, proizlazi iz autorova promišljanja i profesionalnog usmjerenja u području sigurnosti i zaštite građana i imovine, te svakodnevnog sudjelovanja u aktivnostima koje iziskuju veliki angažman ljudskih resursa za uspješno izvršavanje takvih zadataka. Stoga je kontinuirano promišljanje dovelo do ideje o, u prvom redu, isptivanju mogućnosti metoda dubokog učenja za detekciju i raspoznavanje osoba na termovizijskim slikama i videozapisima, što implementirano u područje sigurnosti, posljedično može dovesti do povećanja razine zaštite građana i imovine, posebice u otežanim vremenskim uvjetima i na teškim terenima. Naime, jasno je samo po sebi da implementacija ovakvog sustava u velikoj mjeri može sniziti troškove, posebice u kontekstu potrebe za angažmanom ljudskih resursa (vojnika, policajaca i sl.). Nadalje, sustav temeljen na metodama dubokog učenja i neuronskim mrežama može uvelike ubrzati detekciju prisutnosti neovlaštenih osoba u nadziranim i štićenim područjima, a ukoliko takav sustav kao ulazni senzor koristi termovizijsku kameru, njegove mogućnosti se protežu na mogućnost kontinuiranog nadzora 24/7, po danu i po noći, kao i u svim vremenskim uvjetima (kiša, magla, snijeg). Kao što je navedeno, ovakva promišljanja su bila polazna točka za ovo nevjerojatno putovanje koje se zove doktorski studij, a rezultate prikazane u doktorskom radu može se, u nekom prenesenom smislu, smatrati putopisom, koji je primarno rezultat ideje temeljene na poznavanju svojstava i specifičnosti termovizijske kamere koja kao samostalan instrument uvelike proširuje ljudske sposobnosti koji po svojoj biologiji po noći i otežanim vremenskim uvjetima imaju gotovo zanemarive sposobnosti detekcije osoba. Stoga se, na određen način, nameće potreba za razvojem sustava koji će ljudima omogućiti ono što priroda nije, a to je da vide u mraku.

1.3 Cilj, hipoteze i znanstveni doprinosi istraživanja

Kao što je prethodno već navedeno, cilj istraživanja u okviru doktorskog rada je ispitivanje mogućnosti primjene metoda dubokog učenja, primarno razvijenih za detekciju i klasifikaciju objekata na RGB slikama, na termovizijskim slikama, s konačnim ciljem razvoja autonomnog sustava za nadzor štićenih područja i prostora, što u konačnici dovodi povežanja razine sigurnosti i zaštite osoba i imovine, posebice kada je riječ o državnoj granici ili infrastrukturnim objektima od nacionalnog značaja za pojedincu državu.

U tom kontekstu postavljene su sljedeće hipoteze čija se istinitost želi potvrditi ovim istraživanjem:

H1: *Primjena infracrvenih termovizijskih sustava omogućava povećanu zaštitu i sigurnost štićenih prostora*

H2: *Korištenje infracrvene termovizije omogućava pouzdanu detekciju i raspoznavanje osoba u otežanim vremenskim uvjetima*

H3: *Ove funkcije mogu se dostići u realnom vremenu, te time omogućavaju smanjenu potrebu za neprekinutim ljudskim nadzorom*

Očekivani znanstveni doprinos je:

- baza slika osoba snimljenih termovizijskom kamerom u otežanim vremenskim uvjetima uz anotacije i slike pogodne za usporedbu i raspoznavanje,
- postupak detekcije osoba na termovizijskim slikama korištenjem temperaturnih pragova i biometrijskih značajki,
- arhitektura sustava za detekciju i klasifikaciju osoba u realnom vremenu, koja se temelji na sustavu infracrvenih termovizijskih kamera,
- postupak ocjenjivanja uspješnosti primjene autonomnih infracrvenih termovizijskih sustava u području sigurnosti.

1.4 Struktura rada

Preostali dio ovog doktorskog rada koncipiran je na sljedeći način:

U drugom poglavlju opisane su osnove infracrvene termovizije, odnosno termalnog snimanja, pregled povijesnog razvoja termalnih detektora, kao i negativni utjecaji na termalno snimanje.

U trećem poglavlju prezentiran je pregled područja istraživanja i baze termalnih slika koje se uobičajeno koriste za učenje modela za detekciju i raspoznavanje osoba.

U četvrtom poglavlju napravljen je pregled neuronskih mreža, od početnih modela, povijesnog razvoja, gradivnih elemenata koji čine umjetne neuronske mreže do aktualno naraširenijih platformi koje se koriste za detekciju i raspoznavanje objekata.

U petom poglavlju napravljen je detaljniji pregled konvolucijskih neuronskih mreža i predstavljene su verzije YOLO detektora koji je korišten u eksperimentalnom dijelu rada.

U šestom poglavlju prezentirani su rezultati istraživanja i eksperimentalni dio koji se odnosi na detekciju i raspoznavanje osoba u otežanim vremenskim uvjetima, raspoznavanje aktivnosti, te mogućnosti učenja modela na malom skupu slika.

U sedmom poglavlju predstavljena je i opisana moguća arhitektura sustava za detekciju i raspoznavanje osoba primjenom termovizije i detektora objekata temeljenog na neuronskoj mreži.

U osmom poglavlju izvedeni su zaključci istraživanja i eksperimentalnog dijela.

U devetom poglavlju prezentirani znanstveni doprinosi i perspektiva budućih istraživanja u području detekcije osoba primjenom termovizije, odnosno općenito moguća nova istraživanja i primjena termovizije u području sigurnosti.

2 Osnove infracrvene termovizije

Temelj infracrvene termovizije čini infracrveno (*engl. Infrared - IR*) ili toplinsko (termalno) zračenje koje zrače sva tijela koja imaju temperaturu višu od apsolutne nule. Infracrvena termovizija odnosno termalno ili termovizijsko snimanje su postupci snimanja u kojima se upotrebljavaju termovizijske, odnosno termalne kamere koje imaju mogućnost snimanja infracrvenog dijela elektromagnetskog (EM) spektra koji nije izravno vidljiv ljudskom oku.

2.1 Elektromagnetski spektar i infracrveno zračenje

Elektromagnetsko zračenje sastoji se od oscilirajućih električnih i magnetskih polja koja se kroz prostor šire brzinom svjetlosti. Zračenje ima svojstva vala i stoga ne zahtijeva medij za širenje. Nadalje, sinusni elektromagnetski valovi su funkcije vremena i položaja te posjeduju određenu frekvenciju i valnu duljinu. Ti valovi formiraju elektromagnetski spektar odnosno radio i TV zračenje, mikrovalove, vidljivu svjetlost, infracrveno (IR) i ultraljubičasto zračenje, X-zrake i gama zrake. Kako se učestalost zračenja povećava, valna duljina se skraćuje i energija fotona (E) povećava:

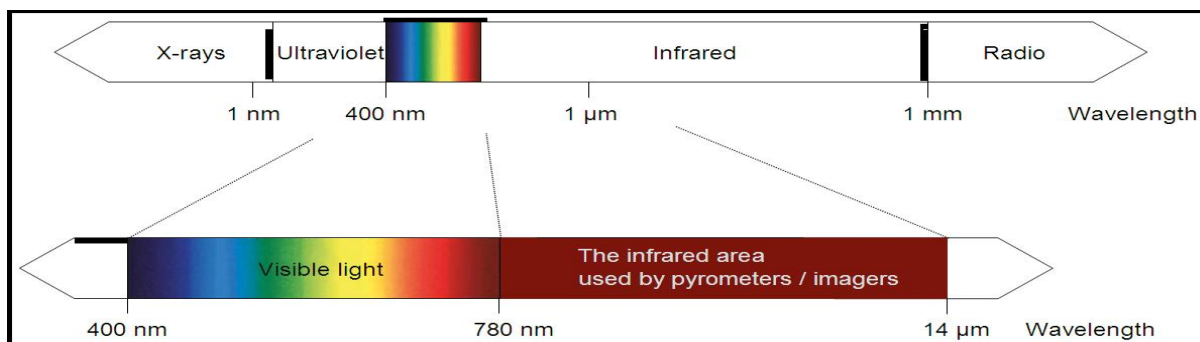
$$E = hf = \frac{hc}{\lambda} \quad (1)$$

Navedena jednadžba pokazuje energiju fotona pri čemu je:

- h - Planckova konstanta koja iznosi $h = 6,626 \times 10^{-34} \text{ Js}$,
- f - frekvencija zračenja,
- c - brzina svjetlosti, $3 \times 10^8 \text{ m/s}$,
- λ - valna duljina zračenja

[29].

Elektromagnetski valovi svrstani su u elektromagnetski spektar koji se proteže od valova najmanje frekvencije i najveće valne duljine do valova najveće frekvencije i najmanje valne dužine (Slika 1). Energija valova, veća je što je veća frekvencija valova i što je kraća valna duljina. [30].



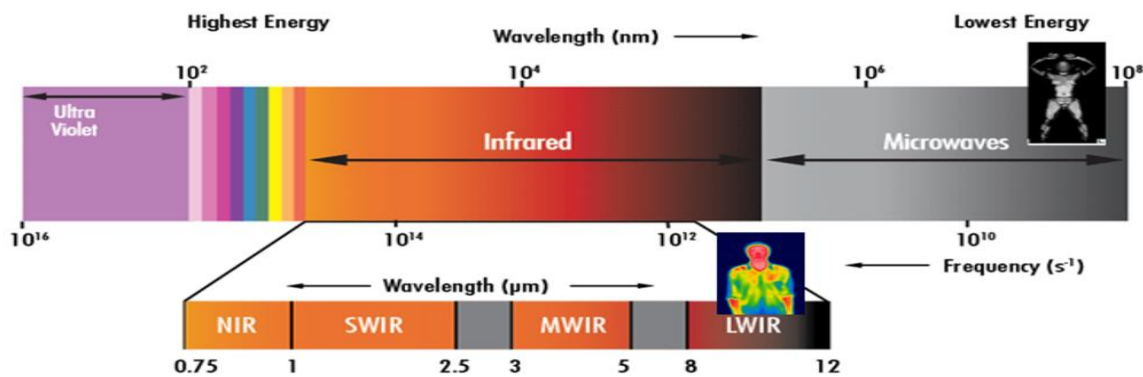
Slika 1. Elektromagnetski spektar [31]

U okviru teme ovog doktorskog rada, bit će napravljen osvrt samo na infracrveni (IR) dio elektromagnetskog spektra, s obzirom da je infracrveno zračenje temelj termovizijskog snimanja.

Svi objekti temperature veće od apsolutne nule emitiraju toplinsku energiju koja je proporcionalna zagrijanosti objekta odnosno njihovoj temperaturi. Isto vrijedi i za toplokrvna živa bića, sva emitiraju toplinsku energiju. Međutim, toplinska energija se emitira u IR dijelu EM spektra zbog čega nije vidljiva ljudskom oku, a tek nakon što su tijela zagrijana na temperaturu višu od 100°C, počinju emitirati valove iz vidljivog dijela EM spektra, te njihova zagrijanost postaje vidljiva i ljudskom oku.

Formalno otkrivanje IR zračenja pripisuje se sir Wiliamu Herschel (1738 – 1822), koji 1800. godine izveo eksperiment kako bi utvrdio koliko različito obojene zrake iz sunčeve svjetlosti raspršene staklenom prizmom mogu različito zagrijavati površinu. [32].

Naknadno su, oko 1900.g., Planck, Stefan, Boltzman, Wien i Kirchoff precizno definirali EM spektar te utvrdili kvalitativnu i kvantitativnu korelaciju za opis infracrvenog zračenja [33]. IR dio EM spektra nalazi se ispod posljednjeg segmenta vidljivog dijela EM spektra, odnosno počinje nakon završetka crvene boje i dodatno se dijeli na podpodručja prema valnoj duljini, počev od područja najbližeg vidljivom dijelu spektra (slika 2).

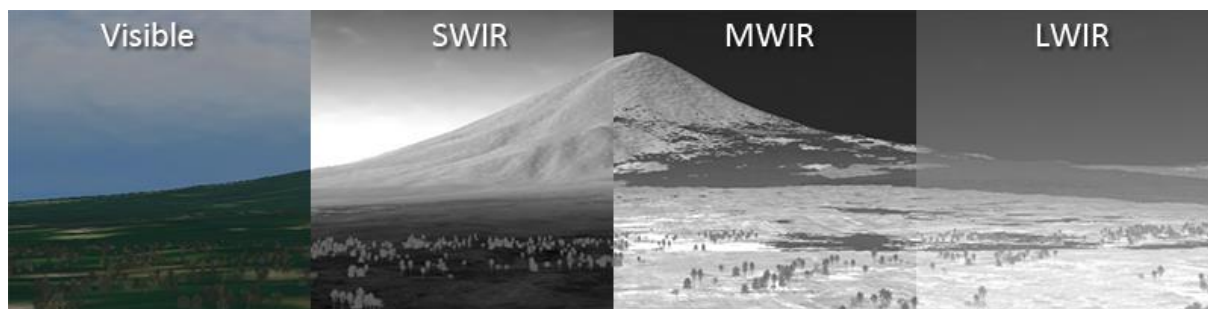


Slika 2. Podjela IR segmenta elektromagnetskog spektra [34]

IR spektar elektromagnetnog zračenja nije vidljiv ljudskom oku zbog veće valne duljine, a samim tim i manje frekvencije. IR spektar pokriva područje valnih duljina od 0,7 do 300 μm , što odgovara frekvencijskom pojasu od 1 do 430 THz, zbog čega je valna duljina veća od spektra vidljivog svjetla, a kraća od terahercih valnih duljina mikrovalova. Infracrveni dio spektra primarno se dijeli u tri podpodručja: IR-A, od 0,78 do 1,4 μm , IR-B, od 1,4 do 3 μm , te IR-C, od 3 do 1000 μm [35].

Preciznija podjela IR spektra glasi:

- Područje blizu vidljivog dijela EM spektra (engl. *Near-Infrared* – NIR), u rasponu od 0,7 to 1 μm ,
- Kratkovalni IR (engl. *Short-Wave infrared* – SWIR), u rasponu od 1 do 3 μm ,
- Srednjevalni IR (engl. *Mid-Wave infrared* – MWIR), u rasponu od 3 do 5 μm ,
- Dugovalni IR (engl. *Long-Wave infrared* – LWIR), u rasponu od 8 do 14 μm ,
- Dugo-Dugovalni IR (engl. *Very Long-Wave infrared* – VLWIR), u rasponu većem od 14 μm . [36]



Slika 3. Prikaz područja snimljenog u vizualnom i različitim dijelovima IR spektra [37]

Svaki dio IR spektra ima svoje specifičnosti i mogućnosti primjene u pojedinim područjima života i granama industrije, te samim tim snimke iz svakog pojedinog dijela

IR spektra daju drugačiji prikaz snimljenog objekta ili prostora, te omogućuje drugačiji uvid.

Kao što je ranije navedeno, sva tijela koja imaju temperaturu iznad apsolutne nule zrače toplinsku energiju, a što je viša temperatura objekta ili tijela, to je i zračenje toplinske energije veće. Toplina je dio unutrašnje energije nekog tijela koja prelazi iz područja više temperature u područje niže temperature. Taj prijelaz se odvija na tri načina: vođenjem ili kondukcijom (engl. *Conduction*), strujanjem ili konvekcijom (engl. *Convection*), zračenjem ili radijacijom (engl. *Radiation*).

Zračenje ili radijacija je proces kada se toplinska energija pretvara u elektromagnetsko zračenje koje tijelo emitira u okolni prostor, odnosno zračenje predstavlja prijenos energije na daljinu EM valovima bez posredovanja materije, a ovaj oblik prijelaza toplinske energije je ujedno i najvažniji za termalno snimanje. Naime, riječ je o zračenju u IR području EM spektra te se još naziva i IR zračenje, a s obzirom da je riječ o prijenosu topline, naziva se i toplinsko ili termalno zračenje. [38].

Količina emitiranog toplinskog zračenja definira se prema zračenju crnog tijela (engl. *Black body*), koje je zamišljeno kao idealno tijelo koje apsorbira svo toplinsko zračenje koje je emitirano prema njemu i reemitira energiju karakterističnu samo za taj specifični sustav koji zrači. [39]. U fizici je toplinsko zračenje opisano kroz Planckov zakon, Wienov zakon, Stefan-Boltzmanov zakon, Wienov zakon pomaka, Lambertov zakon [40], kao i Kirchhoffov zakon toplinskog zračenja [41]. Termografija se temelji na Kirchhoffovu zakonu toplinskog zračenja [42], koji glasi

$$E + T + R = 1 \quad (2)$$

gdje je:

- E – emitirana energija,
- T - prenesena energija,
- R – reflektirana energija.

a u većini slučajeva vrijedi i jednakost da je

$$E + R = 1 \quad (3)$$

kao i odnos da je "*dobar emiter loš reflektor i dobar reflektor loš emiter toplinske energije*".

Sukladno navedenom, termografija ovisi o emitivnosti, reflektivnosti i propusnosti, u svezi čega je potrebno istaknuti da je:

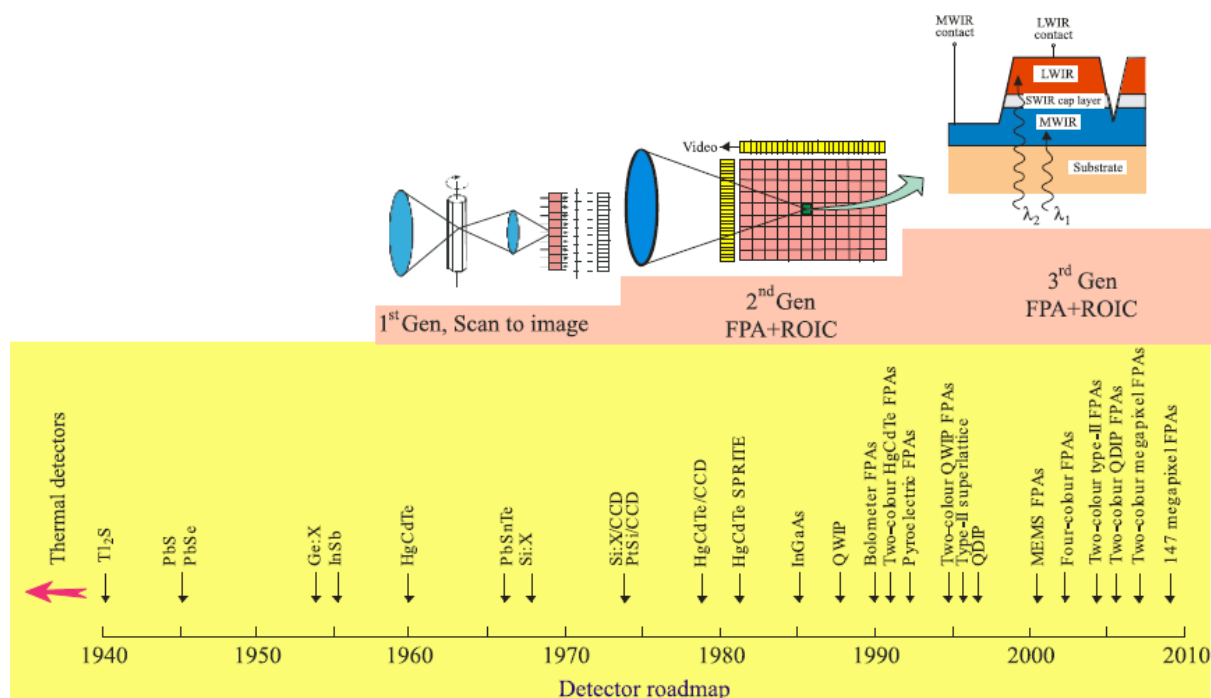
- **Emitivnost** (engl. *Emittance*) - količina toplinske energije koju emitira tijelo, pri čemu je crno tijelo savršen emiter s vrijednošću energije $E = 1$, dok su sva ostala tijela u prirodi obojena tijela koja imaju vrijednost $E < 1$. U tom kontekstu, kao i u kontekstu teme ovog doktorskog rada treba istaknuti da je vrijednost energije koju emitira ljudsko tijelo, odnosno ljudska koža $E = 0,98$. Također treba dodati da tijelo koje dobro apsorbira toplinsku energiju istovremeno dobro i emitira toplinsku energiju.
- **Reflektivnost** (engl. *Reflectance*) - količina svjetla koja se reflektira od površine tijela, gdje primjerice aluminijska vrpca ima vrijednost $R = 0,98$ dok je vrijednost $E = 0,02$, što znači da tijela koja dobro reflektiraju toplinsku energiju loše emitiraju, odnosno apsorbiraju toplinu.
- **Propusnost** (eng. *Transmittance*) - količina toplinske energije koja prolazi kroz tijelo (objekt) u smjeru kamere. U tom kontekstu treba istaknuti da većina tijela slabo propušta toplinsku energiju, iako primjerice tanji slojevi plastike, tekstila i sličnih materijala dobro propuštaju toplinsku energiju, dok armirani beton, drvo i slični materijali slabo propuštaju toplinsku energiju, iako postoje i drugi fizikalni parametri o kojima ovisi propusnost pojedinih tijela, kao što su količina emitirane energije, vrijeme izloženosti, uvjeti prijenosa topline i sl. [43]

IR zračenje omogućuje snimanje objekata ovisno o njihovoj reflektivnosti ili emitivnosti zračenja infracrvenog dijela EM spektra. Emisija IR zračenja proporcionalna je razini zagrijanosti promatranog objekta zbog čega se snimanje u ovom spektru zračenja još naziva i termalno snimanje (ili termografija), a snimka se zove termogram [44].

2.2 Detekcija IR zračenja

Nakon otkrića IR zračenja, prirodan slijed je doveo do pitanja kako vizualizirati IR zračenje, međutim u to razdoblju isto nije bilo moguće zbog nepostojanja odgovarajućih uređaja, odnosno senzora, do čega je došlo tek 30-ak godina kasnije, u svezi čega je šire napisano u [45]. U kontekstu detekcije i vizualizacije, vrlo je značajna 1880. godina kada je Samuel P. Langley razvio bolometar, kasnije nazvan po njemu - Langleyjev bolometar, koji se temeljio na vrpcoma platine, a imao je mogućnost

detekcije promjene temperature čak do razine 0.0001°C [46]. Razvoj bolometra i drugih senzora za detekciju IR zračenja postupno je doveo i do razvoja prve IR kamere.



Slika 4. Pregled povijesnog razvoja IR detektora i sustava. [45]

IR detektore je bazično moguće podijeliti u dvije kategorije:

- Detektori fotona
- Termalni detektori.

Detektori fotona generiraju slobodne električne nositelje kroz interakciju fotona i vezanih elektrona. Četiri glavne vrste fotodetektora su intrinzični, vanjski, fotoemisijski i noviji QWIP. Termalni detektori reagiraju na promjene temperature koje nastaju od emitiranog IR zračenja kroz promjene fizičkih i električnih svojstava. Termalni detektori se mogu klasificirati kao termometri, termospojevi, termopili, bolometri te noviji mikrobolometri i mikrokantileveri, koji se ponašaju slično kao detektori fotona [47]. Pored navedenog, detektori IR zračenja tradicionalno se također mogu podijeliti u tri kategorije: točkasti detektori, linijski senzori i termalne kamere. U kontekstu ovog doktorskog rada najvažnije su termalne kamere o kojima će više riječi biti u nastavku. IR kamere se načelno mogu podijeliti prema mogućnostima: kamere bez mogućnosti snimanja, samo za promatranje, kamere s mogućnošću snimanja, za promatranje i snimanje te ostali specijalni instrumenti [48]

Prema pristupu mjerenjima i obradi rezultata termografija se dijeli na aktivnu i pasivnu, te na kvalitativnu i kvantitativnu.

Aktivna se termografija zasniva na promatranju dinamičkog ponašanja površine objekta izloženog toplinskoj pobudi. Toplinska se pobuda može postići na različite načine, a svima je zajednički cilj poslati određeni paket energije u promatrani objekt i promatrati njegov odziv na tu toplinsku pobudu - vremenski razvoj površinske temperaturne raspodjele. S druge strane, pasivna termografija promatra objekte u stacionarnom stanju, a razlike u iznosima IR zračenja koje dolaze s površine objekta su ili posljedica razlika u temperaturi ili razlika u svojstvima promatrane površine [49]. U kontekstu ovog doktorskog rada, fokus će biti na termalnim detektorima, odnosno termalnim kamerama koje pasivno, neinvazivno detektiraju IR zračenje temeljem kojeg u daljnjoj obradi stvaraju vidljivu sliku koja se prikazuje na zaslonu i snima na odgovarajući medij.

2.3 Snimanje IR zračenja – Termalne kamere

Prvi termografski uređaji bili su tzv. linijski skeneri koji su pretraživali objekt samo po jednoj liniji pa ih je, za dobivanje dvodimenzionalne slike, trebalo micati uzduž promatranog objekta. Takav način snimanja bio je uglavnom neupotrebljiv za civilne potrebe, ali je zadovoljavao mnoge vojne potrebe. Zato je za stacionarne termografske sustave bilo potrebno razviti poseban uređaj za skeniranje koji će u nepokretnoj kameri stvoriti dvodimenzionalnu sliku. Linijski skeneri su imali značajne nedostatke koji su naknadno riješeni razvojem mikroelektronike i tzv. FPA (engl. *Focal Plane Array*) detektora. Tako FPA-detektor "vidi" cijelo vidno polje objektiva kamere odjednom, pa nije potreban poseban sustav za skeniranje. FPA detektori također donose znatne prednosti u području termografskih sustava: povećana brzina osvježavanja slike (oko 30 Hz), povećana razlučivost (320 x 240 piksela), te minijaturizacija kamere. Osim spomenutih, u termografiji se koriste i tzv. toplinski detektori: bolometri koji su vrlo osjetljivi otpornički termometri i pri promjeni temperature mijenjaju električni otpor, piroelektrični osjetnici koji pod utjecajem IR – zračenja mijenjaju polarizaciju, što se mjeri kao razlika napona. Daljnja poboljšanja u području termografije postignuta su razvojem mikrobolometra čiji je razvoj započeo tijekom 70-ih godina 20. stoljeća [45].

2.4 Termalna kamera i snimanje slike/videoa

Snimanje slike odnosno videa početni je korak u kontekstu detekcije i raspoznavanja osoba u otežanim vremenskim uvjetima, a temelji se na korištenju određene opreme odnosno *hardware*-a za snimanje. U slučaju predstavljenog sustava za detekciju i raspoznavanje osoba u otežanim vremenskim uvjetima kao ulazni senzor odabrana je termovizijska ili termalna kamera s obzirom na mogućnosti snimanja u svim vremenskim uvjetima kao i u potpunom mraku. Ranije je u 2. poglavlju već opisana fizikalna osnova IR zračenja kao i pregled povijesnog razvoja otkrivanja IR zračenja i prvih primjena u civilne i vojne svrhe, dok će u ovom poglavlju primarno biti riječi o termovizijskoj kameri i načinu na koji funkcionira, odnosno načinu na koji bilježi snimke promatranog prostora i objekata. Također će biti pojašnjena razlika između različitih tipova uređaja za snimanje u noćnim uvjetima, kao i utjecaj pojedinih vremenskih te drugih okolišnih uvjeta na snimanje termalnom kamerom.

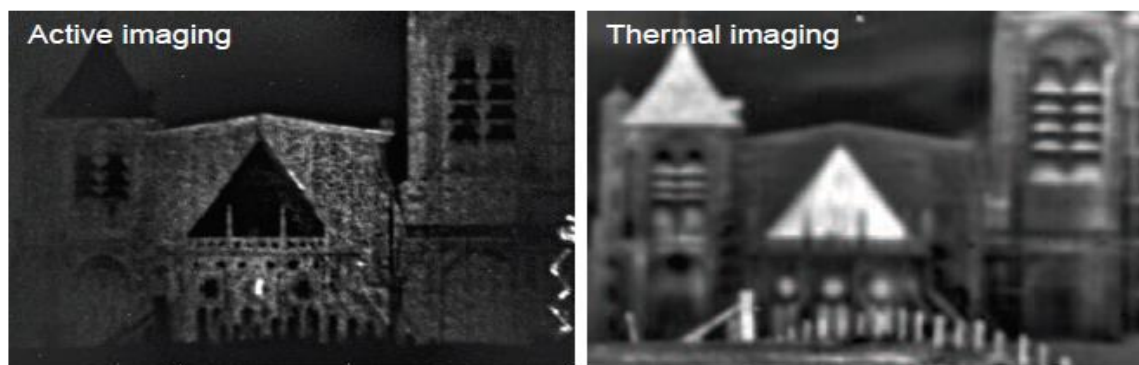
S obzirom na temu rada, u ovoj disertaciji naglasak će biti stavljen na termovizijske kamere koje snimaju u LWIR segmentu IR spektra, odnosno od 8 do 14 μm .

Uvodno je potrebno definirati razliku između aktivnog i pasivnog, odnosno termalnog snimanja. Naime, aktivno IR snimanje je proces pri kojemu je u uređaj implementiran izvor svjetla čiji je zadatak dodatno osvijetliti promatrani objekt, te će kamera zabilježiti takvu sliku. Kod većine aktivnih IR sustava, izvor svjetla je usmjereni laser kratkih pulseva. Kada su kamera i laser u potpunosti sinkronizirani, tada je moguće promatranje i snimanje objekta koji se nalazi na određenoj udaljenosti. Pozadinski raspršena svjetlost se eliminira iz konačne slike, te je moguće stvaranje i 3D ili pseudo 3D slike. Metoda aktivnog IR snimanja je neovisna o prirodnom osvjetljenju i moguće ju je koristiti i tijekom dana i tijekom noći. Aktivno IR snimanje obično koristi valno područje blizu IR spektra (engl. *Near Infrared* – NIR), ali i kratkovalno IR područje (eng. *Shortwave IR* – SWIR), iako je poželjnije korištenje SWIR dijela spektra zbog zaštite očiju. U konačnici, objekt je vidljiv temeljem svojih reflektivnih sposobnosti.

S druge strane, termalno snimanje je proces koji obuhvaća primjenu termalne kamere koja detektira toplinsko zračenje emitirano od strane promatranih objekata. Za termalno snimanje nije potreban vanjski izvor osvjetljenja, ali je potrebno hlađenje nekih senzora na određenu temperaturu. Termalno snimanje funkcionira u području srednjevalnog (engl. *Midwave IR* – MWIR) i dugovalnog (engl. *Longwave IR* - LWIR)

područja IR spektra. Kod ovih područja IR spektra, objekt postaje vidljiv temeljem vlastite emisije toplinskog zračenja i vlastite temperature. Moguća je primjena i tijekom dana, a posebno je praktična primjena u noćnim uvjetima.

Aktivno IR snimanje obično se temelji na kraćim valnim duljinama nego termalno snimanje, a snimke imaju bolju rezoluciju. S druge strane, kvalitetni dalekometni sustavi su ograničeni određenim atmosferskim turbulencijama koje mogu zamutiti snimke te utjecati na homogenost osvjjetljenja.



Slika 5. Primjer snimanja istog objekta primjenom aktivnog IR snimanja (lijevo) i termalnog snimanja (desno). Fotografija prikazuje katedralu Saint Etienne u Bourgesu (Francuska) [50]

2.5 Termalna kamera i detektori IR zračenja

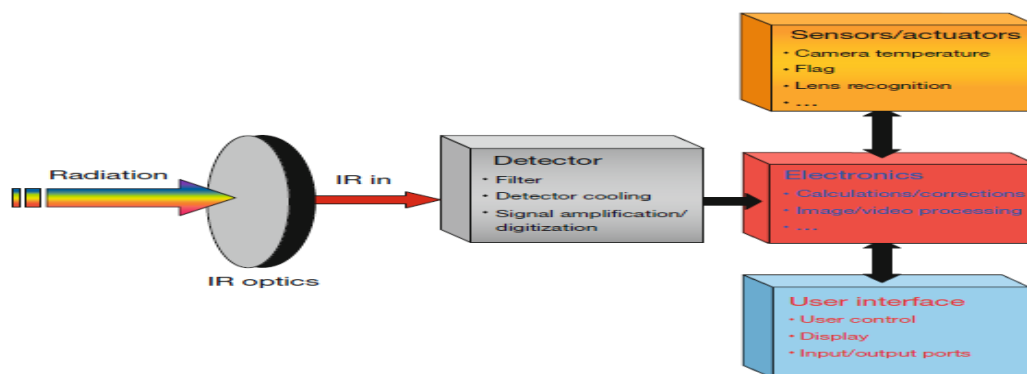
Termovizijske ili termalne kamere su opto-elektronički uređaji za snimanje IR zračenja u određenom području IR segmenta EM spektra, a uobičajeno se odnose na MWIR i LWIR područje IR spektra. Proces snimanja termalnom kamerom zove se termalno snimanje ili termografija [51]. U širem smislu pojam termalno snimanje također se odnosi i na sustave koji ne prikazuju termalnu snimku, nego je ista snimljena i strojno obrađena u cilju mjerenja određenih parametara ili u cilju detekcije određenog objekta, te su naknadno prikazani samo rezultati tog procesa. Termalne kamere IR zračenje snimaju ili bilježe tako da specijalne leće na objektivu fokusiraju IR zračenje koje emitiraju promatrani objekti [52].

Glavne komponente termalnih kamera su (slika 6):

- Optički sustav koji ima mogućnost snimanja prostora ili objekata temeljem zračenja u valnim duljinama termalnog zračenja (uobičajeno se pojam termalnog odnosi na MWIR i LWIR područje IR spektra, nap.a.). Temelj optičkog sustava čine objektiv termalne kamere i leće objektiva koje su napravljene od materijala koji blokiraju prolaz valnih duljina iz vidljivog dijela

spektra, a propuštaju samo valne duljine iz IR termalnog dijela spektra. Za izradu leća objektiva termalnih kamera, najčešće se koristi germanij (Ge).

- Jedan ili više detektorskih elemenata ili senzora koji imaju mogućnost pretvaranja termalnog zračenja u električni signal proporcionalan količini zračenja koje dolazi do njega. Detektor IR zračenja je jedan od najvažnijih elemenata termalne kamere, s obzirom da određuje mogućnosti i rezoluciju snimanja termalne kamere. Postoje dva tipa detektor IR zračenja: prvi, koji ovisi o IR zračenju koje zagrijava detektorski element, pri čemu je rezultat rast temperature što dovodi do aktivacije drugih fizikalnih mehanizama koji mjere količinu IR zračenja koje dolazi do detektora i uobičajeno se naziva termalni detektor, dok je drugi tip detektora je onaj kod kojega fotoni, kao posljedica emitiranog zračenja, reagiraju na atomskoj ili molekularnoj razini s materijalom od kojeg je detektor napravljen kako bi došlo do stvaranja nositelja naboja koji potom generira napon na detektorskom elementu ili kroz promjenu električnog otpora. Ovaj mehanizam uobičajeno obuhvaća elektrone koji apsorbiraju fotone, a rezultat je promjena razine kvantne energije s jedne razine na drugu. Ovaj tip detektora se uobičajeno naziva fotonski ili kvantni detektor,
- Pojedini sustavi zahtijevaju mehanizam za skeniranje koji skenira termalnu sliku na detektoru u određenim uzorcima, iako gotovo sve suvremene termalne kamere ne zahtijevaju mehanizam za skeniranje, s obzirom da koriste velike detektorske matrice koje u potpunosti pokrivaju područje gledanja (engl. *Field of view* - FOV),
- Elektronički procesor koji ima mogućnost obrade izlaznih signala sa detektorskog elementa, te dobiveni signal može konvertirati u video signal,
- Zaslona ili monitor koji kreira vizualni prikaz slike dobiven iz video signala [52].



Slika 6. Blok dijagram glavnih komponenti IR termovizijske kamere [53]

Proces termalnog snimanja počinje prolaskom IR zračenja kroz leće objektiva termalne kamere, nakon čega fokusirano zračenje dolazi do detektorskog elementa na kojem dolazi do stvaranja električnih impulsa. Nastali električni impulsi se šalju u jedinicu za obradu signala, koja prevodi podatke iz elemenata u podatke za prikaz. Jedinica za obradu signala šalje podatke na zaslon, gdje se pojavljuje slika prikazana u različitim bojama ovisno o intenzitetu IR emisije (uobičajeno tamnije nijanse predstavljaju hladna područja, dok svjetlije, prema bijeloj, predstavljaju toplija područja, iako je kod pojedinih kamera obratno označavanje, dok je kod suvremenih kamera moguće u postavkama odrediti način prikaza, nap.a.) [53].

IR detektori odnosno termalne kamere bazno se mogu podijeliti na hlađene i nehlađene detektore, a razlika je, osim u potrebi za hlađenjem samog detektora i u načinu detekcije IR zračenja, kao i u načinu snimanja IR zračenja.

Prema tome, IR kamere mogu biti napravljene kao uređaji za skeniranje snimajući samo jednu točku ili jedan red slike u jednom trenutku ili uz upotrebu dvodimenzionalne infracrvene žarišne ravnine (engl. *Infrared Focal Plane Array* - IRFPA), koji sve elemente slike snima odjednom sa svakim elementom detektora u žarišnoj ravnini. Ovaj tip tehnologije (IRFPA) je dominantan u današnje vrijeme s obzirom da nema pokretnih dijelova zbog čega je snimanje brže, te ima bolju razlučivost u odnosu na uređaje za skeniranje [54]. Općenito, detektore koji se koriste u termalnim kamerama moguće je podijeliti na dva osnovna tipa: na detektore fotona ili termalne detektore [55]. Detektori fotona apsorbirano EM zračenje izravno pretvaraju u promjenu distribucije energije elektrona u poluvodiču od strane promjena koncentracije slobodnog punjenja. S druge strane, termalni detektori apsorbirano EM zračenje pretvaraju u toplinsku energiju uzrokujući porast temperature u detektoru. Tada je električni izlaz toplinskog senzora uzrokovan odgovarajućom promjenom kod nekih fizikalnih svojstava materijala, npr. temperatura koja je ovisna o električnom otporu u bolometru. Detektori fotona obično rade u MWIR segmentu IR spektra, gdje postoji visok toplinski kontrast, čineći ga vrlo osjetljivim na male razlike u temperaturi promatranog objekta ili prostora. Trenutna tehnologija detektora fotona također omogućuje veću brzinu snimanja (engl. *higher frame rate*), međutim glavni nedostatak ove vrste detektora je potreba za hlađenjem. Detektor fotona treba biti ohlađen na temperaturu ispod 77 K kako bi se smanjila toplinski šum. Takvo hlađenje je moguće tekućim dušikom, iako se često koriste i različiti tipovi krio-hladnjaka. S obzirom na tehnologiju izrade i potrebu za

redovitim održavanjima, ali i druge stavke, početna cijena detektora fotona je viša u odnosu na termalne detektore.

Termalni detektori detektiraju radijaciju u MWIR i LWIR segmentu IR spektra i mogu koristiti različite tipove detektora.

Nehlađeni termalni detektori uglavnom su razvijeni na temelju dva tipa detektora, kao fero-električni detektori i kao bolometri [56, 57]. Feroelektrični detektori imaju prednost zbog faze feroelektričnog prijelaza kod .

S druge strane, mikrobolmetar je specifični tip otpornika, a najčešće korišteni materijali u izradi mikrobolometra su Vanadium Oksid (VOx) i amorfni silikon (a-Si). IR zračenje mijenja električni otpor materijala koji potom može biti konvertiran u električni signal i prikazan kao slika. U današnje vrijeme mikrobolometri imaju značajne prednosti u odnosu na feroelektrične detektore, s obzirom da tehnologija temeljena na VOx materijalima ima značajniji tržišni udio. U prvom redu to je zbog toga što mikrobolometri imaju veću osjetljivost. Naime, šum ekvivalentan razlici temperature (engl. *Noise Equivalent Temperature Difference* - NETD), koji pokazuje minimum moguće detektirane temperaturne razlike za VOx detektore iznosi 0,039 K. Pored navedenog, mikrobolometri također imaju manju veličinu piksela na detektoru što omogućuje veću rezoluciju. [58].

Za izradu leća najčešće koristi germanij, a riječ je o bijelo-sivom metaloidnom elementu koji je gotovo u potpunosti propusan za IR zračenje, te također gotovo potpuno odbija vidljivi dio EM spektra. S obzirom na visoku cijenu germanija, veličina leća također predstavlja poseban značaj. F-broj optičkog sustava predstavlja omjer fokalne duljine leća i promjera irisa ili otvora blende. Ovo znači da veći F-broj snižava cijenu leća, međutim istovremeno se smanjuje promjer leće što smanjuje i količinu IR zračenja koje će doći do detektora u kameri. Također je bitno naglasiti da bi se postigla viša razina osjetljivosti, poželjan je što niži F-broj. Za hladene kamere prihvatljiviji je viši F-broj s obzirom da je kod njih moguće povećati vrijeme ekspozicije kako bi zadržala ista razina IR zračenja koje prolazi kroz leću. S obzirom na izravnu ovisnost promjera leće objektiva i F-broja, te količine IR zračenja koje će proći kroz objektiv do detektora, za veću učinkovitost kamere, a samim tim i veću osjetljivost poželjne su leće većih promjera, što dovodi i do visokih cijena termalnih kamera, ali i dodatnih objektiva.

Upravo zbog navedenog, kod kamera s jako velikim promjerom leća, hladene kamere će imati nižu cijenu u odnosu na nehlađene [58].

Osim leća i tipa detektora termalne kamere, značajan dio predstavljaju i drugi elementi kao što su mogućnosti snimanja u određenom video standardu i kompresija video signala, te mogućnosti snimanja slike i kompresije slike, kao i različiti izlazi koji omogućavaju spajanje kamere na vanjsko napajanje, izlazi za spajanje na računalo za daljnju obradu signala ili video snimače za pohranu snimke. Glede navedenog, u današnje vrijeme je uobičajeno da je kamera opremljena USB, mrežnim (*Ethernet*), FireWire i drugim tipovima konektora za spajanje s drugim uređajima. Vezano za video i slikovne standarde, uobičajena je mogućnost kompresije slika u JPEG format, dok je za video uobičajen standard kompresije H264 (ili AVC) ili MPEG. Također su značajni parametri kao što je termalna osjetljivost koja za nehlađene kamere iznosi 40 mK, dok je za hladene uobičajena termalna osjetljivost od 20 mK. Snimke snimljene termalnim kamerama uobičajeno se prikazuju u sivim nijansama, iako se često prikazuju i u pseudoboja kako bi uočavanje toplih objekata bilo lakše. Snimke su obično snimljene u 8 ili 16-bit po pikselu, a kut gledanja iznosi od 1 do 58°. Glede rezolucije snimanja, tržišno dostupne kamere imaju mogućnost snimanja u rezolucijama od 160 x 120 pa sve do 1280 x 1024 piksela, iako pojedini proizvođači promoviraju i termalne kamere koje imaju mogućnost snimanja u tzv. *Full HD* formatu, odnosno u rezoluciji 1920 x 1080 piksela, iako je njihova dostupnost za sada ograničena uglavnom na vojne sustave s obzirom na izrazito visoku cijenu. Unatoč tomu, u novije vrijeme termalne kamere su postale dostupne i za vrlo široki krug korisnika s obzirom da su ugrađene i u pojedine pametne telefone kao što CAT S60 ili CAT S61, čija je cijena pristupačna za širi krug korisnika, iako su mogućnosti značajno slabije (primarno zbog malog promjera leće), u odnosu na standardne termalne kamere koje se koriste u sustavima nadzora i kontrole šticećenih prostora i objekata.

Iako je mikrobolometar među najraširenijim termalnim detektorima, njegova kompleksna proizvodnja i nadalje održava visoku cijenu u kontekstu profesionalnih termalnih kamera te na taj način otežava pristup istraživačima u korištenju profesionalne opreme.

Kao zamjena za mikrobolometar, 2006. godine predstavljena je tehnologija optičko-termalnog snimanja koja je razlikuje od tehnologije mikrobolometra. Za razliku od mikrobolometra koji se oslanja na promjene u električnom otporu kako bi mjerio promjene u temperaturi, optičko-termalna tehnologija se temelji na promjeni u optičkim

svojstvima kad se izloži promjenama temperature. Ključne razlike između mikrobolometra i optičko-termalne tehnologije se očituju u tome što nije riječ o elektroničkom elementu, nego se radi o potpuno pasivnom sloju tankog optičkog filma na staklu, što značajno smanjuje troškove i tehnologiju proizvodnje, itd. Pored navedenog, optičko – termalna tehnologija pruža i niz drugih prednosti u odnosu na mikrobolometar, kao što su niži troškovi proizvodnje, veća skalabilnost, manja potrošnja električne energije, te veliki potencijal u povećanju rezolucije snimaka. [59] Pored navedenog, koliko je područje IR termovizije zanimljivo i atraktivno za istraživanje govori i podatak o velikom broju autora koji su svoje radove posvetili ovom području, odnosno fokus je na termalnim kamerama i njihovoj primjeni u različitim segmentima posla i života općenito.

Tako je u [45] predstavljen detaljan pregled povijesnog razvoja IR detektora. U [60], predstavljena je metodologija testiranja te autor donosi svoje iskustvene spoznaje dvadesetogodišnje karijere bavljenja elektro-optičkim uređajima i istraživanjima u tom području, s posebnim osvrtom na IR termoviziju. U prilog činjenici da je termovizija i njezina primjena već duže vrijeme predmet istraživanja govori i rad autora [47], objavljen 2003. godine, koji donosi pregled tadašnjeg stanja i noviteta u području IR detektora i termovizije. S druge strane, rad objavljen 2017. godine [61], također donosi pregled stanja i trendova u području IR termovizije. Usporedba između prethodno spomenutog i potonjeg rada donosi vidljivu razliku u proširenju područja primjene IR termovizije. Tematika IR termovizije te IR kamera predstavljena je i u [62].

2.6 Termalno snimanje i negativni utjecaji na termografiju

Termalne kamere, zbog mogućnosti detekcije i snimanja IR zračenja, imaju veliku prednost u odnosu na standardne (optičke) kamere koje snimaju u vidljivom dijelu spektra. Prednosti se primarno očituju u mogućnosti snimanja u mraku te u svim vremenskim uvjetima. Unatoč tomu, pojedine atmosferske prilike imaju i određen negativan utjecaj na snimanje termalnom kamerom, a isto se uglavnom očituje u slabijem kontrastu između promatranog objekta i okoline, djelomičnoj apsorpciji toplinske energije koju emitiraju promatrani objekti, zamućenosti slike i sl.

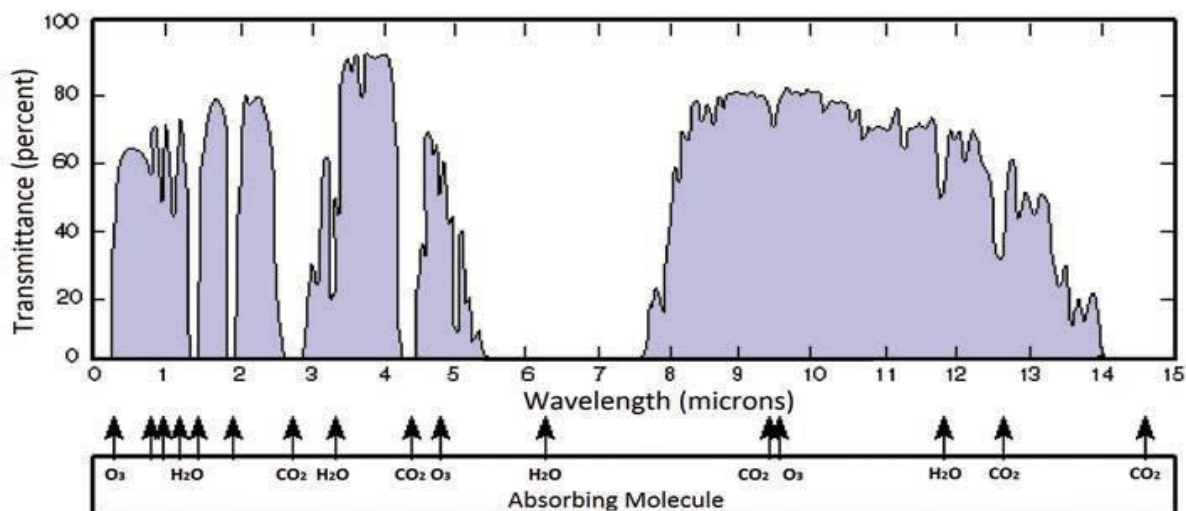
Na kvalitetu snimanja IR termalnom kamerom mogu utjecati različiti uvjeti iz okoline, kako uvjeti koji se odnose na operatera koji upravlja kamerom, tako i različiti utjecaji ovisno o mjestu gdje je kamera postavljena, odnosno da li se kamera nalazi u ruci operatera, da li je montirana na fiksno postolje ili je montirana na vozilo (automobil, letjelicu, borbeno vozilo, tenk ili sl.) koje se kreće. Ovisno o smještaju i položaju kamere, potrebno je voditi računa o šumovima koji se mogu javiti zbog vibracija vozila, neravnina terena ili nestabilnoj ruci operatera, što će dovesti do određenih izobličenja na slici i povećati razinu šumova do razine da promatrani objekt postane u potpunosti neprepoznatljiv na snimci, pa samim tim niti snimka neće biti kvalitetna za obradu primjenom metoda računalnog vida u cilju detekcije i prepoznavanja promatranih objekata. Pored navedenog, na kvalitetu termalne snimke (termograma), utjecaj mogu imati i atmosferski i drugi okolišni uvjeti. Primjerice čestice vodene pare, vodene kapljice, molekule ugljičnog dioksida (CO_2) i ozona, prašina mogu u manjoj ili većoj mjeri utjecati na disperziju toplinske energije koju emitiraju promatrani objekti. S obzirom da apsorpcija ovisi o gustoći plinova kroz koje prolazi toplinsko zračenje, efekt nije moguće konstantno mjeriti i gotovo ga je nemoguće kompenzirati. Nadalje, temperatura zraka u području snimanja izravno utječe na temperaturu kamere (i druge opreme) kao i njihove performanse te učinkovitost, međutim IR oprema ovaj utjecaj može kompenzirati interno. Utjecaj na snimanje također ima i količina toplinskog zračenja emitirana s površine promatranog objekta. Jak vjetar će pojačati prijenos toplinske energije s površine te veći konvektivni gubici energije mogu sniziti površinsku temperaturu. Zbog navedenog, mjerenja ne bi trebalo provoditi u vjetrovitim uvjetima ukoliko je brzina vjetra veća od pet metara u sekundi. S druge strane, jak vjetar može olakšati detekciju objekata koji se nalaze u blizini otvora za prozore i vrata. Utjecaj na snimanje imaju još i kut snimanja i udaljenost, s obzirom da udaljenost i kut snimanja

mogu imati određene utjecaje na interpretaciju termalne snimke. Rezolucija i kvaliteta snimke se smanjuje s povećanjem udaljenosti s obzirom da svaki piksel predstavlja veću površinu. Toplinsko zračenje emitirano s te površine je uprosječno te se gube određeni detalji. Slično je i u slučaju kuta snimanja, odnosno promatranje objekta iz manjeg kuta smanjuje količinu podataka u odnosu na promatranje objekta pod pravim kutom [63].

Ranije je navedeno da se toplina prenosi kondukcijom odnosno provođenjem, konvekcijom i radijacijom odnosno zračenjem, na čemu se zapravo temelji i IR termalno snimanje, odnosno termografija. Prijenos topline zračenjem ili radijacijom razlikuje se od provođenja i konvekcije s obzirom da za prijenos topline zračenjem nije potreban temperaturni gradijent, nije potreban medij za prijenos, a prijenos topline zračenjem moguć je i u vakuumu i to brzinom svjetlosti. S druge strane provođenje i konvekcija zahtijevaju prisutnost medija za prijenos topline koji može biti obliku čvrstih tvari, plinova ili fluida. Kao što je ranije navedeno, svi objekti i tijela čija je temperatura iznad apsolutne nule (0 K) zrače toplinu, što je točka na kojoj se zaustavljaju sva molekularna gibanja, za snimanje termalne slike objekti moraju biti zagrijani na bilo koju temperaturu iznad apsolutne nule te zračiti toplinsku energiju, s obzirom da je radijacija jedini izvor energije da bi termalna kamera mogla snimati promatrane objekte i kreirati sliku. Kao što je već rečeno, svi objekti zrače toplinsku energiju koja se potom reflektira, može biti apsorbirana ali i reemitirana od drugih objekata iz okoline. Ovaj proces može se ponavljati beskonačno, osim ako postoji način da se sustav izolira od svih vanjskih utjecanja (npr. solarnog zračenja). Ako se sustav, sastavljen od n objekata na temperaturi $T_1, T_2, \dots, T_n$ može toplinski izolirati, na temperaturi T_e doći će do temperaturnog ekvilibrija, ali će unatoč tomu sustav nastaviti zračiti toplinu jer temperatura T_e nije apsolutna nula. Termalne kamere objekte mogu detektirati samo kad je njihova snaga emitirane topline P_o , različita od snage emitirane topline koju emitira pozadina i objekti iz pozadine P_B , a ta razlika je dovoljno velika i nalazi se u području osjetljivosti termalnog detektora. Za shvaćanje načina funkcioniranja termalne kamere i termalnog snimanja nužno je temeljito poznavanje mehanizama koji utječu na emitiranu toplinu koju emitiraju objekti i pozadina s obzirom da emitirana toplina ovisi o relativnoj temperaturi objekta, ali i temperaturi pozadine. Općenito, temperature objekta i pozadine se konstantno mijenjaju, a odabir trenutka kada je njihova međusobna razlika najveća je jedna od ključnih stvari za definiranje

adekvatnog plana za terenska snimanja. Postoji veliki broj parametara koji mogu izravno utjecati na jačinu pozadinskog zračenja, a osim prevladavajućih atmosferskih uvjeta, količina detektiranog pozadinskog toplinskog zračenja i utjecaj na snimanje termalnom kamerom uvelike ovisi i dobu dana, lokaciji, nadmorskoj visini kao i smjeru promatranja ovisno o površini Zemlje. Ukoliko je objekt jednake temperature kao i pozadina, u tom slučaju ne postoji temperaturna razlika između objekta i pozadine, te gotovo nemoguća detekcija objekta. Unatoč tomu, termalna kamera može zabilježiti određenu temperaturnu razliku zbog toga što mjeri količinu toplinskog zračenja, a s obzirom da emisivnost objekta nije jednaka emisivnosti pozadine, ipak je moguće je djelomično bilježenje temperaturne razlike [64].

Utjecaj atmosferskih prilika na termalno snimanje vidljiv je i na sljedećem grafičkom prikazu.



Slika 7. Atmosferski "prozori" za snimanje u vizualnom i IR spektru [65]

Utjecaj atmosferskih prilika na termalno snimanje vidljiv je i obliku različitih mehanizama atmosferskog prigušenja toplinskog zračenja, u prvom redu zbog apsorpcije i raspršivanja EM zračenja, odnosno toplinskog zračenja zbog prisutnosti plinovitih molekula ili aerosolnih čestica, kao i zbog prisutnosti vodenih kapljica koje su prisutne u atmosferi. EM, odnosno toplinsko zračenje apsorbiraju agensi kao što su vodena para, ugljični dioksid, dušični oksid, ozon, molekularni i atomski kisik, te dušik. Apsorpcija je općenito zanemariva u vidljivom dijelu EM spektra, kao i u nekoliko atmosferskih prozora unutar IR segmenta EM spektra (Slika 7). S druge strane, atmosfersko raspršivanje je proces u kojem EM zračenje preusmjeravaju plinovite

molekule i čestice aerosola koje postoje u atmosferi. Procesom raspršenja upravlja odnos između radijusa molekula raspršenja ili čestica i valne duljine toplinskog zračenja. Atmosferska apsorpcija i raspršivanje stvaraju dvostruki učinak na osvjetljenje i kontrastnu transmisiju. Prvo, EM odnosno toplinsko zračenje koje emitira promatrani objekt kao i njegova najbliža pozadina je progresivno raspršeno s linije vidljivosti. Osim toga, dio toplinskog zračenja će biti apsorbiran na putu te neće stići do senzora, odnosno termalnog detektora. Prigušenje EM zračenja u atmosferi slijedi zakon eksponencijalnosti u homogenom zraku.

Zraka EM zračenja na objektu udaljenosti R sadrži tok odnosno fluks F_0 , međutim do senzora će doći samo rezidualni fluks F koji je određen prema izrazu:

$$F = F_0^{[-(b+k)R]} = F_0^{(-\sigma_e R)} \quad (4)$$

gdje je:

- R – udaljenost objekta,
- F_0 – svjetlosni tok ili fluks,
- F – rezidualni svjetlosni tok ili fluks,
- b i k - koeficijenti raspršenja i apsorpcije,
- σ_e – koeficijent nestajanja atmosfere.

Nadalje, EM zračenje koje ne dolazi izravno od promatranog objekta ili njegove neposredne okoline rasprši se na putu do senzora.

Govoreći o izravnim utjecajima atmosferskih prilika na termalno snimanje, odnosno ključni utjecaj na prigušenje toplinskog zračenja i izravan utjecaj na smanjenje performansi senzora, kako u vidljivom tako i u IR spektru imaju kiša, oblaci, magla i izmaglica ili smog. Prva tri spomenuta stanja rezultat su prisutnosti vodene pare koja istovremeno ima najveći utjecaj, ali i je i najvarijabilnija. Relativna i apsolutna vlažnost su dva okolišna parametra izravno povezana s prisutnošću vodene pare u atmosferi.

Utjecaj kiše je načelno teško procijeniti s obzirom na varijacije u veličini kišnih kapljica, njihovoj distribuciji, funkcije stanja, oblika kišnih kapi, kao i drugih efekata koji indeks loma vode (vodenih kapi) ima na različite dijelove EM spektra. Kiša prigušuje EM zračenje kroz apsorpciju i raspršivanje, a relativna količina ovisi o omjeru radijusa kišnih kapi prema valnoj duljini. U vidljivom i IR dijelu EM spektra, kišno prigušenje je neovisno o valnoj duljini s obzirom da je radijus kišnih kapi (koji uobičajeno iznosi od

0,5 do 5 mm) mnogo veći od valne duljine. Koeficijent nestajanja, koji mjeri koliko je EM zračenje apsorbirano od strane pojedinog medija (u ovom slučaju kiše), te pokazuje da je u slučaju jače kiše i apsorpcija veća.

Slično kao i kod kiše, utjecaj oblaka je teško procijeniti s obzirom na varijacije u veličini čestica. Prigušenje je ovisno o tipu oblaka i postotku vodene pare. Pored toga, u samom oblaku, prigušenje EM zračenja javlja se kako zbog apsorpcije tkao i zbog raspršenja.

Kada je riječ o magli, ona na smanjenje vidljivost utječe zbog toga što kapljice vodene pare djelomično apsorbiraju i raspršuju EM zračenje, a količina smanjenja vidljivosti izravno ovisi o veličini i distribuciji kapljica [65].

Nadalje, utjecaj plinova iz atmosfere i vremenskih uvjetia odnosno njihove utjecaje na termalno snimanje istražio je i Holst [66], koji je utvrdio da plinovi u atmosferi apsorbiraju i raspršuju zračenje na putu od promatranog objekta do termalne kamere, a prigušenje zračenja predstavlja transmisivnost atmosfere. S obzirom da su plinovi iz atmosfere također zagrijani na određenu temperaturu, obično na temperaturu zraka odnosno atmosfere, također zrače određenu količinu toplinske energije. Zračenje koje se nakuplja na liniji promatranja naziva se putno ili stazno zračenje, odnosno zračenje na liniji promatranja. U vidljivom dijelu spektra, zračenje na liniji promatranja nastaje zbog raspršivanja svjetla iz okoline. Raspršenje svjetla nam onemogućava promatranje objekata na velikim udaljenostima ili tijekom magle i sumaglice. Sličan efekt se javlja i kod vidljivog i kod IR segmenta EM spektra, te u konačnici smanjuje kontrast. Kada se uzmu u obzir atmosferski utjecaji, ta izlazna vrijednost detektora postaje:

$$V_{DETECTOR} = kR_D(T_A M_T + M_A) \quad (5)$$

gdje je

- T_A – transmisivnost atmosfere,
- M_A – zračenje na putu promatranja,

te u konačnici zračenje koje dolazi do detektora odnosno termalne kamere predstavlja zbroj prigušenog zračenja promatranog objekta i zračenja na liniji promatranja (snimanja). Magnituda transmisivnosti i zračenja na putu ovisi o prigušnim elementima iz atmosfere, dužini puta kao i spektralnoj osjetljivosti termalne kamere. S obzirom da

se stanje atmosfere kontinuirano mijenja, elementi prigušenja mogu varirati svaki sat, ali i dnevno, sezonski ili ovisno o lokaciji (urbana sredina, primorje i sl.). Gotovo je nemoguće jednoznačno odrediti transmisivnost ili putno zračenje. Te vrijednosti je potrebno mjeriti ili barem procijeniti prilikom provođenja specifičnih mjerenja. U cilju provođenja studija učinkovitosti, kao reprezentativne, obično se uzimaju vrijednosti za pojedinu lokaciju (prema uobičajenim vojnim scenarijima), ali unatoč tomu takve studije daju samo prosječne vrijednosti učinkovitosti [66].

Također značajan faktor pri snimanju termalnom kamerom je koeficijent nestajanja zračenja na putu od promatranog objekta do termalne kamere. Prema Beer-Lambertovu zakonu [67] transmisivnost atmosfere je predstavljena jednadžbom:

$$T_A = e^{-\gamma^R} \quad (6)$$

gdje je:

- R – dužina puta koji prijeđe zračenje od promatranog objekta do termalne kamere,
- γ - koeficijent nestajanja.

S obzirom da su raspršenje i apsorpcija neovisni parametri, koeficijent nestajanja može biti odvojen u svoju raspršujuću komponentu σ_A i apsorpcijsku komponentu k_A . Ovo omogućuje odvojeno proučavanje ovih dvaju fenomena, te u konačnici kombiniranje njihovih rezultata u jednadžbi:

$$\gamma = \sigma_A + k_A \quad (7)$$

U vojnoj primjeni se, umjesto koeficijenta nestajanja, koristi nešto drugačiji koeficijent prigušenja u atmosferi koji je definiran kao:

$$\tau_A \approx e^{-\gamma} \quad (8)$$

iz čega slijedi:

$$T_A = \tau_A^R = e^{-\gamma^R} \quad (9)$$

S tim u svezi napomenuto i da terminologija ovisi o pojedinom autoru. Međutim, kada se govori o koeficijentu nestajanja (γ) ili koeficijentu prigušenja (τ_A), uvijek se mjerenje odnosi na udaljenost od jednog kilometra (1 km). Općenito, koeficijent nestajanja (γ) gotovo uvijek ima malu vrijednost koja se kreće u rasponu od 0,001 do 0,4, dok koeficijent prigušenja (τ_A) ima veće vrijednosti, obično u rasponu od 0,65 do 0,95, dok je u potpunoj prozirnoj atmosferi $\gamma = 0$, dok je $\tau_A = 1$.

S obzirom da svaki plin i aerosolni agent ima svoju specifičnu spektralnu transmisivnost, koeficijent nestajanja postaje:

$$\gamma(\lambda) = \sigma_A(\lambda) + k_A(\lambda) \quad (10)$$

i

$$T_A(\lambda) = e^{\gamma(\lambda)R} \quad (11)$$

Uobičajena krivulja transmisivnosti atmosfere prikazana je na slici 7., međutim u obzir je uzeta samo apsorpcija. Glavni element apsorpcije u području od 3 do 5 μm je ugljični dioksid (CO_2) koji apsorbira valnu duljinu od 4,3 μm , a ova apsorpcija je vidljiva već nakon nekoliko metara udaljenosti. Vodena para određuje gornje i donje ograničenje kako za MWIR tako i za LWIR područje IR spektra. Voda, odnosno vodena para su također dominantni apsorpcijski elementi u LWIR području. Kako se povećava koncentracija vodene pare, tako se transmisivnost za LWIR valne duljine smanjuje, čak i brže nego za MWIR. Ovo govori da su MWIR termalne kamere prikladnije za područja s izraženijim postotkom vodene pare u atmosferi kao što su to tropska ili primorska područja. Nadalje, utjecaj na termalno snimanje ima i udaljenost odnosno put koji zračenje prijeđe od promatranog objekta do termalne kamere. Prema ranijoj jednadžbi razvidno je da je transmisivnost funkcija udaljenosti, te je sukladno tomu razvidno i da je atmosferska transmisivnost veća na manjim udaljenostima, dok je na većim udaljenostima ona manja. S obzirom da je transmisivnost atmosfere funkcija valne duljine, nužno je taj efekt opisati i matematički uz pomoć integrala, s obzirom da integral izračunava površinu koja se nalazi ispod krivulje. Stoga, prigušeno zračenje emitirano od strane promatranog objekta proizvodi:

$$V_{\text{DETECTOR}} = k \int_{\lambda_1}^{\lambda_2} R_D(\lambda) T_A(\lambda) M(\lambda, T_T) d\lambda \quad (12)$$

Izlaz ovisi o pojedinom korištenom detektoru. LWIR termalne kamere koriste HgCdTe detektore koje će dati drugačiji izlaz nego ne-hlađene kamere koje koriste mikrobolometar. Slično je u slučaju korištenja MWIR kamera gdje će Pt:Si detektor dati drugačiji rezultat negko InSb detektor. Pregled transmisivnosti otkriva da atmosferska stanja imaju drugačiji utjecaj na MWIR i na LWIR kamere, pa je tako manja udaljenost snimanja pod manjim utjecajem nego veća udaljenost, s tim da rezultat snimanja ovisi i o temperaturi promatranog objekta, kao i spektralnoj osjetljivosti detektora. Za male udaljenosti (manje do 10 m), atmosferska transmisivnost se može smatrati jedinstvenom za MWIR i LWIR područje.

Na termalno snimanje izravno utječe koncentracija ozona, vodene pare, ugljičnog dioksida i ugljičnog monoksida. Visoki postotak apsolutne vlage u atmosferi smanjuje transmisivnost, a to je posebno izraženo u primorskim područjima gdje je u atmosferi visoka prisutnost i čestica morske soli. Iako vidljivost može biti dobra, učinkovitost termalne kamere može biti umanjena zbog apsorpcije od strane čestica vodene pare. Aerosolni agenti kao što su prašina, mali organizmi, morska sol, kapljice vodene prašine (sumaglica ili magla), dim te drugi umjetni aerosolni agenti (smog), također mogu imati utjecaja na učinkovitost termalnog snimanja. Relativna koncentracija navedenih čestica je različita o pojedinom području snimanja. Primjerice, prisutnost ugljičnih spojeva (CO ili CO_2) je veća u urbanim sredinama nego u šumskim ili planinskim područjima. Negativan utjecaj na termalno snimanje očituje se u povećanoj apsorpciji toplinskog zračenja od strane plinova prisutnih u atmosferi, a s druge strane prisutnost aerosolnih agenata, osim na povećanu apsorpciju utječe i na povećano raspršivanje toplinskog zračenja. Potrebno je napomenuti i da je meteorološki domet mjera koeficijenta raspršenja toplinskog zračenja.

Vodena para ili postotak apsolutne vlage u atmosferi je glavni apsorber IR zračenja. Nažalost, transmisivnost nije linearno povezana sa apsolutnom vlažnosti zraka. Apsolutna vlažnost zraka uobičajeno se ne mjeri kvantitativno, ali se izračunava prema meteorološkim standardima, npr. temperatura i relativna vlažnost. Prisutnost vodene pare u zraku može biti od nekoliko g/m^3 u pustinji, pa do 40 do 50 g/m^3 u tropskim prašumama ili u blizini površine mora. Kako se apsolutna vlažnost povećava, tako se transmisivnost smanjuje, a apsolutna vlažnost se kontinuirano mijenja, kako po danu tako i po noći, a ovisi i o godišnjim dobima.

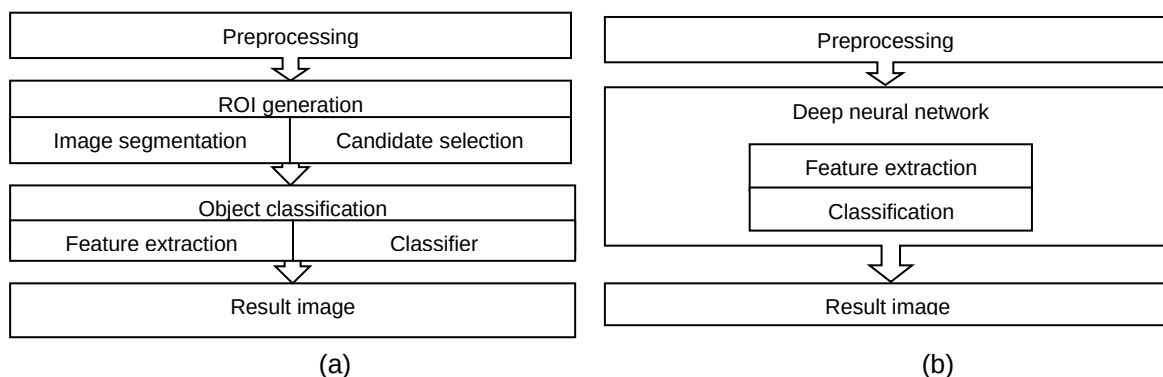
Osim vodene pare, na prigušenje termalnog zračenja u atmosferi određene utjecaje mogu imati i različiti aerosolni agenti koji su prisutni u atmosferi u području snimanja. Magnituda raspršivanja ovisi o odnosu između promjera prisutnih čestica i valne duljine. Kada je taj omjer visok (veliki promjer prema valnoj duljini), raspršivanje može biti značajno. S druge strane, kada je taj omjer mali, tada je i raspršivanje minimalno. Promjer čestica aerosolnog agenta pri liganj izmaglici iznosi oko 1 μm , što znači da ukoliko je valna duljina manja od 1 μm (standardne optičke kamere), tada je raspršivanje značajno, te je i rezultat snimanja uvelike smanjen. Vidljivost se mjeri prema meteorološkoj vidljivosti te ovisi o vizualnoj procjeni prisutnosti odnosno koncentraciji aerosolnih agenata. Tako se procjenjuje da je pri čistom vremenu vidljivost i do 50 km, dok je pri izuzetno gustoj magli vidljivost smanjena na tek 50-ak

metara. Nadalje, kako se povećavaju čestice aerosolnih agenata, tako se povećava i koeficijent raspršenja pri svim valnim duljinama s obzirom da se transmisivnost smanjuje, dok se istovremeno spektralno područje s konstantnim koeficijentom povećava. Govoreći o termalnom snimanju, treba uzeti u obzir da su valne duljine termalnog snimanja 5 do 20 puta veće od valnih duljina vidljivog svjetla, zbog čega je i manji utjecaj aerosolnih čestica na termalno, nego na standardno optičko snimanje. Posljedično, može se reći da termalne kamere, s minimalnim gubicima, mogu detektirati prisutnost objekta pri slaboj izmaglici, što bi odgovaralo meteorološkoj udaljenosti od 4 do 10 km. Pri nešto gušćoj magli, mogućnost detekcije se smanjuje na 2 do 4 km, s tim da je utjecaj raspršenja veći na MWIR nego na LWIR područje, što znači da bolji učinak ima LWIR nego MWIR kamera. Pri gustoći magli kada se čestice vodene pare značajno povećavaju u dimenziji, ali i gustoći, kao i u slučaju kiše, pri meteorološkoj udaljenosti do 500 m, utjecaj raspršenja je podjednak na MWIR i na LWIR područje.

Osim koncentracije vodene pare, odnosno vodenih kapi, različitih plinova te drugih aerosolnih čestica u atmosferi, utjecaj na termalno snimanje može imati i drugo toplinsko zračenje koje se pojavljuje u području snimanja između promatranog objekta i termalne kamere. IR putno zračenje primarno se javlja zbog atmosferske samoemisije toplinskog zračenja. Ono je neovisno o temperaturi promatranog objekta i prisutno je čak i kada kada je promatrani objekt skriven ili nije prisutan u vidnom polju kamere. Magnituda ovog pozadinskog zračenja može varirati ovisno o smjeru promatranja, nadmorskoj visini, lokaciji, dobu dana te meteorološkim uvjetima. Putno zračenje ovisi o prosječnoj temperaturi u atmosferi, a predstavlja prosjek temperature koja se dobije uprosječivanjem temperatura na području snimanja, odnosno na pravcu od promatranog objekta do termalne kamere. Prosječne temperature ovise o kutu podizanja. Kada je promatranje pravocrtno, tada se nebo prikazuje kako vrlo hladno. U neposrednoj blizini tla, gdje je ujedno i veća koncentracija aerosolnih čestica, pri izuzetno malom kutu podizanja (u odnosu na tlo), aerosolne čestice će imati temperaturu gotovo jednaku temperaturi tla odnosno površine Zemlje. S obzirom da ovakvo pozadinsko zračenje ne sadrži nikakve informacije o promatranom objektu, ono se može smatrati šumom. Međutim, utjecaj ovog zračenja na termalno snimanje postaje značajan samo kada je riječ o snimanju na izuzetno velikim udaljenostima kada je transmisivnost atmosfere jako niska [66].

3 Pregled dosadašnjih istraživanja

Zadatak detekcije osoba je utvrditi prisutnost osoba na slikama ili videu. U nekim slučajevima, posebice kada je angažiran operater za pregled slika ili videa, detekcija osoba može biti svedena samo na razinu detekcije kretanja, te eventualno praćenja s obzirom da ne postoji potreba za klasifikacijom objekata čije kretanje je detektirano. Detekcija kretanja u tom slučaju je dovoljna kao pomoć operateru za detekciju prisutnosti osoba, odnosno objekta na slikama ili videu, posebice kada je riječ o pregledavanju snimki u realnom vremenu za područje štićenih prostora i objekata, te u situacijama kada nadzorni sustav služi samo kao pomoćno sredstvo, bez potrebe za automatskom detekcijom, odnosno klasifikacijom osoba jer će u slučaju utvrđivanja prisutnosti osobe na videu, operater aktivirati potrebne alarme. Nadalje, metode detekcije osoba na slikama ili video zapisu primarno možemo podijeliti na klasične koje se temelje na starijoj generaciji algoritama i ranim metodama strojnog učenja, te suvremene metode detekcije osoba koje temelje na konvolucijskim neuronskim mrežama. Nadalje, unatoč napretku u razvoju tehnologije, odnosno termalnih kamera, detekcija raspoznavanje osoba na termalnim slikama, pogotovo na većim udaljenostima još uvijek predstavlja izazovno područje istraživanja. Termalne slike, pogotovo osoba, odnosno objekta snimljenih na većim udaljenostima snimljene objektu prikazuju relativno malim brojem piksela što otežava detekciju osobe odnosno objekta. Pored toga, izazov kod detekcije osoba na termalnim slikama je i činjenica da kada su snimana osoba, odnosno objekt iste ili slične temperature kao i pozadina, mali je kontrast između osobe, odnosno objekta i okoline. Navedeni problem uvelike mogu negativno utjecati na rezultat detekcije i raspoznavanja osoba, odnosno objekata na termalnim slikama ili video snimkama [68].



Slika 8. Tipičan model (a) klasične arhitekture sustava za nadzor koji se temelji na termoviziji i (b) sustava temeljenog na neuronskoj mreži [69]

3.1 Klasični pristupi detekcije i raspoznavanja osoba

Jednostavni sustavi za nadzor sastoje se od termalne kamere koja može detektirati male promjene u temperaturi između promatranog objekta, odnosno osobe i okoline. Termalna kamera ima mogućnost detekcije IR zračenja koje se reflektira ili koje emitiraju objekti na promatranom području temeljem čega stvara sliku ili video na temelju podataka o razlikama u temperaturi. S obzirom da objekti na promatranom području najčešće imaju različitu temperaturu od okoline, termalna kamera ih može uočiti. Detektirani objekti će biti prikazani kao različite temperaturne regije na termalnoj snimci. Nadalje, termovizijske snimke mogu biti crno-bijele, odnosno prikazane u sivim tonovima kod kojih su hladni objekti prikazani tamnim gotovo crnim nijansama, dok su topli objekti prikazani svijetlim nijansama, do bijele, nijanse od crne do bijele boje prikazuju razinu zagrijanosti objekta. Osim toga, mogu biti prikazani i u boji pri čemu je raspon od crne ili tamnoplave za hladne objekte do svijetložute ili bijele za tople objekte [70, 71].

Tipična arhitektura sustava klasičnog sustava za nadzor, odnosno detekciju prema [69] sastoji se od faze pretprocesiranja, generiranja regije interesa (engl. ROI - *Region of Interest*) prema kojima se provodi segmentacija slike ili videa te se odabire ona s najvećom vjerojatnosti da sadrži objekt, a nakon toga slijedi klasifikacija objekata s ekstrakcijom značajki iz odabrane regije interesa, te se na kraju provodi klasifikacija temeljem ranije naučenog klasifikatora, a krajnji rezultat je detekcija i/ili raspoznavanje objekata (slika 8). U slučajevima kada je potrebno utvrditi i točnu lokaciju objekta na slici, detekcija se provodi temeljem podataka o klasi objekta i njegovoj lokaciji na slici. S druge strane, moguće je i da cijela slika bude označena nazivom objekta koji je prepoznat i detektiran na slici.

3.1.1 Termovizijski sustavi za detekciju osoba temeljeni na oduzimanju pozadine

Jednostavan sustav koji su predstavili [70] sastoji se od termovizijske kamere, računala i algoritma za detekciju osobe. Autori navode da su ostvarili prilično visoku stopu detekcije (oko 83,3% na 1.000 slika) primjenom ovog, prilično jednostavnog sustava.

U [72] predstavljen je sustav za detekciju osoba koji se sastoji od termalne kamere, prijenosnog računala i jeftinog mikro-kontrolera sa softverskom podrškom koja je razvijena u MATLAB-u, a detekcija je ostvarena gotovo u stvarnom vremenu. Sustav funkcionira tako da u fazi pretprocesiranja slike detektira područja s višim temperaturama primjenom histogramskog izjednačavanja, a potom slijedi segmentacija slike primjenom dogovarajućeg praga. Također su implementirane i morfološke operacije kako bi se uklonio pozadinski šum na slici, a pored toga korištene su i određene karakteristike vezane za ljudsko tijelo kao što su omjer ograničenja, filteri lokalnog histograma kao i morfološki model. Za detekciju osoba, autori su koristili omjer visine i širine ljudskog tijela te su na taj način automatski uklonili sve objekte koji ne predstavljaju osobe. Postigli su 90 % točnost detekcije uz 10 % prisutnost lažno pozitivnih detekcija (engl. *False Positives*), što je značajno visoka točnost s obzirom na jednostavnu opremu koja je korištena. Glavna prednost ovog termovizijskog sustava je upravo upotreba jednostavnog i jeftinog mikrokontrolera koji je omogućio *online* i *offline* obradu podataka, izračun značajki i odlične rezultate u izoliranim područjima kao što su šumska ili planinska područja. S druge strane, primijenjeni algoritam nije moguće primijeniti za kompleksniju analizu slike te istovremenu detekciju više osoba. Pored toga, glavni nedostatak je nemogućnost detekcije osoba u različitim položajima tijela kao što je ležanje, pogureno hodanje, čučanje, osoba koje se djelomično preklapaju ili osoba koje su djelomično skrivene.

Autori [73] predstavili su metodu za detekciju osoba na termalnim slikama odnosno videu niske rezolucije. Istražili su set značajki koji se temelji na deskriptorima histograma gradijenta (engl. HOG - histogram of gradients) [6]. Prisutp se temelji na ekstrakciji histograma gradijenta iz rekurzivno generiranih okvira (engl. *patch*) te slijednog računanja omjera histograma između njih. Svaki set je definiran u pojmovima relativnog položaja unutar pomičnog okvira, te je potom rekurzivno primijenjen na manje okvire. Omjeri histograma gradijenta između okvira postaju vektori značajki, a

potom je primijenjen linearni stroj s potpornim vektorima (engl. *SVM – Support Vector Machine*) za klasifikaciju osoba. Eksperiment je proveden na skupu slika OTCBVS [73], kao i na preko 600 termalnih slika iz različitih skupova podataka namijenjenih učenju SVM-a, odnosno stroja s potpornim vektorima.

Pristup detekciji osoba temeljen na klasičnim metodama na MWIR i LWIR termalnim slikama također je predstavljen u [74]. Autori su testirali nekoliko metoda za pretprocesiranje slika kako bi utvrdili koja je regija na slici najbolja, odnosno koja eventualo sadrži osobu, te su u testovima eksperimentirali sa filtriranjem prema prosjeku i medijanu, kombinirali su razlike između piksela i rubova s pozadinom i ulaznom slikom uzimajući u obzir dimenzije te omjer visine i širine objekta, smjer snimanja kamere i projekciju perspektive. Također su koristili podatke o horizontalnom i vertikalnom histogramu detektirane površine te su spajali dijelove slike s dvije ili više detektiranih osoba. U eksperimentalnoj fazi koristili su MWIR i LWIR slike iz OTCBVS skupa podataka [73], kao i proizvoljno kreirani skup slika. Spomenuti OTCBVS skup se sastoji od sekvenci razvrstanih u kategorije podijeljene u kategorije prema dijelu dana kada je snimanje provedeno (jutro, dan i sl.), te vremenskim uvjetima (kiša, sunce). Rezolucija slika je 320 x 240 piksela, a dimenzije osoba na slikama variraju od 20 do 68 piksela (širina), te od 34 do 103 piksela (visina). Proizvoljni skup podataka sastoji se od sekvenci termalnih slika koje obuhvaćaju različite oblike ponašanja, odnosno načine hoda (hodanje, trčanje, mirno stajanje, sjedenje). Svaka sekvenca sadrži od 64 do 144 sličice rezolucije 320 x 240 piksela, s ukupno 768 sličica koje su snimljene ICI 7320 termalnom kamerom. Za procjenu učinkovitosti detekcije, primjenjen je Paskalov kriterij [75] koji mjeri postotak preklapanja sličice s detekcijom i bazne sličice (engl. *Ground Truth Frame*).

U [76] predstavljen je multidirekcijski termalni sustav za detekciju uljeza i onesviještenih osoba. Predloženi sustav koristi visoko-rezolucijske termalne kamere spojene na IR hiperbolično zrcalo kako bi mogao nadzirati prostor u 360°. Korištena je kamera FLIR Thermovision A-20M, koja ima temperaturnu osjetljivost u koracima od 0,1 u rasponu od 20 to 350°C, s mogućnošću detekcije od 35 do 40°C, i rezolucijom od 320x240 piksela te alarmnim modulom (kao što su zvučnici za računalo), dok su razvoj algoritma za detekciju glave uljeza izveli u MATLAB-u.

Za mapiranje, slike su snimljene iz više smjerova sve do panorame, a log-polarno mapiranje slike korišteno je za postizanje šireg kuta vidljivosti. Scene nadziranog područja su potom prenete u sustav za obradu slika, a slike snimljene preko hiperboličnog zrcala su mapirane u panoramsku sliku s tim da su korištene jedino regije interesa kako bi se slika konvertirala u sive nijanse, tako da je područje koje odgovara postavljenom temperaturnom pragu postavljeno na vrijednost 1, dok je za sve ostalo postavljena vrijednost 0. Autori su naveli da je postotak detekcije uljeza 81,38%, a vrijeme potrebno za aktivaciju alarma je 2,27 sekundi.

Nadalje, u [77] autori su, za klasifikaciju termalnih slika, koristili CENTRIST značajke [78] i linearni SVM klasifikator [79]. U svrhu eksperimenta snimljeno je pet sati materijala pri 25 FPS (engl. *Frame per Second*) korištenjem terlane kamere NEC-C200 s rezolucijom od 320x240 piksela. Nakon snimanja video je smanjen na 5 FPS a rezolucija povećana na 640x480 piksela, te je u konačnici ekstrahirano 2.000 slika na kojima se nalaze osobe (pješaci). Snimanje je provedeno u blizini ceste tako da su slike obuhvaćale cestu, osobe i bicikliste. Rezultati testiranja pokazali su poboljšanje u brzini i točnosti raspoznavanja objekata kad su korištene CENTRIST značajke zajedno s HOG značajkama. Na primjer, za testiranje s HOG značajkama bilo je potrebno 1.407 sekundi, dok je za testiranje s korištenjem samo CENTRIST značajki bilo potrebno samo 463 sekundi. Dodatni izvori podataka kao što je kretanje ili samosličnost nisu korišteni s CENTRIST značajkama.

Sustav za klasifikaciju objekata (osobe – pješaci, biciklisti i automobili) na videima niske rezolucije predstavljen je u [80]. Klasifikacija je provedena na simuliranim i atmosferski stvarno degradiranim video snimkama rezolucije 640x480 piksela, a cilj je bio usporediti različite utjecaje različitih kontroliranih uvjeta na učinkovitost klasifikacije primjenom gemoetrijskih značajki i tekstura objekta. Za klasifikaciju su korišteni standardni klasifikatori kao što su KNN i SVM. Za oba korištena klasifikatora pokazalo se da postoji evidentan utjecaj atmosferskih uvjeta na detekciju i klasifikaciju pokretnih objekata na termalnim snimkama. Glede detekcije pokretnih objekata, pokazalo se da konvencionalne metode detekcije, kao što je oduzimanje pozadine, trebaju prilagodbu za detekciju pragova kada su u pitanju atmosferski uvjeti.

3.1.2 Sustavi termalnog nadzora temeljeni na praćenju

U [81] predstavljena je modificirana metoda detekcije i praćenja osoba na termalnim videozapisima na temelju vremenskih i prostornih značajki u slikama (engl. *mRTCT – modified Real-Time Tracking Compressive*). Za klasifikaciju vremenskih značajki korišten je SVM kako bi se automatiziralo pokretanje RTCT algoritma, s tim da je navedeni algoritam dodatno modificiran kako bi zadovoljavao zahtjeve za istovremenom evaluacijom vremenskih i prostornih značajki tijekom praćenja. Za potrebe testiranja kreirali su novi skup 12 i 16 bitnih slika u rezoluciji 640x512 piksela pri 25 FPS. Snimke su dodatno obrađene kako bi maksimalno odgovarale realnim uvjetima snimanja s različitih udaljenosti. Zaključno su istaknuli kako su rezultati testiranja pokazali da modificirani RTCT algoritam daje bolje rezultate nego bazni RTCT.

Metoda temeljena na klasičnom pristupu detekciji osoba predstavljena je i u [82] a temelji se na statičkoj analizi za detekciju osoba, nakon čega slijedi dinamička analiza Lukaas-Kanade optičkog toka [83]. Njihov algoritam poravnava dvije slike s ciljem maksimalnog preklapanja te pokušava utvrditi postojanje pokreta između dvije slike nakon čega se slike slijedno uspoređuju upotrebom Newton-Raphson [84] iteracija za izračun grešaka u gradijentu.

3.1.3 Sustavi termalnog nadzora temeljeni na prostorno-vremenskim značajkama

Metoda za detekciju predstavljena u [85] odnosi se na detekciju tijekom dana upotrebom FIR (engl. *FIR - Far Infra Red*) valne duljine i primjenom hijerarhijskih SURF značajki (engl. *SURF - Speeded Up Robust Features*) [85]. Treba napomenuti da su autori koristili nešto drugačiju nomenklaturu glede valnih duljina unutar IR spektra te se spomenuti FIR zapravo odnosi na LWIR valne duljine (od 7 – 14 μm). Metoda je primarno razvijena za detekciju pješaka u prometu za inteligentna osobna vozila kao pomoć u vožnji. Glavni doprinos metode očituje se u iskorištavanju specifičnosti FIR, odnosno LWIR termalnih snimki u cilju razvoja brzog, robusnog i na skaliranje neosjetljivog detektora. Predloženi model sastoji se od tri modula, a svaki se temelji na ubrzanom robusnom podudaranju značajki. Prvi modul omogućuje generiranje ROI, s obzirom da na FIR snimkama mjerilo u kojem se prikazuju osobe može uvelike varirati, iako se glava gotovo uvijek pojavljuje u svjetlijim područjima. ROI se detektiraju s visokom stopom odziva sa hijerarhijskim SURF značajkama na dijelu koji prikazuje

glavu. Drugi modul služi za klasifikaciju cijelog tijela upotrebom SVM klasifikatora, koji omogućuje povećanje preciznosti s niskim računalnim zahtjevima. U trećem modulu autori su kombinirali algoritam pomaka prosjeka sa SURF značajkama za praćenje neosjetljivima na skaliranje unutar slike kako bi se poboljšala robusnost sustava. Osim već spomenutog iskorištavanja određenih specifičnosti LWIR termalnih snimki, autori su kao novost istaknuli i način kreiranja algoritma za automatsko generiranje odgovarajućih značajki za opis glave koje su dovoljno prilagodljive da toleriraju važne promjene u prikazu osobe. Još jedna novost je definiranje hijerarhijskog prikaza koji omogućava ne samo detekciju i praćenje glave osobe, nego i cijelog tijela, te iskorištavanje hijerarhijske strukture značajki za daljnje unaprijeđivanje vremenskog procesiranja, detekcije i učinkovitosti klasifikacije. Novost je i prijedlog vremenskih SURF značajki za podudaranje osoba na sljedećim slikama koje se uzimaju za obradu. Nadalje, svoje pretpostavke su testirali na Tetravision skupu slika [86] koji je pripremljen u Laboratoriju za umjetni vid i inteligentne sustave na Sveučilištu u Parmi, a nastale su snimanjem stereokamerom u vidljivom i FIR spektru tijekom dana, te prikazuju stanje na cesti. Skup se sastoji od 3.900 slika, s tim da se u skupu za učenje nalazi 987 ručno anotiranih primjera i korišteni su za generiranje HC i treniranje SVM-a. Ostatak skupa, (2.092 primjera) iskorišten je za optimizaciju strukture HC-a i potvrdu SVM parametara. Glede postignutih rezultata, autori su istaknuli da najbolji rezultat daje kombiniranje lokalnih i globalnih SURF značajki upotrebom RBF-SVM klasifikatora, a testiranje je omogućilo i iskorištavanje hijerarhijske strukture za poboljšanje SVM vektora značajki, kao i klasifikacijskih rezultata. Drugim testiranjem pokazali su da predloženi pristup ima potencijal i za rješavanje problema detekcije na srednjim i velikim udaljenostima kao i problema sa djelomičnom skrivenošću osoba. Također su pokazali da se kombiniranjem detekcije osoba i klasifikacije s praćenjem u jedinstveni sustav, povećava pouzdanost kao i da se u značajnoj mjeri smanjuje postotak lažnih detekcija. Zaključno je istaknuto da predloženi sustav čak ima potencijal i za rješavanje potpune skrivenosti glave osobe, jer je u tom slučaju moguće generirati novi skup značajki za ostatak tijela.

Autori [87] su također predstavili metodu za detekciju osoba na termalnim snimkama koja se temelji na karakteristikama pojedinih dijelova tijela osoba. Njihova metoda temelji se na izdvajanju područja koja moguće prikazuju osobu na temelju histograma prostornog modela intenzitet - udaljenost kako bi se zaobišla prepreka kada se

promatra samo područje visokog intenziteta. Autori su istaknuli da su rezultati postignuti testiranjem na spomenutom OTCBVS skupu slika obećavajući.

Nadalje, u [88] predstavljeno je korištenje značajke pod nazivom TPIHOG ili T π HOG (engl. *Thermal-Position-Intensity-Histogram Of Oriented Gradient*) odnosno termalna pozicija intenziteta histograma orijentiranih gradijenata. Autori su kombinirali T π HOG i AKSVM (engl. *Additive Kernel SVM*) za noćnu detekciju pješaka. Predložena metoda obuhvaća detaljne podatke o lokaciji gradijenta te samim tim ima veću snagu nego standardni HOG. Zaključno su autori istaknuli da kombinacija T π HOG-AKSVM daje dobre rezultate noćne detekcije osoba s brzim računalnim odzivom, te da je bolja i brža od SVM-a s linearnom jezgrom i drugih klasičnih metoda. Predložena metoda testirana je na KAIST skupu slika [89].

3.1.4 Sustavi termalnog nadzora s pokretnim kamerama

U novijim istraživanjima detekcije osoba uz pomoć termalnih kamera one više nisu statične nego pokretne tako da su montirane na različite tipove vozila, robota ili letjelica (dronova). Također, niža cijena toplinskih senzora omogućila je ugradnju u pametne telefone i učinila termalne slike još dostupnijim istraživačima i široj javnosti za eksperimentiranje. U [90] predstavljen je sustav za detekciju osoba temeljen na Android pametnom telefonu s ugrađenim jeftinim termalnim modulom (engl. *TIM – Thermal Imaging Module*). Rezultati testova su pokazali da ovakav sustav ima sposobnost detekcije na udaljenostima od oko 29 metara. Još jedan trend u području detekcije osoba je korištenje bespilotnih letjelica odnosno dronova (engl. *UAV – Unmanned Aerial Vehicle*). Detekcija i praćenje osoba na slikama koje su snimljene upotrebom drona predstavlja veliki izazov s obzirom na nisku rezoluciju slika, kretanje drona, nestabilnost slike te relativno male dimenzije objekta [91]. Autori su predložili sustav za detekciju i praćenje osoba koji najprije filtrira značajke gradijenta pojedine regije i geometrijska ograničenja za detekciju bloba koji moguće prikazuje osobu, a potom slijedi klasifikacija detektiranih okvira primjenom linearnog SVM-a s hibridnim deskriptorom koji kombinira HOG i diskretnu kosinusnu transformaciju (engl. *DCT – Discrete Cosine Transformation*) kako bi postigao učinkovitu detekciju. Autori u [92] primijenili su filtriranje po medijanu na ulaznim termalnim snimkama, a potom su sliku podijelili na blokove veličine 15x15. Računali su omjer entropije, energije kao i kombinaciju ovih kriterija sa pojedinih blokova na cijelu sliku. Sukladno tomu, ako je omjer bio veći od jedan, taj blok je detektiran kao osoba. Također su pokazali da

primjena kombinacije energije i entropije daje bolji rezultat nego primjena samo jednog od tih parametara. Metoda koju su predstavili, osobu ne može dobro prepoznati u otežanim vremenskim uvjetima kao što je magla. Također dolazi do zamjene pojedinih objekata sa slike za osobu (primjerice, stup električne rasvjete i automobile prepoznaje kao osobu), ali jednostavnost metode je njezina značajna prednost. Autori u [93] su svoju metodu temeljili na detekciji regije interesa (ROI). Kako bi to postigli, koristili su adaptivnu statističku metodu koja je orijentirana prema projekciji intenziteta, a koja sadrži horizontalne i vertikalne projekcije. Fleksibilni prag je korišten za eliminaciju šuma i kompenzaciju slabe osvjetljenosti piksela. ROI su rastavljene dvodimenzionalnom kontinuiranom transformacijom valića dvostruke gustoće dvostrukog stabla (engl. *2D DD DT CWT - 2-Dimensional Double Density Dual Tree Continuous Wavelet Transform*) na tri razine, dok je entropija 96 visokofrekventnih podsegmenata uzeta kao značajka za učenje SVM klasifikatora. Nedostatak metode je niska preciznost detekcije osoba za vrijeme vrućih dana i magle, dok je visoka preciznost u dobrim vremenskim uvjetima i njezina generalizacija prema drugim skupovima podataka prednost ove metode. Nadalje, metodu temeljenu na klasičnom pristupu prezentirali su i u [94] koja se temelji na adaptivnom oduzimanju pozadine na slikama. Kako bi ovo postigli, autori su najprije odredili prvi vrh histograma za svaku sliku (P_e), a potom su računali intenzitet za svaki piksel. Ukoliko je intenzitet svakog piksela bio manji od P_e pretvoren je u nulu, a vrijednost drugih piksela postaje njihova razlika u odnosu na P_e . Rezultat ovog postupka je snižavanje svjetline na slici. Nakon toga su koristili *high-boost* filter za povećanje intenziteta na slici, te je tako dobivena binarna slika koja je rezultat računanja praga iz varijance, entropije i prosjeka vrijednosti piksela. U konačnici, bijele regije na binarnoj slici smatraju se prikazom osoba. Glede učinkovitosti metode, testovi su pokazali nisku preciznost u otežanim vremenskim uvjetima tijekom vrućih dana, što je i nedostatak ove metode, dok je adaptivno dobivanje pozadine jedna od prednosti. Wei Li et al. [95] su koristili HOG za detekciju osoba na termalnim slikama. U njihovoj metodi, slika je podijeljena na manje blokove veličine 20x25, a potom je svaki od tih blokova dodatno podijeljen na manje blokove dimenzija 4x5. SVM klasifikator je treniran sa HOG značajkama vektora za svaki mali blok. Koristili su tri geometrijske značajke uključujući prosjek kontrasta, standardnu devijaciju i omjer svijetlih prema ukupnom broju piksela kako bi dobili bolji rezultat. Nedostatak metode je niska preciznost detekcije osoba pri otežanim

vremenskim uvjetima, dok je pri dobrim vremenskim uvjetima postignuta visoka preciznost, što je jedna od prednosti.

Xinyue Zhao et al. [96] su koristili značajke histograma distribucije oblika (engl. *SDH - Shape Distribution Histogram*) kao i mape kontura (engl. *CSM - Contour Saliency Map*) kako bi detektirali osobe na termalnim slikama. U predstavljenoj metodi najprije se oduzima pozadina na slici primjenom prosjeka intenziteta na slikama za učenje. Nakon toga se izračunava CSM za prednji plan slike. Na kraju, većina istaknutih kontura određuje se primjenom OTSU metode za određivanje praga [97, 98]. Kod ove metode primjenjuju se morfološki operatori kako bi se smanjio šum, a spojene regije iz prednjeg plana se ekstrahiraju i poravnavaju s kvadratnim okvirom (engl. *Bounding-box*). Prednost metode je visoka preciznost u svim atmosferskim uvjetima s korištenjem nekoliko pozitivnih primjera za učenje, međutim metoda ne daje dobar rezultat kada je riječ o nepokretnim osobama i slikama koje su znatno zasićene šumom.

Nadalje, metodu detekcije osoba temeljenu na klasičnim metodama predstavili su i Olmeda et al. [100], a temelji se na upotrebi jedne termalne kamere i nepromijenjenog Kalmanovog filtera (engl. *UKF - Unscented Kalman filter*), a modul za detekciju pješaka namijenjen je za vozila, odnosno detekciju pješaka u sklopu naprednog sustava za pomoć vozaču (engl. *ADAS – Advanced Driver Assistance Systems*).

3.2 Metode detekcije i raspoznavanja osoba temeljene na CNN

Primjena konvolucijskih neuronskih mreža (CNN) odnosno neuronskih mreža općenito umjesto klasičnih algoritama za detekciju i klasifikaciju je najnoviji trend u području detekcije i raspoznavanja osoba, pa je tako i u [100] korištena neuronska mreža za detekciju osoba. Nadalje, Chunwei Yang et al. [101], su za detekciju regija interesa (ROI), a u cilju povećanja brzine detekcije osoba, primijenili BING (engl. *BING - Binarized Normed Gradients*) [102]. Zatim su regije interesa podijelili u preklapajuće blokove, a potom su svaki blok podijelili u ne-preklapajuće ćelije. Neki od ovih preklapajućih blokova smatrani su kao pozitivni (pješaci, osobe), dok su drugi smatrani negativnima (nisu osobe). Na kraju su iz ovih primjera izračunate značajke prikaza osobe i HOG značajke za svaku ćeliju pojedinačno. Za ekstrakciju značajki prikaza, odnosno pojavnosti osobe, sive vrijednosti svakog piksela kandidata kojemu je promijenjena dimenzija su normalizirane s vrijednosti 2. Tako normalizirani pikseli su preoblikovani u matricu s jednim redom koja je korištena kao značajka prikaza kandidata. Nakon toga je FFNN (engl. *FFNN - Feedforward Neural Network*) s jednim skrivenim slojem trenirana na ovim značajkama pri čemu je korišten ELM (engl. *ELM – Extreme Learning Machine*) [103] kao metoda za učenje FFNN. ELM je relativno nova metoda učenja za neuronske mreže s jednim skrivenim slojem kod koje se neuroni u skrivenom sloju mogu nasumično generirati neovisno o podacima za učenje i okolini primjene. Prema navodima autora, ELM je pokazao dobru sposobnost u kontekstu nelinearne regresije i klasifikacije [103], te je savladao izazove kao što je sporo učenje i niska računalna skalabilnost. Neodstatak predstavljene metode je niska preciznost detekcije osoba u otežanim vremenskim uvjetima kao i za vrijeme dana s visokim temperaturama, dok je s druge strane prednost metode u detekciji nepokretnih osoba.

Također, u [104 – 106] istraženo je koliko su metode temeljene na dubokom učenju i neuronskim mrežama uspješne za detekciju i raspoznavanje osoba na RGB termovizijskim slikama. U istraživanju je korišten YOLO algoritam temeljen na neuronskoj mreži [13 – 16]. Cilj istraživanja je bio ispitati učinkovitost ovog algoritma za detekciju osoba na termalnim videozapisima koji su snimljeni tijekom zimskog razdoblja, u noćnim uvjetima i pri različitim vremenskim uvjetima i na različitim udaljenostima, od 30 do 215 metara [107]. Autori su naveli da je originalni YOLO model treniran na COCO RGB skupu slika postigao loše rezultate detekcije (oko 23%),

međutim nakon treniranja na vlastitom skupu slika, preciznost detekcije je iznosila oko 97%. Za detekciju osoba na istom skupu slika, korišten je YOLOv3 [15, 104]. Učinkovitost, nakon dodatnog treniranja na slikama koje su snimljene u drugačijim vremenskim uvjetima (vedro, kiša, magla) se razlikuje ovisno o vremenskim uvjetima pa je tako najbolji rezultat postignut po kišnom vremenu (98,8%), a zatim po vedrom vremenu i maglovitom vremenu (97,85%). Eksperiment je također pokazao da je relativno mali skup slika za učenje modela neuronske mreže dovoljan za postizanje vrhunskih rezultata detekcije osoba na termalnim slikama, što je svakako poticajno za daljnja istraživanja u kontekstu implementacije CNN modela za detekciju i raspoznavanje osoba na termalnim slikama.

U [108] predložena je učinkovita metoda za detekciju osoba na termalnim slikama u vanjskom okolišu koja se prilagođava izmjeni vremenskih uvjeta. U prvom koraku se procjenjuju vremenski uvjeti temeljem globalnih značajki termalne slike. Nakon toga, za svaku situaciju odgovarajući algoritam se primjenjuje za detekciju osoba. Kako bi se ovo postiglo, termalne slike su podijeljene u tri kategorije: dobro kao što je vedro vrijeme, loše kao što je kišovito vrijeme i vruće kao što je ljetno razdoblje kada su osobe tamnije od okoline. Nakon toga se primjenjuje dvodimenzionalna dvostruka gustoća dualnog stabla diskretne valične transformacije na tri razine koja se dobija iz ulaznih slika, dok se koeficijenti energije niskih frekvencija dobivaju na trećoj razini, a računaju se kao diskriminantna značajka za identifikaciju vremenskih uvjeta. FFNN se potom trenira na ovom vektoru značajki kako bi se odredila kategorija vremenskih uvjeta. Na kraju se primjenjuje predodređeni odgovarajući algoritam za pojedinu kategoriju vremenskih uvjeta za detekciju osoba. Predložena metoda ima visoku stopu detekcije tako da je detekcija osoba na dva često korištena skupa (OSU thermal pedestrian database i OSU Color-Thermal database) podataka veća od 99%.

Heo et al. [109] su se u navedenom radu fokusirali na detekciju osoba u realnom vremenu na termalnim slikama snimljenim u noćnim uvjetima. U radu su predstavili ABMS metodu (engl. *Adaptive Boolean Map Based Saliency*) kako bi dodatno istaknuli osobu u odnosu na pozadinu ovisno o dijelu godine. Za raspoznavanje osoba odnosno pješaka koristili su YOLO algoritam baziran na konvolucijskoj neuronskoj mreži koja se ponešto razlikuje od standardnih metoda temeljenih na konvolucijskim neuronskim mrežama. Za razliku od originalnog YOLO algoritma, navedeni autori su YOLO kombinirali s mapom kontura koja je kreirana primjenom spomenute ABMS metode

koja je primijenjena kao jezgra, a temelji se na prethodnom znanju da je osoba odnosno pješak istaknutiji u odnosu na pozadinu. Predloženu metodu su testirali na termalnim slikama koje su snimljene iz vozila u pokretu odnosno na skupu slika KMUPD [110], a rezultati su bolji nego kod drugih sličnih suvremenih metoda.

Herrmann et al. [111] su predložili metodu koja se temelji na primjeni konvolucijskih neuronskih mreža koje su prethodno trenirane na standardnim slikama snimljenima u vizualnom spektru. Kako bi to postigli, najprije su primijenili pretprocesiranje termalnih slika koje ih je maksimalno približilo RGB domeni. Navedeni postupak je omogućio da RGB značajke budu učinkovite u novoj domeni (termalnoj). U drugom koraku su preostali jaz svladali finim podešavanjem CNN koja je prethodno trenirana na ograničenom skupu termalnih podataka. Također su istražili različite aspekte jaza između domena termalnih i RGB slika, a primjeri obuhvaćaju dinamički raspon, zamućenje i kontrast, a testiranja su proveli na KAIST bazi slika. Nadalje, Zhang et al. [68] su predstavili novi, praktičan sustav nadzora temeljen na termalnoj kameri, odnosno video snimkama, koji se temelji na poboljšanoj rezoluciji te automatiziranoj detekciji i raspoznavanju objekta (engl. *ATD/R - Automatic Target Detection/Recognition*) koji može imati primjenu kako u civilnom tako i vojnom sektoru. S obzirom da kod detekcije i raspoznavanja objekata na termalnim slikama postoji problem povezan s malim brojem piksela na objektu, posebice na velikim udaljenostima, autori su primijenili metodu super-rezolucije kako bi se povećao termalni trag objekata, odnosno rezolucija objekta te su optimizirali osnovnu kvalitetu ulaznih slika za detekciju objekata. Kako bi riješili izazov detekcije kod objekata izuzetno male rezolucije (malog broja piksela), koristili su konvolucijsku neuronsku mrežu temeljenu na *faster R-CNN* neuronskoj mreži upotrebom LWIR skupa slika koji su sami pripremili i anotirali. Sustav su testirali u različitim vremenskim uvjetima korištenjem dva skupa podataka u kojima su prikazane osobe, odnosno pješaci te šest različitih tipova motornih vozila. Autori su u navedenom radu pokazali da njihov sustav može detektirati objekte izuzetno male rezolucije s vrlo visokim rezultatima, posebice u odnosu na klasične metode detekcije.

Problem detekcije, odnosno klasifikacije objekata na termalnim slikama upotrebom konvolucijskih neuronskih mreža obradili su i Rodger et al. [112], zatim Rodin et al. [113], Shahid et al. [114].

3.3 Baze slika

Skupovi slika koji se koriste za učenje modela i testiranje u području termalne detekcije i raspoznavanja osoba igraju važnu ulogu kako za definiranje modela tako i za evaluaciju njegove učinkovitosti i pouzdanosti. Termalni skupovi podataka odnosno slika značajno se razlikuju ovisno IR segmentu EM spektra (NIR, SWIR, MWIR, LWIR), ali prema tipu opreme koja je korištena za snimanje, udaljenostima snimanja, broju snimljenih osoba i drugih objekata, ciljanim dijelovima tijela (lice ili cijela osoba). Pored toga razlikuju se i prema scenarijima snimanja, uvjetima (unutra/vani) varijacijama osvjetljenja, promjenama položaja tijela, izraza lica, itd. Osim navedenog, snimanje i kreiranje termalnog skupa podataka zahtijeva vrlo sofisticiranu opremu koja je značajno skuplja od kamera koje snimaju u vizualnom dijelu spektra (iako je u novije vrijeme dostupna i jeftinija oprema, poput termalnih modula implementiranih u pametne telefone ili dodataka za dronove), za snimanja na većim udaljenostima i otežanim vremenskim uvjetima ipak je nužna skuplja, profesionalna oprema koja ima veću osjetljivost na toplinsko zračenje koje emitiraju osoba ili drugi snimani objekti i koja imaju veću rezoluciju snimanja. Uz sve navedeno, snimanje za termalni skup slika za učenje modela dubokog učenja u području detekcije osoba zahtijeva i dodatni angažman kako istraživača, tako i volontera koji će sudjelovati u istraživanju kao osobe koje će biti snimane, a to se u prvom redu očituje u zahtjevnosti snimanja u noćnim uvjetima vani kada se volonteri moraju kretati po različitim tipovima terena bez osvjetljenja (ukoliko se želi simulirati stvarne uvjete), kao i dodatna izloženost atmosferskim uvjetima ukoliko se snimanje provodi po kiši, magli, snijegu ili niskim temperaturama.

Glede skupova podataka koji su dostupni za istraživanje postoje oni kao što je OTCBVS (eng. Object Tracking and Classification Beyond the Visible Spectrum) Benchmark Database [115], KAIST Multispectral Pedestrian Dataset [89], CASIA Night Gait Dataset [116] and CASIA Infrared Night Gait Dataset [117] koji su postali svojevrsno mjerilo te se uzimaju kao relevantna vrijednost za termalnu detekciju osoba.



Slika 9. Primjeri slika iz nekoliko skupova: a) OTCBVS (fusion of thermal and color imagery), b) KAIST pedestrian dataset c) CASIA infrared night gait dataset

Kada se koriste postojeće baze podataka, cilj istraživanja je potrebno prilagoditi dostupnim podacima, za razliku od situacije kada se baza podataka, odnosno skup slika kreira ispočetka, sukladno vlastitim potrebama i ciljevima istraživanja. Na primjer, CASIA skup [118] sastoji se od snimki 153 (130 muškaraca i 20 žena), snimljenih u rezoluciji 320x240 piksela pri 25 FPS. Sve snimane osobe su na četiri snimke hodale normalno, zatim su dva puta hodale usporeno, dva puta usporeno i ubrzano, a nakon deset sekvenci je za svaku osobu dobiveno je ukupno 1.530 sekvenci hoda.

S druge strane, [107] kreiran je skup slika koji simulira realne uvjete i situacije kada se neovlaštene osobe kreću u nadziranom području kao što je državna granica ili neki štićeni objekt (vojna infrastruktura ili sl.). Osobe su snimane u noćnim uvjetima na različitim udaljenostima od kamere sve do maksimalne udaljenosti na kojoj su osobe bile vidljive (po vedrom vremenu na 110, 110 do 165 m, po magli od 30 do 50 m, te po jakoj kiši od 30 do 215 m), s tim da su snimane osobe mijenjale brzinu kretanja i položaje tijela (od puzanja, skrivanja, normalnog uspravnog hoda, pogurenog hodanja do trčanja uspravno i pogurenog trčanja). Snimanje je provedeno termalnom kamerom FLIR ThermaCam P10 LWIR uz korištenje standardnog i tele objektivu (P/B series), po vedrom vremenu, gustoj magli i jakoj kiši.

Nadalje, u [119] predstavljen je skup slika koji je kreiran za svaki dio IR segmenta EM spektra s različitim postavkama i u različitom vremenu. SWIR slike su snimljene u kontroliranim i polukontroliranim uvjetima, u unutarnjem i vanjskom prostoru, na udaljenostima od 60 do 400 m. MWIR slike su snimljene u unutarnjem prostoru korištenjem FLIR kamere na udaljenostima od oko 1,95 m kroz razdoblje od 20 dana. Snimanje za pojedine skupove podataka provedeno je i na manjim udaljenostima, npr. na udaljenosti od 62 cm što je korišteno u [120] za snimanje 17 osoba koje su sjedile

na stolici, a snimljene su FLIR 7 termalnom kamerom. Nadalje, u [121] skup se sastoji od 1.000 LWIR slika od 50 osoba snimljenih ThermoVisionA40 kamerom.

U tablici 1., prikazan je pregled termalnih baza korištenih za istraživanja u području detekcije osoba, s kratkim opisom snimaka koje sadrže, IR segmenta koje uključuju i referencu na istraživanja u kojem su korištene.

Tablica 1. Pregled baza termovizijskih slika za detekciju osoba [107]

IR band	Dataset	Ref.
LWIR	OTCBVS Benchmark Database; OSU Color-Thermal Database; image resolution 240x320	[115]
LWIR	CASIA Night Gait Dataset	[116]
LWIR	CASIA Infrared Night Gait Dataset; human silhouette image is normalized to the resolution of 129x130 NLPR Database - daytime	[117]
LWIR	Controp Fox 720 thermal camera, video sequence resolution 640 x 480 pixels at distance 3.000 m	[122]
LWIR	520 images captured with FLIR ThermoVision A320 InfraRed camera; resolution 320x240	[123]
LWIR	OTCBVS datasets	[73]
MWIR LWIR	OTCBVS; 360 x 240 pixels; 768 images captured with ICI 7320 thermal camera; resolution 320x240	[124]
LWIR	Video sequence 25 FPS at resolution 640x512	[125]
LWIR	OTCBVS dataset 01; OSU Thermal Pedestrian Database\00001\00001; OSU Thermal Pedestrian Database\00002\00002; Terravic Motion Infrared Database	[126]
LWIR	Tetravision image database	[85]
LWIR	ANTID dataset; Resolutions range from 320x240 to 920x480 both 8 and 16 bit pixel; PNG file type. Images are recorded with various FLIR cameras and images representing different type of objects: humans, animals, cars, static objects...	[127]
LWIR	IR Tau2 640x512, 13mm f/1.0 (HFOV 45°, VFOV 37°) FLIR BlackFly (BFS-U3-51S5C-C) 1280x1024, Computar 4-8mm f/1.4-16 megapixel lens (FOV set to match Tau2). Conditions: Day (60%) and night (40%) driving, streets and highways, clear to overcast weather. Frame Annotation Label Totals: 10.228 total frames and 9.214 frames with bounding boxes: Person (28.151), Car (46.692), Bicycle (4.457), Dog (240), Other Vehicle (2.228); Video Annotation Label Totals: 4.224 total frames and 4.183 frames with bounding boxes.: Person (21.965), 2. Car (14.013), Bicycle (1.205), Other Vehicle (540).	[128]
LWIR	FLIR ThermoVision A-20M; 10.000 thermal images at resolution 320x240 pixels from height 1,5 m	[129]
LWIR	NEC-C200 infrared thermal camera with 320x240 pixel image resolution up-sampled to 640x480 using cubic kernel; 17.000 images in testing dataset	[130]
LWIR	MoviRobotics S.L. indoor autonomous mobile platform mSecurit™ with FLIR thermal camera	[82]
LWIR	MTIS - thermal imaging system launched on a mobile phone, resolution 64 x 62; distance: 2 - 9 m, max 29 m	[131]
SWIR	WVU Outdoor SWIR Gait (WOSG) Dataset	[132]
LWIR	1.440 walking samples collected from 8 subjects in different walking conditions at distance 2,6 m	[133]

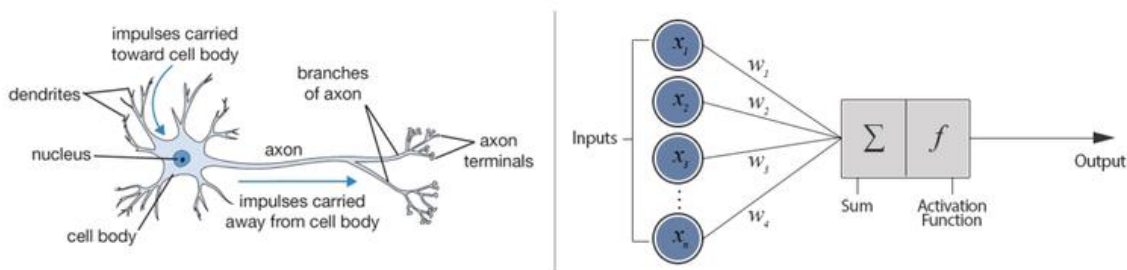
4 Umjetne neuronske mreže (engl. *ANN – Artificial Neural Networks*)

Razvoj računala nemjerljivo je doprinio razvoju društva u cjelini, što je općepoznata i nepobitna činjenica. Paralelno s razvojem računala i porastom njihove snage te mogućnosti u smislu rješavanja sve kompleksnijih zadataka i problema (u sve kraćem vremenu), rasli su i apetiti korisnika u smislu želja za daljnjim povećanjem snage i mogućnosti, posebice u segmentu povećanja automatizacije prepoznavanja i rješavanja zadataka. S vremenom je došlo do postavljanja pitanja da li, i kada, će računala biti u mogućnosti rješavati zadatke ("razmišljati") na isti način kao i ljudi, odnosno da li će posjedovati inteligenciju ljudi. Takva pitanja su dovela i do postupnog usmjeravanja razvoja računala u smjeru pojave i razvoja umjetne inteligencije (engl. *AI - Artificial Intelligence*).

Umjetne neuronske mreže ubrajamo u metode umjetne inteligencije, koja je dio računalnih znanosti u području projektiranja inteligentnih računalnih sustava koji predočavaju karakteristike koje povezujemo s inteligencijom u ljudskom ponašanju. Tijekom evolucije ljudski mozak poprimio je niz karakteristika koje se povezuju s inteligencijom, a neke od karakteristika koje se povezuju s inteligencijom su: paralelno obavljanje više operacija, sposobnost učenja, sposobnost generaliziranja, sposobnost prilagođavanja, suvislo postupanje s informacijama, tolerancija na pogreške i nepotpune informacije. S druge strane, karakteristike računalnih sustava su: brza provedba numeričkih složenih proračuna i rad s velikim količinama podataka.

Prema definiciji, "*Neuronska mreža je međusobno povezani skup jednostavnih procesnih elemenata, jedinica ili čvorova čija se funkcionalnost temelji na biološkom neuronu*". Sposobnost obrade mreže pohranjuje se u međusobnim vezama, odnosno težinama (engl. *weights*), koje su dobivene postupkom prilagođavanja ili učenja iz niza uzoraka za treniranje (učenje) [134]. Druga definicija umjetnih neuronskih mreža glasi: "*Umjetne neuronske mreže su masovno paralelne prilagodljive mreže jednostavnog nelinearnog računanja korištenjem elemenata koji se zovu neuroni, a namijenjeni su apstrahiranju i modeliranju nekih funkcionalnosti ljudskog živčanog sustava u pokušaju djelomičnog zadržavanja njegove računske snage.*" [135]

Biological Neuron versus Artificial Neural Network



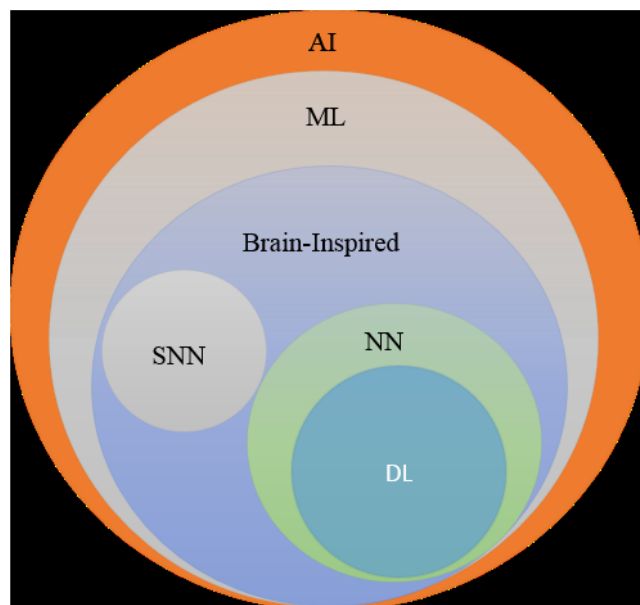
Slika 10. Usporedba biološkog i umjetnog neurona [136]

S obzirom da umjetna neuronska mreža oponaša strukturu i način funkcioniranja mozga, tako je upotrijebljena i analogija za opis osnovnih gradivnih elemenata umjetne neuronske mreže – **neurona**.

Svaki biološki neuron se, kako je prikazano na slici 10., sastoji od tri dijela:

1. **tijela stanice** (soma) – koje sadrži jezgru ili nukleus s informacijama o nasljednim značajkama,
2. **dendrita** – kraćih niti oko stanice, koji prenose signale (impulse) s drugih neurona,
3. **aksona** – dugih i tankih niti, koji prenose signal do drugih neurona pri čemu se grana u vlakna,
4. **Sinapsi** - funkcionalnih jedinica između završetka aksona prethodnog neurona i dendrita sljedećeg neurona koje oslobađaju materijal potreban stanici za prijenos signala, neurotransmiter, pri čemu se odvija elektrokemijska reakcija. Impuls se prenosi preko sinapsi s jednog na drugi neuron. Dendriti pojačavaju ili prigušuju impuls, sumiraju se u jezgri tijela te se putem aksona i sinapsi prenose na druge neurone [137, 138].

S druge strane, unatoč postojanju određene analogije, umjetni neuron ne simulira ljudski mozak na atomskoj ili molekularnoj razini, nego predstavlja apstrakciju biološkog neurona u smislu njegovih osnova te nelinearne integracije signala, plastičnosti ili konkurentnosti. Kao i biološki, umjetni neuron prihvaća signale kroz umjetne dendrite, obrađuje taj signal u umjetnoj somi ili tijelu neurona i kao izlaz prema drugim povezanim neuronima daje obrađeni signal. Umjetni neuron oponaša ovu aktivnost tako što prihvaća signale, dodjeljuje im nasumične vrijednosti težina prema određenim težinskim parametrima, objedinjuje sve težine signala kroz svoju aktivacijsku funkciju koja oponaša prostorno-vremensku obradu na hridi aksona kod biološkog neurona, a potom prosljeđuje signale drugim spojenim neuronima [138].

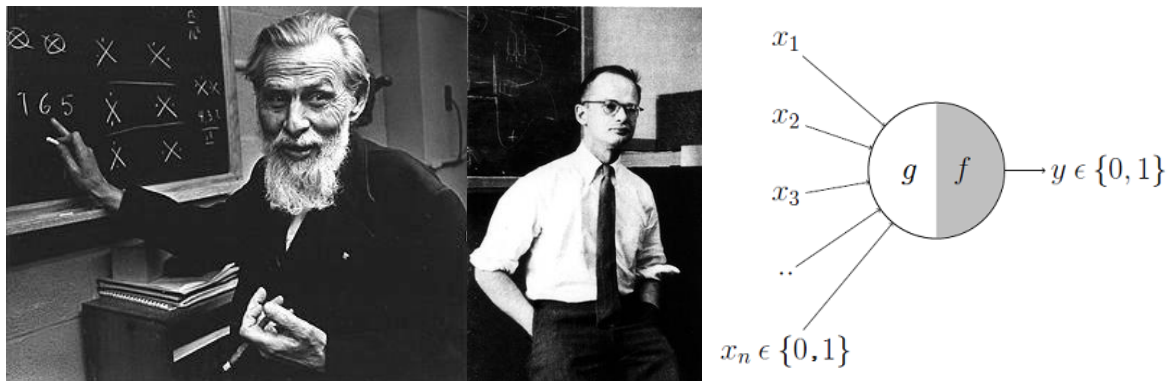


Slika 11. Međuodnosi umjetne inteligencije (AI), strojnog učenja (engl. ML - Machine Learning), umjetnih neuronskih mreža, dubokog učenja (engl. DL – Deep Learning) i Spiking Neural Networks (SNN) [139]

Prvi koraci u smjeru razvoja umjetnog koncepta koji će biti u stanju oponašati rad ljudskog mozga, odnosno umjetne neuronske mreže napravljen je 1943. godine, kada su Warren McCulloch i Walter Pitts [140] pokušali razumjeti kako mozak može stvoriti vrlo složene uzorke pomoću velikog broja međusobno povezanih osnovnih stanica, te su u svom radu dali vrlo pojednostavljen model neurona. Njihov model neurona, koji se uobičajeno naziva MCP neuron (po prvim slovima njihovih imena, nap.a) i koji modelira ključne značajke bioloških neurona dao je važan doprinos razvoju umjetnih neuronskih mreža. Izvorni MCP neuroni imali su ograničenja, te su dodane značajke koje su im omogućile "učenje". Njihov model neurona se naziva i TLU (engl. TLU - Threshold Logic Unit). Iako je prvi umjetni neuron kao osnovna jedinica neuronske mreže predstavljen prije gotovo 80 godina, postavlja se pitanje zašto su se one popularizirale tek prije nekoliko godina. Jedan od osnovnih razloga je činjenica da su neuronske mreže računalno jako zahtjevne, a tada postojeća računala nisu bila dorasla takvom zadatku. Kroz kasnije godine, tek je pojavom velikih strukturiranih baza podataka, ali i značajno dostupnijih računalnih resursa, postala moguća šira primjena neuronskih mreža.

4.1 Povijesni pregled razvoja neuronskih mrež, tipovi NN mreža i konvolucijske neuronske mreže

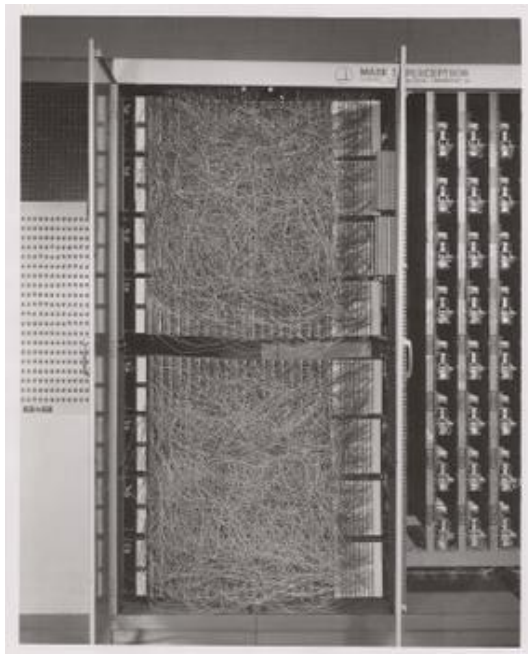
Ranije je već spomenuto da je prvi koncept neuronskih mreža predstavljen 1943. godine, međutim zbog tada ograničenih računalnih resursa, nije bio primjenjiv u praksi.



Slika 12. Warren McCulloch (lijevo), Walter Pitts (sredina) i njihov model umjetnog neurona (desno) [141, 142]

Naknadno je Alan Turing, poznati britanski matematičar, posebno poznat po razbijanju Enigme u II. Svjetskom ratu, 1947. godine predvidio daljni razvoj neuronskih mreža i strojnog učenja s osvrtom na mogućnost da će imati značajan utjecaj i na zapošljavanje. Turing je 1950. godine u svom radu postavio je pitanje "Može li stroj misliti?", glede čega je sam Turing naveo da je riječ o "igri oponašanja" te je postavio i opisao test kasnije prozvan Turingov test koji dokazuje da li "stroj može misliti" [143]. Šire o Turingovu testu, te objavama u kojima se tvrdi da su računala položila isti navedeno je u [144] i [145]. Pored navedenog, postoje i određena predviđanja da će računala u potpunosti biti sposobna imitirati čovjeka oko 2029. godine [146]. Također, značajan napredak u razvoju neuronskih mreža, odnosno umjetne inteligencije općenito, od znanstvene zajednice je zahtijevao da postavi nove standarde za dokazivanje da računalo uistinu oponaša ljudski mozak, pa je tako predstavljen novi tip testa za umjetnu inteligenciju nazvan "Multimodal Turing test" [147]. Kronološki gledano, sljedeći značajniji napredak u razvoju neuronskih mreža, odnosno dubokog učenja napravljen je 1952. godine kada je Arthur Samuel nastavio kreirati prve programe računalnog učenja, koji su bili namijenjeni igranju igre dame. Njegov program je bio jedinstven po tome što je svaki put, kada se igrala dama, računalo bilo bolje ispravljaјуći pogreške iz ranijih igara i pronalazeći bolje načine za pobjedu na temelju tih podataka. Ovaj primjer automatskog učenja je jedan od prvih primjera strojnog učenja [148]. Prvi pravi temelji razvoja neuronskih mreža zapravo su postavljeni tek 1957. godine kada je Frank Rosenblatt objavio da je moguće

"konstruirati elektronički ili elektromehanički sustav koji će naučiti prepoznati sličnosti ili identitete između obrazaca optičkih, električnih ili tonalnih informacija na način koji može biti usko analogan percepcijskim procesima biološkog mozga" [149].



Slika 13. Mark I Perceptron Machine [141]

U okviru svog istraživanja Rosenblatt je predstavio prvi perceptron koji je u osnovi MCP neuron, gdje ulazi najprije prolaze kroz "preprocesore", koji se nazivaju asocijacijske jedinice. Te jedinice za pridruživanje otkrivaju prisutnost određenih specifičnih značajki u ulazima. U stvari, perceptron je trebao biti uređaj za prepoznavanje uzoraka, a jedinice za pridruživanje odgovaraju detektorima značajki. Rosenblattovo glavno postignuće pokazalo je da bi, opuštajući neka pravila MCP-a (apsolutnu inhibiciju, jednak doprinos svih ulaza kao i njihovu cijelu prirodu), umjetni neuroni zapravo mogli naučiti iz podataka. Što je još važnije, on je osmislio algoritam za nadzirano učenje za ovaj modificirani model MCP neurona koji je omogućio umjetnom neuronu da sam shvati ispravne težine iz podataka za učenje. U daljnjem razvoju neuronskih mreža i dubokog učenja slijedi 1959. godina kada su otkrivene jednostavne i kompleksne stanice. Naime, te godine su neurofiziolozi i nobelovci David H. Hubel i Torsten Wiesel [150] otkrili dvije vrste stanica u primarnom vidu korteksa: jednostavne i složene stanice. Njihov rad predstavlja prekretnicu za duboko učenje i definitivno je snažno utjecao na daljnja istraživanja s obzirom da su kasnije mnoge umjetne neuronske mreže nadahnute njihovim biološkim istraživanjima. Nadalje, doprinos daljnjem

razvoju neuronskih mreža dao je i Henry J. Kelley 1960. godine koji je objavio rad "Teorija gradijenta optimalnih putanja leta" [151], jedan od glavnih i općepoznatih radova u njegovom području. Iz navedenog rada je proizašlo mnogo njegovih ideja o teoriji upravljanja - ponašanju sustava s ulazima i o tome kako se to ponašanje mijenja povratnim informacijama, a koje su kasnije izravno primijenjene na razvoj umjetnih neuronskih mreža. Njegove postavke su iskorištene za razvijanje osnovnog modela kontinuiranog povratnog zaključivanja (engl. *continuous backpropagation model*) odnosno povratne propagacije grešaka (engl. *backward propagation of errors*) koje se koristi u učenju neuronskih mreža.

Tek kasnije, 1965. godine, predstavljena je prva funkcionalna neuronska mreža za duboko učenje. Alexey Ivakhnenko i V.G. Lapa prvi put su primijenili ono što su do tada bile samo teorije i ideje. Ivakhnenko je razvio grupnu metodu za obradu podataka (engl. GMDH – Group Method of Data Handling), definiranu kao "*obitelj induktivnih algoritama za računalno bazirano matematičko modeliranje višeparametarskih skupova podataka s potpuno automatskom strukturnom i parametarskom optimizacijom modela*" i primijenio je na neuronske mreže [152]. Samo iz tog razloga, mnogi ga smatraju ocem modernog dubokog učenja. Njegovi algoritmi za učenje koristili su duboke višeslojne perceptrone pomoću statističkih metoda na svakom sloju kako bi pronašli najbolje značajke i prosljedili ih kroz sustav. Pomoću GMDH metode, Ivakhnenko je 1971. godine uspio stvoriti duboku mrežu od osam slojeva i uspješno demonstrirao proces učenja u računalnom identifikacijskom sustavu nazvanom Alpha. Nedugo nakon toga, 1974. godine, Paul Werbos je predstavio prvu višeslojnu perceptron mrežu, a riječ je o prvoj verziji mreže s povratnom propagacijom (engl. *Backpropagation network*) koja prevladava nedostatke perceptrona uvođenjem učenja u skrivenom sloju [153].

Sljedeći značajan trenutak u razvoju neuronskih mreža dogodio se 1979./1980. godine kada je Kuniyiko Fukushima, japanski znanstvenik i možda jedan od najpoznatijih inovatora u području neuronskih mreža, predstavio Neocognitron, umjetnu neuronsku mrežu koja je mogla prepoznavati vizualne obrasce [154]. Primarna svrha joj je prepoznavanje rukopisa, ali je mogla prepoznavati različite vrste uzoraka, a čak je imala i mogućnost obrade prirodnog jezika. Na Fukushima rad značajno su utjecali raniji radovi spomenutih Hubela i Wiesel, što je u konačnici dovelo i do razvoja konvolucijskih neuronskih mreža, koje se temelje na organizaciji životinjskog vizualnog korteks. Konvolucijske neuronske mreže su varijacija višeslojnog perceptrona

dizajnirane za upotrebu minimalnih količina predobrada, a kasnije u radu će biti zasebno obrađene.

Nakon Fukushime, u daljnjem kronološkom pregledu razvoja umjetnih neuronskih mreža, značaj doprinos uslijedio je 1982.g., kada je John Hopfield predstavio Hopfieldovu mrežu (engl. *Hopfield Network*) [155]. Hopfieldova mreža je ponavljajuća neuronska mreža (engl. *RNN - Recurrent Neural Network*) [156] koja služi kao memorijski sustav koji može adresirati sadržaj, a i dalje je popularna te je i aktualno popularan implementacijski alat za duboko učenje. Nadalje, jedna od također značajnih točaka u razvoju umjetnih neuronskih mreža je 1985. godina kada je Terry Sejnowski, računalni neuroznanstvenik, iskoristio svoje razumijevanje procesa učenja za stvaranje programa NETtalk koji je bio u stanju naučiti kako izgovarati engleske riječi na isti način na koji to čini dijete, a tijekom vremena je mogao poboljšati pretvaranje teksta u govor [157]. Nadalje, u radu iz 1986. godine, Rumelhart, Hinton i Williams [158] detaljno su opisali povratnu propagaciju greške (engl. *Backpropagation*) te su ujedno i usavršili taj tip umjetne neuronske mreže čime su pokazali da neuronske mreže mogu imati praktičnu primjenu za mnoge zadatke poput prepoznavanja oblika, predviđanja riječi i sl. Nadalje, 1989. godine predstavljena su dva koncepta također značajna u povijesti razvoja dubokog učenja i umjetnih neuronskih mreža. Najprije su Yann LeCun i skupina autora predstavili koncept kombiniranja mreže s povratnom propagacijom greške i konvolucijske neuronske mreže za prepoznavanje ručno napisanih znamenki [159]. LeCun je primijenio algoritam učenja temeljen na povratnoj propagaciji greške na Fukushiminu konvolucijsku neuronsku mrežu. Nakon što je radio na projektu nekoliko godina, LeCun je objavio LeNet-5, prvu modernu konvolucijsku neuronsku mrežu u kojoj je predstavio neke značajne elemente koji se i danas koriste u konvolucijskim neuronskim mrežama. Kao i Fukushima prije njega, LeCun je odlučio svoj koncept primijeniti na raspoznavanje znakova, te je čak kreirao i komercijalni proizvod za automatsko čitanje poštanskih brojeva. Pored toga, njegov je rad rezultirao i stvaranjem MNIST skupa podataka [160], možda najpoznatijeg skupa podataka za strojno učenje.

Zatim je iste godine, 1989., Christopher Watkins, u svojoj doktorskoj disertaciji predstavio koncept Q-učenja (engl. *Q learning*) [161]

Veliki korak naprijed napravljen je 1993. godine kada je Jürgen Schmidhuber riješio zadatak koji je zahtijevao više od 1.000 slojeva RNN mreže [162]. Sljedeći veliki korak napravljen je 1997. godine, kada su spomenuti Schmidhuber i Sepp Hochreiter

predstavili metodu jedinica s kratkotrajnom memorijom (engl. *LSTM – Long Short-Term Memory*). Tom metodom je unaprijeđena efikasnost i praktičnost povratne neuronske mreže (RNN) eliminirajući problem dugoročne ovisnosti [163]. Nakon toga, 1998. godine, također ranije spomenuti LeCun je sa skupinom autora predstavio metodu učenja na bazi gradijenta [160].

Kao što je vidljivo, kroz razdoblje od gotovo 60 godina, neuronske mreže su kontinuirano razvijane, unatoč relativno maloj i kroz to razdoblje relativnoj skupoj računalnoj snazi, međutim nakon predstavljanja LSTM metode 1998. godine, slijedi razdoblje od oko 10 godina tijekom kojeg su neuronske mreže uvelike zanemarene, kao da istraživači nisu bili svjesni njihovih mogućnosti i snage ili nisu bili dovoljno motivirani da bi nastavili kontinuitet u razvoju neuronskih mreža, iako su se tijekom ranijeg razdoblja višestruko dokazale i u praksi. Unatoč tomu, bilo je istraživača koji su nastavili s radom na neuronskim mrežama i polako napredovali. Tijekom razdoblja od 1998. do 2008., odnosno 2009. godine, postala je dostupna sve veća količina podataka, primarno zbog raširenosti kamera ugrađenih u mobitele, ali i zbog sve veće dostupnosti jeftinih digitalnih kamera. Isto tako, tijekom tog razdoblja značajno je rasla računalna snaga, a paralelno s tim snižavala se cijena, te su snažna računala postajala sve dostupnija širokom krugu istraživača. Sve navedeno je dovelo do toga da je postalo nedvojbeno kako će neuronske mreže naći svoje mjesto jer je postalo razvidno da je njihova primjena moguća u sve široj lepezi zadataka. Nakon svojevrsnog zatišja ili bolje rečeno inkubacije, koja je trajala desetak godina, slijedi 2009. godina kada je skupina autora predstavila ImageNet [164], skup podataka za duboko učenje i treniranje neuronskih mreža što je praktički osnova za novi zamah u razvoju i primjeni neuronskih mreža. ImageNet je velika ontologija slika temeljenih na okosnici strukture WordNeta [165]. ImageNet je prilikom predstavljanja imao za cilj popuniti većinu od 80.000 sinteza WordNeta s prosječno 500 – 1.000 slika čiste i pune razlučivosti. To će rezultirati desecima milijuna zabilježanih slika organiziranih semantičkom hijerarhijom WordNeta. Rad Denga i suradnika ponudio je detaljnu strukturu programa ImageNet u njegovom tada izvornom stanju: 12 kategorija i s ukupno 3,2 milijuna slika. U radu je istaknuto da je ImageNet u tom trenutku bio mnogo veći, po razmjerima, ali i mnogo precizniji od tada raspoloživih skupova podataka. U radu je opisana shema prikupljanja podataka pomoću *Amazon Mechanical Turk* sustava [166], te je zaključno ilustrirana korisnost ImageNeta kroz tri jednostavne aplikacije u prepoznavanju objekata, klasifikaciji slika i automatskom grupiranju objekata, što je u konačnici dovelo do novih

moćnosti istraživanja u području računalnog vida i neuronskih mreža. Razvoj i predstavljanje ImageNeta dovelo je i do organizacije natjecanja iz područja računalnog vida koje se redovito održavalo od 2010. do 2017. godine, a samo po sebi natjecanje i ImageNet kao skup podataka predstavljali su mjerilo za testiranje metoda iz područja računalnog vida [167]. Nedugo nakon predstavljanja ImageNet-a, otvorio se cijeli niz novih mogućnosti, a cjelokupna istraživačka zajednica iz domene neuronskih mreža kao da se probudila iz hibernacije te kao da su istraživači iznova otkrili sve (do tada blago zaboravljene) neslućene mogućnosti primjene neuronskih mreža koje su kroz desetak godina bile pomalo zapostavljene. Rezultat iznova probuđenog interesa za neuronskim mrežama je i razvoj novih koncepata. Tako je 2010. godine, predstavljena prva implementacija GPU neuronske mreže koja je funkcionirala unaprijedno i unazad, a implementirana je na NVIDIA GTX 280 grafičkom procesoru i sastojala se od devet slojeva neuronske mreže [168]. Nakon toga je uslijedila prava renesansa za neuronske mreže, posebice nakon što je 2012. godine Alex Krizhevsky predstavio AlexNet, dublju i širu verziju od LeNeta, a primjenom iste pobijedio je na vrlo izazovnom natjecanju iz područja računalnog vida [8]. AlexNet je konvolucijska neuronska mreža koja se sastoji od 60 milijuna parametara i 650.000 neurona, ima pet slojeva, a nakon nekih slijede slojevi maksimalnog sažimanja (engl. *max pool*) i tri potpuno povezana sloja (engl. *fully connected layers*) sa završnim 1.000-kanalnim *softmax* slojem. Za ubrzanje treniranja koristili su i vrlo učinkovitu implementaciju na GPU, a za izbjegavanje pretreniranosti, primijenili su *dropout* metodu regularizacije, što se pokazalo vrlo učinkovitim. Također značajan doprinos je i korištenje ReLU (engl. *Rectified Linear Unit*) aktivacijske funkcije [169]. Uspješna primjena AlexNet konvolucijske neuronske mreže na natjecanju *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) održanom 2012. godine, praktički je pokrenula novu renesansu unutar istraživačke zajednice usmjerene na duboko učenje.

Iste godine je skupina autora [170] izvela eksperiment koji će također ostaviti dalekosežne posljedice na istraživanje i primjenu metoda dubokog učenja, posebice u domeni nenadziranog učenja (engl. *Unsupervised learning*) kod kojega koriste slike bez označenih objekata (za razliku od nadziranog učenja kod kojega se model trenira na prethodno označenim objektima na slikama). U navedenom radu opisali su svoj eksperiment koji je za cilj imao razvoj detektora visoke razine namijenjenog detekciji značajki specifičnih samo za pojedinu klasu, isključivo na temelju neoznačenih podataka odnosno slika (na kojima objekti nisu prethodno označeni). Da bi postigli taj

cilj, autori su trenirali 9-slojni lokalno spojeni autoenkoder s udruživanjem i lokalnom normalizacijom kontrasta na uvećanom skupu slika (model je imao milijardu veza, skup podataka sastojao se od 10 milijuna slika rezolucije 200x200 piksela preuzetih s interneta). Mrežu su trenirali koristeći paralelizam i asinkroni SGD na računalnom klasteru s 1.000 strojeva (ukupno 16.000 jezgara) kroz tri dana. Eksperimentalni rezultati su pokazali da je moguće trenirati detektor lica bez potrebe da se objekti na slikama prethodno označavaju. Kontrolni eksperiment je pokazao da je njihov detektor značajki robustan, ne samo za prevođenje nego i za skaliranje i rotaciju izvan ravnine. Također su otkrili da je mreža osjetljiva i na druge koncepte visoke razine poput lica mačaka i ljudska tijela. Polazeći od ovih naučenih značajki, osposobili su mrežu za postizanje 15,8% točnosti u prepoznavanju 20.000 kategorija objekata iz ImageNeta, što je relativno poboljšanje od 70% u odnosu na prethodne eksperimente na temelju nenadziranog učenja. Uspjeh ovog eksperimenta dodatno je pokrenuo novi trend u istraživanju dubokog učenja, a to je nenadzirano učenje, što je na svojevrsan način bio novi korak u smislu približavanja strojnog učenja procesu prirodnog učenja kao kod biološkog mozga.

Sljedeći značajan korak dogodio se 2014.g., kada je predstavljeno super računalo nazvano DeepFace, a predstavio ga je Facebook, odnosno istraživači okupljeni u Facebookovu timu za umjetnu inteligenciju [171]. Svrha DeepFace-a je detekcija lica na slikama primjenom neuronske mreže. Rezultati su pokazali da DeepFace ima mogućnost detekcije lica sa 97,35 % točnosti, što je poboljšanje od 27% u odnosu na ranije pokušaje. Osim predstavljanja DeepFacea i rezultata koji su postignuti u domeni detekcije lica, iste godine su predstavljene i Generativne suparničke mreže (engl. *GAN – Generative Adversarial Networks*) [172]. Generativne suparničke mreže predstavljaju modele u kojima su suprotstavljene dvije neuronske mreže, generator i diskriminator. Zadatak generatora je modelirati distribuciju ulaznih podataka, a zadatak diskriminatora je odrediti da li podatak potječe iz stvarnog skupa ili je generiran generatorom. Diskriminator minimizira binarnu klasifikacijsku funkciju pogreške (da li je podatak na ulazu stvaran ili generiran), dok generator maksimizira grešku diskriminatora za generirane podatke. Model se uči nenadziranim učenjem, odnosno nisu mu potrebni označeni podaci. Pošto su generator i diskriminator neuronske mreže, uče se algoritmom unazadne propagacije te nisu potrebne vjerojatnosne estimacije podataka kao kod drugih generativnih modela. Za generativne suparničke mreže, ranije spomenuti Yann LeCun, jedan od najutjecajnijih i najznačajnijih

istraživača u području dubokog učenja, ovaj tip neuronskih mreža je opisao kao "*najzanimljiviju ideju u posljednjih 10 godina u strojnom učenju*".

Kao što je već ranije navedeno, novi zamah u istraživanju i razvoju umjetnih neuronskih mreža počeo je 2012. godine nakon predstavljanja AlexNeta, što je potaknulo i ostale istraživače da započnu s razvojem vlastitih projekata u ovom području. Tako je 2013. godine predstavljena ZFNet/Clarifai [173]. ZFNet je koncept koji je 2013. godine osvojio ILSVRC, a temelji se na CNN arhitekturi koja je bila ekstenzija AlexNeta. ZFNet arhitektura je poboljšanje AlexNeta, dizajnirano ugađanjem mrežnih parametara potonjeg. ZFNet koristi 7x7 jezgru (kernel ili filter) umjesto 11x11 s ciljem značajnog smanjenja broja težina. To dramatično smanjuje broj mrežnih parametara i poboljšava ukupnu točnost prepoznavanja. Također treba spomenuti i koncept koji se zove mreža u mreži (engl. *Network in Network*), a predstavljen je 2013. godine. Ovaj se model ponešto razlikuje od prethodnih, s obzirom da je uvedeno nekoliko novih koncepata [174]. Prvi je upotreba višeslojnog perceptrona, gdje se konvolucije izvode pomoću 1x1 filtera koji pomaže u dodatnoj nelinearnosti u modelima. To pomaže povećanju dubine mreže, koja se tada može regulirati primjenom metode slučajnog izostavljanja neurona prilikom učenja (engl. *dropout*). Drugi je koncept korištenje globalnog prosječnog združivanja (engl. *GAP – Global Average Pooling*) kao alternativa potpuno povezanim slojevima. Ovaj pristup pomaže značajno smanjiti broj mrežnih parametara, a primjena GAP-a značajno mijenja i mrežnu strukturu. Također, primjena GAP-a na velikoj mapi značajki uvelike pomaže generiranju manje dimenzionalnog vektora značajki bez redukcije dimenzija mape značajki.

Nadalje, The Visual Geometry Group (VGG) je 2014. godine u sklopu ILSVRC natjecanja predstavila svoj koncept umjetne neuronske mreže namijenjene dubokom učenju [175]. Glavni doprinos se očituje u tome što je pokazano da je dubina mreže ključna komponenta konvolucijskih neuronskih mreža za postizanje boljih rezultata raspoznavanja ili klasifikacije. Arhitektura VGG mreže sastoji se od dva konvolucijska sloja, od kojih oba koriste ReLU aktivacijsku funkciju. Nakon toga slijedi jedan sloj sažimanja maksimalnom vrijednošću (engl. *Max Pooling Layer*) i nekoliko potpuno povezanih slojeva (engl. *Fully Connected Layers*) koji također koriste ReLU funkciju. Zadnji sloj modela je *softmax* sloj za klasifikaciju. U VGG-E veličina konvolucijskog filtera je promijenjena u filter veličine 3x3 sa pomakom 2. Isti autori su također predstavili tri dodatna VGG-E modela (VGG-11, VGG-16 i VGG-19) koji su imali 11, 16 te 19 slojeva.

Pobjednik ILSVRC natjecanja 2014. godine je GoogLeNet [176.], model koji su predstavili Christian Szegedy i skupina autora iz Googlea, a cilj je bio reduciranje računalne kompleksnosti u usporedbi sa standardnim konvolucijskim neuronskim mrežama. Predstavljena metoda sadrži "*Inception* slojeve" koji imaju varijabilna receptivna polja koja su kreirana primjenom kernela različitih veličina. Ova receptivna polja kreiraju operacije koji hvataju raspršene korelacijske uzorke u novu mapu značajki. *Inception* sloj omogućava izvođenje više operacija paralelno, a sastoji se od "mreže unutar sloja mreže", srednje velikih konvolucijskih filtera, jako velikih konvolucijskih filtera i operatora udruživanja. Ovakav pristup omogućava "izvlačenje" jako malih detalja u slici, a pored toga, na svakom konvolucijskom sloju nalazi se i ReLU aktivacijska funkcija koja pomaže poboljšanju nelinearnosti mreže. Nadalje, GoogLeNet je poboljšao preciznost u odnosu, na tada, najpouzdanije i najpreciznije modele primjenkom *Inception* slojeva. Razlika između naivnog *Inception* sloja i zadnjeg *Inception* sloja je u dodatku 1x1 konvolucijskog kernela. Ovi kerneli su omogućili redukcije dimenzionalnosti prije računalno zahtjevnih slojeva. GoogLeNet se sastoji od ukupno 22, što je u svakom slučaju veća mreža od bilo koje ranije predstavljene. S druge strane, broj mrežnih parametara kod ove mreže je značajno manji nego kod ranije predstavljenih modela kao što je AlexNet ili VGG s obzirom da GoogLeNet ima sedam milijuna mrežnih parametara, AlexNet ima 60 milijuna, a VGG-19 ima 138 milijuna.

Sljedeći značajan koncept neuronske mreže je Microsoftova ResNet odnosno Residual Network [177] predstavljena 2015. godine, kada je i pobijedila na spomenutom ILSVRC natjecanju. ResNet je razvijena s ciljem dizajniranja ultra duboke mreže koja neće patiti od problema koji su bili tipični za prethodne koncepte. Mreža je razvijena u više varijanti, ovisno o ukupnom broju slojeva, od 34 pa čak do 1.202, iako je na natjecanju primijenjena mreža sa 152 sloja. Raširena ResNet50 sastoji se od 49 konvolucijskih i jednog potpuno povezanog sloja na kraju mreže. ResNet je prilikom svoje pojave oborila sve rekorde u klasifikaciji, detekciji i lokalizaciji, a pored rekorda u broju slojeva, ResNet ima stopu pogreške od 3,6%, međutim svojevrsna mana je računalna zahtjevnost, te je vrijeme za treniranje ResNet-a duže nego kod ranije predstavljenih AlexNet-a ili VGG, s obzirom da treniranje na osam GPU-a traje oko tjedan dana.

Posljednjih nekoliko godina je praktički prava eksplozija različitih koncepata temeljenih na neuronskim mrežama, pa u tom kontekstu svakako vrijedi spomenuti *DensNet* [178], *FractalNet* [179], *Fast RCNN* [9] i kasnije predstavljeni model *Faster RCNN* [26], *Fully Convolutional Networks – FCN* primarno namijenjena za segmentaciju slika [180], *Deeply-supervised nets* [181], *Ladder network architecture* [182], *Deep networks with stochastic depth* [183], *Xception* [184], *IRCNN* [185], *ViP CNN* [186], *DetectNet* [187], i za kraj ćemo spomenuti *YOLO – You Only Look Once* [13 – 16], koji će biti detaljno obrađen u zasebnom poglavlju s obzirom da je upravo ovaj koncept korišten u eksperimentalnom dijelu disertacije.

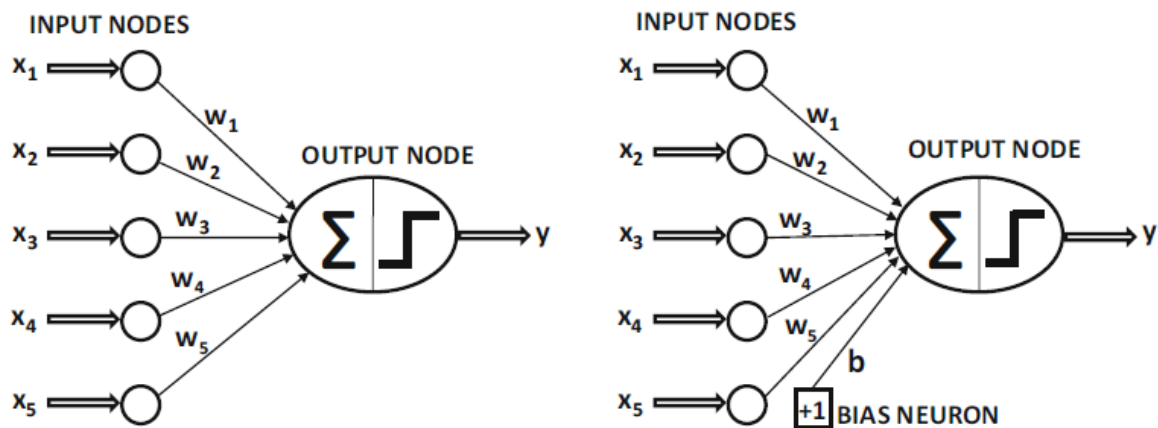
4.2 Elementi neuronskih mreža

Ranije je već navedeno da je osnovni gradivni element umjetne neuronske mreže procesorska jedinica koja se zove neuron po uzoru na biološke neurone u mozgu. Također je spomenuto da različiti načini organizacije neurona tvore različite arhitekture neuronskih mreža te da svaka od tih arhitektura ima svoj značaj u određenom području, pa npr. konvolucijske neuronske mreže daju vrhunske rezultate u području računalnog vida i sl.

Neuron sastoji se od tri osnovna elementa:

1. **Skupa sinapsi ili veza** – od kojih svaku karakterizira njezina vlastita vrijednost ili težina. Preciznije, signal x_j na ulazu sinapse j povezane s neuronom k množi se sa vrijednošću težine sinapse w_{kj} , koja se na početku pokretanja neuronske mreže dodjeljuje nasumično. Kod označene težine w_{kj} , prvi indeks k označava pojedini neuron, dok se drugi indeks j odnosi na ulaz sinapse na koju se odnosi navedena vrijednost težine (slika 14). Za razliku od sinapsi u mozgu, vrijednosti težina na sinapsama umjetnog neurona mogu biti i negativne i pozitivne vrijednosti.
2. **Sumatora** – koji sumira vrijednosti ulaznih signala ponderiranih sa vrijednostima određenih sinapsi neurona, a opisana operacija sadrži linearni kombinator.
3. **Aktivacijske ili prijenosne funkcije** – koja služi za ograničavanje vrijednosti izlaza neurona, te uvođenje nelinearnosti u neuronsku mrežu [188].

Najjednostavnije neuronske mreže su jednoslojne mreže kod kojih je skup ulaza izravno preslikan na izlaz primjenom genralizirane verzije linearne funkcije, a poznata je pod nazivom perceptron [189].



Slika 14. Prikaz osnovne građe perceptrona – lijevo bez prirasta (dodatni prag ili unutarnji pomak neurona), desno – s prirastom [190].

Ovaj tip neuronske mreže sastoji se od jednog ulaznog i jednog izlaznog sloja, a osnovna arhitektura prikazana je na slici 14. Perceptron funkcionira tako da na temelju niza ulaznih varijabli proizvodi jednu izlaznu varijablu $(\bar{X}, y)$, pri čemu svaki $\bar{X} = [x_1, \dots, x_d]$ sadrži d parametara značajki, a izlaz $y \in \{-1, +1\}$ sadrži promatrane vrijednosti varijabli binarne klase, pri čemu je "promatrana varijabla" zadana kao dio skupa za učenje, a cilj je predviđanje klase za sve slučajeve kada nije promatrana, odnosno kada nije dio skupa za učenje. Ulazni sloj sadrži d neurona, a prenosi d značajke $\bar{X} = [x_1, \dots, x_d]$ sa težinama $\bar{W} = [w_1, \dots, w_d]$ prema izlaznom neuronu. Ulazni sloj, sam po sebi, ne izvršava računske operacije, a linearna funkcija $\bar{W} \times \bar{X} = \sum_{i=1}^d w_i x_i$ izvodi se na izlaznom neuronu. Sukladno tomu, predznak ove realne vrijednosti koristi se kako bi se predvidio ovisni parametar $\bar{X}$, pa prema tome izlaz $\hat{y}$ se matematički može prikazati na sljedeći način:

$$\hat{y} = \text{sign}\{\bar{W} \times \bar{X}\} = \text{sign}\{\sum_{j=1}^d w_j x_j\} \quad (13)$$

Signum funkcija preslikava realnu vrijednost u $+1$ ili -1 , što je prikladno za binarnu klasifikaciju. Treba napomenuti da oznaka $(\hat{y})$ iznad parametra y označava da je riječ o predviđenoj, a ne promatranoj varijabli. Greška predviđanja je $E(\bar{X}) = y - \hat{y}$, a riječ je o jednoj od vrijednosti iz zadanog skupa $\{-2, 0, +2\}$. U slučajevima kada je vrijednost greške $E(\bar{X})$ različita od nule, vrijednosti težina je potrebno ažurirati u negativnom smjeru greške gradijenta.

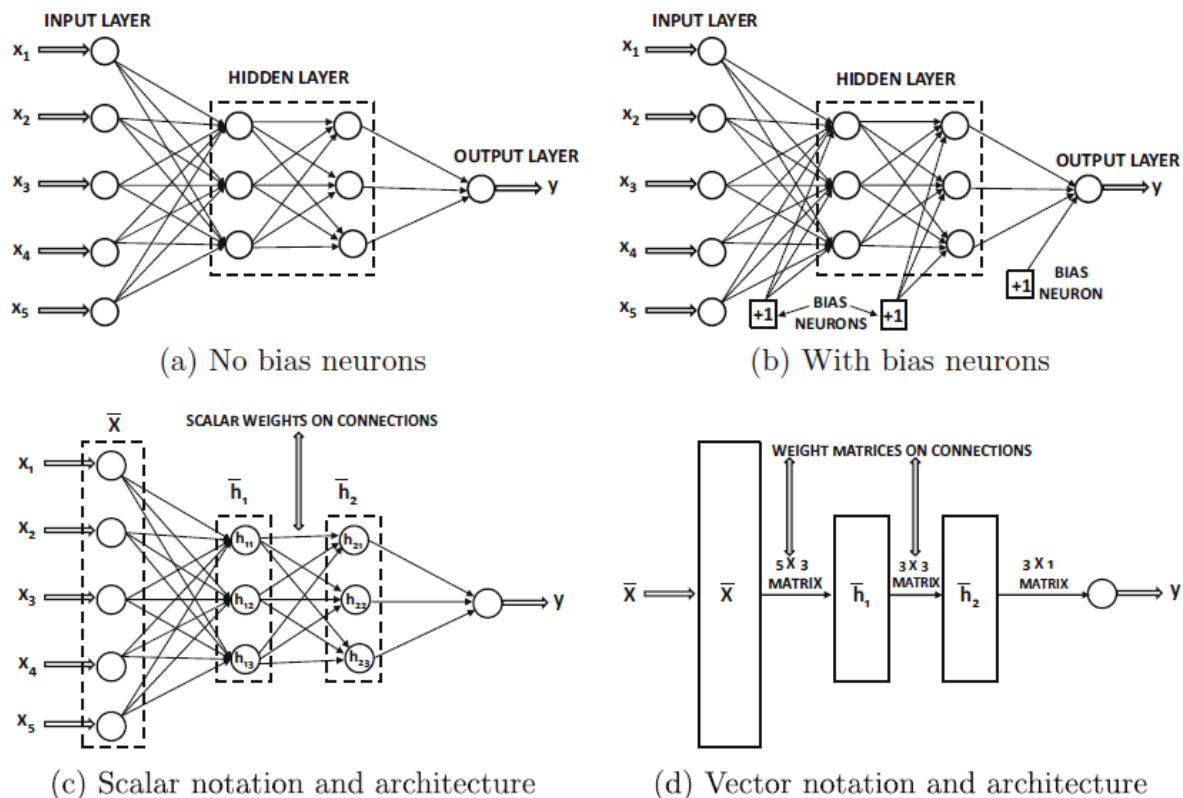
Arhitektura perceptrona prikazana je na slici 14., kod kojeg jedan ulazni sloj samo prenosi značajke prema izlaznom neuronu. Za sumiranje vrijednosti svih ulaza koriste

se i težinske vrijednosti koje definiraju važnost pojedinih ulaznih varijabli. Što je veća težinska vrijednost veći je i značaj ulazne varijable. Posljedično, signum funkcija se primjenjuje kako bi se sumirana vrijednost konvertirala u oznaku klase, dok signum funkcija ima ulogu aktivacijske funkcije. Primjena različitih aktivacijskih funkcija može simulirati različite tipove modela korištenih kod strojnog učenja. Također, većina osnovnih modela strojnog učenja može se prikazati kao jednostavna neuronska mreža. Svakako treba napomenuti da prikazani perceptron ima dva sloja, iako se govori o jednoslojnoj mreži, a to je zato što ulazni sloj ne obavlja računske operacije, nego samo prenosi vrijednosti značajki. Nadalje, u velikom broju slučajeva postoji invarijantni dio predviđanja koji se naziva prirast (na slici 14., desno). Primjerice u situaciji kada su varijable značajki centrirane oko prosjeka, ali je prosjek predviđena binarna klasa u rasponu od $\{-1, +1\}$ različita od 0, što će dovesti do toga da je distribucija binarnih klasa značajno neuravnotežena. U takvom slučaju, ranije opisani pristup nije dovoljan za predviđanje pripadnosti klasi, te je potrebno uvesti dodatnu graničnu vrijednost, prirast, odnosno sklonost perceptrona:

$$\hat{y} = \text{sign}\{\overline{W} \times \overline{X} + b\} = \text{sign}\{\sum_{j=1}^d w_j x_j + b\} \quad (14)$$

Prirast se može uvesti kao granična vrijednost dodavanjem jednog neurona koji prema izlazu uvijek prosljeđuje vrijednost 1, kao što je prikazano na slici 14 [190].

Kod višeslojnih neuronskih mreža, neuroni su organizirani slojevito, a ulazni i izlazni sloj su odvojeni određenim brojem skrivenih slojeva. Ovakva slojevita organizacija zove se unaprijedna neuronska mreža (engl. *FFD - Feed-Forward Network*), slika 15.



Slika 15. Osnovna arhitektura unaprijedne neuronske mreže sa dva skrivena i jednim izlaznim slojem. Iako svaka jedinica sadrži jednu skalarnu varijablu, jedna često predstavlja sve neurone u jednom sloju kao jedan vektor. Vektorske jedinice se često prikazuju kao pravokutnici i spojeni su matricama između njih [190].

Kao što je spomenuto, višeslojne neuronske mreže sadrže više računskih slojeva, te osim ulaznog i izlaznog sloja koji su vidljivi, sadrže i dodatne skrivene slojeve (engl. *hidden layers*) koji nisu vidljivi, odnosno rezultati računskih operacija koje se izvode na skrivenim slojevima nisu vidljivi. Ranije je također spomenuto da se ovaj tip neuronskih mreža naziva unaprijedna neuronska mreža zato što se vrijednosti sukcesivno prenose s jednog sloja na drugi prema naprijed (od početnog sloja prema izlazu). Osnovna arhitektura ove mreže pretpostavlja da su svi neuroni u jednom sloju spojeni sa svim neuronima u sljedećem sloju. Prema tome, arhitektura neuronske mreže je gotovo u potpunosti definirana kada se definira broj slojeva i tip neurona u svakom sloju. Tada još preostaju detalji u svezi funkcije gubitka (engl. *loss function*) koja je optimizirana u izlaznom sloju. Učestala je primjena *softmax* izlaza sa *cross-entropy* gubitkom za diskretna predviđanja i linearni izlazi sa kvadratnim gubitkom za slučajeve predviđanja u realnim vrijednostima. Kao i kod jednoslojnih mreža, neuroni prirasta mogu biti korišteni u skrivenim i izlaznim slojevima. Primjeri višeslojnih mreža sa i bez prirasta (biasa) prikazani su na slici 15. U svakom slučaju, višeslojna neuronska mreža sastoji

se minimalno od tri sloja, iako se ulazni sloj uobičajeno ne uračunava u arhitekturu mreže s obzirom da samo prosljeđuje podatke i ne izvodi računske operacije.

Ukoliko neuronska mreža sadrži $p_1 \dots p_k$ jedinica u svakom od k slojeva tada su u vektorskom (stupac) zapisu ovi izlazi označeni kao $h_1 \dots h_k$ te imaju dimenzionalnost $p_1 \dots p_k$, odnosno broj jedinica (neurona) u svakom sloju odnosi se na dimenzionalnost tog sloja [190].

Prethodno je opisana građa umjetnog neurona i osnovnih računskih operacija koje se izvode o ulaza do izlaza neurona, kao i jednoslojna te višeslojna neuronska mreža. U nastavku će detaljnije biti opisane aktivacijske funkcije s obzirom na njihov kritični značaj i utjecaj na vrijednost izlaznog neurona.

4.3 Aktivacijske funkcije

Aktivacijske funkcije su iznimno važan element neuronskih mreža s obzirom da, između ostalog, "odlučuju" da li će određeni neuron biti aktiviran ili ne. Sam koncept aktivacije akcijskog potencijala (engl. *firing action potential*) preuzet je iz biološkog neurona i ugrađen u umjetni neuron. Na apstraktnoj razini, aktivacija je predstavljena aktivacijskom funkcijom, koja predstavlja čvor koji se dodaje na kraju izlaza svakog sloja neuronske mreže, a poznata je još i kao funkcija prijenosa. S obzirom da svaki neuron u nekom sloju ima istu aktivacijsku funkciju, ona je definirana globalno. Osim korištenja postojećih, moguće je definirati i vlastite aktivacijske funkcije. Značaj aktivacijskih funkcija je u tome što uvode nelinearnost u linearni izlaz neurona. Naime, aktivacijske funkcije obični zbroj produkata ulaznih parametara, dodanih težina i prirasta, pretvaraju u nelinearni rezultat s obzirom da zbrajanje produkata ulaznih parametara i dodanih težina pojedinog neurona uvijek daje linearni rezultat koji omogućuje klasifikaciju isključivo linearno odvojivih podataka, dok aktivacijske funkcije sumirani zbroj produkata ulaznih vrijednosti neurona pretvaraju u nelinearni rezultat koji omogućuje i klasifikaciju linearno neodvojivih podataka. Aktivacijske funkcije mogu biti linearne i nelinearne, a moguća je i primjena jednostavnih aktivacijskih funkcija čija je svrha samo "odlučiti" da li će pojedini neuron biti aktiviran ili ne [191].

Prema tome, svrha aktivacijskih funkcija u neuronskim mrežama je izračunavanje ponderirane (vagane) sume ulaza i biasa (prirasta). Aktivacijska funkcija manipulira predstavljenim podacima pomoću neke gradijentske obrade, obično gradijentnog spusta, a potom izračunava vrijednost izlaza za neuronsku mrežu. Da li će aktivacijska funkcija biti linearna ili nelinearna, ovisi o području odnosno zadatku i tipu podataka

koji se obrađuju kroz umjetnu neuronsku mrežu, a svakako treba istaknuti da korištenje pojedine aktivacijske funkcije izravno utječe na rezultate neuronske mreže (u pozitivnom ili negativnom smislu). [192]

Kod linearnih modela, linearno preslikavanje ulazne funkcije na izlaz, izvodi se u skrivenim slojevima prije konačnog predviđanja pripadnosti klasi, što u većini slučajeva daje afina transformacija. Transformacija ulaznog vektora x određena je sa

$$f(x) = w^T x + b \quad (15)$$

gdje je:

- x = ulaz,
- w = težine,
- b = bias (sklonost ili pristranost).

Nadalje, neuronske mreže linearne rezultate preslikavanja računaju iz jednadžbe (15) zbog čega se javlja potreba za aktivacijskom funkcijom, prvo za pretvaranje ovih linearnih izlaza u nelinearne za daljnje računanje, posebice za učenje uzoraka u skupu podataka. Ovi izlazi definirani su kao:

$$y = (w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b) \quad (16)$$

Ovi izlazi svakog sloja prenose se na sljedeći sloj višeslojne neuronske mreže kao što je duboka neuronska mreža, sve dok se ne postigne konačna vrijednost izlaza, ali unatoč tomu, po definiciji su linearni. Očekivana vrijednost izlaza određuje koja će aktivacijska funkcija biti korištena u neuronskoj mreži. Nadalje, s obzirom da je vrijednost izlaza, po prirodi linearna, potrebna je nelinearna aktivacijska funkcija kako bi se ove linearne vrijednosti izlaza pretvorile u nelinearne. Ove aktivacijske funkcije su prijenosne koje se primjenjuju na izlazne vrijednosti linearnih modela kako bi se transformirale u nelinearnu vrijednost izlaza koja je spremna da daljnju obradu. Nelinearna vrijednost izlaza nakon primjene aktivacijske funkcije definirana je kao:

$$y = \alpha(w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b) \quad (17)$$

gdje je α aktivacijska funkcija.

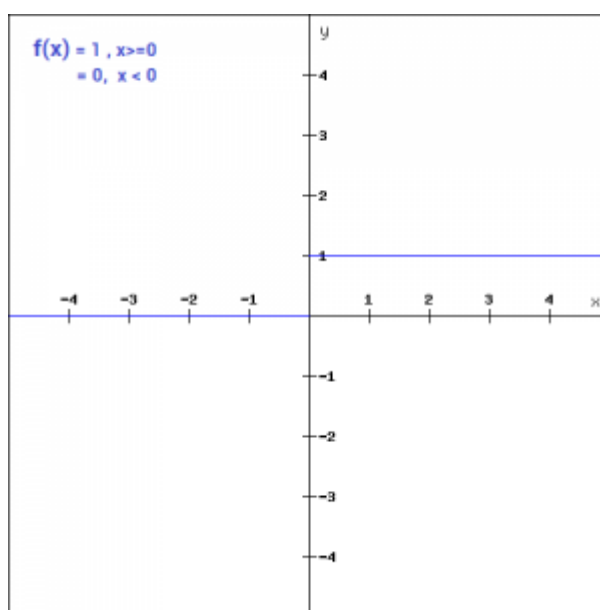
Potreba za aktivacijskim funkcijama proizlazi iz potrebe za pretvaranjem linearnih ulaznih vrijednosti (signala) u nelinearne izlazne vrijednosti što pomaže učenju polinoma višeg reda, jer u suprotnom slučaju je riječ o učenju polinoma prvog reda. Posebno svojstvo nelinearnih aktivacijskih funkcija je u tome što su derivabilne, jer inače ne bi radile tijekom povratne propagacije kod neuronskih mreža. [192]

Jednostavnije rečeno, kada se ne bi koristile aktivacijske funkcije, tada bi izlazni signal bio jednostavna linearna funkcija, koja je samo polinom prvog reda. U ovom slučaju, linearnu jednadžbu je lako riješiti, ali one su ograničene u svojoj složenosti i imaju manje snage za učenje složenih preslikavanja funkcija iz zadanih podataka. Neuronska mreža bez aktivacijske funkcije jednostavno je model linearne regresije, koji ima ograničenu snagu i većinu vremena ne postiže dobre rezultate, a krajnji cilj je da neuronska mreža ne samo nauči i izračuna linearnu funkciju, već i nešto složenije od toga. Također, bez aktivacijske funkcije neuronska mreža ne bi mogla učiti i modelirati druge kompleksne vrste podataka kao što su slike, videozapisi, audio, govor itd. Zato se koriste tehnike umjetne neuronske mreže kao što je duboko učenje da bi se dobio smisao kod kompleksnih uzoraka ili nelinearnih skupova podataka velikih dimenzija, gdje model ima veliki broj skrivenih slojeva te ima vrlo kompleksnu arhitekturu koja pomaže da se shvate i izvuku znanja iz tako složenih velikih skupova podataka.

Zbog navedenog, potrebne su nelinearne funkcije koje imaju viši red i imaju zakrivljenost kada se nelinearna funkcija nacрта. S obzirom da se neuronske mreže smatraju univerzalnim aproksimatorima funkcija, to znači da praktički mogu izračunati i naučiti bilo koju funkciju, te se posljedično gotovo svaki proces, koji se može zamisliti, može predstaviti kao računanje funkcija u neuronskoj mreži. Sukladno tomu, sve se svodi na to da je potrebno primijeniti aktivacijsku funkciju $f(x)$ kako bi mreža bila snažnija i dobila sposobnost da nauči složenije i kompleksnije uzorke iz skupa podataka te prikaže složeno proizvoljno nelinearno preslikavanje funkcija ulaza i izlaza, odnosno koristeći nelinearnu aktivacijsku funkciju moguće je generirati nelinearna preslikavanja od ulaza do izlaza. Pored toga, važno obilježje aktivacijske funkcije je to što bi trebala biti derivabilna. Ovo je potrebno kako bi bilo moguće izvesti strategiju optimizacije povratne propagacije za izračunavanje gradijenta pogreške (gubitka) u odnosu na težine, a zatim u skladu s tim optimizirati težine koristeći gradijentni spust ili bilo koju drugu tehniku optimizacije kako bi se smanjila pogreška. Prema tipu, aktivacijske funkcije je moguće podijeliti na funkcije s pragom osjetljivosti (engl. *Step functions*), linearne i nelinearne funkcije koje mogu poprimiti različite oblike, ali se najčešće koriste sigmoidne, hiperbolične i harmoničke funkcije. Također je važno napomenuti da je poželjno da su aktivacijske funkcije derivabilne, zbog toga što je, prilikom ažuriranja krivulje funkcije, potrebno znati u kojem smjeru i za koju vrijednost treba promjena na nagibu krivulje.

4.3.1 Funkcija s pragom osjetljivosti (engl. *Step funkcija*)

Binarna funkcija s pragom osjetljivosti je, kao što joj i samo ime kaže, temeljena na određenoj vrijednosti praga (slika 16). Stoga, ukoliko je ulazna vrijednost iznad ili ispod postavljenog praga, neuron se aktivira te identičan signal prosljeđuje prema sljedećem sloju. Nedostatak ovog tipa funkcije je što ne podržava klasificiranje ulaznih vrijednosti u jednu od više kategorija, iako može biti iskorištena u razvoju binarnog klasifikatora, s obzirom da su izlazne vrijednosti iznad ili ispod zadanog praga 0 ili 1. S obzirom na to, ovaj tip funkcije je odličan izbor kada je potrebno odgovoriti na pitanje da li nešto pripada nekoj klasi ili ne, s obzirom da će ili neće aktivirati određeni neuron unutar sloja neuronske mreže. Nadalje, gradijent ove funkcije jednak je nuli, te ujedno sve vrijednosti reducira prema nuli, pa praktički ne dolazi do poboljšanja modela. [194]



Slika 16. Funkcija s pragom osjetljivosti (engl. *step function*) [195]

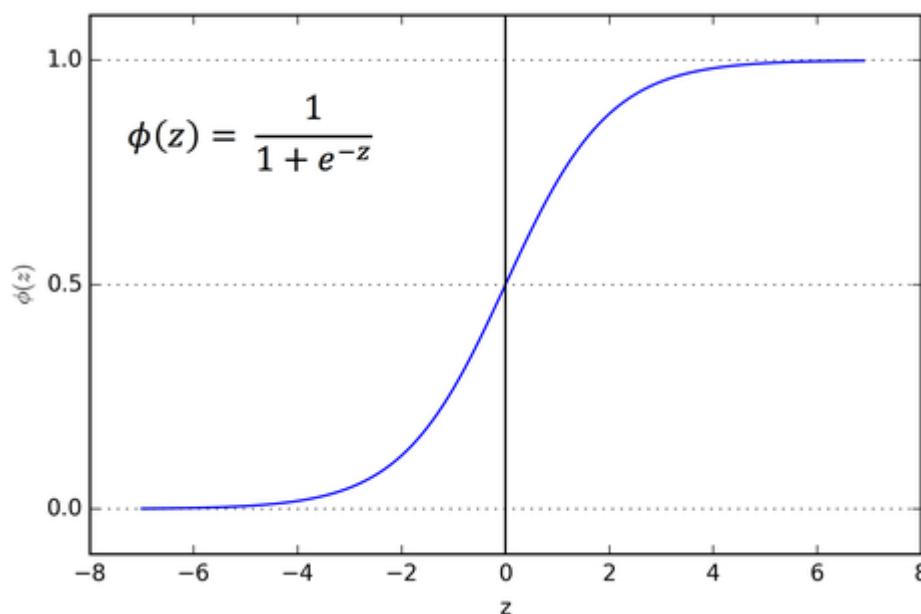
4.3.2 Linearna i nelinearna funkcija

Linearna funkcija također može biti korištena kao aktivacijska funkcija, međutim problem kod linearne funkcije je to što će rezultat uvijek biti linearna transformacija ulaznih podataka, neovisno o dubini neuronske mreže, tako da će rezultat na izlazu neuronske mreže uvijek biti linearna transformacija ulaza. Kao što je i ranije navedeno, neuronska mreža s linearnom aktivacijskom funkcijom je samo model linearne regresije koji ima ograničenu snagu i mogućnost da se nosi s kompleksnošću različitih parametara iz ulaznih podataka, iako je njihova primjena moguća na ulaznom i izlaznom sloju, dok se na skrivenim slojevima uglavnom primjenjuju nelinearne funkcije. [194]

U suvremenim neuronskim mrežama primjenjuju se nelinearne aktivacijske funkcije, s obzirom da omogućavaju da model kreira kompleksna preslikavanja između ulaza i izlaza neuronske mreže, što je ključno za učenje i modeliranje kompleksnih podataka kao što su slike, video zapisi, audio zapisi te nelinearni skupovi podataka ili skupovi podataka velikih dimenzija. Gotovo svaki process koji je moguće zamisliti, moguće je prikazati kao izračun funkcija unutar neuronske mreže, naravno ukoliko se primjenjuje nelinearna funkcija. Za razliku od linearnih, nelinearne funkcije omogućavaju povratnu propagaciju s obzirom da su derivabilne, te omogućavaju slaganje većeg broja slojeva neurona kako bi bilo moguće učenje kompleksnih skupova podataka sa visokim stupnjem točnosti. Vezano za nelinearne funkcije, pojedinačno gledano najraširenije su: sigmoidna ili logistička, tanh ili tangens hiperbolna funkcija i ReLU – Rectified Lineary Units, ali i razne izvedenice navedenih aktivacijskih funkcija. [192]

4.3.3 Sigmoidna funkcija

Sigmoidna funkcija je jedna od najraširenijih aktivacijskih funkcija u umjetnim neuronskim mrežama. Ovo je glatka funkcija i kontinuirano je derivabilna, a najveća prednost joj je to što je nelinearna. To u osnovi znači da je, kada više neurona imaju sigmoidnu funkciju kao aktivacijsku funkciju, izlaz također nelinearan. Funkcija se kreće od 0 do 1 i ima oblik slova S (slika 17).



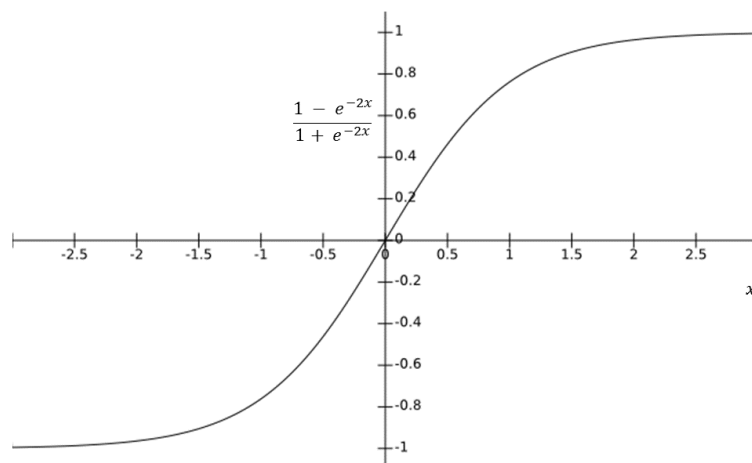
Slika 17. Sigmoidna funkcija [193]

Kod sigmoidne funkcije također je značajno to što male promjene na apscisi izazivaju velike promjene na ordinati, što u konačnici znači da funkcija pokušava vrijednosti na ordinati odvesti prema jednom ili drugom kraju krivulje, što je ujedno i dodatna kvaliteta ove funkcije kada je riječ o klasifikaciji odnosno razvrstavanju u klase. Sigmoidna funkcija obično se, kod arhitektura dubokog učenja, primjenjuje u izlaznim slojevima koji se koriste za predviđanje vjerojatnosti temeljeno na vrijednosti izlaza, te se mogu uspješno primijeniti kod binarne klasifikacije, modeliranja logističke regresije, ali i u drugim područjima primjene neuronskih mreža [194]. Nadalje, Neal je [196] istaknuo glavnu značajku to što je relativno jednostavna i razumljiva, te se koristi u neuronskim mrežama s manjim brojem slojeva, odnosno "plićim" neuronskim mrežama. S druge strane, u [197] je navedeno da bi ovu funkciju trebalo izbjegavati kada se neuronska mreža pokreće korištenjem malih nasumičnih vrijednosti. Osim toga, sigmoidna funkcija ima i drugih nedostataka, a jedan od najizraženijih je *problem iščezavajućeg gradijenta* [198]. Problem iščezavajućeg gradijenta (engl. *vanishing gradient problem*)

je fenomen koji se javlja pri učenju neuronskih mreža čije se učenje temelji na gradijentnim metodama i algoritmu propagiranja unatrag (engl. *backpropagation algorithm*). Ažurirajući težinske koeficijente neuronske mreže, promjene su proporcionalne parcijalnoj derivaciji funkcije greške u odnosu na trenutnu težinu u svakoj iteraciji učenja. U nekim slučajevima se dogodi da je gradijent iščezavajuće mali, pri čemu efektivno onemogućava promjenu težina. Zbog jako malih vrijednosti (blizu nule), učenje se značajno usporava, a u najgorem slučaju se potpuno zaustavlja [199]. Postoji više podtipova sigmoidne funkcije, a detaljnije su pojašnjenje u [192].

4.3.4 Funkcija hiperbolnog tangensa ili Tanh funkcija

Funkcija hiperbolnog tangensa (slika 18) je još jedan tip često korištene aktivacijske funkcije, a često se označava kao *tanh*. Ova funkcija je slična sigmoidnoj funkciji s tim da je zaglađenija, centrirana oko ishodišta, te skalirana u smislu da joj raspon leži od -1 do 1. Ovaj tip aktivacijske funkcije postao je poželjniji za primjenu s obzirom da postiže bolje rezultate učenja za višeslojne neuronske mreže. S druge strane, niti ova funkcija nije rješavala problem iščezavajućeg gradijenta, međutim glavna prednost je to što se rezultat centrira oko ishodišta što pomaže procesu povratnog širenja. Još jedna značajka ove funkcije je to što vrijednost nagiba može biti 1 samo kada je vrijednost ulaza 0, odnosno kada je $x = 0$. Ovo svojstvo dovodi do pojave neaktivnih neurona tijekom računskog procesa. Ograničenja ove funkcije dovela su do daljnjih istraživanja što je rezultiralo razvojem ReLU funkcije (engl. Rectified Linear Unit). Funkcija hiperbolnog tangensa najčešće se koristi za obradu i prepoznavanje prirodnog govora, kao i za binarnu klasifikaciju. Pored standardne *tanh*, postoji i *hardtanh* aktivacijska funkcija koja se također koristi za obradu i prepoznavanje govora [192].

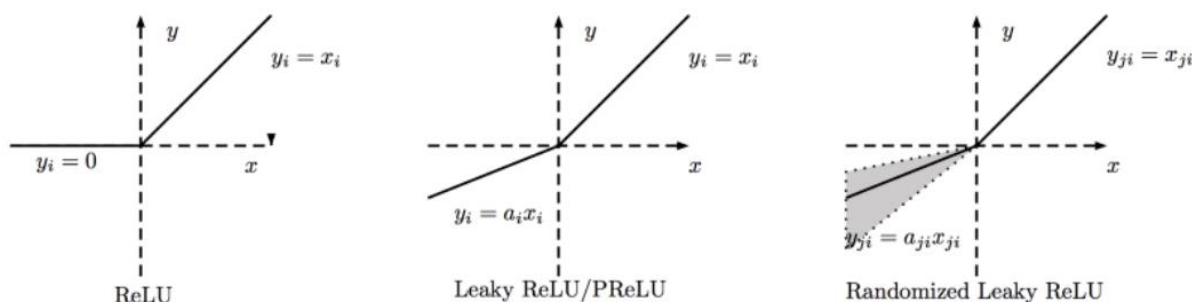


Slika 18. Hiperbolna tangens – Tanh funkcija [200]

4.3.5 ReLu funkcija

ReLU (engl. *Rectified Linear Unit*) funkcija je jedna od nakorištenijih aktivacijskih funkcija u novije vrijeme (slika 19), s obzirom da je korištena u svim konvolucijskim neuronskim mrežama i općenito u dubokom učenju, a prvi put je opisana u [169]. Raspon ove funkcije ide od 0 do beskonačno, uključujući 0. ReLU funkcija je do pola ravna, a zatim raste linearno. Moguće ju je promatrati kao kombinaciju funkcije s pragom osjetljivosti i linearne. Problem ove funkcije je što loše radi sa negativnim ulazima, pa postoje neke poboljšane verzije kao što su propusna ReLU (engl. *LReLU - leaky ReLU*) ili randomizirani ReLU (engl. *RReLU – Randomized ReLU*). Ovaj tip funkcije uči brže i pruža bolju generalizaciju u odnosu na sigmoidnu ili tanh funkcije. S obzirom da ReLu funkcija jako podsjeća na linearnu funkciju, zadržava svojstva linearnog modela što ga čini jednostavnim za optimizaciju metodama gradijentnog spusta. Ova funkcija ispravlja vrijednosti ulaza manjih od nule tako da ih usmjerava prema nuli čime eliminira problem iščezavajućeg gradijenta koji je postojao kod prethodno opisanih aktivacijskih funkcija. ReLU funkcija se koristi unutar skrivenih slojeva zajedno s drugim aktivacijskim funkcijama koje se koriste na izlaznim slojevima neuronske mreže, a tipičan primjer primjene je kod klasifikacije objekata i prepoznavanja govora. Glavna prednost primjene ove funkcije je da osigurava brže izvođenje računskih operacija s obzirom da ne računa s eksponencijama i dijeljenjem, te u konačnici povećava ukupnu brzinu. Još jedan nedostatak ove funkcije to što za vrijednosti manje od nule vraća nulu, zbog čega dolazi do pojave "mrtvih" neurona. Zbog toga je moguće da tijekom procesa ažuriranja težina one budu ažurirane tako da su za pojedine neurone ulazi vrijednosti uvijek manje od nule, što u konačnici znači da

su skirvene vrijednosti ovih neurona uvijek jednake nuli te neće doprinijeti procesu učenja. Međutim, rješenje ovog problema bilo je uvođenje spomenute propusne ReLU funkcije (LReLU). Ovaj tip funkcije uči brže i pruža bolju generalizaciju u odnosu na sigmoidnu i tanh funkciju. LReLU funkcija, za razliku od ReLU, za vrijednosti ispod nule dodaje malu linearnu vrijednost, što blago mijenja nagib krivulje ispod nule, čime se izbjegava pojava nul-gradijenta, što u konačnici rješava problem "mrtvih", odnosno neaktivnih neurona [192].

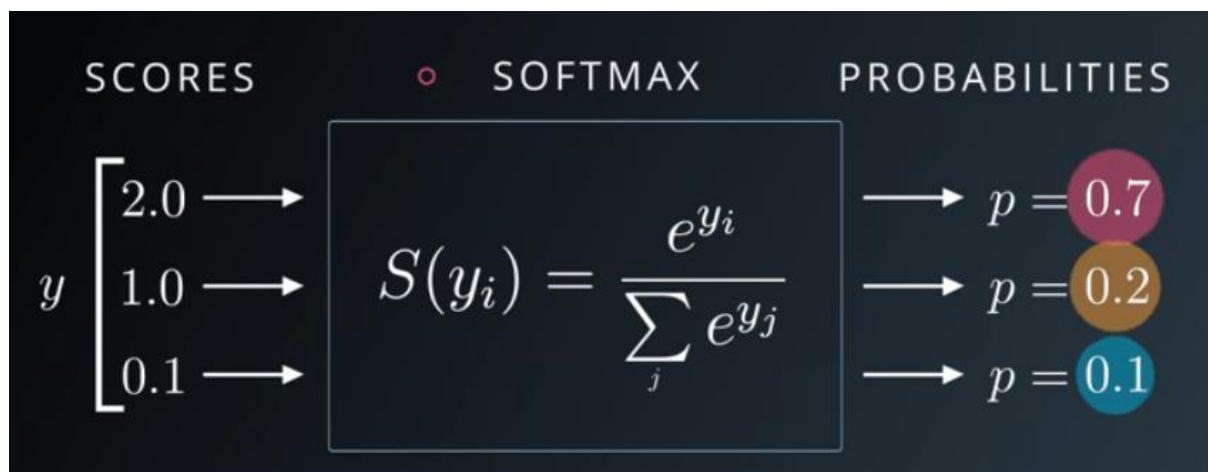


Slika 19. Usporedba ReLU, LReLU i RLReLU aktivacijskih funkcija [201]

4.3.6 Softmax funkcija

Softmax funkcija izračunava raspodjelu vjerojatnosti događaja preko n različitih događaja. Drugim riječima, ovom funkcijom izračunava se vjerojatnost ciljne klase nad svim ostalim mogućim ciljnim klasama. Glavna značajka ove funkcije je raspon vjerojatnosti izlaza, sve vrijednosti izlaza bit će između 0 i 1, a njihov ukupan zbroj treba biti 1. Ukoliko se funkcija koristi na više klasifikacijskom modelu, funkcija vraća vjerojatnost svake klase, a ciljna klasa bi trebala imati najveću vjerojatnost. Također, ova funkcija se uglavnom pojavljuje u gotovo svim izlaznim slojevima neuronskih mreža za duboko učenje. Nadalje, kada je riječ o klasifikaciji u više klasa, te klase su međusobno isključive, što znači da svaki vektor značajki x pripada samo jednoj klasi. Ukoliko vektor značajki predstavlja sliku osobe, nije moguće dodijeliti 50% vjerojatnosti pripadnosti klasi "osoba", a 50% klasi "životinja", nego ovaj vektor značajki mora predstavljati klasu "osoba" s vjerojatnošću od 100%. Nadalje, u slučaju međusobno isključivih klasa, zbroj vjerojatnosti pripadnosti pojedinim klasama mora biti 1, jer samo na ovaj način neuronska mreža može prikazati pravilnu distribuciju vjerojatnosti. Suprotan primjer je kada neuronska mreža sliku osobe predstavi s 80 % vjerojatnosti da pripada klasi "osoba", a sa 60 % vjerojatnosti da pripada klasi "životinja". Na sreću,

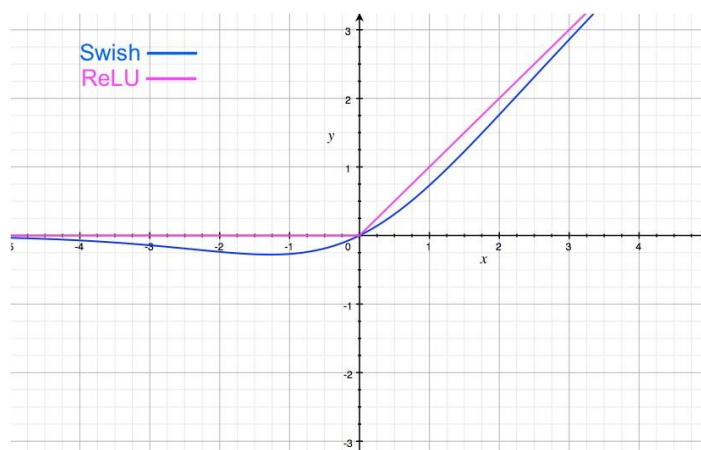
softmax funkcija, ne samo da forsira izlaz u raspon između 0 i 1, nego osigurava i to da će zbroj svih vjerojatnosti na izlazima biti jednak 1 [202].



Slika 20. Prikaz računanja Softmax funkcije [203]

4.3.7 SWISH aktivacijska funkcija

Relativno nova aktivacijska funkcija je SWISH (slika 21) [204 – 206].



Slika 21. Usporedba ReLU i SWISH aktivacijskih funkcija [207]

Riječ je o samo-usmjerenoj funkciji koja je razvijena u Googleu, a matematički, SWISH funkcija je predstavljena kao

$$f(x) = x \cdot \text{sigmoid}(x) \quad (18)$$

Prema autorima [204] ova aktivacijska funkcija postiže bolje rezultate nego ReLU sa sličnom razinom računalne učinkovitosti. Na eksperimentima koji su provedeni na ImageNet skupu podataka za identične modele za ReLU i SWISH, potonja funkcija je ostvarila maksimalnu učinkovitost bolju za 0,9% kod Mobile NASNet-A [208] i 0,6% za Inception-ResNet-v2 [209]. Autori su također istaknuli da jednostavnost SWISH funkcije kao i njezina sličnost sa ReLU funkcijom omogućava njezinu jednostavnu

primjenu u smislu da se ReLU funkcija zamijeni SWISH funkcijom u bilo kojoj neuronskoj mreži.

Osim navednih, postoji još cijeli niz aktivacijskih funkcija koje je moguće primijeniti u neuronskim mrežama, a dio ih je prikazan na slici 22.

Name	Plot	Equation	Derivative (with respect to x)	Range
Identity		$f(x) = x$	$f'(x) = 1$	$(-\infty, \infty)$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$	$\{0, 1\}$
Logistic (a.k.a. Sigmoid or Soft step)		$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$ ^[1]	$f'(x) = f(x)(1 - f(x))$	$(0, 1)$
Tanh		$f(x) = \tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$	$f'(x) = 1 - f(x)^2$	$(-1, 1)$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2 + 1}$	$(-\frac{\pi}{2}, \frac{\pi}{2})$
ElliotSig ^{[9][10][11]} Softsign ^{[12][13]}		$f(x) = \frac{x}{1 + x }$	$f'(x) = \frac{1}{(1 + x)^2}$	$(-1, 1)$
Inverse square root unit (ISRU) ^[14]		$f(x) = \frac{x}{\sqrt{1 + \alpha x^2}}$	$f'(x) = \left(\frac{1}{\sqrt{1 + \alpha x^2}} \right)^3$	$(-\frac{1}{\sqrt{\alpha}}, \frac{1}{\sqrt{\alpha}})$
Inverse square root linear unit (ISRLU) ^[14]		$f(x) = \begin{cases} \frac{x}{\sqrt{1 + \alpha x^2}} & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} \left(\frac{1}{\sqrt{1 + \alpha x^2}} \right)^3 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$(-\frac{1}{\sqrt{\alpha}}, \infty)$
Square Nonlinearity (SQNL) ^[11]		$f(x) = \begin{cases} 1 & : x > 2.0 \\ x - \frac{x^2}{4} & : 0 \leq x \leq 2.0 \\ x + \frac{x^2}{4} & : -2.0 \leq x < 0 \\ -1 & : x < -2.0 \end{cases}$	$f'(x) = 1 \mp \frac{x}{2}$	$(-1, 1)$
Rectified linear unit (ReLU) ^[15]		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$[0, \infty)$
Bipolar rectified linear unit (BReLU) ^[16]		$f(x_i) = \begin{cases} ReLU(x_i) & \text{if } i \bmod 2 = 0 \\ -ReLU(-x_i) & \text{if } i \bmod 2 \neq 0 \end{cases}$	$f'(x_i) = \begin{cases} ReLU'(x_i) & \text{if } i \bmod 2 = 0 \\ -ReLU'(-x_i) & \text{if } i \bmod 2 \neq 0 \end{cases}$	$(-\infty, \infty)$
Leaky rectified linear unit (Leaky ReLU) ^[17]		$f(x) = \begin{cases} 0.01x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0.01 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$(-\infty, \infty)$
Parametric rectified linear unit (PReLU) ^[18]		$f(\alpha, x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(\alpha, x) = \begin{cases} \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$(-\infty, \infty)$ ^[2]
Randomized leaky rectified linear unit (RReLU) ^[19]		$f(\alpha, x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$ ^[3]	$f'(\alpha, x) = \begin{cases} \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$(-\infty, \infty)$
Exponential linear unit (ELU) ^[20]		$f(\alpha, x) = \begin{cases} \alpha(e^x - 1) & \text{for } x \leq 0 \\ x & \text{for } x > 0 \end{cases}$	$f'(\alpha, x) = \begin{cases} f(\alpha, x) + \alpha & \text{for } x \leq 0 \\ 1 & \text{for } x > 0 \end{cases}$	$(-\alpha, \infty)$
SoftPlus ^[24]		$f(x) = \ln(1 + e^x)$	$f'(x) = \frac{1}{1 + e^{-x}}$	$(0, \infty)$
Bent Identity		$f(x) = \frac{\sqrt{x^2 + 1} - 1}{2} + x$	$f'(x) = \frac{x}{2\sqrt{x^2 + 1}} + 1$	$(-\infty, \infty)$
Sigmoid Linear Unit (SiLU) ^[25] (AKA SiLU ^[26] and Swish-1 ^[27])		$f(x) = x \cdot \sigma(x)$ ^[5]	$f'(x) = f(x) + \sigma(x)(1 - f(x))$ ^[6]	$[\approx -0.28, \infty)$
SoftExponential ^[28]		$f(\alpha, x) = \begin{cases} -\frac{\ln(1 - \alpha(\pi + \alpha))}{\alpha} & \text{for } \alpha < 0 \\ x & \text{for } \alpha = 0 \\ \frac{e^{\alpha x} - 1}{\alpha} + \alpha & \text{for } \alpha > 0 \end{cases}$	$f'(\alpha, x) = \begin{cases} \frac{1}{1 - \alpha(\pi + \alpha)} & \text{for } \alpha < 0 \\ e^{\alpha x} & \text{for } \alpha \geq 0 \end{cases}$	$(-\infty, \infty)$
Soft Clipping ^[29]		$f(\alpha, x) = \frac{1}{\alpha} \log \frac{1 + e^{\alpha x}}{1 + e^{\alpha(x-1)}}$	$f'(\alpha, x) = \frac{1}{2} \sinh\left(\frac{p}{2}\right) \text{sech}\left(\frac{px}{2}\right) \text{sech}\left(\frac{p}{2}(1 - x)\right)$	$(0, 1)$
SinusoId ^[30]		$f(x) = \sin(x)$	$f'(x) = \cos(x)$	$[-1, 1]$
Sinc		$f(x) = \begin{cases} 1 & \text{for } x = 0 \\ \frac{\sin(x)}{x} & \text{for } x \neq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x = 0 \\ \frac{\cos(x)}{x} - \frac{\sin(x)}{x^2} & \text{for } x \neq 0 \end{cases}$	$[\approx -0.217234, 1]$
Gaussian		$f(x) = e^{-x^2}$	$f'(x) = -2xe^{-x^2}$	$(0, 1]$

Slika 22. Pregled aktivacijskih funkcija ^[210]

Koju aktivacijsku primijeniti, odnosno koja aktivacijska funkcija je najbolja, ovisi o samom zadatku koji neuronska mreža rješava. Na primjer, ukoliko neuronska mreža treba predvidjeti vrijednosti veće od 1, tada je dobro koristiti sigmoidnu ili tanh aktivacijsku funkciju, međutim ne u izlaznom sloju, u kojem je polježno primijeniti ReLU. S druge strane, ukoliko su očekivane vrijednosti u rasponu $[0,1[$ ili $[-1,1]$ tada ReLU nije dobro rješenje za izlazni sloj, te je poželjna primjena sigmodine ili tanh funkcije. Za klasifikaciju i razvrstavanje prema međusobno isključivim klasama, tada je u izlaznom sloju, preporučljiva primjena softmax funkcije i sl.

4.4 Učenje umjetnih neuronskih mreža

Prema definiciji *"Učenje je proces kojim se slobodni parametri neuronske mreže adaptiraju kroz kontinuirani proces stimulacije od okoline u kojoj se mreža nalazi, a vrsta učenja je određena načinom na koji se parametri mijenjaju"* [211, 188]. Iz navedene definicije proizlazi da okolina stimulira neuronsku mrežu zbog čega dolazi do promjene u neuronskoj mreži, što je rezultat stimulacije te zbog nastalih promjena neuronska mreža prema okolini reagira na novi način. [188]

Ranije je već spomenuto da umjetne neuronske mreže imaju sposobnost učenja koje oponaša učenje u ljudskom mozgu, iako se proces uvelike razlikuje, polazište je jednako u smislu postojanja određenih ulaznih podatka temeljem kojih se kroz proces učenja stvara model temeljem kojeg moguće prepoznavanje uzoraka ili rezultata u budućnosti. Sličnost učenja između umjetnih neuronskih mreža i ljudskog mozga je i u tome što je naučeni model moguće koristiti za rješavanje budućih zadataka, ali isto tako moguće ga je dopuniti novim vrijednostima kako bi bio prilagodljiviji i primjenjiviji za rješavanje budućih zadataka na što je moguće širem spektru zadataka. Naravno, za razliku od ljudskog mozga, koji je savršen generalizator te ima mogućnost rješavanja izuzetno širokog spektra zadataka, također ima i mogućnost automatskog ažuriranja naučenih modela i to čak i temeljem minimalnog broja ulaznih varijabli. Najbolji primjer učenja ljudskog mozga su djeca u ranim fazama odrastanja kada tek počinju biti svjesna svijeta oko sebe. Ono što odraslim osobama predstavlja svakodnevne uobičajene predmete ili značajke pojedinih predmeta, aktivnosti i sl., (ulazne vrijednosti) koje mozak odrasle osobe prepoznaje potpuno automatski, neovisno o broju varijabli koje definiraju taj predmet (oblik, boja, volumen, da li je riječ o slici predmeta ili je pak riječ o stvarnom predmetu, miris...), djeca moraju proći proces učenja jedne po jedne varijable za svaki predmet koji ih okružuje. Proces učenja je

zahtjevan, iterativan je i zahtijeva kontinuirano ažuriranje ulaznih vrijednosti, a samim tim i konačnog modela. Ono što je kod ljudskog mozga specifično je to što ima mogućnost samostalnog učenja na temelju iskustva, kao i učenja s učiteljem. U slučaju samostalnog učenja veća je vjerojatnost pojave greške pri zaključivanju, dok je proces učenja s učiteljem znatno jednostavniji jer je učitelj taj koji prenosi znanja o svakoj pojedinoj varijabli bilo kojeg promatranog predmeta. Slično je kod procesa učenja umjetnih neuronskih mreža, s tim da je jedna od temeljnih razlika to što je svaka umjetna neuronska mreža načelno specijalizirana samo rješavanje jedne vrste zadataka ili grupe sličnih zadataka te je jako teško postići da npr. neuronska mreža razvijena za identifikaciju osoba temeljem lica, s jednakom točnošću i pouzdanošću može prepoznavati osobe prema govoru, iako postoje primjeri vrlo uspješnog prijenosa znanja (engl. transfer learning) između pojedinih vrsta zadataka. [212, 213]

Iako je takvu neuronsku mrežu moguće prilagoditi i za rješavanje drugih zadataka, osim onih za koje je primarno razvijena, taj proces je zahtjevan te i nakon prilagodbe zahtijeva učenje na novom skupu podataka, a istovremeno s tim slijedi iterativno optimiziranje parametara učenja u cilju minimizacije pogreške. S obzirom da je umjetna neuronska mreža računalni program koji ima sposobnost učenja iz podataka definicija navodi da *"računalni program može učiti iz iskustva E (engl. Experience) s obzirom na klase zadataka T (engl. Task) i mjere učinkovitosti P , (engl. Performance measures) ukoliko se njegov učinak na zadacima T , mjeren prema P , poboljšava prema iskustvu E ".* [214]

U tom kontekstu zadatak, odnosno izvršavanje zadatka T je krajnji cilj učenja neuronske mreže, dok se sam proces učenja odnosi na razvoj mogućnosti da se izvrši predviđeni zadatak. Primjerice, ukoliko je zadatak naučiti robota hodanju, to je moguće izvesti na način da se napiše izravan program za hodanje robota ili je moguće robota naučiti hodanju na temelju postojećih primjera. Sukladno tomu, proces učenja neuronskih mreža provodi se temeljem primjera. Primjer, u tom smislu je skup značajki nekog objekta ili događaja koje su kvantitativno opisane, a za koji želimo da ga neuronska mreža nauči procesirati. Primjer se obično prikazuje kao vektor $x \in R^n$ gdje svaki ulaz vektora x_i predstavlja drugu značajku. Primjerice, kod slika, značajke su vrijednosti piksela. Neuronske mreže mogu rješavati jako širok raspon različitih tipova zadataka, a neke od najčešćih su klasifikacija, klasifikacija s nedostajućim vrijednostima, regresija, transkripcija, strojno prevođenje, detekcija anomalija,

strukturirani podaci, sinteza i uzorkovanje, uklanjanje šuma i sl. Nadalje, da bismo procijenili sposobnosti algoritma strojnog učenja, potrebno je razviti kvantitativne mjere, a obično je ta mjera uspješnosti P specifična za zadatak T koji izvodi sustav. Za zadatke kao što su klasifikacija, klasifikacija s nedostajućim ulazima i transkripcija, uobičajeno se koristi točnost modela (engl. *accuracy*). Točnost je samo proporcija primjera za koje model daje točan izlaz. Osim toga, ekvivalentne informacije moguće je dobiti i mjerenjem stope pogreške (engl. *error rate*), koja predstavlja postotak primjera za koje model daje netočan izlaz. Stopa pogreške često se naziva očekivani gubitak $0 - 1$. Gubitak $0 - 1$ na konkretnom primjeru je 0 ako je primjer točno klasificiran, a 1 ako nije. Međutim, za neke zadatke poput procjene gustoće, takva mjera nema smisla, stoga je potrebno koristiti drukčiju metriku za mjerenje točnosti koja modelu daje rezultat s kontinuiranom vrijednosti za svaki primjer. U tom smislu, kao najčešća mjera, koristi se prosječna log-vjerojatnost koju model dodjeljuje pojedinim primjerima. Glede učenja neuronske mreže, cilj je utvrditi koliko dobro radi na podacima koji ranije nisu bili prezentirani neuronskoj mreži jer to određuje i njezinu učinkovitost u realnom svijetu. Stoga se za mjerenje učinkovitosti ili točnosti neuronske mreže koriste podaci odvojeni od podataka za učenje, odnosno neuronsku mrežu se uči na jednom, a testira na drugom skupu podataka. Iako se izbor mjere uspješnosti može činiti kao jednostavan zadatak, ponekad je teško odabrati onu koja će najbolje mjeriti točnost neuronske mreže, a to ponajviše ovisi o samom zadatku koji neuronska mreža treba riješiti. Spomenuto je da neuronske mreže odnosno općenito algoritmi strojnog učenja mogu učiti iz iskustva, odnosno iz primjera E , a u široj definiciji učenje neuronskih mreža moguće je kategorizirati kao nenadzirano ili nadzirano, ovisno o tome temeljem kakvog iskustva im je "dozvoljeno" učiti, odnosno da li uče iz primjera na kojima su prethodno označeni primjeri ili algoritmi uče bez prethodno označenih primjera u skupu za učenje. [214]

4.4.1 Tipovi učenja neuronskih mreža

Tipovi učenja ili treniranja neuronskih mreža, koji se, ovisno o izabranom pristupu i konceptu načelno dijele na:

- nadzirano učenje,
- polunadzirano učenje,
- nenadzirano učenje,
- duboko podržano učenje. [215, 216]

Nadalje, Haykin [188], u kontekstu paradigme učenja, definira tri osnovne vrste učenja: nadzirano, nenadzirano i podržano učenje.

4.4.1.1 Nadzirano učenje

Nadzirano učenje (engl. *Supervised Learning*) je tehnika strojnog učenja koja koristi prethodno označene podatke. U slučaju nadziranog učenja, okolina ima skup ulaza i odgovarajućih izlaza. Na primjer, za određeni ulaz inteligentni agent ili neuronska mreža predviđa određen izlaz, a agent ili mreža će dobiti i određenu vrijednost gubitka. Zatim će neuronska mreža iterativno izmijeniti mrežne parametre radi boljeg približavanja željenim izlazima. Nakon uspješnog učenja, neuronska mreža će moći dobiti točne odgovore na upite iz okoline. Postoje različite metode nadziranog učenja, a načelno se za regresijske i klasifikacijske zadatke. Širi pregled metoda nadziranog učenja predstavljen u [217, 218]. Pored toga, u domeni nadziranog učenja primjenjuju se i različiti tipovi neuronskih mreža [arhitekture neuronskih mreža koje se mogu koristiti kod nadziranog učenja korištenjem dubokih neuronskih mreža (engl. DNN - Deep Neural Networks) [219], odnosno prema pojedinom tipu neuronske mreže vrlo često se koriste: konvolucijske neuronske mreže (CNN ili ConvNet) [160, 220], rekurentne ili ponavljajuće neuronske mreže (RNN) [221], rekurentne mreže s kratkotrajnom memorijom (LSTM) [163], propusne povratne jedinice (engl. GRU - Gated Recurrent Units) [222] i sl.

4.4.1.2 Polunadzirano učenje

Polunadzirano učenje (engl. *Semi-supervised learning*) je tip učenja koje se izvodi na temelju djelomično označenih skupova podataka, a često se naziva i podržano učenje. U nekim slučajevima duboko podržano učenje (engl. DRL – *Deep Reinforcement Learning*) [223, 224] i generativne suparničke mreže (GAN) također se koriste kao polunadzirano učenje [225].

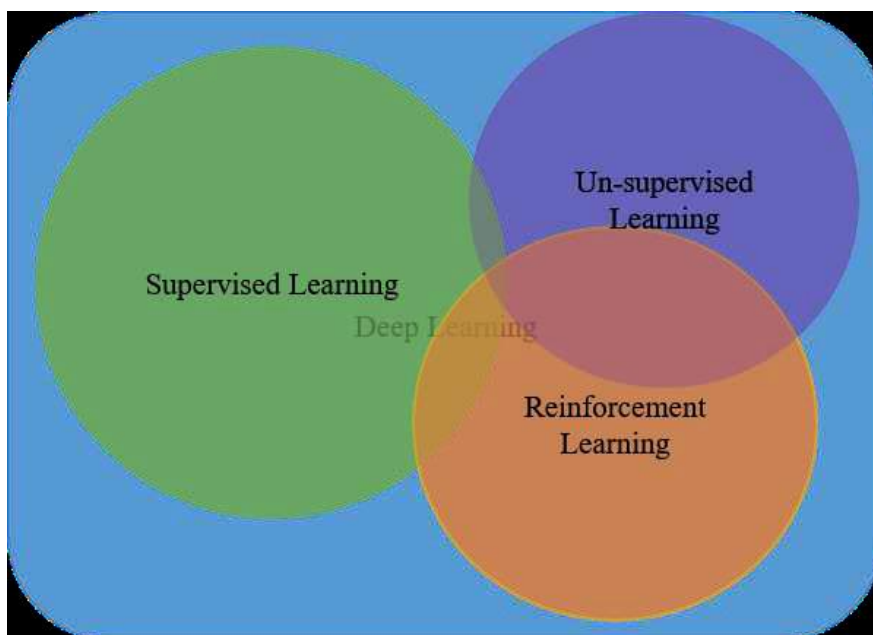
4.4.1.3 Nenadzirano učenje

Nenadzirano učenje (engl. *Unsupervised Learning*) je tip učenja koje se izvodi na potpuno neoznačenim skupovima podataka. U ovom slučaju, uči se unutarnja struktura, pravilnosti ili važne značajke kako bi se otkrili nepoznati odnosi ili strukturu unutar ulaznih podataka. Kao metode nenadziranog učenja često se smatraju grupiranje, smanjenje dimenzija i generativne tehnike. Nekoliko metoda dubokog učenja pokazale su se uspješne u grupiranju tako i u nelinearnom smanjenju dimenzija, uključujući Automatske enkodere (AE – *Auto-Encoders*) [226, 162], ograničeni Boltzmanov stroj (engl. RBM – *Restricted Boltzmann Machines*) [227, 228] i ranije spomenuti GAN. Pored toga, RNN modeli kao što su spomenuti LSTM i DRL (engl. DRL – *Deep Reinforcement Learning*), u mnogim se područjima primjene također koriste za nenadzirano učenje.

4.4.1.4 Duboko podržano učenje

Duboko podržano učenje (DRL) je tip učenja neuronskih mreža kod kojih se učenje provodi na temelju obavljenih akcija u nekom okruženju, a koristi se u nepoznatom okruženju. Ovaj tip učenja može smatrati i kao nenadzirano i kao nadzirano učenje, a može se razmatrati i kao zasebna kategorija. [216, 223, 224]

Međuodnosi pojedinih tipova strojnog odnosno dubokog učenja prikazano je na slici 23.



Slika 23. Vrste i međuodnos vrsta dubokog učenja [215]

4.4.2 Algoritmi učenja neuronskih mreža

Algoritam učenja neuronske mreže predstavlja skup pravila za rješavanje problema učenja, ne postoji jedinstveni algoritam učenja nego mnogo različitih algoritama od kojih svaki ima prednosti i mane. U osnovi, algoritmi učenja se međusobno razlikuju po tome kako prilagođavaju vrijednosti težina na sinaptičkim vezama neurona. Još jedan značajan faktor koji treba uzeti u obzir je kako neuronska mreža, koja je sastavljena od međusobno povezanih neurona, ostvaruje odnos prema svojoj okolini. Ukupno postoji stotinjak algoritama za učenje neuronskih mreža, a pet najčešćih algoritama ili pravila učenja kod neuronskih mreža su: učenje korekcijom pogreške (engl. *Error correction learning*), učenje zasnovano na memoriranju (engl. *Memory-based learning*), Hebbovo učenje (engl. *Hebbian learning*), kompetitivno učenje (engl. *Competitive learning*), Boltzmanovo učenje (engl. *Boltzman learning*).

Ukratko rečeno, učenje korekcijom pogreške temelji se na optimizaciji filtriranja, učenje zasnovano na memoriranju temelji se na eksplicitnom memoriranju podataka iz skupa za učenje. Nadalje, Hebbovo i kompetitivno učenje inspirirani su neurobiološkim temeljima, dok je Boltzmanovo učenje nešto drugačije s obzirom da se temelji na idejama koje su posuđene iz statističke mehanike. [188]

4.4.2.1 Učenje korekcijom pogreške

Učenje korekcijom pogreške je tipičan primjer nadziranog učenja. Učenje korekcijom pogreške funkcionira tako da ulazni vektor $x(n)$ aktivira neuron k , a dobiven je procesiranjem ulaznog vektora kroz prethodne skrivene slojeve mreže. Argument n predstavlja vremensku oznaku, odnosno vremenski korak u ponavljajućem procesu učenja mreže. Izlaz promatranog neurona označen je s $y_k(n)$, a kada se izlazni signal $y_k(n)$ usporedi s željenim izlaznim signalom $d_k(n)$ generira se signal pogreške $e_k(n)$, pri čemu je

$$e_k(n) = d_k(n) - y_k(n). \quad (19)$$

Generirani signal pogreške je korektivni mehanizam ovog algoritma, te je ujedno i temelj za korekciju težinskih faktora neurona k s krajnjim ciljem smanjenja pogreške koju će generirati mreža obrađujući sljedeći skup vrijednosti ulaznog vektora na ovom neuronu. Algoritam učenja teži minimum funkcije gubitka (engl. *Loss function*) odnosno $C(n)$ koja je definirana signalom pogreške pri čemu se iteracija zaustavlja, a mreža dostiže stabilno stanje. Matematički zapis troškovne funkcije je:

$$C(n) = \frac{1}{2} e_k^2(n) \quad (20)$$

Na temelju minimiziranja funkcije gubitka nastalo je delta pravilo koje se još naziva i Widrow-Hoff pravilo [229]. Matematički zapis delta pravila prikazan je slijedećom jednadžbom:

$$\Delta w_{kj}(n) = \mu e_k(n) x_j(n) \quad (21)$$

gdje:

- $w_{kj}(n)$ predstavlja težinski faktor neurona k koji je potaknut djelovanjem elementa $x_j(n)$ ulaznog vektora $x(n)$ u iteracijskom koraku n ,
- $w_{kj}(n)$ predstavlja prilagodbu težinskog faktora $w_{kj}(n)$ u n -toj iteraciji i upravo je proporcionalno tzv. parametru stope učenja μ .

Nadalje, izvorna definicija Widrow-Hoff pravila glasi: "*Korekcija sinaptičke težine neurona proporcionalna je produktu signala greške i ulaznog signala odnosne sinapse.*" [229]

Delta pravilo temelji se na mogućnosti izravnog mjerenja signala pogreške i u osnovi je sustav zatvorene povratne veze. Stabilnost ovakve veze određena je parametrima koji je direktno određuju, a jedan od njih je i parametar učenja μ . Odabir veličine ovog parametra ima odlučujuću ulogu u preciznosti i pouzdanosti sustava nastalog procesom učenja. [188]

4.4.2.2 Učenje zasnovano na memoriranju

Učenje zasnovano na memoriranju je tip učenja neuronske mreže kod kojega je znanje potrebno za primjenu ovog algoritma pohranjeno u fizičku memoriju sustava, a matematički se može predstaviti kao skup uređenih parova, gdje je prvi član para vektor ulazne veličine x_i , a drugi član je izlazna vrijednost d_i . Ako je u memoriju sustava pohranjeno N slučajeva, matematički zapis memorije je:

$$\{(x_i, d_i)\}_{i=1}^N \quad (22)$$

Algoritam učenja prihvaća vektor ulaznih vrijednosti x_j , a zadatak mu je klasificirati promatrani slučaj i odrediti ciljnu veličinu d_j . Također treba napomenuti da postoje dvije verzije algoritma: metoda najbližeg susjeda i metoda k -najbližih susjednih klasifikacija, koje se razlikuju po kriteriju specificiranja najbližih susjeda i pravilu učenja koje se

primjenjuje na njih. Prva metoda definira najbližeg susjeda tako da pronalazi vektor iz skupa pohranjenog u memoriji sustava koji ima minimalnu Euklidovu distancu s vektorom ulaznih veličina promatranog slučaja koji se želi klasificirati. Označe li se s x_t ulazni podaci slučaja koji se testira i s x_A najbliži vektor susjed, tada najbliži susjed zadovoljava slijedeću jednadžbu:

$$\min_i d(x_i, x_t) = d(x_A, x_t) \quad (23)$$

u kojoj $d(x_i, x_t)$ predstavlja Euklidsku distancu vektora x_i i x_t . Promatranom slučaju će biti dodijeljena izlazna vrijednost tako da ista preuzme od određenog najbližeg susjeda. Metoda k-najbližih susjeda funkcionira na sličan način. Prethodna metoda je određivala jednog susjeda (najbližeg) i preuzimala klasifikaciju od njega, a ova određuje k-najbližih susjeda i klasificira promatrani slučaj na osnovu njihove učestalije klasifikacije. [188]

4.4.2.3 Hebbovo učenje

Hebbovo učenje je tip učenja neuronskih mreža koji je definiran s neurobiološkog polazišta stavljajući u odnos sinaptičku aktivnost povezanih neurona, odnosno ako akson stanice A potakne stanicu B i trajno sudjeluje u ovom poticanju, tada se u stanici A odigrava određeni proces tako da se povećava njezina aktivnost. U kontekstu učenja neuronskih mreža, pravilo definira povećavanje težinskih faktora sinapse na čijim krajevima se nalaze neuroni koji se sinkrono aktiviraju. Matematičkim izrazom, generalno pravilo Hebbovog učenja može se predstaviti kao:

$$\Delta w_{kj}(n) = F(y_k(n), x_j(n)) \quad (24)$$

Promjena sinaptičke težine neurona k je funkcija dva argumenta: presinaptičkog signala (vektor ulaznih veličina x_j) i postsinaptičkog signala y_j (izlazna veličina neurona k) u vremenskom koraku n . Najjednostavniji oblik Hebbovog pravila definira stopu učenja μ . Stopa učenja je pozitivna konstanta i definira odnos između promjene sinaptičke težine neurona i produkta ulaznog i izlaznog signal neurona:

$$\Delta w_{kj}(n) = \mu y_k(n) x_j(n) \quad (25)$$

Kasnije je izvedena i modificirana verzija gdje je originalna Hebbova hipoteza definirana u tzv. hipotezu kovarijance. Pri tome se Hebbovoj originalnoj hipotezi

mijenja ulazni i izlazni signal s promjenom njihove razine u odnosu na srednje vrijednosti ovih signala za određeni vremenski period. [188]

4.4.2.4 Kompetitivno učenje

Kompetitivno učenje je tip učenja neuronskih mreža koji se temelji na koncepciji mreže u kojoj samo jedan od izlaznih neurona može biti aktivan u bilo koje vrijeme, dok su svi ostali izlazni neuroni neaktivni, a svrha neuronske mreže koja koristi ovaj algoritam je klasifikacija predložaka iz ulaznog skupa podataka. Mehanizam učenja izlaznim neuronima osigurava međusobno natjecanje tako da samo jedan od njih pobjeđuje. Kako bi se osiguralo natjecanje izlaznih neurona mreže, neophodno je postojanje veze među njima tako da svaki izlazni neuron ostalim izlaznim neuronima pruža informaciju o vrijednosti vlastitog lokalnog induciranog polja. Jednostavna arhitektura mreže koja uči po ovom algoritmu je npr. jedan ulazni sloj s tri čvora i samo jedan izlazni sloj mreže. Algoritam kompetitivnog učenja prikazuje sposobnost neuronskih mreža da provode klaster analizu. Svakako, preduvjet uspješnosti ovih analiza je postojanje predložaka ulaznih vektora koji gravitiraju dovoljno jasno distribuiranim klasterima. U suprotnom, mreža se ponaša nestabilno i ne može jasno odgovarati odgovarajućim signalima izlaznih neurona na poticaje vektora ulaznih signala. [188]

4.4.2.5 Boltzmanovo učenje

Boltzmanovo učenje je tip učenja neuronskih mreža koji se temelji na principima termodinamike i statističke mehanike. Neuronsku mrežu zasnovanu na ovom algoritmu, tzv. Boltzmanov stroj, čine neuroni raspoređeni u vidljive i skrivene slojeve mreže s povratnim vezama. Pretpostavka mreže je da neuroni ne mogu imati povratnu vezu na samoga sebe i da djeluju binarno, odnosno da imaju dva moguća stanja: 1 ili -1. Stanje mreže opisuje tzv. energetska funkcija određena stanjem svih neurona i pripadajućih težinskih faktora. Algoritam učenja, slučajnim odabirom, bira pojedinačni neuron mreže i mijenja mu stanje s tim da je nužno postojanje određene razine temperature, pri čemu temperatura nije stvarna fizička, nego pseudotemperatura stroja, međutim upotrebljavaju se navedeni izrazi s obzirom da su pojmovi energije i temperature preuzeti iz termodinamičke teorije na kojoj se temelji ovaj algoritam učenja. Konstantnim ponavljanjem ovog procesa, mreža mijenja razinu svoje energetske funkcije i dostiže tzv. termalni ekvilibrijum, odnosno stanje u kojem, s određenom vjerojatnošću, ispravno odgovara na poticaje iz okoline.

Osim navedenih, jedan od najčešće korištenih je algoritam povratnog širenja greške (engl. Back-propagation algorithm). [188]

4.4.2.6 Algoritam "propagacije unazad"

Algoritam "propagacije unazad" (engl. *Back Propagation Algorithm*) jedan je od najčešće korištenih algoritama kod praktične primjene neuronskih mreža. Predložili su ga Widrow i Hoff [229], a popularizirali su ga Rumelhart, Hinton i Williams. [158] Algoritam se sastoji od sljedećih koraka:

- označavanja ulaznih čvorova,
- punjenja ulaznim varijablama u odgovarajućem redu i propagiranje unaprijed kroz slojeve mreže ka izlaznom sloju,
- produciranja i zapisivanja vrijednosti izlaznih čvorova i računanja pogreške svakog izlaznog čvora,
- propagiranja greške unazad počevši od izlaznog sloja uz modificiranje sinaptičkih težina kako bi se minimizirala očekivana pogreška.

Algoritam se temelji na spomenutom delta pravila, pa ga se još naziva i generalizirano delta pravilo. Delta pravilo pojačava vezu između dva neurona koristeći vrijednost izlaza prvog neurona za potencijal drugog neurona za redukciju pogreške, a relativno s željenom ciljnom vrijednošću. Algoritam se koristi za prilagodbu sinaptičkih težina tijekom faze testiranja mreže. U radu se oslanja na pravilo gradijentnog spusta, optimizacijsku tehniku koja koristi pravac maksimalne promjene (gradijent) dužinom višedimenzijske nelinearne funkcije kako bi se inkrementalno došlo do globalnog optimuma. Spomenuta inkrementalna veličina bira se od strane korisnika. Pravilo može biti prvog reda (najstrmije opadanje) ili drugog reda (konjugirani gradijent). U neuronskim mrežama uporaba ovog pravila se ogleda u konstantnom smanjivanju izlazne greške korištenjem malih prilagodbi težinskih faktora mreže. Gradijent daje relativnu promjenu promatrane greške za izvršenu promjenu težinskog faktora. Cilj je kretanje najstrmijim pravcem u prostoru težinskih faktora izvedeći simultane izmjene svih ulaznih težinskih faktora za fiksnu izlaznu grešku pojedinog neurona. Koristeći testni skup podataka postupak se ponavlja i trebao bi rezultirati sve boljom i boljom predikcijom izlaznih veličina. Vrijeme potrebno za potpuno učenje relacije između ulaznih i izlaznih veličina mreže, naziva se vrijeme učenja ili vrijeme konvergencije. Preveliko ili premalo vrijeme učenja indicira neadekvatnost topologije mreže ili pak

grešku u odabiru odgovarajućeg algoritma učenja, aktivacijske funkcije, transformacije podataka i sl. Sljedeći problem može izazvati pre-trenirana mreža (engl. *Over-trained network*) u fazi korištenja. Potrebno je naime i sam proces učenja zaustaviti u određenoj fazi. Indeks performansi mreže može se koristiti za zaustavljanje učenja na odgovarajućoj razini naučenosti mreže. Postoji veliki broj verzija ovog algoritma, a neke od njih su: algoritam s promjenjivom veličinom koraka (engl. BPVS – *Backpropagation with Variable Step Size*), online adaptivni BP (engl. OABP – *Online Adaptive Backpropagation*), koeficijent učenja i momentum adaptivni algoritam (engl. *Learning rate and momentum adaptive BP*), elastični algoritam (engl. *Resilient BP*), delta-bar-delta algoritam (DBD), algoritam za učinkovito učenje uz ograničenu optimizaciju (engl. ALECO – *Algorithm for Learning Efficiently with Constrained Optimisation*), RLS BP (engl. *Recursive Least Square BP*). [188]

4.5 Umjetne neuronske mreže i duboko učenje

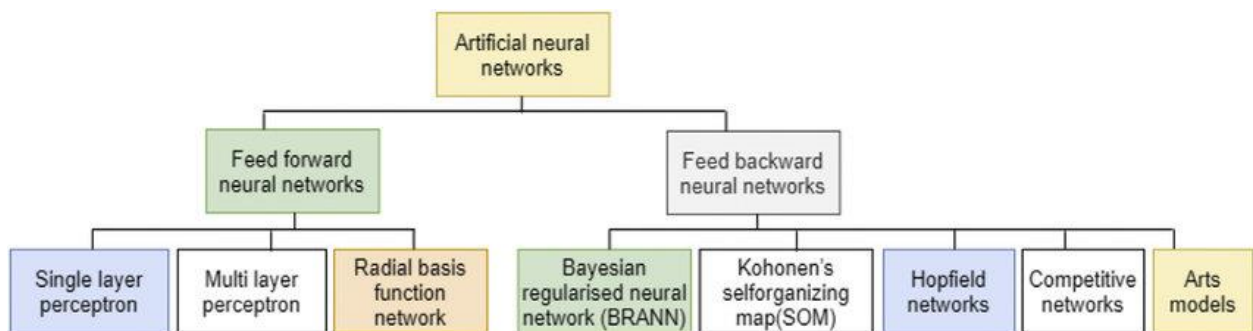
Ranije je već spomenuto da umjetne neuronske mreže nastoje oponašati ljudski mozak kako bi se postigla razina za rješavanje vrlo kompleksnih zadataka.

Ugrubo, umjetne neuronske mreže moguće je podijeliti na unaprijedne neuronske mreže i rekurentne ili ponavljajuće neuronske mreže i njihove različite varijante [230]. Klasifikacija umjetnih neuronskih mreža također je predstavljena i u [231] na sljedeći način: potpuno povezane neuronske mreže (engl. *Fully Connected Networks* - FCN), slojevite neuronske mreže (engl. *Layered Networks*), acikličke neuronske mreže (engl. *Acyclic networks*), unaprijedne neuronske mreže (engl. *Feedforward Networks* - FFN) te modularne neuronske mreže (engl. *Modular Neural Networks*).

Nadalje, zbog velikog broja tipova neuronskih mreža teško ih je sustavno klasificirati. Sa strukturnog gledišta neuronske mreže dijele se na statičke unaprijedne (engl. *feedforward*) i dinamičke (povratne, engl. *feedback*), ovisno o modelu neurona od kojeg su građene te po načinu prolaska signala kroz mrežu. Kao zasebne strukture izdvajaju su neizrazite neuronske mreže (engl. *fuzzy neural networks* – FNN) kod kojih je naglašeno integriranje koncepcija neizrazite logike i neuronskih mreža. Kao nestandardne strukture navode se neuronske mreže kod kojih neuroni imaju histereznu karakteristiku, te CMAC mreže (engl. *Cerebellar Model Articulation Controller* - CMAC) koje aproksimiraju nelinearne funkcije po načelu kodiranja. [232] Za primjenu u identificiranju i vođenju nelinearnih dinamičkih procesa najčešće se koriste višeslojne statičke neuronske mreže. Od dinamičkih neuronskih mreža

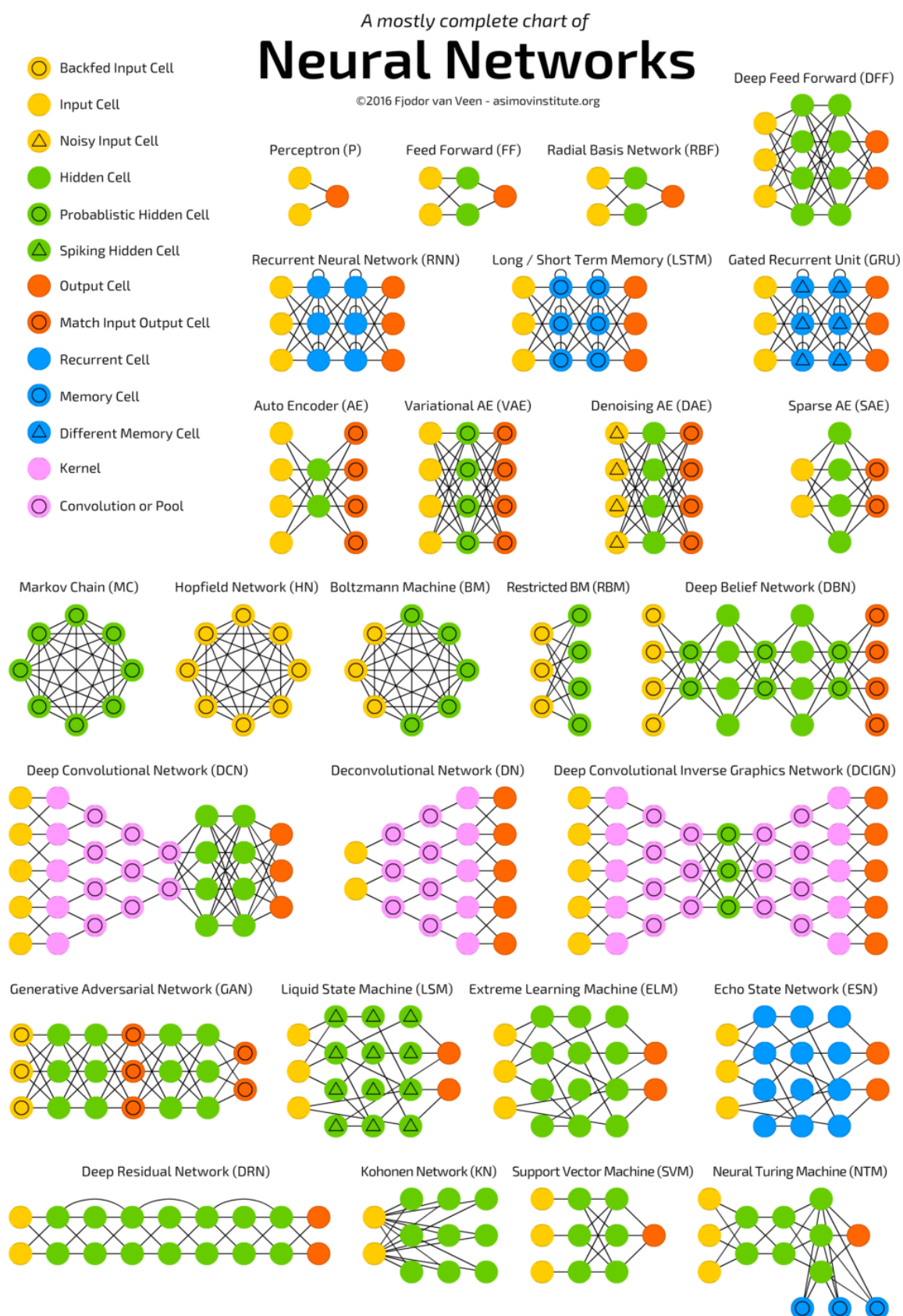
uglavnom se koriste višeslojne neuronske mreže s elementima zadržke (engl. *Time Delay Neural Networks* - TDNN). U okviru *BigData* platformi razvijaju se neuronske mreže koje imaju po nekoliko stotina skrivenih slojeva, što se danas naziva duboko učenje. Također, neuronske mreže moguće je klasificirati prema određenim kriterijima kao što je broj slojeva, tip učenja, tip veze između neurona, veza između ulaznih i izlaznih podataka, ulazne i prijenosne funkcije, sigurnost ispaljivanja, vremenske karakteristike, vrijeme učenja [214, 233].

Prema arhitekturi, neuronske mreže moguće je podijeliti u tri specifične kategorije:



Slika 24. Klasifikacija neuronskih mreža [234]

Niže su također grafički prikazane različite arhitekture neuronskih mreža.



Slika 25. Tipovi neuronskih mreža [235]

Kako je vidljivo na slici 25., rezultat dosadašnjih istraživanja je veliki broj tipova i modela neuronskih mreža od kojih svaka ima svoje specifičnosti, prednosti i nedostatke, ovisno i o primjeni same neuronske mreže. Sukladno prethodnoj slici, najrašireniji tipovi neuronskih mreža su Unaprijedne neuronske mreže (engl. FFNN - *Feed-Forward Neural Networks*) [234, 236 – 238], Rekurentne ili ponavljajuće neuronske mreže (engl. Recurrent Neural Networks – RNN) [239, 230], Mreže s dugim kratkoročnim memorijama (engl. LSTM – *Long Short-Term Memory*) [163], Mreže s povratnom propusnom ćelijom (engl. GRU – *Gated Recurrent Unit*) [240, 241], Simetrično povezane neuronske mreže (engl. SCN – *Symmetrically Connected Networks*) [155, 242 – 245], Duboka probabilistička mreža (engl. DBN – *Deep Belief Network*) [246, 196], Autoenkoderi (engl. *Autoencoders*) [247, 248], Generativne suparničke mreže (engl. GAN – *Generative Adversarial Network*) [225], te Konvolucijske neuronske mreže (engl. CNN ili CONVnet – *Convolutional Neural Nets*) [160], o kojima će više riječi biti kasnije.

4.6 Aktualno naraširenije platforme

U kontekstu razvoja aplikacija beziranih na neuronskim mrežama, u samom početku postoji dilema - da li razvijati vlastiti koncept "od nule", kodirati vlastite biblioteke (engl. *library*) ili koristiti gotovu platformu (engl. *framework*). Iako razvoj vlastite platforme garantira veću mogućnost da bude razvijena i prilagođena vlastitim željama i potrebama, korištenje već gotovih rješenja značajno skraćuje vrijeme razvoja, te samim tim na raspolaganju ostaje više vremena za testiranje aplikacije ili drugog rješenja na kojem se radi. Pozitivna strana gotovih rješenja je i to što pružaju velike mogućnosti prilagodbe vlastitim specifičnim potrebama. Nakon 2012. godine i renesanse u razvoju neuronskih mreža, krenuo je i razvoj gotovih rješenja za razvoj i istraživanje u području neuronskih mreža. U nastavku će biti predstavljeni aktualno najraširenije i najpopularnije platforme, od kojih su neke razvijene od strane velikih tvrtki, dok su druge rezultat razvoja neovisnih timova.

4.6.1. TensorFlow (TF)

TensorFlow (TF) [249] je platforma koju su razvili istraživači i inženjeri s *Googlea*, a predstavljena je 2015. godine [250], a većina tehnologija koje koristi Google već se temelji na ovoj platformi. Primjerice, *Googlev* prevoditelj [251] koristi TensorFlow kao processor prirodnog jezika, klasifikaciju i sumrizaciju teksta, govor, raspoznavanje

rukopisa i slika te označavanje. TensorFlow se temelji na programskom jeziku Python, otvorenog koda je, a još uvijek ga podržava i Google. TensorFlow je opširan paket koji se dobro povezuje s *third-party* platformama za duboko učenje. TF podržava regresiju, klasifikaciju i neuronske mreže, a moguće ga je izvoditi i na CPU i na GPU.

4.6.2. Keras

Keras [252] je platforma otvorenog koda (engl. *open-source framework*) za duboko učenje koji se naširoko koristi. Keras je razvijen na temeljima TensorFlowa (također je napisan u programskom jeziku Python), s tim da je postigao nešto što je kod TF-a predstavljalo određen problem, a to je da je Keras jednostavan za korištenje. Upravo je zbog toga Keras naknadno implementiran u TF. Keras zathijeva relativno malo pisanja programskog koda te je jedna od najjednostavnijih platformi za korištenje, posebno za početnike. Tijekom 2019. godine, predstavljena je nova verzija TF-a (TensorFlow 2.0) koji je uvelike preuzeo i nahvalio sintaksu Kerasa, što ujedno ukazuje na značaj koji Keras ima unutar zajednica istraživača i razvijatelja.

4.6.3. Caffe

Caffe [253] je također platforma otvorenog koda namijenjena dubokom učenju, a posebno je poznata po svojoj brzini s obzirom da može obraditi više od 60 milijuna slika u danu, što ju posebno pogodnom čini za raspoznavanje slika. Radi s C, C++, Python, MATLAB, te CLI. Arhitektura ju čini pogodnom za treniranje neurnoskih mreža bez potrebe za kompleksnim programiranjem što svakako ohabruje u pogledu aktivnog razvoja. S druge strane, Caffe nema određene mogućnosti kao što ih ima TensorFlow ili Microsoft Cognitive Toolkit, zbog čega postavljanje kompleksnih tipova slojeva i nije baš jednostavno, te je to potrebno izvesti u jezicima nižeg ranga. Iako je Caffe doprinijela razvoju pojedinih, veličanstvenih projekata kao što je Google DeepDream [254], njezino vrijeme je na neki način prošlo.

4.6.4 PyTorch

PyTorch [255] je realitvno nova platforma, ali joj popularnost kontinuirano raste. Razvijena je od strane *Facebooka*, a poznata je po jednostavnosti, fleksibilnosti i prilagodljivosti. U posljednje vrijeme je jako dobro prihvaćena unutar istraživačke zajednice dubokog učenja te se na odrđen način smatra svojevrsnim suparnikom TF-u. PyTorch ima vrlo čistu arhitekturu što proces treniranja i razvoja modela dubokog učenja čini jednostavnim za učenje i i izvođenje. S obzirom da Py u nazivu PyTorch

označava Python, to ujedno znači da praktički bilo tko s osnovnim poznavanjem Pythona može započeti s razvojem vlastitih modela dubokog učenja.

4.6.5 Microsoft Cognitive Toolkit

Microsoft Cognitive Toolkit [256] je platforma koja je ranije bila poznat pod nazivom CNTK (engl. *Cognitive Toolkit*), a također je riječ o platformi otvorenog koda koja je namijenjena dubokom učenju. Moguće ju je koristiti za treniranje CNN i RNN neuronskih mreža za gotovo svaki zadatak iz domene dubokog učenja, uključujući slike, govor, analizu teksta i slično. Kao i ranije spomenuti Caffe, podržava Python, C++ i CLI. Također osigurava bolje performanse i skalabilnost, pogotovo kada se usporedi s TF-om dok se izvodi na više strojeva [257] .

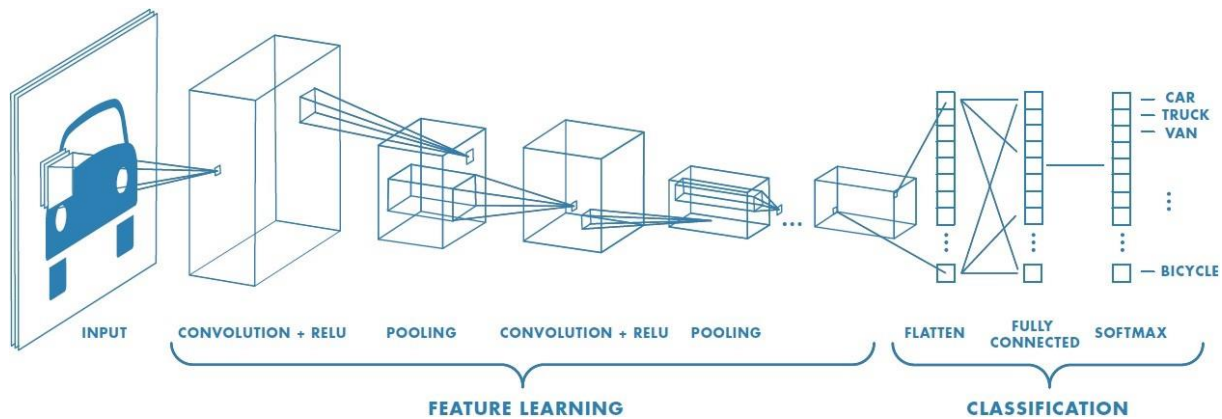
Osim navedenih, postoji još cijeli niz platformi namijenjenih dubokom učenju, a neke od popularnijih su: Apache Spark [258], Scikit-Learn [259], Apache Mahout [260], Apache Singa [261], Amazon Machine Learning [262], Accord.NET [263].

5 Konvolucijske neuronske mreže

Konvolucijske neuronske mreže (CNN ili ConvNet) [264, 265] su neuronske mreže [233, 266, 134] koje koriste najmanje jedan konvolucijski sloj umjesto potpuno povezanog sloja. Cijela mreža predstavlja proces gdje je u slučaju rješavanja problema vizualnog prepoznavanja ulaz slika (ili video), odnosno niz vrijednosti piksela na slici, dok su na drugoj strani (izlaz) vrijednosti pripadnosti određenoj klasi. CNN također ima funkciju gubitka (engl. *loss function*) na zadnjem sloju te sve što vrijedi za neuronske mreže vrijedi i za konvolucijske neuronske mreže.

Kao što je navedeno, neuronske mreže primaju ulazne parametre koje potom transformiraju kroz niz skrivenih slojeva na zadnji sloj. Osnovni problem s neuronskim mrežama je da se ne skaliraju dobro na cijele slike (pogotovo velike slike) jer za cijelu sliku treba jako velik broj težinskih vrijednosti. Potpuno spojeni slojevi neuronskih mreža kod ovakvih problema, s velikim brojem ulaznih podataka, bi vrlo lako doveli do prevelike prilagodbe ulaznim podacima za treniranje (engl. *overfitting*). Da bi bili precizniji, neuron je u prvom skrivenom sloju spojen na samo jedan dio ulazne slike, npr. na regiju veličine 5x5 piksela. Ova regija u ulaznoj slici zove se lokalno receptivno polje (engl. *local receptive field*).

Slojevi CNN-a imaju tri dimenzije: visinu, širinu i dubinu, kako je prikazano na slici 26.



Slika 26. Arhitektura konvolucijske neuronske mreže [267]

Neuroni su u određenom sloju spojeni s malom regijom slojeva prije njih, umjesto da je svaki sloj spojen sa svim ostalim slojevima. Ne samo da su neuroni trodimenzionalni, nego i izlazni vektor ima tri dimenzije. CNN predstavlja niz slojeva i svaki sloj pretvara volumen vrijednosti preko derivabilne funkcije. Postoje tri osnovna tipa slojeva CNN-a:

- konvolucijski sloj (engl. *convolutional layer*),
- sloj udruživanja ili sažimanja (engl. *pooling layer*) i
- potpuno povezani sloj (engl. *fully connected layer*).

Tipična CNN arhitektura redosljedno uključuje sljedeće elemente: **ulazni sloj** (INPUT) → **konvolucijski sloj** (engl. *Convolutional layer* - CONV) → **aktivacijska funkcija**, a najčešće se koristi ReLU (engl. *Rectified Linear Unit* - RELU) → **sloj sažimanja** (engl. *Pooling layer* – POOL) → **potpuno povezani sloj** (engl. *Fully Connected Layer* - FC). Prema zadaći koju obavlja svaki od navedenih elemenata:

- **ulazni sloj** (INPUT) – predstavljaju ulazni podaci (npr. slika) koja se sastoji od niza piksela određene visine i širine te broja kanala (obično tri, ako je RGB, odnosno jedan ako je crno-bijela),
- **konvolucijski sloj** (CONV) – računa izlazne neurone koji su spojeni na lokalne regije u ulaznom nizu podataka. Ovaj sloj može se sastojati od $[x \times y \times k]$, gdje k predstavlja broj korištenih filtera, a veličina filtera ovisi o tome koje značajke filter može detektirati.
- **aktivacijska funkcija** (ReLU ili neka druga) – sloj koji će primijeniti aktivacijsku funkciju kao što je $\max(0, x)$ gdje se granične vrijednosti stavljaju na 0. Ovaj sloj ne mijenja volumen,

- **sloj sažimanja** (POOL ili Pooling layer) – sloj koji će izvršiti uzorkovanje prema prostornim dimenzijama smanjujući volumen npr. na $[x^2 \times y^2 \times 12]$,
- **potpuno povezani sloj** (FC) – potpuno povezani sloj izračunava rezultate svih klasa, koji rezultira volumenom veličine $[1 \times 1 \times c]$, a vrijednost ovisi o broju klasa. [264, 265]

Na ovaj način CNN transformira originalnu sliku, sloj po sloj od piksela do klase kojoj taj piksel pripada. Neki od navedenih slojeva ne samo da primjenjuju određene aktivacijske funkcije već i očekuju određene parametre (težinske vrijednosti i prirast) kao što su CONV i FC, dok drugi slojevi kao što su ReLU i POOL primjenjuju fiksnu funkciju. Parametri CONV i FC slojeva se prilagođavaju uz pomoć metode spuštanja stupnja da bi rezultati klasa bili konzistentni s oznakama u skupu za učenje u svakoj slici.

Prilikom primjene konvolucijskog sloja postoje tri parametra koji utječu na veličinu izlaznog volumena, a to su dubina (engl. *depth*), korak ili pomak (engl. *stride*) i popunjavanje nulama (engl. *zero padding*). Dubina označava broj filtera ili kernela koje želimo primijeniti. Različiti neuroni kroz dimenziju dubine mogu se aktivirati ukoliko se pojavi različito orijentiran rub ili blok određenih boja. Korak ili pomak određuje pomicanje filtera po ulaznoj slici - ako je korak 1 onda se konvolucijski prozor pomiče za jedan piksel kroz određenu sliku. Ponekad je potrebno sliku popuniti i nulama, a jako dobro svojstvo popunjavanja nulama je mogućnost kontroliranja veličine izlaznog volumena. Veličina izlaznog volumena može se izračunati uz pomoć ulaznog volumena W , veličine receptivnog polja F , s primijenjenim korakom S i količine primijenjenog popunjavanja nulama P . Formula za izračunavanje broja neurona koji "odgovaraju" je sljedeća:

$$\frac{W-F+2P}{S} + 1 \quad (26)$$

Svaki neuron u skrivenom sloju ima sklonosti i težinske vrijednosti koje su spojene s njegovim lokalnim receptivnim poljem, a sve te vrijednosti jednake su za sve neurone u skrivenom sloju, što znači da svi neuroni u prvom skrivenom sloju detektiraju jednake značajke, samo na različitim lokacijama. Uz pretpostavku da određene težinske vrijednosti i sklonosti određuju okomiti rub na slici, isto svojstvo može biti korisno i na vodoravnom rubu na toj slici. Drugim riječima, konvolucijske neuronske mreže su jako otporne na translaciju. Upravo iz ovih razloga mapa dobivena iz ulaznog sloja često se

naziva mapom značajki (engl. *feature map*). Struktura mreže, naravno, nema samo jednu mapu značajki već više njih. [264, 265]. Ono što konvolucijsku neuronsku mrežu razlikuje od klasične neuronske mreže je postojanje spomenutog konvolucijskog sloja, koji se najčešće sastoji od dva dijela, s tim da je u prvom dijelu, ulazna aktivacijska mapa, koja u prvom sloju mreže predstavlja sliku te ulazi u konvoluciju s određenim brojem filtera koji prepoznaju različite značajke. U drugom dijelu, na dobiveni rezultat primjenjuje se aktivacijska funkcija, s tim da se kao jedna od najučinkovitijih, za različite tipove zadataka, pokazala ReLU aktivacijska funkcija. Svaki konvolucijski sloj sadrži unaprijed odabrani broj filtera, odnosno jezgri fiksne dimenzije, koji pohranjuju vrijednosti težina w . Konvolucija filtera s mapom značajki odvija se samo na prostornim koordinatama širine i visine mape značajki. Ukoliko se radi o RGB slici, za svaki kanal boje uglavnom se primjenjuje isti filter, te se rezultati zbrajaju po kanalima mape značajki. Konvolucijski slojevi koji su na početku mreže, uče prepoznati detalje i strukture, pa npr. prvi sloj mreže može naučiti prepoznati rubove objekata na slikama. Što se konvolucijski sloj nalazi dublje u neuronskoj mreži, on uči prepoznavati kompleksnije značajke. Tako drugi sloj mreže može prepoznavati jednostavne geometrijske oblike, dok posljednji sloj može prepoznavati osobe, pse, automobile i slično, odnosno objekte prema unaprijed definiranim klasama. Vrijednosti težina koje su pohranjene u filterima predstavljaju parametre modela te ih mreža može naučiti. Uobičajeni hiperparametri modela koji se definiraju u konvolucijskom sloju su broj filtera d (engl. *number of filters*), veličina filtera f (engl. *filter size*), dopunjavanje ili popuna p (engl. *padding*), korak s (engl. *stride*), sažimanje (engl. *pooling*), stopa učenja (engl. *learning rate*), moment (engl. *momentum*), a svaki od navedenih hiperparametara ima različit učinak na izlazni rezultat detekcije. Konfiguriranje hiperparametara konvolucijske neuronske mreže je kompleksno s obzirom da je učenje mreže vremenski zahtjevno, ali i zbog postojanja velikog broja parametara koje je moguće konfigurirati. Unatoč tomu, kao četiri najvažnija parametra izdvajaju se: veličina jezgre (kernela ili filtera), broj filtera, dopunjavanje (engl. *padding*) i korak (engl. *stride*), a koji su prethodno u tekstu šire pojašnjeni.

Osim navedenih moguće je konfigurirati i druge hiperparametre kao što su: stopa učenja, momentum, broj epoha (engl. *epochs*), veličina mini grupe slika (engl. *batches*), aktivacijske funkcije i drugi, a bilo koji od navedenih također može imati manji ili veći utjecaj na konačni rezultat učenja.

- Stopa učenja (engl. *learning rate*) – najčešće se odabire u rasponu od 0,01 do 0,0001, i jedan je od ključnih hiperparametara mreže o kojem izravno ovisi konvergencija mreže. Stopa učenja se može i dinamički mijenjati sa svakom iteracijom. Stopa učenja kontrolira kako se ažuriraju težine u optimizacijskom algoritmu. Stopa učenja se može postaviti kao fiksna, kao postupno smanjujuća stopa učenja, stope učenja temeljene na momentumu ili adaptivne stope učenja, te ovise i o izboru optimizatora kao što je SGD, Adam, Adagrad, AdaDelta, RMSProp i drugi. [268]
- Broj epoha (engl. *number of epochs*) – Epoha je jedan prolaz svih uzoraka kroz neuronsku mrežu za vrijeme procesa učenja. Nakon svake epohe može se vidjeti pogreška koju mreža daje. Manjim brojem epoha se skraćuje vrijeme učenja, ali i točnost naučene mreže, stoga bi se broj epoha trebao povećavati sve dok se ne uoči mala razlika između greške u testiranju i greške u treniranju.
- Veličina mini grupe podataka (engl. *batch size*) - o ovom hiperparametru ovisi koliko će uzoraka odjednom proći kroz mrežu. Dobar izbor veličine mini grupe znatno ubrzava učenje mreže, ne gubeći previše na njenoj točnosti. Na temelju veličine jedne mini grupe može se odrediti i ukupan broj mini grupa. Ako se koristi algoritam gradijentni spust s mini grupama, težine će se ažurirati nakon svake mini grupe. Korištenje mini grupa se obično preporučuje kod učenja konvolucijskih neuronskih mreža, a za početak je dobar izbor u rasponu od 16 do 128. Također treba istaknuti da je konvolucijska neuronska mreža osjetljiva na veličinu mini grupe.
- Aktivacijska funkcija (engl. *activation function*) – aktivacijske funkcije su prethodno već šire pojašnjenje u 4.3, te će ovdje biti samo napomenuto da je glavni zadatak aktivacijske funkcije uvođenje nelinearnosti u linearni produkt u neuronskoj mreži, a kad je riječ o konvolucijskim neuronskim mrežama, uobičajeno se koristi ReLU ili neka od verzija navedene aktivacijske funkcije koje daju dobre rezultate. Osim navedene moguća je primjena sigmoidne, Tanh ili neke druge funkcije što primarno ovisi o problemu koji se rješava. [214]

5.1 Metode evaluacije točnosti i pouzdanosti rezultata detekcije i klasifikacije

Točnost klasifikacije kod neuronskih mreža može se procijeniti računanjem broja točno prepoznatih primjera pojedine klase odnosno stvarno pozitivnih rezultata (engl. *TP True Positive*), broja ispravno prepoznatih primjera koji ne pripadaju klasi, odnosno stvarno negativnih primjera (engl. *TN – True Negatives*) i primjera koji su pogrešno dodijeljeni klasi, odnosno lažno pozitivnih (engl. *FP – False Positives*), te primjera koji nisu prepoznati kao primjeri klase, odnosno lažno negativnih (engl. *False Negatives – FN*).

Nadalje, ovisno da li je riječ o binarnoj ili multiklasnoj klasifikaciji koriste se ponešto različite mjere evaluacije.

Kada je riječ o binarnoj klasifikaciji koriste se:

- točnost (engl. *Accuracy*) koja pokazuje ukupnu efektivnost treniranog modela, a računa se prema formuli:

$$\frac{TP+TN}{TP+FN+FP+TN} \quad (27)$$

- Preciznost (engl. *Precision*) koja pokazuje broj točno klasificiranih primjera s obzirom na ukupan broj primjera klasificiranih u neku klasu, a računa se prema formuli:

$$\frac{TP}{TP+FP} \quad (28)$$

- Prosječna srednja preciznost (engl. *mAP – mean Average Precision*) koji pokazuje učinkovitost modela, kao što je onaj korišten na PASCAL VOC 2012 natjecanju [269]. Da bi se dobio mAP, računaju se srednje vrijednosti prosječne preciznosti (engl. *AP – Average Precision*) za sve klase, s tim da je u ovom eksperimentu u obzir uzeta samo jedna klasa – Osoba (engl. *Human/Person*). Dobiveni rezultati su uspoređeni sa baznim rezultatima (engl. *Ground truth*). Detekcija se računa kao pozitivna ako je postotak presjeka unije (engl. *IoU – Intersection-over-Union*) detektiranog pokaznog okvira veći ili jednak 50%. Primjer pozitivne i negativne detekcije prema IoU kriteriju prikazan je na slici 45. Metrika mAP je standardni postupak za ocjenu performansi sustava za detekciju objekata. Svaka metrika, pa tako i ova, mora imati dostupne podatke za usporedbu s dobivenim rezultatima, a ti podaci se nazivaju "referentni podaci" (engl. *Ground truth*). Kod problema detekcije objekata referentni podaci uključuju koordinate svakog objekta na slici te pripadnu klasu korespondentnog objekta.

Metrika mAP računa se na sljedeći način:

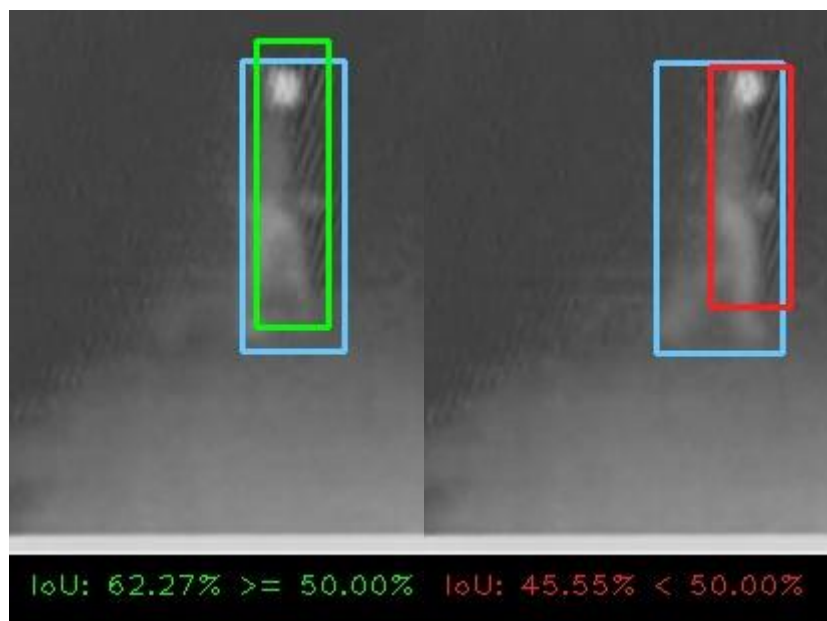
1. Odabere se klasa za koju se želi utvrditi točnost detekcije,
 2. Inicijaliziraju se cjelobrojne varijable TP, FP i FN na nulu,
 3. Ide se slijedno po slikama i traže se objekti koji pripadaju odabranoj klasi,
 4. Kada je pronađen objekt pripadajuće klase, pokrene se sustav za detekciju objekata,
 5. Pregledavaju se podaci o dobivenim rezultatima i traži se taj isti objekt,
 6. Ako je objekt pronađen među rezultatima, računa se IoU. IoU pokazuje u kojoj se mjeri pronađeni objekt i stvarni objekt poklapaju, kao što je prikazano na slici 27. Ako je vrijednost IoU veća 50% (slika 27 lijevo), tada se rezultat uzima u obzir i bilježi kao točna detekcija te se varijabla TP povećava za jedan. Ako objekt nije pronađen ili je vrijednost IoU manja od 50% (slika 27 desno), onda se kaže da se pojavila greška tipa 2, odnosno da objekt postoji na slici, ali ga sustav za detekciju nije pronašao. U tom slučaju se varijabla FN povećava za jedan,
 7. Ako među rezultatima postoji objekt kojeg nema na originalnoj slici, kaže se da se dogodila pogreška tipa 1, te se povećava varijabla FP za onaj broj koliko postoji objekata među rezultatima kojih nema na stvarnoj slici,
 8. Na temelju rezultata nakon pregleda svih slika provodi se postupak crtanja krivulje preciznosti/odziva (engl. *PR Curve – Precision/Recall curve*). Važno je napomenuti da prema [269] integral krivulje preciznosti/odziva odgovara prosječnoj točnosti detekcije dotične klase,
 9. Postupak se ponavlja za sve klase te se uzima prosjek integrala kao mjera mAP.
- Odziv ili osjetljivost (engl. *Recall or Sensitivity*), pokazuje broj točno klasificiranih primjera u odnosu na sveukupni broj primjera neke klase, a računa se prema formuli:

$$\frac{TP}{TP+FN} \quad (29)$$

- F-mjera (engl. *F-score*) pokazuje odnos između stvarno pozitivnih primjera i onih koje je dodijelio klasifikator kao pozitivne, a računa se prema formuli:

$$\frac{TP}{TP+FN} \quad (30)$$

Ukoliko je isti objekt detektiran više puta, kao pozitivna računa se samo jedna detekcija.



Slika 27. Vizualni pozitiv (lijevo) i negativ (desno) koji prikazuje IoU kriterij

5.2 YOLO

Algoritmi za detekciju objekata na slikama ili videozapisima omogućuju razvoj i primjenu naprednih tehnologija i primarni su fokus istraživanja za mnoga područja, od transporta, zdravstvene zaštite, kao i sigurnosti s obzirom kolika se pažnja u današnje vrijeme pridaje sigurnosti, posebice kada je riječ o zaštiti osoba i imovine. Postoji cijeli niz algoritama za detekciju objekata, a svaki ima svoje prednosti i mane. Algoritmi za detekciju objekata uglavnom su podijeljeni u dvije skupine, odnosno prema načinu na koji obavljaju svoje zadatke. Prvu skupinu čine algoritmi temeljeni na klasifikaciji i rade u dvije faze. Najprije odabiru dijelove slike od interesa (ROI), a zatim klasificiraju objekte unutar tih regija pomoću CNN-a. Ova grupa koja obuhvaća rješenja poput R-CNN-a obično je prespora da bi se primjenjivala u stvarnom vremenu. Drugu skupinu čine algoritmi temeljeni na regresiji – skeniraju cijelu sliku i predviđaju lokaliziranje, prepoznavanje i klasificiranje objekata unutar slike. Ovoj skupini pripadaju i algoritmi kao što je YOLO (engl. *You Only Look Once*) koji je korišten za izvođenje eksperimenata u okviru ovog rada. Algoritmi iz ove skupine su brži i mogu se koristiti za otkrivanje objekata u stvarnom vremenu.

5.2.1. YOLOv1

Za razliku od mnogih drugih detektora temeljenih na konvolucijskim neuronskim mrežama koji koriste više faza ili prolaza u arhitekturi kako bi prvo odredili potencijalne regije na slici gdje su prisutni određeni objekti, a zatim ih klasificirali u klase objekata, YOLO u jednom prolazu kroz neuronsku mrežu obavlja oba zadatka, zbog čega je i dobio naziv koji u doslovnom prijevodu s engleskog znači "pogledaj samo jedanput". Autori [13] su zadatak detekcije objekata definirali kao regresijski problem, od slikovnih piksela do koordinata detekcijskih okvira objekata i pridruženih vjerojatnosti klasa objekta.

Arhitektura izvornog YOLO detektora (YOLOv1) sastoji se od 24 konvolucijska sloja u kombinaciji sa slojem maksimalnog sažimanja, nalik prilično standardnom CNN-u koji izvode ekstrakciju značajki, a zatim dva potpuno povezana sloja koji predviđaju lokacije potencijalnih objekata koji su označeni detekcijskim okvirima kojima su pridruženi rezultati pouzdanosti da detekcijski okvir doista sadrži objekt i uvjetne vjerojatnosti pripadnosti za svaku pojedinu klasu objekta za detekcijski okvir koji sadrži objekt. Dodani su i redukcijski slojevi (engl. *bottleneck*) s jezgrom (kernelom, filtrom) 1×1 kako bi smanjili dimenzionalnost značajki prethodnih slojeva, nakon čega dolaze

konvolucijski slojevi s jezgrom 3×3 . Prilikom treniranja mreže na novom skupu podataka koristi se inicijalizacija težina dobivena učenjem na ImageNetu. Na taj način se ubrzava učenje i izbjegava se prenaučenos modela (engl. *overfitting*) ako je mali skup podataka. Završni sloj koristi linearnu aktivacijsku funkciju, dok svi ostali slojevi koriste propusnu (engl. *leaky*) aktivacijsku funkciju:

$$\phi(x) = \begin{cases} x, & \text{ako je } x > 0 \\ 0, & \text{inače} \end{cases} \quad (30)$$

Takva mrežna struktura i formulacija problema omogućuje "s kraja na kraj" (engl. *end-to-end*) učenje mrežnih parametara, izravno koristeći performanse detekcije kao kriterij optimizacije. Treniranje duboke neuronske mreže "s kraja na kraj" znači optimizirati težine simultano, u jednom prolazu, uzimajući u obzir piksele na ulazu neuronske mreže te njihovo direktno mapiranje na izlazu. Neuronska mreža koja nema mogućnost treniranja "s kraja na kraj" zahtijevat će nekoliko zasebnih postupaka učenja prije nego što se dobiju konačani rezultati. Ulazna slika prvo se dijeli na $S \times S$ rešetku ($S = 7$ u tipičnoj konfiguraciji), gdje svaka ćelija mreže predviđa B potencijalnih okvira za detekciju objekta, tako da se za svaku ćeliju mogu otkriti najviše dva objekta. Tijekom procesa učenja, vrijednost pouzdanosti predviđenog detekcijskog okvira odgovara rezultatu odnosa presjeka i unije (engl. *IoU – Intersection over Union*) tog detekcijskog okvira i detekcijskog okvira bazne vrijednosti, tako da detekcijski okvir koji ne sadrži nijedan objekt ima dodijeljenu vrijednost pouzdanosti jednaku nuli. Rezultat pouzdanosti da objekt pripada određenoj klasi računa se prema:

$$\text{Pr(Object)} \times \text{IOU}_{\text{truth}} \quad (31)$$

Vjerojatnosti klase povezane su s mrežom ćelija, tako da svi detekcijski okviri koje predviđa jedna ćelija dijele iste pouzdanosti klase. Predviđanje detekcijskog okvira definira pet parametara (centroidne x i y koordinate, širina, visina i vrijednosti pouzdanosti detekcije objekta) i povezano je s ćelijskim predviđanjem vrijednosti pouzdanosti C za svaku klasu C objekata, koja se efektivno sastoji od $5 + C$ predviđanja. Točka (x, y) predstavlja centar okvira, dok parametri w i h relativnu širinu i visinu u odnosu na cijelu sliku. Svaka ćelija predviđa samo jedan set vjerojatnosti klase neovisno o broju okvira B . Zbog svih ovih parametara, YOLO za svaku sliku kreira tenzor veličine $S \times S \times (B \times 5 + C)$, gdje je $S \times S$ broj ćelija, dok je $B \times 5$ broj predviđenih okvira po ćeliji i 5 parametara za svaki okvir, a C je broj klasa. Originalni

YOLO u radu koristi vrijednosti $S = 7$ i $B = 2$. Model postavlja snažna ograničenja s obzirom da svaka ćelija predviđa samo dva okvira koji mogu pripadati samo jednoj klasi. Zbog toga je dosta bitna veličina okvira jer se javlja poteškoća ako su objekti manji i nalaze se jedan pokraj drugoga, što dovodi do loših predikcija za manje objekte koji se nalaze u grupama. Ako objekt zauzima više od jedne ćelije, ćelija koja sadrži središte detekcijskog okvira odabrana je kao nositelj predviđanja za taj objekt odnosno odgovorna je za njegovu detekciju, dok su detekcijski okviri predviđeni od strane drugih ćelija za isti objekt zanemareni od strane funkcije gubitka učenja (engl. *loss function*). YOLO predviđanje objekata izvodi u završnom sloju na temelju klasifikacijskih i lokalizacijskih pogrešaka detekcijskog okvira. Radi lakše optimizacije na izlaznom sloju, funkcija gubitka izražena je kvadratnom pogreškom za računanje funkcije gubitka, a to su pogreške između točne i predviđene lokacije centra detekcijskog okvira (x, y) , relativne širine w , relativne visine h , rezultata pouzdanosti $p(c)$ i klasifikacije objekata C . Međutim, ona nije savršeno usklađena s ciljem da se poveća prosječna preciznost modela jer lokalizacijske pogreške izjednačuje s klasifikacijskim pogreškama što nije idealno. Također, mnogo ćelija ne sadrži objekte zbog čega je rezultat klasifikacije negativan, odnosno jednak je nuli. Puno negativnih potiskuje pozitivne rezultate što dovodi do neuravnoteženosti modela i treniranje divergira brzo. Kako bi spriječio ovaj problem, YOLO model povećava vrijednost pogreške dobivene predviđanjem lokalizacije okvira, a smanjuje pogrešku klasifikacije gdje je rezultat pouzdanosti jednak nuli, odnosno u kojoj detekcijski okvir ne sadrži objekt. To radi tako da u funkciju gubitaka uvodi dva parametra λ_{coord} i λ_{noobj} . Lokalizacijsku pogrešku množi faktorom $\lambda_{coord} = 5$, a klasifikacijsku pogrešku koja predviđa da nema objekta u okviru s $\lambda_{noobj} = 0,5$.

5.2.2. YOLOv2

U drugoj verziji YOLO detektora (YOLOv2) [14], napravljena je značajna promjena u načinu na koji se generiraju predviđanja detekcijskog okvira u cilju bolje detekcije malih objekata te povećanja odziva (engl. *Recall*). Umjesto izravnog predviđanja koordinata detekcijskog okvira za svaku ćeliju koja koristi potpuno povezani sloj, u YOLOv2 potpuno povezani slojevi su uklonjeni i zamijenjeni konvolucijskim slojem koji predviđa relativne pomake unaprijed definiranih sidrenih okvira (engl. *anchor boxes*) rešetkastih ćelija koje čine izlaznu mapu značajki. Struktura mreže je na taj način donekle izmijenjena i sastoji se od 19 konvolucijskih slojeva slojeva i pet slojeva sa

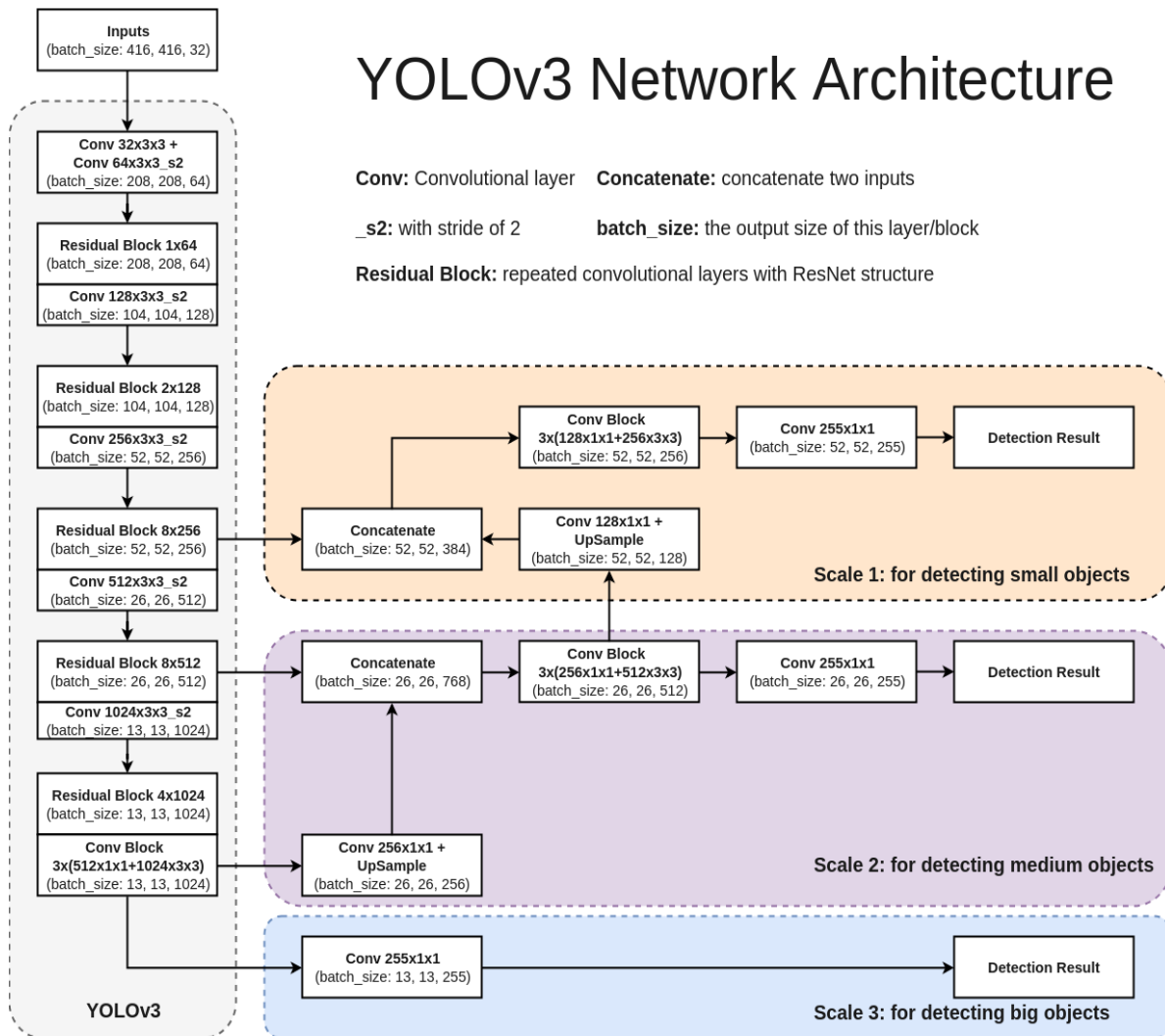
maksimalnim sažimanjem (engl. *max pooling layers*), bez potpuno povezanih slojeva. U usporedbi s YOLOv1, gdje su obično predložena dva detekcijska okvira za svaku ćeliju u mreži 7×7 , izlazna mapa značajki u YOLOv2, kod tipične konfiguracije, je 13×13 s pet predloženih detekcijskih okvira po ćeliji. U YOLOv1, detekcijski okviri iz iste ćelije dijele iste vjerojatnosti klase, dok u YOLOv2 svaki detekcijski okvir ima zasebnu vjerojatnost za pojedinu klasu, tako da jedna ćelija može predvidjeti objekte iz različitih klasa. Sidreni okviri definirani su iz podataka koristeći *k-means* grupiranje detekcijskih okvira iz baznih vrijednosti.

U YOLOv2 treniranje se provodi u dvije faze, prvo se mreža osposobljava za zadatak klasifikacije na skupu podataka ImageNet, a zatim se zadnji konvolucijski sloj u mreži zamijeni s tri 3×3 konvolucijska sloja i jednim 1×1 konvolucijskim slojem, a mreža se uči za zadatak otkrivanja na skupu podataka VOC. Proces učenja je izveden na različitim ulaznim veličinama podataka (slika), kako bi se povećala robusnost detekcije za različite veličine objekta. Osim navedenih, u YOLOv2 je uveden i cijeli niz drugih promjena koje su detaljno opisane u izvornom radu. [14]

5.2.3. YOLOv3

U trećoj verziji YOLO detektora (YOLOv3) [15] koristi se dublja mreža za ekstrakciju značajki nego u prethodnim verzijama, a sastoji se od 53 konvolucijska sloja s filterima 3×3 i 1×1 , (slika 28), s nekim prečacima spojeva između slojeva (rezidualnih blokova), umjesto mreže s 19 slojeva kao što je kod YOLOv2. YOLOv3 mijenja način predviđanja vjerojatnosti da se objekt nalazi unutar okvira tako da koristi logističku regresiju. Vjerojatnost bi trebala biti 1 ako apriorni okvir preklapa originalni objekt više od drugih okvira. Ako apriorni okvir nije najbolji ali preklapa objekt više od nekog praga, njegova predikcija se ignorira. U originalnom radu koristi se prag 0,5. Ovaj pristup je sličan eliminaciji nemaksimalnih elemenata u prethodnim verzijama te njime sustav primjenjuje gubitak samo na apriornom okviru koji najbolje preklapa objekt, odnosno svaki objekt može imati samo jedan okvir. Korištenje logističke regresije je pogodno ako apriorni okvir nije dodijeljen objektu, zato što on ne uzrokuje nikakav gubitak za predviđanje lokalizacije i klasifikacije. Sljedeće se uvodi višestruka klasifikacija u predviđanju objekata unutar okvira. Izbacuje se *softmax* funkcija za klasifikaciju pošto se pokazalo da nije nužna za dobre performanse modela te se umjesto višerazredne klasifikacije (engl. *multi-class*) koristi klasifikacija oznaka (engl. *multilabel*). Tijekom treniranja koristi se gubitak entropije za predviđanje klasa. Ove promjene pomažu kod

klasifikacije kompleksnijih skupova podataka. *Softmax* funkcija pretpostavlja da svaki detekcijski okvir može imati samo jednu klasu što često nije slučaj. Višeklasni pristup bolje predstavlja podatke. YOLOv3 radi predikcije na temelju tri skale, odnosno koriste se tri mape značajki različitih dimenzija za predviđanje objekata. Predikcije se rade na mapama čija je dimenzija umanjena rezolucijom ulaza za faktore 32, 16 i 8. Ako je ulaz u mrežu 416×416 predikcije se rade na slojevima dimenzija 13×13 , 26×26 i 52×52 . Metoda koja se ovdje spominje je dekonvolucija. YOLOv3 na izlazni sloj YOLOv2 modela dodaje novi sloj koji je povećan faktorom 2 i spojen s prethodnom mapom značajki jednakih dimenzija. Nakon njega dodano je još nekoliko konvolucijskih slojeva za obradu kombinirane mape značajki nakon kojih se vrši detekcija. Ovaj postupak se ponavlja još jednom samo za veću skalu. YOLOv3 koristi devet apriornih okvira i to tri za svaku od skala. Korištenjem dekonvolucije dobiva se više korisnih semantičkih informacija što omogućuje bolju detekciju manjih objekata. Nova mreža koju koristi YOLOv3 je hibrid između prethodne mreže u YOLOv2 modelu Darknet-19 i ResNet mreže. Ona ima uzastopne 3×3 i 1×1 konvolucijske slojeve, sadrži rezidualne veze i znatno je veća. Mreža ima 53 konvolucijska sloja, pa je zbog toga nazvana Darknet-53. Njezina struktura prikazana je na slici 27. Ova nova mreža je znatno snažnija od Darknet-19, ali nema toliko slojeva kao ResNet mreža zbog čega je i dalje zadržala brzinu.



Slika 28. Arhitektura YOLOv3 [270]

5.2.4. YOLOv4

YOLOv4 [16] je posljednja verzija YOLO detektora i jedina na čijem razvoju nije radio izvorni autor J. Redmon koji je nakon treće verzije odustao od daljnjeg razvoja zbog određenih etičkih dvojbi glede moguće vojne primjene YOLO detektora. Unatoč tomu, razvoj je nastavio A. Bochkovskiy sa svojim timom, u sklopu Darknet platforme. Prema navodima autora, YOLOv4 na MS COCO skupu podataka postiže najbolje rezultate brzinom u stvarnom vremenu s 43,5% AP koji radi na 65 FPS na GPU Tesla V100. Da bi postigli ove rezultate, autori su kombinirali neke značajke poput ponderiranih rezidualnih veza (engl. WRC - *Weighted-Residual-Connections*), *Cross-Stage-Partial-Connections* (CSP), *Cross mini-Batch Normalization* (CmBN), Self-adversarial-training (SAT) i Mish aktivacijske funkcije [271], povećanja mozaičkih podataka, regulacije *DropBlock*-a i gubitak *Clou*-a. Navedene značajke nazivaju se univerzalnim

značajkama jer bi trebale dobro funkcionirati neovisno o zadacima računalnog vida, skupovima podataka i modelima. Autori su također naveli da su kroz istraživanje nastojati pronaći tzv. "brzi" detektor objekata čije će učenje biti jednostavno i koji se može lako primijeniti u proizvodnim sustavima. Istaknuli su da im je glavni cilj bio optimizirati neuronske mreže detektora za paralelno računarstvo te su konačnici, nakon pažljivog proučavanja učinaka različitih performansi detektora razvijenih u prošlosti predložili nekoliko različitih arhitektura. Nakon istraživanja velikog broja poboljšanja u odnosu na nekoliko najsuvremenijih modela i komponenata, njihov konačni model YOLOv4 na kraju se sastojao od CSPDarknet53 kao okosnice koja se koristi za ekstrakciju značajki kada se učenje provodi na GPU, SPP i PAN modela kao vrata, s tim da se korišteni PAN ponešto razlikuje od originala i YOLOv3 kao glave. Unutar tih komponenti, novi model uključuje mnoštvo već dokazanih značajki kao što su CutMix i Mosaic povećanje podataka, DropBlock regularizacija, Mish aktivacija, Self-adversarial training i mnoge druge. Modeli su testirani pomoću skupova podataka za detekciju objekata ImageNet i MS-COCO, a utjecaj različitih značajki proučavan je za sve različite modele. Rezultati procjena pokazuju da YOLOv4 nadmašuje sve ostale metode i u brzini i u preciznosti izmjerenoj u mAP (engl. *mean Average Precision* – mAP).

Za potpunu vizualizaciju pojedinog modela neuronske mreže moguće je koristiti NETRON aplikaciju koja je dostupna na GitHub platformi [272] u koju je dovoljno učitati konfiguracijsku (.cfg) datoteku mreže koju se želi vizualizirati.

Kao što je prethodno navedeno, YOLO je, kao i druge konvolucijske neuronske mreže sastavljen od slojeva i prvi sloj je ulazni koji uzima sliku kao ulaznu vrijednost i prosljeđuje ju dalje prema ostalim slojevima neuronske mreže, a izlaz je vektor koji predstavlja označavajući okvir (engl. *bounding box*) i pripadnost klasi (engl. *class prediction*). Arhitektura YOLO detektora slična je standardnim CNN mrežama, te sadrži konvolucijske i slojeve maksimalnog sažimanja (engl. *CONV i Max pooling layers*), nakon kojih slijede dva potpuno povezana sloja. Kao i drugi CNN algoritmi, YOLO koristi slojeve za stvaranje potpuno povezane CNN. Standardne CNN mreže sastavljene su od različitih slojeva pri čemu svaki sloj ima drugu ulogu, a rezultat obrade na pojedinom sloju prosljeđuje se sljedećem sloju. U tom kontekstu, konvolucijski sloj je "srce" konvolucijske neuronske mreže koji koristi konvolucijsku jezgru kao filter za prijelaz preko ulazne slike. Vrijednosti svakog piksela lokalno

koreliranih podataka unutar filtera se množe i dodaju kao rezultat konvolucije. Ovaj tip pravila se zove konvolucija s obzirom da omogućuje ekstrakciju značajki slike pomoću definirane konvolucijske jezgre. Pored toga, metoda filtriranja različitih dijelova slike istom jezgrom (engl. *kernel*) zove se dijeljenje težine (engl. *shared weights*), što omogućuje da neutralne ćelije u istoj značajki mogu biti prepoznate i klasificirane za isti tip objekta. Svaki filter je prostorno relativno mali (definiran je širinom i visinom) ali se proteže u ukupnu dubinu ulazne slike. Na primjer, tipičan filter na prvom sloju CNN mreže ima dimenzije 5x5x3, što znači da su visina i širina veličine pet piksela, a dubina je tri piksela s obzirom da slika ima dubinu tri, što označava broj kanala boja, odnosno u ovom slučaju je riječ o slici u boji (RGB), dok će crno-bijela slika imati dubinu jedan jer ima samo jedan kanal koji označava vrijednost intenziteta osvijetljenosti pojedinog piksela. U procesu prolaza unaprijed, svaki filter prolazi, odnosno konvoluirá, preko širine i visine ulaznih dimenzija i računa skalarni produkt između ulaza filtera i ulaza na lokaciji na kojoj se u tom trenutku nalazi. Kako filter prolazi preko ulazne slike, počevši od gornjeg lijevog kuta, kreirat će se dvodimenzionalna aktivacijska mapa koja odgovara filteru na bilo kojoj lokaciji u prostoru, odnosno na slici. Intuitivno, mreža će u prvom sloju naučiti filtere koji se aktiviraju kada "vide" neki tip vizualnih značajki kao što je orijentacija određenog ruba objekta, skup piksela iste vrijednosti boje ili one koji prepoznaju složenije oblike i geometrijske oblike na dubljim slojevima mreže. U konačnici će postojati skup filtera u svakom konvolucijskom sloju i svaki od njih će rezultirati stvaranjem zasebne dvodimenzionalne aktivacijske mape. [273]

Glavna uloga aktivacijskog sloja mreže je uvesti nelinearnost u rezultat konvolucijskog sloja i riješiti problem nestajućeg gradijenta koji se javlja zbog podtreniranosti. U ovom sloju se primjenjuju ranije opisane aktivacijske ili prijenosne funkcije, a kao najpopularnija u posljednje vrijeme se pokazala propusna ReLU funkcija (engl. *Leaky ReLU*), s obzirom da brže konvergira nego sigmoidna ili Tanh, te izračun provodi na jednostavniji i učinkovitiji način i ne sadrži "mrtve zone". Sažimajući sloj se uobičajeno primjenjuje za smanjivanje dimenzionalnosti rezultata konvolucijskog sloja, a stavlja se između dva konvolucijska sloja.

Glede sažimanja, uobičajeno se govori o tri tipa sažimanja: generalizirano sažimanje, (engl. *general pooling*), preklapajuće sažimanje (engl. *overlapping pooling*) i prostorno-piramidalno sažimanje (engl. *Spatial Pyramid Pooling (SPP)*) [274]. Međutim, YOLOv3 ne koristi sloj sažimanja, nego umjesto toga koristi konvolucijski sloj s korakom 2 za

smanjivanje dimenzionalnosti mape značajki, a zatim značajke invarijantne na veličinu prosljeđuje naprijed, što pomaže u sprečavanju gubitka značajki nižeg ranga [15]. Potpuno povezani slojevi su uobičajeno posljednji slojevi CNN mreže, a njihova zadaća je prenijeti rezultat na izlaz, kao i pojednostaviti te ubrzati obradu podataka. [274]. Za razliku od standardnih CNN mreža, YOLOv3 ne koristi potpuno povezani sloj kao posljednji sloj mreže, nego koristi *softmax* funkciju koja pretpostavlja da svaki pokazni okvir označava samo jednu klasu, što često nije slučaj, te se kao bolji pristup pokazuje multiklasni pristup [15]. YOLO je invarijantan na veličinu ulazne slike, te svaku ulaznu sliku skalira na unaprijed definiranu veličinu koja standardno iznosi 416x416 piksela, međutim ista se može mijenjati ovisno o potrebama i dimenzijama slika u skupu za učenje, stoga je preporučljivo u skupu za učenje imati sve slike jednakih dimenzija jer se u postupku implementacije mogu pojaviti različiti problemi ukoliko su velike oscilacije među dimenzijama slika, posebno ukoliko su znatno manjih dimenzija od standardno predviđene veličine. Jedan od većih problema koji se pojavljuje je to što, ukoliko postoji potreba za obradom slika u podskupovima ili miniskupovima (engl. *batches*), koji se mogu paralelno obrađivati na GPU što će značajno ubrzati proces učenja, sve slike trebaju biti jednakih dimenzija. Ovo je potrebno zbog ulančavanja slika u velike podskupove. YOLO detektor dimenzije ulazne slike smanjuje prema vrijednosti faktora koji se zove korak (engl. *stride*), pa primjerice ukoliko je vrijednost koraka 32, to znači da će ulazna slika dimenzija 416x416 biti smanjena na dimenzije 13x13. Općenito, korak u bilo kojem sloju mreže jednak je faktoru za koji je izlazna dimenzija slike manja od dimenzija ulazne slike. Vezano za spomenuti korak kao hiperparametar, potrebno je istaknuti njegovu važnost. Naime, prema definiciji, korak je broj piksela za koji se filter matrica, odnosno jezgra pomiče preko ulazne matrice, odnosno ulazne slike s obzirom da je slika zapravo matrica. Kada korak ima vrijednost jedan tada se filter pomiče za jedan piksel, kada je vrijednost dva tada filter preskače dva piksela kako prelazi preko ulazne slike. Veća vrijednost koraka stvara manju mapu značajki, odnosno mapu značajki manjih dimenzija [220]. S tim u svezi je bitno pojasniti i ulogu filtera odnosno kernela. Filter ili kernel je jedna od glavnih komponenti CNN mreže, a riječ je o kvadratnoj matrici dimenzija $n_K \times n_K$, kod koje je n_K obično relativno mali broj kao što je 3 ili 5 [275]. Jedan od velikih nedostataka konvolucije je mogućnost gubitka podataka koji se nalaze uz rub slike, zbog čega se uvodi postupak dopunjavanja (engl. *padding*), a provodi se tako da se, prije samog procesa konvolucije matricu okruži nulama uz svaki rub. Riječ je o jednostavnom, ali

učinkovitim postupku koji sprječava gubitak informacija koje se nalaze na rubu slike odnosno ulazne matrice, kako bi filter tijekom prolaza preko slike obuhvatio i rubove slike, a postupak ujedno omogućava kontrolu veličine mape značajki. [220]

6 Detekcija i raspoznavanje osoba u otežanim vremenskim uvjetima

Uvodno je istaknuto da su termalne kamere u današnje vrijeme sve raširenije i dostupnije, a samim tim i njihova primjena u području sigurnosti i zaštite osoba i imovine, odnosno nadzora štićenih prostora sve je učestalija. Pored navedenog, aktualno je upotreba termalnih kamera proširena i na kontrolu tjelesne temperature u kontekstu aktualne pandemije koronavirusa, te je samim tim povećana i potreba za novim, suvremenijim sustavima automatiziranog nadzora koji će djelovati samostalno ili kao dodatna pomoć pripadnicima vojske, policije i drugih državnih ili međunarodnih agencija zaduženih za sigurnost i zaštitu građana i imovine. Značaj termalnih kamera u spomenutim područjima primjene posebno dolazi do izražaja zbog mogućnosti njihove primjene po noći i u svim vremenskim uvjetima, što je posebno važno jer su u takvim uvjetima standardne optičke kamere gotovo beskorisne. S obzirom na razlike između značajki na standardnim (optičkim) RGB i termalnim slikama, u kontekstu istraživanja za doktorski rad cilj je bio istražiti učinkovitost metoda dubokog učenja u području detekcije i raspoznavanja osoba na termalnim snimkama, posebice u otežanim vremenskim uvjetima kao što su magla, kiša i snijeg. Nažalost, zbog određenih objektivnih otegotnih okolnosti, tijekom snimanja scenarija za eksperimentalni dio nije bilo moguće provesti testiranja po snježnim uvjetima, ali su, uz snimanje po vedrom vremenu, uspješno provedena i snimanja po gustoj magli i jakoj kiši, a iz dobivenih video zapisa kreirana je baza slika koja je korištena za treniranje i testiranje modela dubokog učenja. S obzirom na osnovni zadatak istraživanja, kreiran je i vlastiti skup podataka, odnosno spomenuta baza slika i video zapisa, a kroz scenarije snimanja pokušalo se maksimalno simulirati ilegalne migracije preko državne granice ili nedozvoljeno kretanje unutar drugih štićenih prostora, odnosno za detekciju i raspoznavanje osoba. Kreirani skup slika korišten je za učenje konvolucijske neuronske mreže u cilju dobivanja pouzdanog modela čija će primjena biti moguća u realnim uvjetima. Video zapisi su snimljeni u različitim terenskim uvjetima – šuma, livada, brisani prostor, zatim u različitim vremenskim uvjetima - vedro, kiša, magla, a osobe su prilikom snimanja nastojale simulirati kretanje ilegalnih migranata i drugih neovlaštenih osoba koje se kreću kroz štićeno, odnosno nadzirano područje.

Tako su se osobe kretale različitim brzinama (hodanje, trčanje), pri različitim položajima tijela (uspravno, pogureno, četveronoške, puzanje), a navedeni parametri brzine i načina hoda su međusobno kombinirani. Skup slika je kreiran i u cilju treniranja modela dubokog učenja za raspoznavanje osoba i životinja, odnosno da li je ili nije riječ o osobi (*engl. Human/Non-Human*), kao i u cilju ispitivanja mogućnosti modela dubokog učenja za raspoznavanje aktivnosti osoba, odnosno načina kretanja i položaja tijela prilikom kretanja. Za snimanje je korištena termalna kamera sa standardnim i teleobjektivom, sve s ciljem ispitivanja utjecaja opreme na kvalitetu termalnih snimki u različitim vremenskim uvjetima i na različitim udaljenostima od kamere. Iz video zapisa su najprije ekstrahirane sve sličice, a nakon vizualne analize, odabran je manji broj slika iz svakog pojedinog skupa – za vedro, kišno te maglovito vrijeme, kao i varijante skupova po udaljenostima, položajima tijela i sl. Za prve eksperimente kreiran je skup koji se sastoji od slika snimljenih po svim vremenskim uvjetima i na svim udaljenostima koji je sadržavao ukupno 70.189 slika, međutim kako je hod repetitivna aktivnost, prilikom vizualne analize slika postalo je razvidno da nije potrebno anotirati objekte (osobe) na cjelokupnom skupu, te su izabrane slike koje najbolje prikazuju različitosti između pojedinih sekvenci koraka, položaja tijela i sl., te je tako kreiran skup slika koji sadrži 7.412 ručno anotiranih slika, a snimljene su termalnom kamerom koja snima video zapise dugovalnog dijela IR spektra (LWIR). Ovaj skup uključuje slike snimljene u svim vremenskim uvjetima (vedro, kiša, magla), na svim udaljenostima na kojima je snimanje provedeno, sve položaje tijela, kao i slike snimljene standardnim i teleobjektivom, te je prikladno označen kao *All_In_One*, dok je u fazi treniranja i testiranja označen sa *t_All*. Pored navedenog, sveobuhvatnog skupa, kreirani su i zasebni skupovi slika za vedro vrijeme koji obuhvaća slike po vedrom vremenu, na svim udaljenostima i pri svim položajima tijela *Clear_All*, (označen kao *t_Clear*), kao i skupovi slika za kišno (*Rain_All*, odnosno *t_Rain*) te maglovito vrijeme (*Fog_All*, odnosno *t_Fog*). Također su kreirani skupovi slika za treniranje modela u cilju raspoznavanja, odnosno razlikovanja da li je riječ o čovjeku ili životinji, odnosno *Human-Non-Human* skup slika koji sadrži ručno anotirane slike na kojima se nalazi osoba, odnosno osoba i pas (*t_hNh*), te skup slika koji je korišten za treniranje modela za raspoznavanje aktivnosti (*Activity_All*, odnosno *t_Act*), koji se sastoji od slika na kojima su ručno označene različite aktivnosti osoba, odnosno brzine kretanja i položaji tijela, a o navedenim skupovima više će riječi biti u nastavku.

6.1 Snimanje i obrada slika

Kao što je već spomenuto, skup slika kreiran je za treniranje modela dubokog učenja za detekciju osoba u realističnim uvjetima kakvi se pojavljuju u kontekstu nadzora štićenih prostora i područja. Svi scenariji snimanja su provedeni po noći tijekom zimskog razdoblja po vedrom, kišnom i maglovitom vremenu. Snimane osobe su simulirale namjerno, ali ilegalno prelaženje kroz nadzirani prostor ili područja (državna granica, područje značajnih infrastrukturnih objekata – elektrane, zračne i pomorske luke, vodocrpilišta i sl.), zbog čega su tijekom snimanja mijenjali brzinu kretanja (hodanje, trčanje) i položaj tijela (uspravno, pogureno, četveronoške, puzanje). Baza slika prigodno je nazvana po Sveučilištu u Rijeci - **UNIRI_Thermal Dataset for Person Detection** ili skraćeno **UNIRI_Thermal**, odnosno u fazi testiranja označen je kao UNIRI-TID, a već je predstavljen i opisan u [107,106].

6.1.1 Kriteriji za kreiranje baze slika

Za kreiranje baze slika, bilo je potrebno zadovoljiti određene uvjete koji u najboljoj mjeri simuliraju realne uvjete u različitim vremenskim uvjetima i pri različitim brzinama i načinima kretanja.

6.1.1.1. Vedro vrijeme

Snimke snimljene po vedrom vremenu su primarno korištene za utvrđivanje referentne udaljenosti i općenito mogućnosti korištene termalne kamere i dodatne opreme za snimanje osoba. Vizualnom analizom, odnosno pregledom snimki za različite udaljenosti utvrđeno je da je udaljenost od 110 m optimalna te je postavljena za referentnu udaljenost, a zatim je snimanje nastavljeno na udaljenosti od 165 metara. Videozapisi, iz kojih su kasnije ekstrahirane slike, snimljeni po vedrom vremenu primarno su korišteni za treniranje modela za detekciju i raspoznavanje osoba, ali i raspoznavanje aktivnosti (trčanje, pogureno trčanje, pogureni hod, normalan hod, četveronoške, puzanje). Osobe su snimane u uspravnom i pogurenom hodu i pri različitim brzinama kretanja – trčanje i hodanje.

6.1.1.2. Gusta magla

Osnovni cilj snimanja po gustom magli bio je simuliranje otežanih vremenskih uvjeta te ocjena mogućnosti termovizijske kamere za snimanje po maglovitom vremenu s ciljem utvrđivanja maksimalne udaljenosti na kojoj je osoba na snimci vidljiva golim okom. Nakon početnih probnih snimki bilo je razvidno da gusta magla ima značajno izraženiji

negativni učinak na termalno snimanje nego što su noćni uvjeti pri vedrom vremenu. S tim u svezi treba naglasiti da na primarno određenoj referentnoj udaljenosti od 110 m osoba na snimci nije bila vidljiva golim okom, čak niti uz korištenje teleobjektiva, te su udaljenosti postupno smanjivane temeljem čega je utvrđeno da je referentna udaljenost za gustu maglu između 30 i 50 m. Slike i video zapisi snimljeni po gustoj magli korišteni su za testiranje YOLOv3 detektora, ali i kao dio skupa t_All , koji je korišten za treniranje modela koji će bolje generalizirati nego model treniran samo na slikama snimljenima po vedrom vremenu. Prilikom snimanja po magli, osobe su snimane u uspravnom i pogurenom hodu i pri različitim brzinama kretanja – trčanje i hodaње, a bio je prisutan i pas, te su snimke na kojima se nalaze i osoba i pas korištene za kreiranje skupa slika, te treniranje i testiranje modela za raspoznavanje Osoba – Ne-Osoba, odnosno *Human – Non-Human*.

6.1.1.3. Jaka kiša

Snimanjem po jakoj kiši također je bio cilj utvrditi maksimalnu udaljenost na kojoj kamera može snimiti osobu, a da bude vidljiva golim okom. Pokusne snimke su pokazale da kiša ima manji do gotovo nikakav utjecaj na termalno snimanje, te je korištena referentna udaljenost kao i po vedrom vremenu, a kako je promijenjena lokacija snimanja, bilo je moguće i povećati udaljenosti pa je tako snimanje provedeno na udaljenostima od 30 do maksimalnih 215 m, što je bila najveća udaljenost koju je dozvoljavala lokacija. Osobe su snimane u uspravnom i pogurenom hodu i pri različitim brzinama kretanja – trčanje i hodaње.

6.1.2 Prikupljanje podataka

Za snimanje je korištena dugovalna (LWIR) termalna kamera FLIR ThermaCam P10 prikazana na slici 28. [276], a snimanje je provedeno korištenjem standardnog objektiva kamere, ali i FLIR 7° teleobjektiva (P/B series) [277] prikazanog na slici 28.



Slika 29. Korištena termalna kamera i teleobjektiv

Kamera je bila postavljena na tronožac na visini od oko 140 cm. Rezolucija kamere je 320 x 240 piksela, zbog čega je korišten vanjski digitalni snimač koji je imao mogućnost automatskog povećanja rezolucije snimke koja je u tom postupku povećana na 1280 x 960 piksela, bez značajnih gubitaka na snimkama. Za mjerenje udaljenosti korištena je mobilna aplikacija ViewRanger [278] instalirana na pametnom telefonu CAT S60 [279]. Naknadno je snimljena ruta kretanja za vrijeme kretanja prenesena u aplikaciju Google Earth [280].

Ovdje je potrebno istaknuti da snimanje nije proteklo kako je prvotno planirano i to uglavnom zbog tehničkih ograničenja raspoložive opreme. Naime, korištena termalna kamera je model star 15-ak godina, zbog čega je i izvorna rezolucija relativno niska. Starost kamere je utjecala i na trajnost baterija, ali i samu funkcionalnost. Naime, na raspolaganju su bile dvije baterije, međutim zbog starosti njihov kapacitet je bio značajno smanjen te je sa obje baterije na raspolaganju bilo samo oko pola sata za snimanje, nakon čega je slijedilo punjenje više od 12 sati, a u međuvremenu su se značajno promijenili vremenski uvjeti. Osim toga, postojao je i problem sa zapisom na SD karticu, zbog čega je bilo nužno snimanje na vanjski digitalni snimač, te je usput napravljeno i povećanje rezolucije, ali zbog toga nije bilo mogućnosti za snimanje u originalnoj rezoluciji, te samim tim niti naknadno izvođenje usporedbe detekcije na originalnoj i poboljšanoj rezoluciji snimke. Pored navednog, otegotna okolnost su bili i vremenski uvjeti u kojima je snimanje izvedeno, što je bilo posebno izraženo za snimanje po magli, koja je na lokaciji snimanja (Hrvatska, Pula, Marsovo polje) bila prisutna tek oko sat vremena, te se u takvom obliku i gustoći naknadno nije pojavljivala više od godinu dana, a u međuvremenu je kameru trebalo vratiti vlasnicima s obzirom da je ista posuđena. Slično je bilo s kišnim uvjetima, s obzirom da je primarni cilj bio

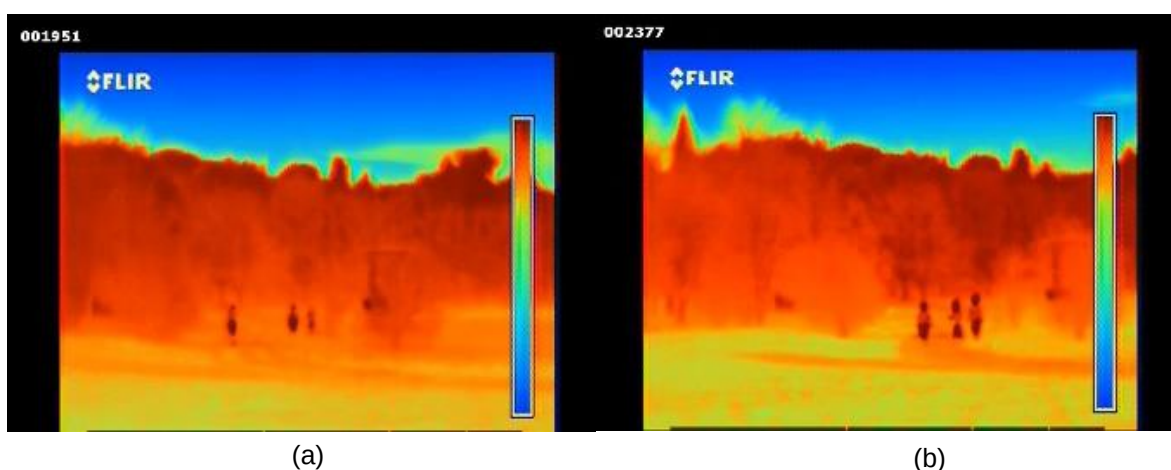
provesti snimanje po jakom pljusk u noćnim uvjetima, snimanje je uspješno provedeno do maksimalnog trajanja baterija. U planu, odnosno predviđenim scenarijima, bilo je i snimanje u snježnim uvjetima, međutim na lokaciji snimanja snijeg je rijetka pojava, a tijekom razdoblja raspoloživosti kamere, zbog relativno tople zime snijeg je bio rijetka pojava i u kontinentalnim krajevima Hrvatske, zbog čega se u konačnici odustalo od simulacije i snimanja u snježnim uvjetima. Unatoč tomu, snimljeni video zapisi predstavljali su dovoljnu količinu materijala za daljnju obradu i pripremu baze slika za izvođenje eksperimenata.

6.1.2.1. Snimanje po vedrom vremenu

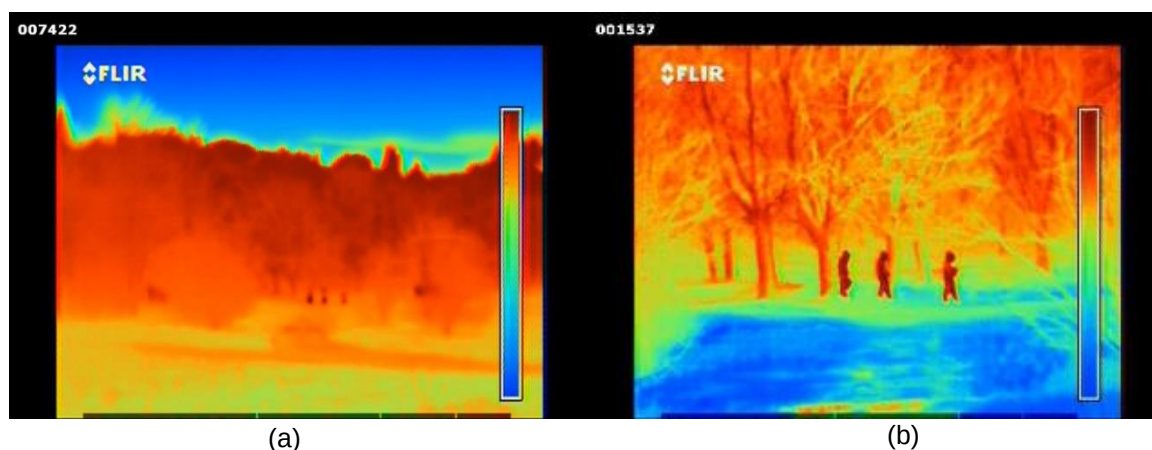
Snimanje po vedrom vremenu provedeno je na lokaciji Marsovo polje u Puli (slika 30), a riječ je o livadi s malo šume i niskog raslinja, zbog čega je bilo moguće simulirati uvjete kada se osobe skrivaju, a cilj je, između ostalog, bio i "koliko termalna kamera može vidjeti kroz šumu i grmlje". Dužina lokacije je oko 180 m, temperatura zraka je bila oko 2°C, a snimanje je provedeno po noći uz dobru vidljivost. Lokacija je, između ostalog, izabrana i zbog činjenice da nije "zagađena" uličnom rasvjetom što je predstavljalo dodatnu pogodnost u cilju simulacije realnih uvjeta na teško pristupačnim područjima. Najprije je provedeno snimanje osobe na udaljenosti od 110 m pri normalnom hodu, pojedinačno i u grupi od tri osobe (slika 30), korištenjem standardnog objektiva kamere. Nakon toga je slijedilo snimanje jedne i tri osobe prilikom njihovog udaljavanja od kamere (od 50 do 110 m), a zatim prolaz kroz vidno polje kamere, na udaljenostima od 110 i 165 m. U ovom scenariju, snimane osobe su mijenjale položaj tijela (uspravno, pogureno), a zatim samo uspravno na udaljenostima od 110 do 165 m. Osim korištenjem standardnog objektiva, snimanje je provedeno i korištenjem ranije opisanog teleobjektiva. Korištenjem standardnog objektiva osobe su snimane na udaljenosti od 110 m, pri normalnom hodu, pojedinačno i u grupi od tri osobe, kao i prilikom trčanja. U istom scenariju, osobe su snimane i korištenjem teleobjektiva. Nakon toga snimanje je nastavljeno na udaljenosti od 165 m, korištenjem standardnog objektiva prilikom prolaska kroz vidno polje kamere pri normalnoj brzini hoda uspravno, trčeći i pogureno (normalan hod i trčanje). Ukupno je snimljeno oko 20 minuta video materijala, 25 FPS, odnosno nakon ekstrakcije sličica na raspolaganju je bilo oko 24.448 slika za daljnju obradu.



Slika 30. Prikaz lokacije i kretanja prilikom snimanja po vedrom [106]



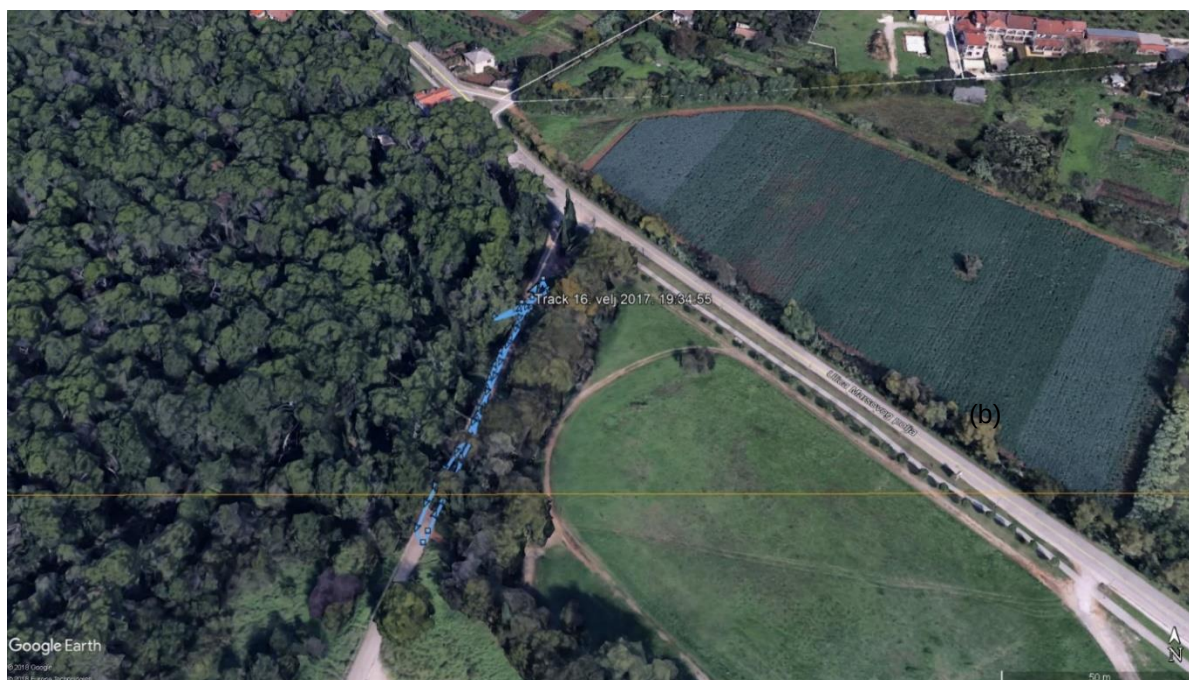
Slika 31. Usporedni prikaz osoba snimljenih termalnom kamerom po vedrom vremenu na udaljenosti od 110 m: (a) standardni objektiv, (b) teleobjektiv



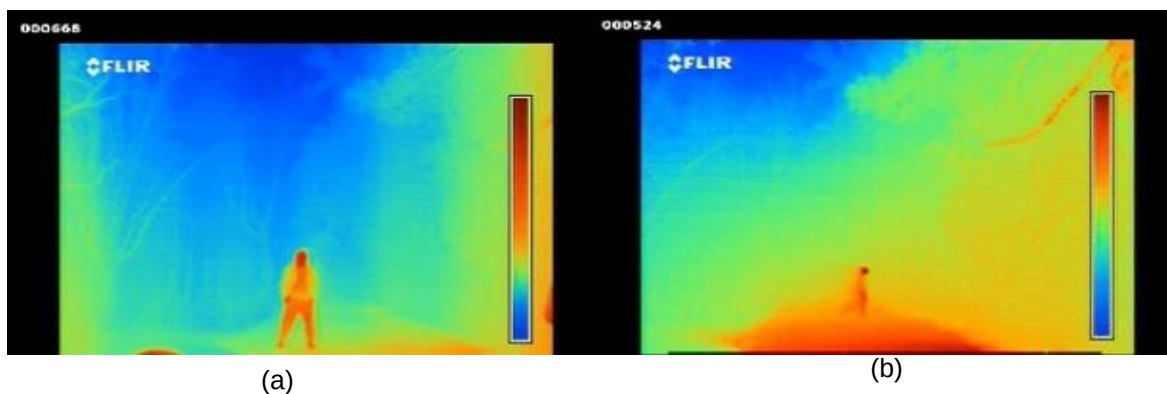
Slika 32. Usporedni prikaz osoba snimljenih termalnom kamerom po vedrom vremenu na udaljenosti od 165 m: (a) standardni objektiv (b) teleobjektiv

6.1.2.2. Snimanje po gustoj magli

Snimanje po gustoj magli provedeno je na istoj lokaciji kao i snimanje po vedrom vremenu, osim što je izmijenjena putanja kojom su se kretale snimane osobe (slika 33). Zbog gustoće vodenih kapljica koje čine maglu, dugovalno toplinsko zračenje je značajno disperzirano zbog čega je i vidljivost za termalnu kameru znatno smanjena u odnosu na druge vremenske uvjete (vedro, kiša). Snimanje po gustoj magli je bilo ujedno i najzahtjevnije, kao što je opisano i u [281, 66]. Zbog takvih vremenskih uvjeta bilo je nužno odustati od primarno predviđenog scenarija i snimanja na referentnoj udaljenosti od 110 m, kao po vedrom vremenu, jer isto nije bilo moguće s raspoloživom opremom. Naime, osobe na udaljenostima većim od 50 m uopće nisu bile vidljive na snimkama, zbog čega su nakon prethodne vizualne analize, odabrane udaljenosti za snimanje od 0 do 30 m i na 50 m, isključivo korištenjem teleobjektiva (slika 33). Snimane su do tri osobe koje su mijenjale smjer kretanja i položaj tijela. Snimanje je provedeno u šumovitom predjelu na asfaltnoj cesti, s obzirom da zbog blata i guste magle, volonteri nisu mogli hodati po prethodnoj predviđenoj lokaciji, iako je ista u neposrednoj blizini. Temperatura zraka također je bila oko 2°C, a zbog guste magle vidljivost golim okom je bila minimalna (manje od pet metara). Ukupno je snimljeno oko 15 minuta video zapisa pri brzini 25 FPS, odnosno nakon ekstrakcije bilo je ukupno oko 22.737 slika.



Slika 33. Prikaz lokacije i putanje kretanja za snimanje po gustoj magli



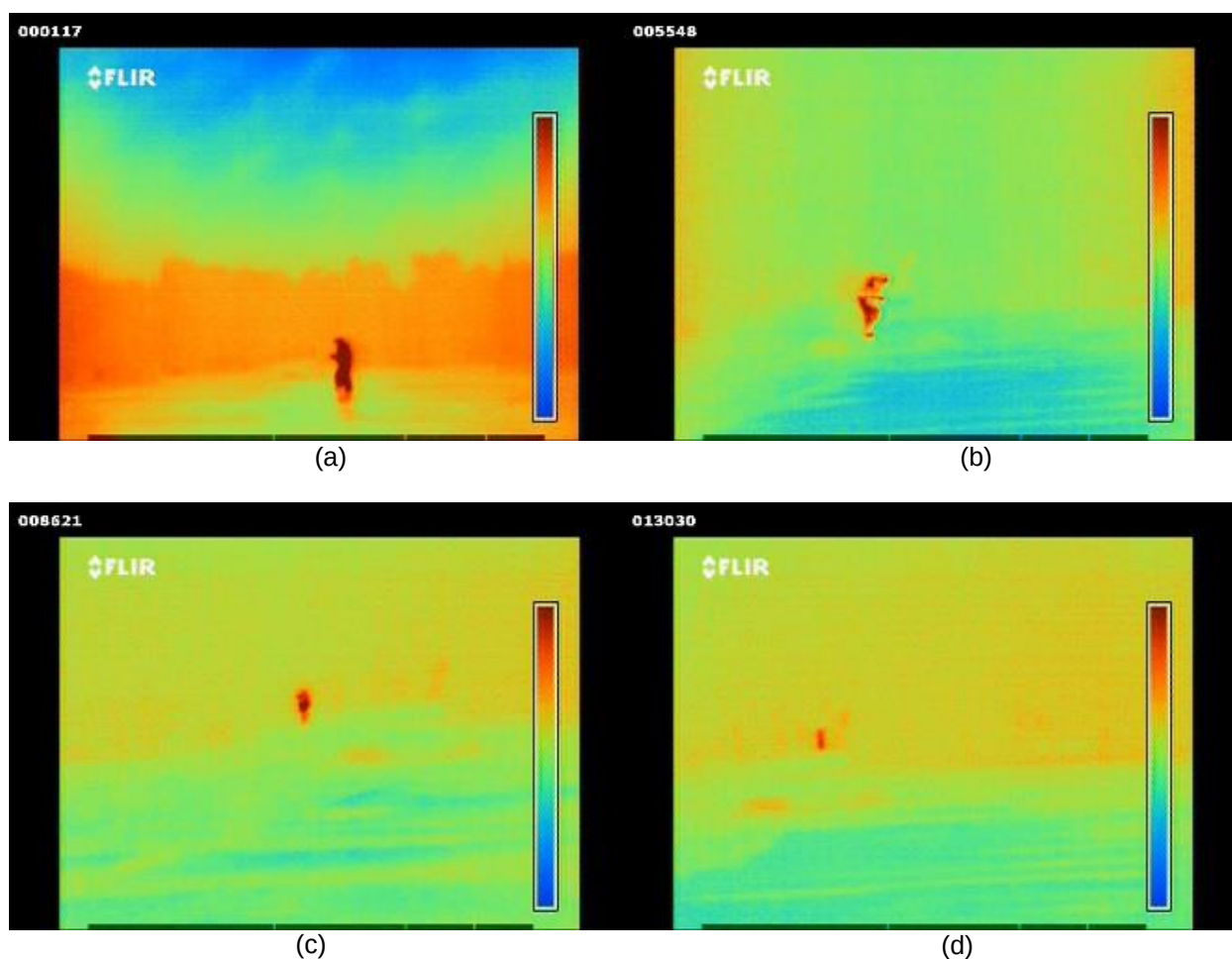
Slika 34. Termalna snimka osobe u gustom magli: (a) teleobjektiv na 30 m, (b) teleobjektiv na 50 m

6.1.2.3. Snimanje po jakoj kiši

Za snimanje po jakoj kiši odabrana je nova lokacija (Hrvatska, Pula, Hidrobaza), na kojoj je bilo moguće snimanje na udaljenostima većim od 165 metara (slika 35). S obzirom da je jaka kiša potrajala, bilo je moguće izvesti snimanje na različitim udaljenostima, iznad i ispod referente udaljenosti od 110 m. Stoga je snimanje izvedeno na udaljenostima od 30, 70, 110, 140, 170, 180 i 215 m (slika 35). U ovom scenariju su snimane dvije osobe na svim udaljenostima, pojedinačno ili zajedno pri normalnom hodu i trčanju uspravno i pogureno. Zbog izrazito velike količine vode na tlu, nije bilo moguće simulirati četveronoški hod i puzanje. Snimljeno je ukupno oko 15 minuta video materijala, pri 25 FPS, odnosno ukupno 23.004 slike su bile na raspolaganju za daljnju obradu.



Slika 35. Prikaz lokacije za snimanje po jakoj kiši

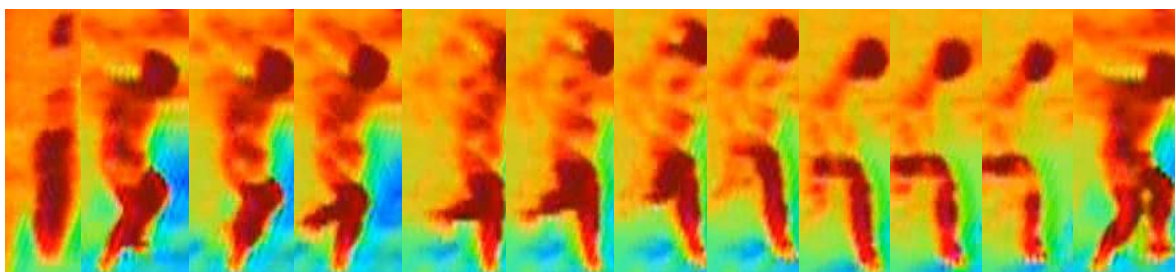


Slika 36. Termalna snimka osobe po jakoj kiši: (a) teleobjektiv na 30 m, (b) teleobjektiv na 70 m (pogureno); (c) teleobjektiv na 140 m; (d) teleobjektiv na 215 m

6.2 Pretporcesiranje snimljenog materijala

Nakon snimanja po ranije opisanim scenarijima bilo je potrebno napraviti vizualnu analizu video materijala te kroz predobradu izdvojiti odgovarajuće dijelove videa koji će u najboljoj mjeri prikazivati simulaciju stvarnih uvjeta kakvi se mogu očekivati na terenu odnosno teško pristupačnim područjima i drugim štićenim prostorima. Na raspolaganju je bilo ukupno oko 20 minuta video zapisa snimljenih po vedrom vremenu, oko 15 minuta po gustoj magli te oko 15 minuta po jakoj kiši. Osim izrezivanja odgovarajućih dijelova video zapisa, materijal je trebalo dodatno obraditi u cilju pripreme baze slika za treniranje modela za detekciju i raspoznavanje osoba. Za pregled i izrezivanje video zapisa korišten je besplatni softverski alat VSDC Video Editor's [282]. Nakon izrezivanja odgovarajućih dijelova video zapisa, slijedila je ekstrakcija odgovarajućih sličica (engl. *frames*), za što je korišten MATLAB [283]. Konačni rezultat je bio ukupno 13.127 sličica rezolucije 1280 x 960, za snimke po vedrom vremenu na svim udaljenostima, u svim položajima tijela te brzinama kretanja.

Iz video zapisa snimljenih po gustoj magli, dobiveno je ukupno 5.189 slika za snimke na obje udaljenosti i sve položaje tijela. Za kišne snimke, dobiveno je ukupno 7.031 sličica za sve udaljenosti i sve položaje tijela. Za treniranje modela kod nadziranog učenja, prethodno je potrebno anotirati objekte na sličicama i dodijeliti im odgovarajuću klasu, a to je potrebno učiniti kako za trening skup, tako i za testni skup. Stoga je sljedeća faza bila anotacija objekata odnosno svih osoba na svakoj slici koja je odabrana za skup za učenje. Ova faza je spora i zahtjevna jer je riječ o ručnom označavanju svakog objekta, odnosno u ovom slučaju, svake osobe koja se nalazi na slici. Treba također napomenuti da nisu sve sličice iz pripremljenih video zapisa korištene za pripremu skupa za učenje jer nema potrebe za tim s obzirom da je hod repetitivna aktivnost koja se svodi na jedan ciklus koraka (slika 37).



Slika 37. Ciklus koraka prilikom trčanja, vedro vrijeme, udaljenost 165 m, snimljeno korištenjem teleobjektiva.

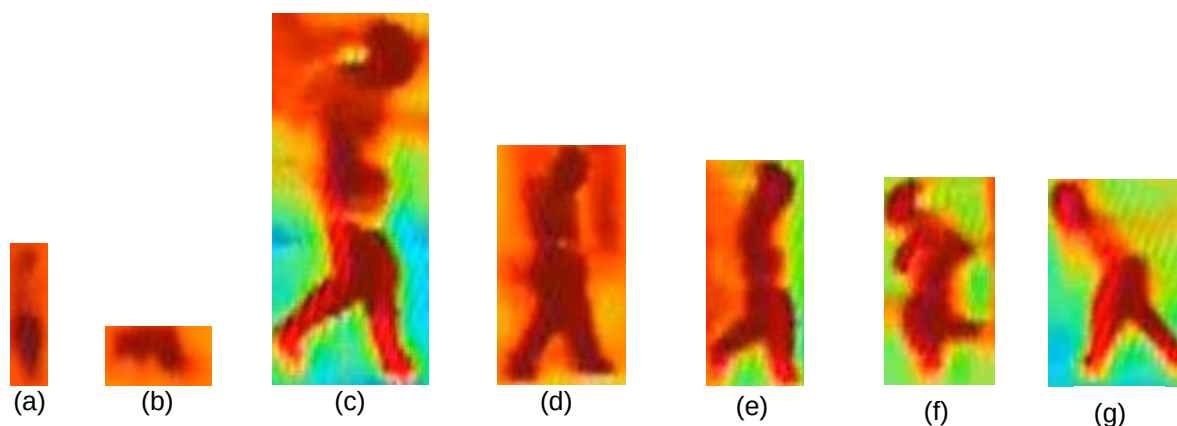
U konačnici su kreirani skupovi slika, a detaljni podaci o svakom prikazani su u tablici 2. Nadalje, osim bazne verzije skupa slika koja sadrži samo RGB slike, kreirane su i dvije dodatne baze slika koje sadrže slike iz baznog skupa konvertirane u slike sivih nijansi i to u dvije verzije. Prva verzija topla područja prikazuje svijetlim nijansama prema bijeloj, a druga topla područja prikazuje tamnim nijansama, odnosno prema crnoj boji. Potonje baze su kreirane za potrebe usporednog testiranja treniranih modela za testiranje na vanjskim skupovima slika koji sadrže slike sivih nijansi, a ovisno o pojedinom skupu, u nekima su topla područja prikazana svijetlim nijansama, a u nekima tamnim.

Tablica 2. Broj slika u kreiranoj bazi slika za treniranje modela dubokog učenja

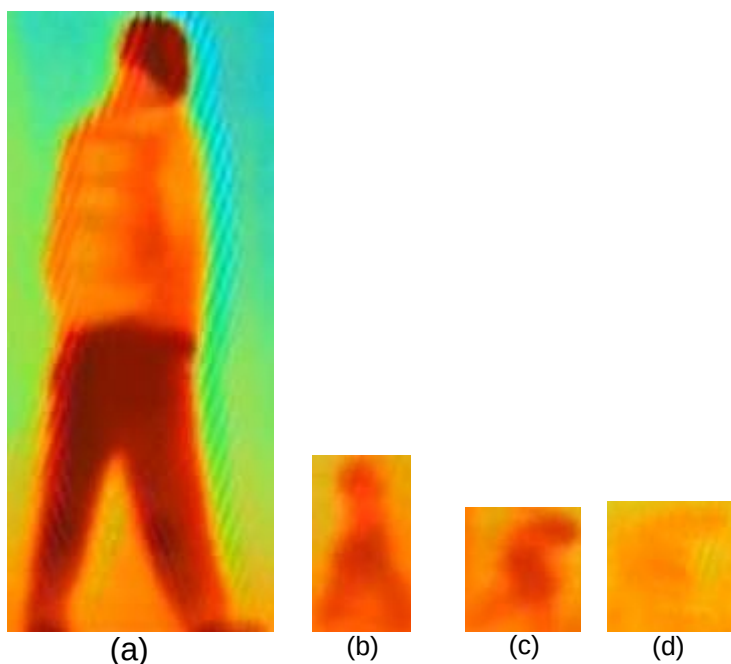
Naziv skupa	Broj slika (RGB), izravno ekstrahirano iz video zapisa	Broj slika (RGB), odabrano za skup za učenje	Broj slika (RGB + Grey-scale, obje verzije)
All_In_One (t_All)	23.835	6.340	19.008
Clear weather (t_Clear)	11.900	2.700	15.980
Foggy weather (t_Fog)	4.905	1.136	3.408
Rain (t_Rain)	7.030	2.541	7.623
Human-NonHuman (t_HnH)	5.419	5.419	16.257
Activity recognition (t_Act)	5.946	5.946	17.838

Upravo zbog toga su izabrane slike koje će najbolje prikazivati razlike među pojedinim fazama u ciklusu koraka u odnosu na položaj tijela, nogu i ruku. U tom kontekstu, za izbor odgovarajućih sličica najprije je provedena vizualna analiza svih slika, a potom su odabrane određene slike prema prethodno opisanom kriteriju. Nakon ove faze, bazni skup odabranih slika za ručno označavanje objekata, odnosno osoba na slikama smanjen je na ukupno 6.340 slika, od čega su 2.700 slika snimljene po vedrom vremenu, 1.136 po gustoj magli i 2.541 po jakoj kiši. Također su kreirana i dva dodatna skupa, jedan u cilju treniranja modela za raspoznavanje da li je riječ o osobi ili životinji (engl. *Human – Non-Human*) koji sadrži slike snimljene po vedrom vremenu i magli, na svim udaljenostima, svim položajima tijela i brzinama kretanja, te se sastoji od ukupno 5.419 slika, od čega je na 2.685 slika prikazana samo osoba, na 1.496 slika su prikazani osoba i pas, a dodano je i 1.238 negativnih primjera, odnosno slika koje ne sadrže objekte. Skup slika za raspoznavanje aktivnosti sadrži slike po svim vremenskim uvjetima, pri svim položajima tijela i brzinama kretanja, te se sastoji od 5.946 slika, na kojima su prikazane osobe označene različitim klasama aktivnosti: normalan hod (engl. *Human Normal Walk*), trčanje (engl. *Human Running*), pogureni hod (engl. *Human Hunched*), ležanje na tlu (engl. *Human Lying On The Ground*), te četveronoški hod (engl. *Human Fourleg Walking*). Ovdje treba napomenuti da je u prvoj verziji skupa bila i zasebna klasa "pogureno hodaње", međutim kako je položaj tijela prilikom pogurenog trčanja i pogurenog hoda gotovo identičan, nije bilo moguće zasebno klasificiranje ova dva načina kretanja, te je sve objedinjeno u jednu klasu "pogureno trčanje".

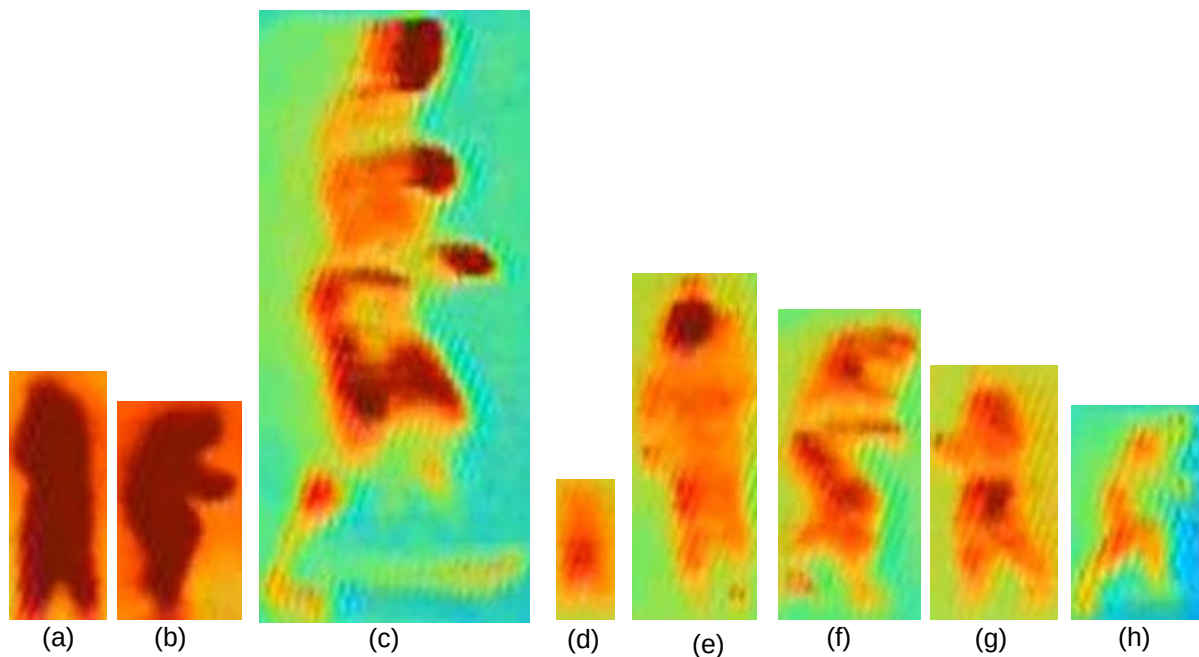
Nadalje, osobe na slikama, odnosno termalne siluete su relativno male, a dimenzije označenih okvira (engl. *bounding box*) za siluete na pojedinim udaljenostima prikazane su na slikama od 37. do 40. Također je važno napomenuti da su označavajući okviri većih dimenzija nego sama termalna silueta s obzirom da obuhvaćaju raspon između najudaljenijih točaka na silueti, kako po visini tako i po širini.



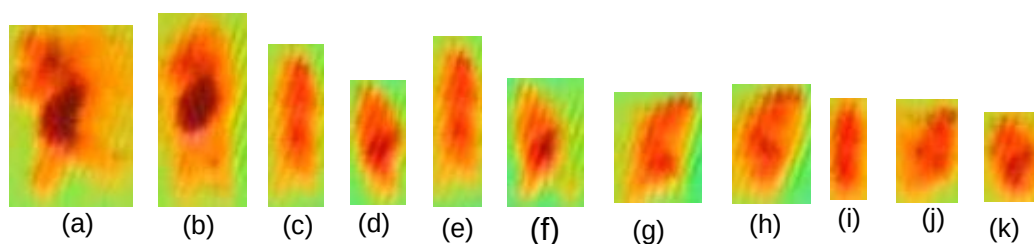
Slika 38. Termalna silueta: (a) udaljenost 110 m, vedro, hodanje, standardni objektiv; (b) 110 m, vedro, standardni objektiv, četveronoški hod; (c) 110 m, vedro, teleobjektiv, trčanje; (d) 165 m, vedro, hodanje, teleobjektiv; (e) 165 m, vedro, teleobjektiv, trčanje; (f) 165 m, pogureno trčanje, teleobjektiv; (g) 165 m, pogureno hodanje, teleobjektiv



Slika 39. Termalna silueta: (a) udaljenost 30 m, magla, hodanje, teleobjektiv; (b) 50 m, magla, teleobjektiv, hodanje; (c) 50 m, magla, teleobjektiv, pogureno hodanje; (d) 50 m, magla, pogureno trčanje, teleobjektiv (magla se dodatno pojačala)



Slika 40. Termalna silueta: (a) udaljenost 30 m, kiša, hodaње, standardni objektiv; (b) 30 m, kiša, standardni objektiv, pogureno hodaње; (c) 30 m, kiša, teleobjektiv, trčanje; (d) 70 m, kiša, hodaње, standardni objektiv; (e) 70 m, kiša, teleobjektiv, hodaње; (f) 70 m, kiša, teleobjektiv, pogureno hodaње; (g) 110 m, kiša, teleobjektiv, hodaње; (h) 110 m, kiša, teleobjektiv, pogureno hodaње



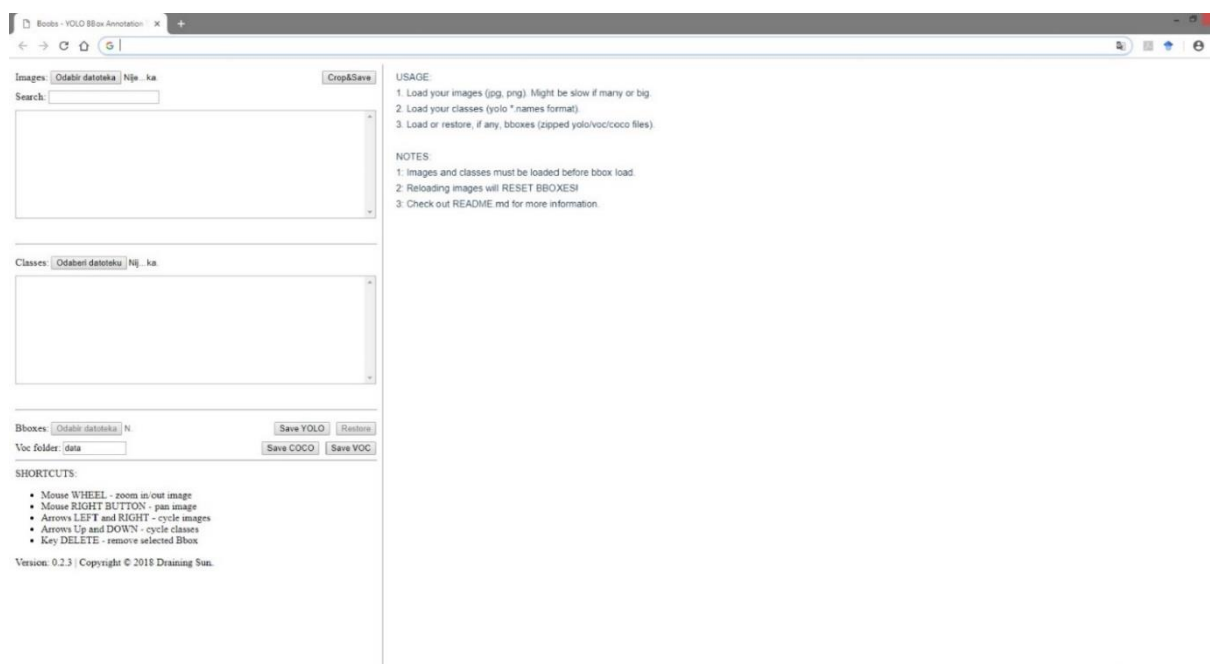
Slika 41. Termalna silueta: (a) udaljenost 140 m, kiša, pogureno hodaње, teleobjektiv; (b) 140 m, kiša, teleobjektiv, hodaње; (c) 170 m, kiša, teleobjektiv, hodaње; (d) 170 m, kiša, pogureno hodaње, teleobjektiv; (e) 180 m, kiša, teleobjektiv, hodaње; (f) 180 m, kiša, teleobjektiv, trčanje; (g) 180 m, kiša, teleobjektiv, pogureno hodaње; (h) 180 m, kiša, teleobjektiv, pogureno trčanje; (i) 215 m, kiša, teleobjektiv, hodaње; (j) 215 m, kiša, teleobjektiv, trčanje; (k) 215 m, kiša, teleobjektiv, pogureno

Tablica 3. Dimenzije označavajućih okvira na slikama ovisno o udaljenosti snimanja i korištenom objektivu kamere

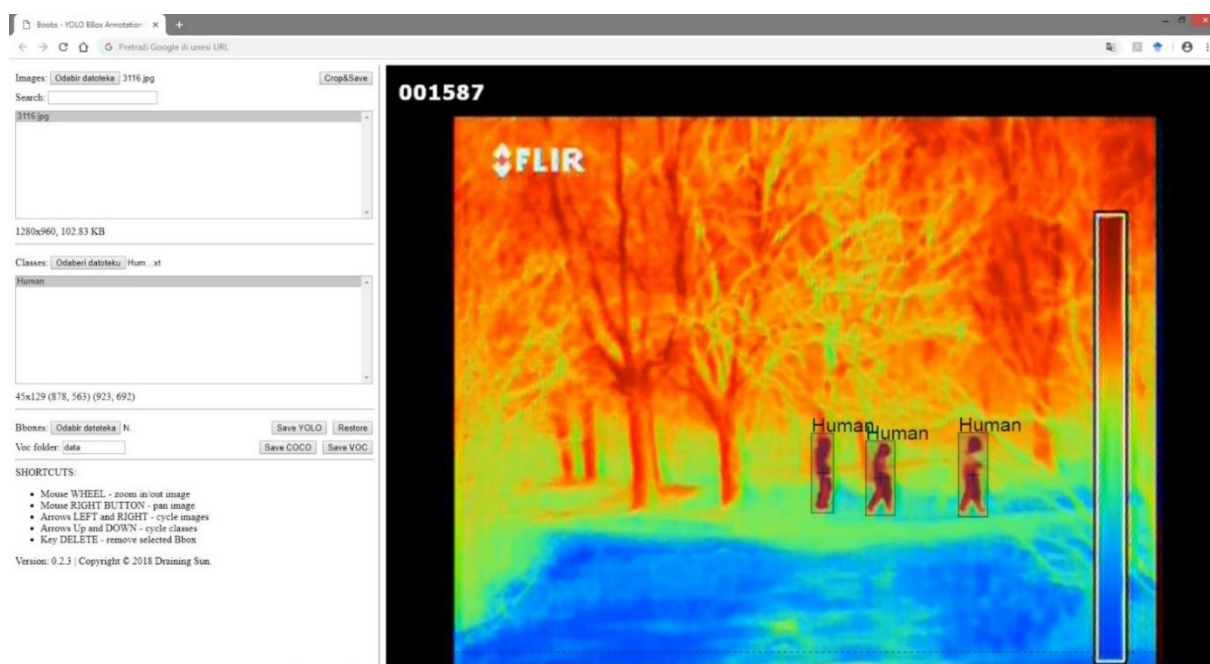
Udaljenost snimanja (m) i vremenski uvjeti	Položaj tijela	Korišteni objektiv	Prosječna veličina termalne siluete za osobu visine 185 cm (izraženo u pikselima)	
Vedro			širina	visina
110	Hodanje	Standardni	od 22 do 30	cca 70
110	Trčanje	Teleobjektiv	od 40 do 62	cca 180
165	Hodanje	Teleobjektiv	od 32 do 50	cca 120
165	Trčanje	Teleobjektiv	od 44 do 60	cca 100
165	Pogureno hodanje	Teleobjektiv	od 43 do 113	cca 100
Magla				
30	Hodanje	Teleobjektiv	od 73 do 121	cca 300
50	Hodanje	Teleobjektiv	od 40 do 60	cca 150
50	Pogureno hodanje	Teleobjektiv	od 43 do 66	cca 65
50	Trčanje	Teleobjektiv	od 48 do 68	cca 100
50	Pogureno trčanje	Teleobjektiv	od 45 do 66	cca 65
Kiša				
30	Hodanje	Standardni	od 54 do 64	cca 140
30	Pogureno hodanje	Standardni	od 64 do 69	cca 116
30	Trčanje	Teleobjektiv	od 97 do 125	cca 290
70	Hodanje	Standardni	od 24 do 29	cca 70
70	Hodanje	Teleobjektiv	od 66 do	cca 160
70	Pogureno hodanje	Standardni	od 23 do 30	cca 58
70	Pogureno hodanje	Teleobjektiv	od 61 do 87	cca 154
110	Hodanje	Teleobjektiv	od 37 do 68	cca 132
110	Pogureno hodanje	Teleobjektiv	od 37 do 59	cca 120
140	Hodanje	Teleobjektiv	od 35 do 40	cca 95
140	Pogureno hodanje	Teleobjektiv	od 42 do 56	cca 90
170	Hodanje	Teleobjektiv	od 26 do 35	cca 80
170	Pogureno hodanje	Teleobjektiv	od 28 do 38	cca 66
180	Hodanje	Teleobjektiv	od 28 do 36	cca 80
180	Pogureno hodanje	Teleobjektiv	Od 30 do 34	cca 70
180	Trčanje	Teleobjektiv	od 26 do 28	cca 73
180	Pogureno trčanje	Teleobjektiv	cca 38	cca 60
215	Hodanje	Teleobjektiv	od 21 do 25	cca 50
215	Trčanje	Teleobjektiv	od 27 do 30	cca 39
215	Pogureno hodanje	Teleobjektiv	od 27 do 30	cca 40

Temeljem podataka prezentiranih u tablici 3., razvidno je da postoji značajna razlika u vidljivosti i dimenzijama termalne siluete ovisno o vremenskim uvjetima, korištenju standardnog ili teleobjektiva, kao i utjecaj udaljenosti snimanja. S tim u svezi treba napomenuti da je najveći utjecaj magle na vidljivost termalne siluete (što je vidljivo na slici 38. (d)), dok je utjecaj kiše gotovo zanemariv u odnosu na snimanje po vedrom vremenu. Utjecaj udaljenosti je vidljiv u tome što se s udaljenošću termalna silueta smanjuje, posebice kada se mijenja položaj tijela, s obzirom da se u tom slučaju, povećanjem udaljenosti umanjuje vidljivost samog položaja tijela s obzirom da se pojedini dijelovi tijela, posebice ekstremiteti i glava stapaju s trupom, zbog čega termalna silueta gubi svoje karakteristike tipične za osobu, te se oblikom približava nepravilnom eliptoidu što svakako može utjecati i na rezultate učenja modela za detekciju i raspoznavanje osoba. Jedan od načina za smanjivanje ovog utjecaja na vidljivost termalne siluete je uporaba teleobjektiva, posebice za snimanje na većim udaljenostima. Također je vidljiv utjecaj magle na vidljivost termalne siluete, što je posebno razvidno na slici 38. (c) i (d), na kojima je razlika u vidljivosti termalne siluete vidljiva i vizualnim pregledom slike. Nažalost, u ovom slučaju, kako je na raspolaganju bio stariji model kamere, unatoč korištenju teleobjektiva, termalna silueta na udaljenosti od 215 m je izuzetno malih dimenzija, gotovo eliptičnog oblika, što otežava raspoznavanje i vizualnom analizom. Prikazane termalne siluete su originalnih dimenzija kao i na izvornim slikama dimenzija 1280 x 960 piksela.

Kao što je već spomenuto, anotacija i dodjela odgovarajućih klasa objektima je vremenski najzahtjevnija faza u pripremi baze slika. Za anotaciju slika postoji široka paleta aplikacija kao što je VGG Annotator [284], koji se pokreće iz internetskog preglednika i nudi nekoliko korisnih mogućnosti, zatim Labelbox [285], LabelMe [286], Annotorious [287]. Odabir alata za anotaciju ovisi o metodi dubokog učenja koja će biti korištena za treniranje modela s obzirom da različite metode koriste različite formate anotacije objekta. U ovom slučaju, odabran je alat dostupan u GitHub repozitoriju koji se zove *Yolo BBox Annotation Tool* [288]. Slika 42., prikazuje korisničko sučelje navedenog alata, a slika 43., prikazuje isto sučelje s učitanim slikom na kojoj su objekti označeni pokaznim okvirom, što su u ovom slučaju osobe na termalnoj slici.



Slika 42. Korisničko sučelje anotacijskog alata Yolo BBox Annotation Tool



Slika 43. Korisničko sučelje anotacijskog alata Yolo BBox Annotation Tool s učitanom slikom i označenim objektima (osobama)

TU SAM STAO Kao i drugi slični anotacijski alati, i ovaj se pokreće unutar internetskog preglednika i podržava sve najraširenije preglednike (Firefox, Chrome, Safari, Opera). Ključni razlog zašto je ovaj alat odabran za anotaciju u ovom slučaju je mogućnost da istovremeno sprema anotacije objekata u tri najraširanija anotacijska formata za duboko učenje - YOLO [13, 289], VOC [290] i MSCOCO [291]. Ova mogućnost štedi vrijeme za kasnije faze jer nije potrebna naknadna konverzija u druge anotacijske

formate u slučaju potrebe za učenje modela na metodama koje ne koriste YOLO format, kao što je u ovom slučaju, te su za svaki slučaj i eventualnu kasniju potrebu, sve anotacije spremljene u sva tri navedena formata, ali je za potrebe ovog doktorskog rada korišten samo YOLO format s obzirom da je za detekciju i rasponavanje osoba korišten YOLOv3 detektor [15], a rezultati su dijelom ranije već predstavljani u [104, 105, 106]. YOLO format anotacije na slikama sprema u kao tekstualnu (.txt) datotetku i to za svaku sliku zasebno tako da na kraju postoji onoliko datoteka s anotacijama koliko je slika u pojediniom skupu, odnosno bazi slika. Anotacija u YOLO formatu sastoji se od pet oznaka s tim da svaki red u .txt datoteci sadrži vrijednosti vezane za svaki pojedini objekt na slici (novi red – novi objekt), slika 44. Prva vrijednost je cijeli broj i označava klasu objekta (engl. *object class*), a ukoliko je više klasa objekata na slici, tada se one označavaju redoslijedno brojevima počevši od 0 (prva klasa je označena sa 0, sljedeća sa 1...). Druga vrijednost u formatu predstavlja koordinate centra označenog objekta po X-osi (engl. *object center in X*). Treća vrijednost predstavlja koordinate centra objekta po Y-osi (engl. *object center in Y*). Četvrta vrijednost predstavlja širinu objekta po X osi (engl. *object width in X*), a peta predstavlja visinu objekta po Y osi (engl. *object width in Y*) [292]. Ukoliko je na slici više anotiranih objekata tada se navedene brojčane vrijednosti za svaki objekt navode u novom redu unutar iste tekstualne datoteke koja je vezana na pojedinu sliku, s tim da, ukoliko objekti pripadaju istoj klasi ispred vrijednosti za svaki objekt ostaje ista brojčana oznaka pripadnosti klase. Na primjer, na slici 44., su prikazane vrijednosti za tri anotirane osobe na slici, a s obzirom da pripadaju istoj klasi (osoba) brojčana oznaka klase nije promijenjena, za razliku od slike... na kojoj su prikazane anotacije za objekte različitih klasa (Human/non-Human).

```
0 0.703515625 0.6541666666666667 0.03203125 0.13125
0 0.5921875 0.6578125 0.034375 0.11770833333333333
0 0.523828125 0.6520833333333333 0.02578125 0.12291666666666666
```

Slika 44. YOLO format anotacija za tri označena objekta na slici koji pripadaju istoj klasi

```
0 0.5015625 0.8255208333333334 0.0515625 0.26145833333333335
0 0.60859375 0.8510416666666667 0.05625 0.2375
1 0.348828125 0.946875 0.11640625 0.1125
```

Slika 45. YOLO format anotacija za tri označena objekta na slici koji pripadaju dvjema različitim klasama (Human – Non-Human)

Nakon završetka procesa anotacije osoba na slikama, sljedeći korak je učenje modela za detekciju i raspoznavanje osoba. Za tu fazu je odabran YOLO detektor, točnije njegova treća verzija [15] kao jedan od novijih detektora baziranih na konvolucijskoj neuronskoj mreži koji postiže vrhunske rezultate u detekciji i klasifikaciji objekata. U ovom slučaju, YOLO detektor je korišten u okviru implementacije Darknet [293]. Osim YOLOv3, korišten je i YOLOv4, s obzirom da je u međuvremenu navedena verzija također postala dostupna, te su međusobno uspoređeni rezultati YOLOv3 i YOLOv4.

6.3 Opis modela, korišteni hiperparametri

Uvodno će biti pojašnjeno zašto je za eksperimentalni dio rada odabran upravo YOLOv3, odnosno YOLOv4 detektor. Naime, rezultati preliminarnog istraživanja odnosno pregleda literature glede odgovarajućih detektora objekata na slikama nedvojbeno su ukazivali da je YOLO detektor najbolji izbor za eksperimentalni dio istraživanja, unatoč tomu prethodno su provedena i probna testiranja primjenom drugih *state-of-the-art* detektora sličnih karakteristika koji se također temelje na neuronskoj mreži. Stoga su trenirani i testirani modeli temeljeni na Faster R-CNN [294], SSD detektor [10] koji se temelji na Inception v2 arhitekturi [295], FCOS [28] koji se temelji na ResNet 50 arhitekturi [296], Cascade R-CNN [297] koji se temelji na ResNet 101 arhitekturi [296] te ranije spomenuti YOLOv3 [15] kojima nije mijenjana originalna arhitektura, a testovi su provedeni na 4.270 termalnih slika iz kreiranog skupa slika. Testiranje je provedeno na 1.841 slici iz testnog skupa, a usporedni rezultati prikazani su u tablici 4. [106]

Tablica 4. Usporedni rezultati za detekciju osoba na termalnim slikama [106]

Model	AP	Inference time	FPS
Faster RCNN	98,86%	0,141	7,097
SSD	94,02%	0,063	15,794
FCOS	97,05%	0,048	20,790
Cascade RCNN	98,80%	0,223	4,480
YOLOv3	97,93%	0,036	27,472

Svi testirani modeli su postigli dobre rezultate gledano prema prosječnoj preciznosti (engl. AP - *Average Precision*), a najbolji rezultat postigli su Faster RCNN, Cascade RCNN i YOLOv3. Međutim, zbog svoje arhitekture, YOLOv3 je značajno najbrži s obzirom da postiže obradu od 27,5 FPS te ima najkraće vrijeme zaključivanja odnosno

detekcije po sličici. Kada se uzme pretpostavka da standardni video zapis ima 24 FPS, razvidno je da samo YOLOv3 ima mogućnost obrade video zapisa u realnom vremenu, dok su drugi testirani modeli ipak sporiji od toga. Zbog svega navedenog, u daljnjim eksperimentima korišten je YOLOv3 model, a postignuti rezultati su dodatno uspoređeni i sa novom verzijom YOLOv4.

Iako su u 5.1. opisane verzije YOLO detektora, za njegovo bolje razumijevanje te rezultata detekcije i raspoznavanja potrebno je detaljnije pojasniti određene hiper parametre koji se konfiguriraju u procesu učenja modela. S tim u svezi treba napomenuti da je YOLO, kao i ostale konvolucijske neuronske mreže sastavljen od slojeva, od kojih svaki ima svoju funkciju, pri čemu je konvolucijski sloj svojevrsni osnovni građevni element konvolucijskih neuronskih mreža te obavlja veliku većinu transformiranja podataka, a u literaturi se vrlo često označava skraćeno kao CONV sloj (engl. *Convolutional layer*).

6.4 Rezultati detekcije i raspoznavanja osoba primjenom YOLOv3 detektora

U eksperimentalnom dijelu korišten je ranije YOLO detektor, i to njegova treća verzija YOLOv3, a cilj je bio utvrditi mogućnosti i pouzdanost sustava za nadzor štićenih prostora i područja temeljen na IR termoviziji.

Za početak je korišten originalni model YOLOv3 mreže, koji je označen sa *bY*, a koji je prethodno treniran na MS COCO RGB bazi slika [298] s ciljem detekcije velikog broja različitih klasa objekata. Ovaj model je korišten kao početni iako se termalne snimke značajno razlikuju od standardnih optičkih (RGB) slika, kako u prikazu boja, tako i prikazu drugih detalja kao što su texture, oblici i sl. Unatoč tomu postavljena je pretpostavka da pojedini slojevi RGB slika liče na dijelove na termalnim slikama, tako da će značajke oblika koje je originalni YOLO model, *bY*, prethodno naučio na RGB slikama biti korisne i za primjenu na termalnim slikama. Treba napomenuti da je u izvornom eksperimentu [13] pokazano da učinkovitost YOLO detektora manje opada nego što je to kod drugih sličnih detektora kada se primijeni na primjerima iz novog područja koji nisu korišteni za učenje, što su prikazali na primjeru detekcije osoba na umjetničkim slikama. Izvedeno je više eksperimenata, te je u nastavku predstavljen redoslijed njihovog izvođenja:

- u prvom eksperimentu su rezultati testiranja korištenjem izvornog YOLOv3 modela uspoređeni sa rezultatima testiranja YOLO detektora koji je naučen na

termalnoj bazi slika, koji je označen sa tY . Proces učenja je proveden na već spomenutoj bazi slika koja je kreirana isključivo u ovu svrhu, a testiranje je provedeno i uspoređeno na termalnim slikama koje su snimljene po različitim vremenskim uvjetima, kao i na neovisnim termalnim skupovima slika. Ovdje također treba istaknuti da je termalni YOLO model učen isključivo na termalnim snimkama koje su snimljene po vedrom vremenu, a testiranje je provedeno na termalnim snimkama koje su snimljene po gustom magli i jakoj kiši, na različitim udaljenostima i pri različitim položajima tijela snimanih osoba.

- u drugom eksperimentu cilj je bio naučiti i testirati učinkovitost naučenog modela u cilju razlikovanja osobe i životinje, u ovom slučaju psa.
- u trećem eksperimentu cilj je bio testirati mogućnosti raspoznavanja aktivnosti osoba (hodanje, trčanje, pogureno hodanje, pogureno trčanje, četveronoški hod te ležanje na zemlji odnosno puzanje).

U svakom eksperimentu, u prvoj fazi korišteni omjer treniranja i testiranja bio je 80:20, odnosno 80% skupa korišteno je za treniranje, a 20% za testiranje. U drugoj fazi testirani su i drugi omjeri, odnosno 10:90 i 20:80 u cilju utvrđivanja pouzdanosti modela koji je treniran na relativno malom broju slika. Naučeni model također je testiran na vanjskim skupovima slika koji nisu korišteni za treniranje modela, s tim da je u tom slučaju napravljena dodatna prilagodba u skupu UNIRI_Thermal (UNIRI_TID) na način da su izvorne slike konvertirane u dvije verzije crno-bijelih slika, na kojima se u jednoj verziji topla područja prikazuju svijetlim tonovima, a u drugoj se topla područja prikazuju tamnim tonovima.

Nadalje, bazni model je korišten i za testiranje detekcije i raspoznavanja osoba na termalnoj snimci koja je preuzeta sa servisa Youtube, a riječ je snimci koja prikazuje snimku vojne vježbe [299], koja korištena u znanstvene svrhe uz odobrenje vlasnika, odnosno Ministarstva obrane Republike Hrvatske (suglasnost i odobrenje u prilogu). S tim u svezi treba napomenuti da je riječ o snimci u sivim nijansama (engl. *grey-scale*), a osim vojnika, prikazuje i oklopna vozila, padobrance, helikoptere, te je zanimljivo promatrati učinkovitost baznog modela u cilju detekcije osoba, kao i učinkovitost modela za raspoznavanje aktivnosti. Za tu svrhu, bazni skup podataka koji obuhvaća sve slike (t_{All}), a kao i u prethodnom slučaju kod vanjskih skupova slika, dodatno je transformiran. Pregled modela naučenih na skupu i podskupovima kreirane baze slika UNIRI-TID predstavljen je u tablici 5.

Tablica 5. Pregled naučenih modela tijekom izvođenja [106]

Model	Additional training on UNIRI-TID	Train images	Testing part of UNIRI-TID	Test images
bY	-		All	1841
tY	All subset	4270	All	1841
tY_clear10	10% of Clear subset	266	90% of Clear subset	2397
tY_clear20	20% of Clear subset	532	80% of Clear subset	2131
tY_clear80	80% of Clear subset	2131	20% of Clear subset	532
tY_clear	70% of clear subset	1862	All	1841
tY_hNh	Human-nonhuman subset	4335	Human-nonhuman subset	1084

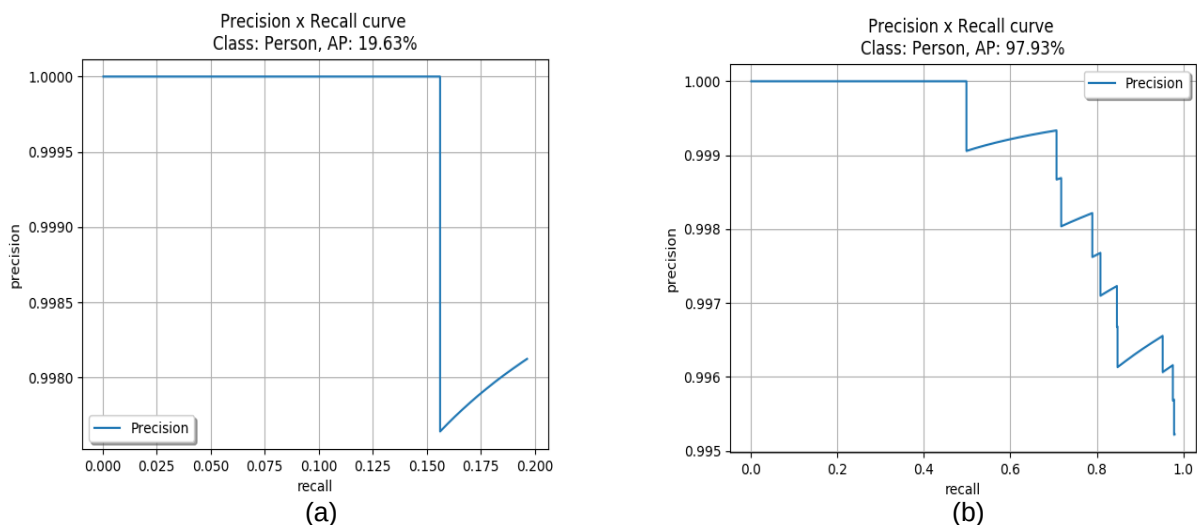
6.4.1 Rezultati evaluacije detekcija osoba na termovizijskim slikama i video snimkama i usporedba rezultata

Kako bi se dobila mAP vrijednost, računa se AP vrijednost za sve klase u skupu slika, međutim u ovom eksperimentu, razmatrana je samo jedna klasa Human/Person (osoba). Promjenom vrijednosti praga detekcije u detektoru, izračunata je krivulja preciznosti/odziva za ciljnu klasu, a AP vrijednost je površina ispod krivulje preciznosti/odziva, što je prikazano na slici 46.

6.4.1.1 Rezultati evaluacije detekcija osoba primjenom bY i tY modela na skupu slika UNIRI-Thermal

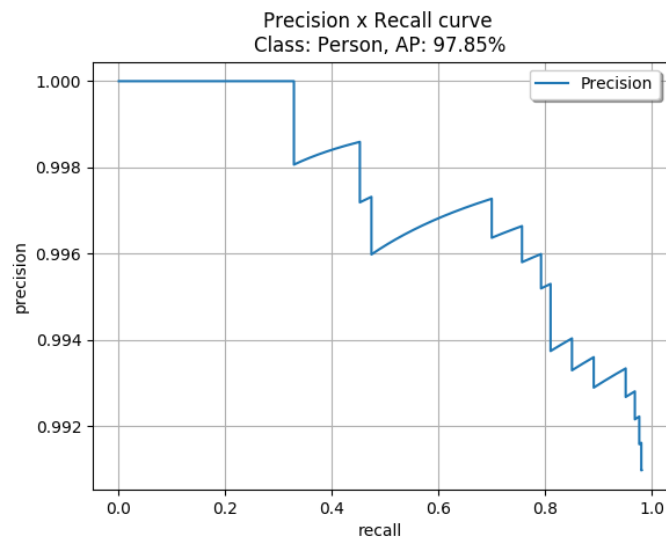
Prvo testiranje provedeno je tako da je pripremljeni skup podataka testiran na originalnom YOLOv3 modelu koji je označen sa *bY*. Slika 46 (a), prikazuje krivulju preciznosti/odziva za klasu Human prilikom testiranja na baznom YOLO modelu *bY*, koji nije treniran na pripremljenom, odnosno termalnom skupu slika UNIRI_Thermal [105], dok slika 46 (b) prikazuje istu krivulju za YOLO model (*tY*) treniran na pripremljenom skupu podataka UNIRI_Thermal, odnosno termalnim slikama, a prikazane krivulje se odnose na cijeli testni skup slika (*All In One*). Iz prikazanog je razvidno da je učenje na termalnom skupu slika značajno poboljšalo rezultate u

odnosu na bazni YOLOv3 model, s obzirom da je AP, odnosno prosječna preciznost modela za *bY* iznosila 19,63%, dok je za *tY* model iznosila 97,93%.

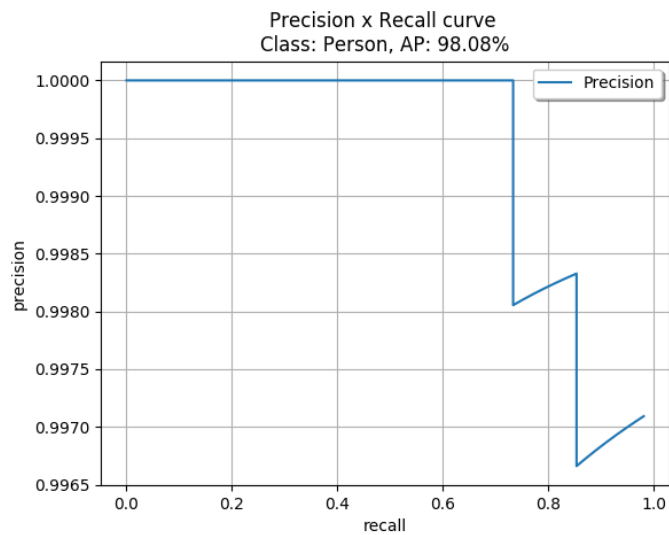


Slika 46. AP rezultat i krivulja preciznosti/odziva za bazni YOLO model, *bY* (a), i za YOLO model, *tY* (b), naučen na pripremljenom termalnom skupu slika [105]

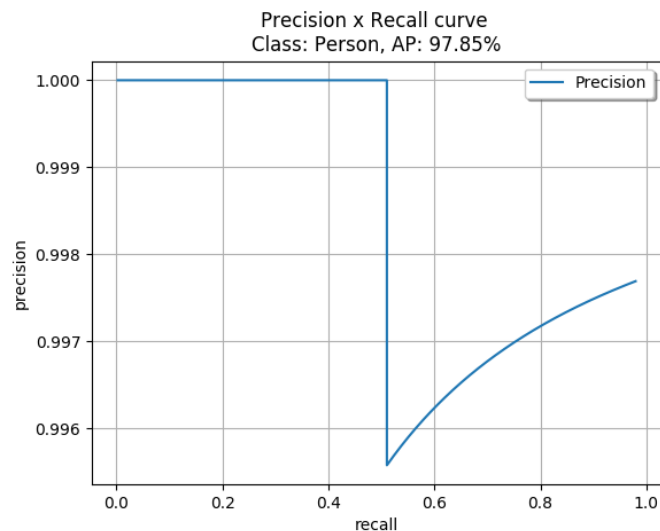
Nadalje, *bY* model postiže 100% preciznost s odzivom od 15,5%, dok *tY* model postiže istu preciznost sa odzivom od oko 50%, što znači da *tY* model detektira više osoba na slikama bez lažno pozitivnih detekcija. Ukoliko se učinkovitost promatra zasebno za različite vremenske uvjete, *tY* model postiže prosječnu preciznost (AP) od 97,85% po vedrom vremenu i gustoj magli, dok je po jakoj kiši taj rezultat čak i bolji s obzirom da iznosi 98,8%. Pri 100% preciznosti, *tY* model ima odziv od 35% po vedrom vremenu, 75% po jakoj kiši te 50% po gustoj magli [105]. Slike 47. – 49., prikazuju krivulje preciznosti/odziva za *tY* model koje su izračunate na pojedinim dijelovima testnog termalnog skupa slika za različite vremenske uvjete.



Slika 47. Krivulja preciznosti/odziva za vedro vrijeme [105]

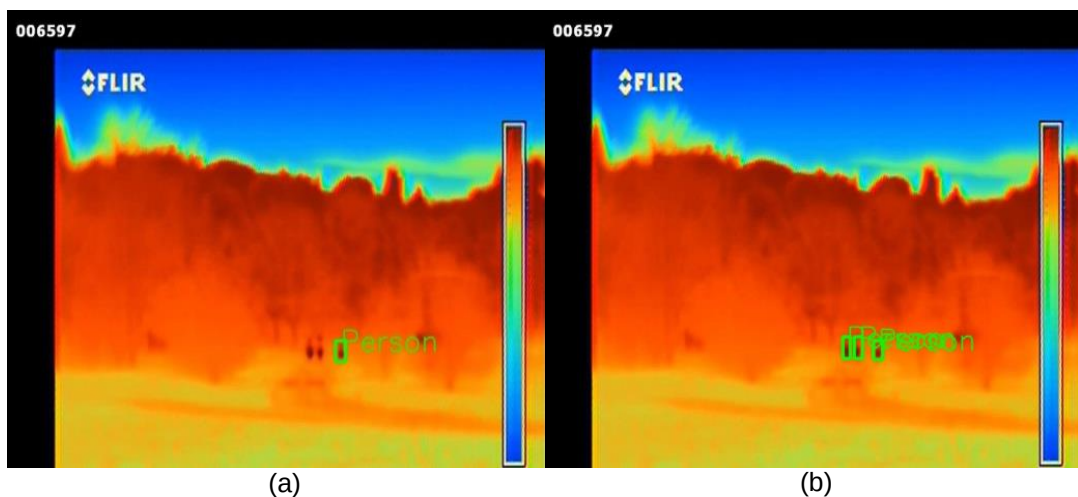


Slika 48. Krivulja preciznosti/odziva za za jaku kišu [105]



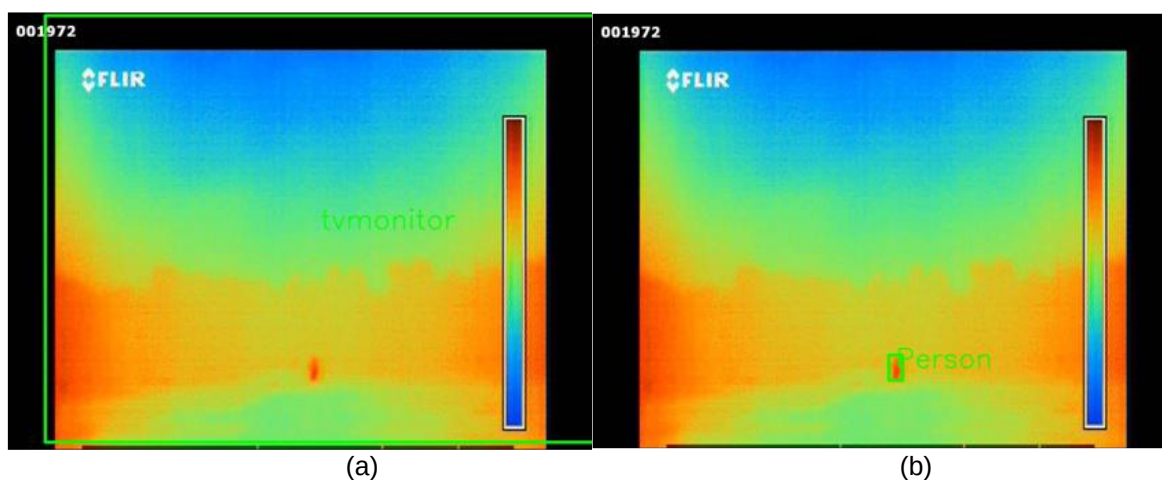
Slika 49. Krivulja preciznosti/odziva gustu maglu [105]

Nekoliko primjera detekcije osoba (Human) primjenom *bY* i *tY* modela prikazano je na slikama 50. i 51., a riječ je o detekcijama na različitim udaljenostima od kamere kao i po različitim vremenskim uvjetima. Na slici 50 (b), *tY* model je točno detektirao tri osobe na slici iako su od kamere udaljene oko 150 m i prikazane su samo s nekoliko piksela, dok s druge strane bazni *bY* model (slika 50 (a)) nije detektirao dvije osobe, odnosno detektirao je samo jednu osobu. Unatoč tomu, ovo je neočekivano dobar rezultat s obzirom da su siluete osoba male i da je mali temperaturni kontrast između osoba i okolne vegetacije.

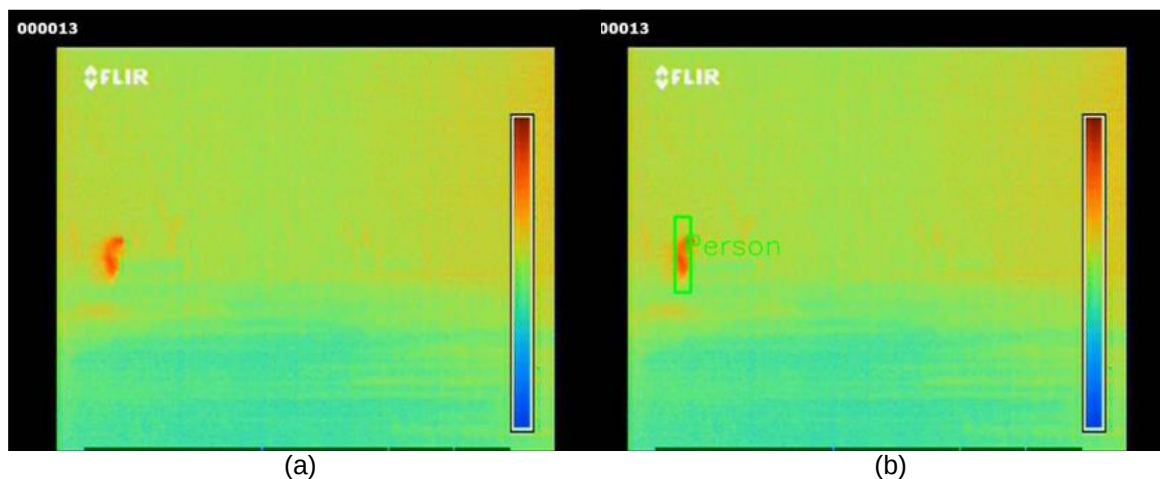


Slika 50. Primjer detekcije osoba primjenom *bY* modela (a) i *tY* modela (b), slike su snimljene korištenjem standardnog objektiva termalne kamere po vedrom vremenu na udaljenosti od ok 150 metara

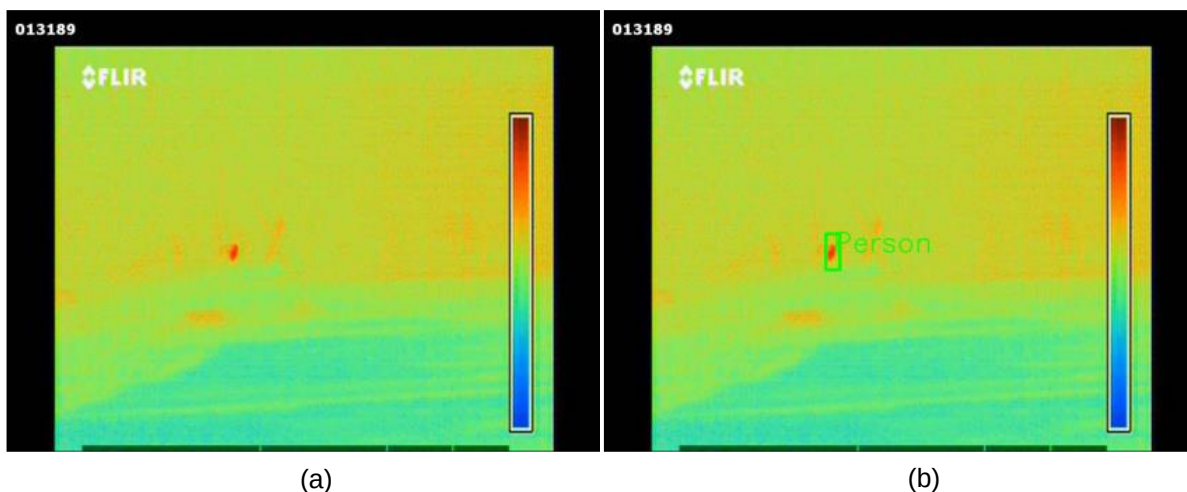
Nadalje, na slikama snimljenima po jakoj kiši (slike 51., 52. i 53.) vidljiva je veća temperaturna razlika između hladnijeg okoliša (prikazano u rasponu od plavih do zelenih tonova) i osobe koja je toplija (prikazana crvenim tonovima) što omogućuje lakšu detekciju osoba, barem vizualnim pregledom slike. Na slici 51.(b), koja je snimljena po jakoj kiši na udaljenosti od oko 70 m, *tY* model je točno detektirao osobu, dok je *bY* model (slika 51.(a)) pogrešno detektirao TV monitor te nije detektirao osobu, iako je temperaturna razlika veća nego na primjeru sa slike 50. Slični rezultati detekcije postignuti su i slučaju pogurenog hoda (slika 52.) i trčanja (slika 53.), koje su snimljene na udaljenosti od oko 100 i 215 metara po jakoj kiši.



Slika 51. Rezultati detekcije osoba (pogureni hod) primjenom *bY* modela (a) i *tY* modela (b). Slike su snimljene korištenjem standardnom objektivu kamere po jakoj kiši na udaljenosti od oko 70 m

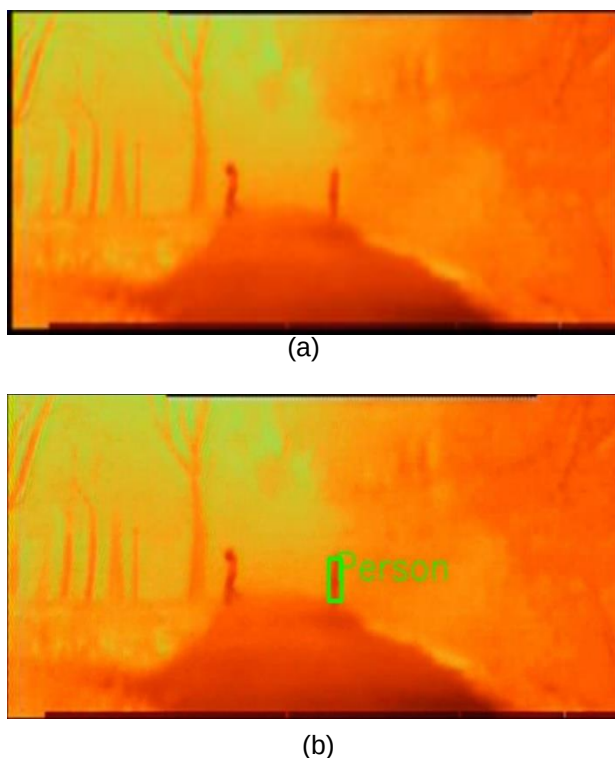


Slika 52. Rezultat detekcije osoba (pogureni hod) primjenom *bY* modela (a) i *tY* modela (b). Slike su snimljene po jakoj kiši korištenjem teleobjektiva na udaljenosti od oko 100 m



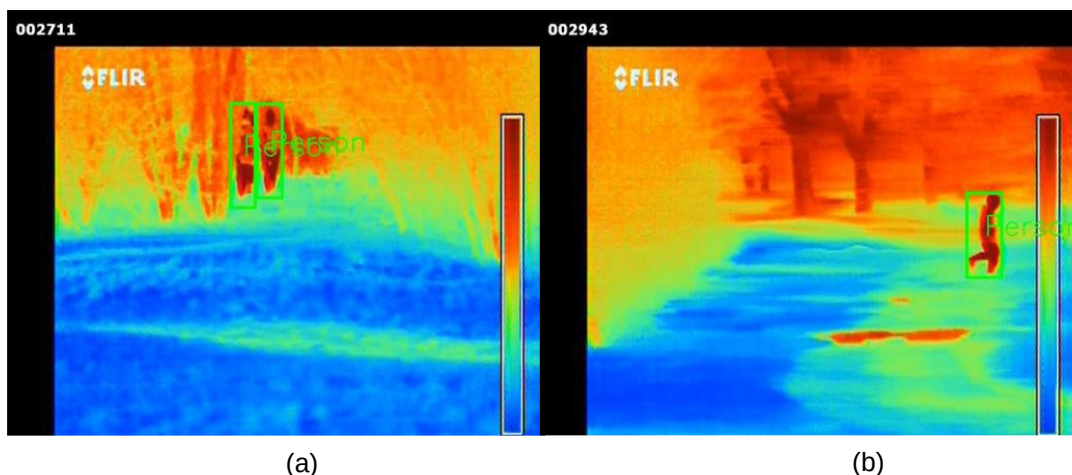
Slika 53. Rezultat detekcije osobe (trčanje) primjenom *bY* modela (a) i *tY* modela (b). Slike su snimljene po jakoj kiši korištenjem teleobjektiva na udaljenosti od oko 215 m

U drugim vremenskim uvjetima, temperaturna razlika između osobe i okoline je često manja zbog čega je detekcija prema temperaturnoj skali i teža, odnosno zahtjevnija, a takav primjer je prikazan na slici 54. U ovom slučaju *tY* model je detektirao jednu od dvije osobe prisutne u polju snimanja, slika 54.(a), dok model *bY* nije detektirao niti jednu osobu, slika 54.(b).

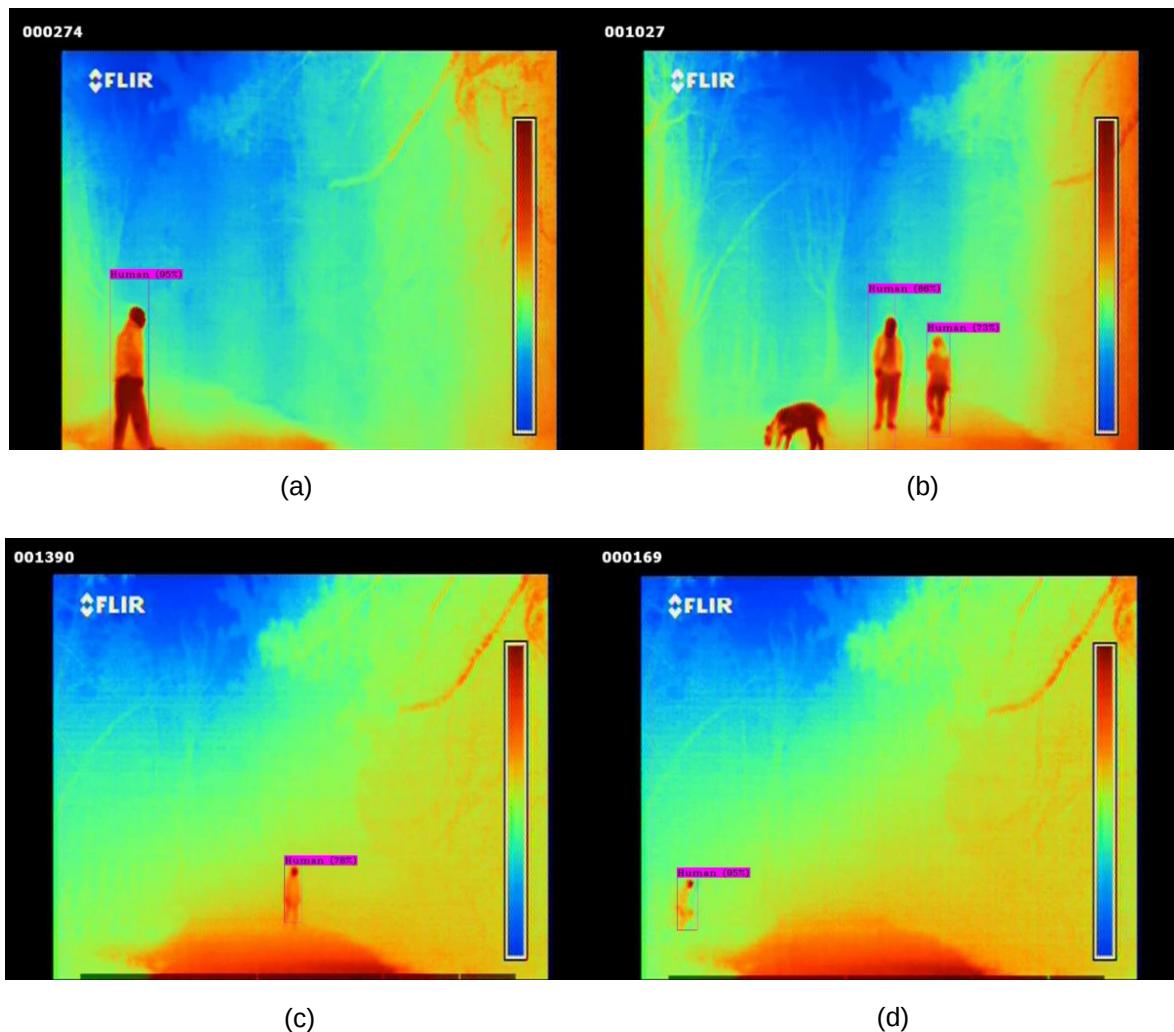


Slika 54. Rezultat detekcije osoba (normalan hod) primjenom *bY* modela (a) i *tY* modela (b). Slike su snimljene po gustom magli korištenjem teleobjektiva na udaljenosti od oko 50 m.

Slike od 55. i 56., dodatno prikazuju primjere detekcije osoba u različitim uvjetima primjenom tY modela. Svakako treba dodati da ovaj model detektira osobe neovisno o načinu kretanja, pa čak i u situacijama kada je termalni kontrast između osobe i okoline jako mali ili kada je riječ o snimanju na relativno velikoj udaljenosti. Pored toga, model je uspješan i u razlikovanju osoba od drugih objekata sličnih kontura ili temperature, kao što su stabla te i u tom slučaju uspješno detektira osobe, bez lažno pozitivnih detekcija (slika 55. (a)).



Slika 55. Rezultat detekcije osoba na slikama koje su snimljene teleobjektivom po vedrom vremenu na udaljenosti od 165 m; normalan hod (a), trčanje (b)

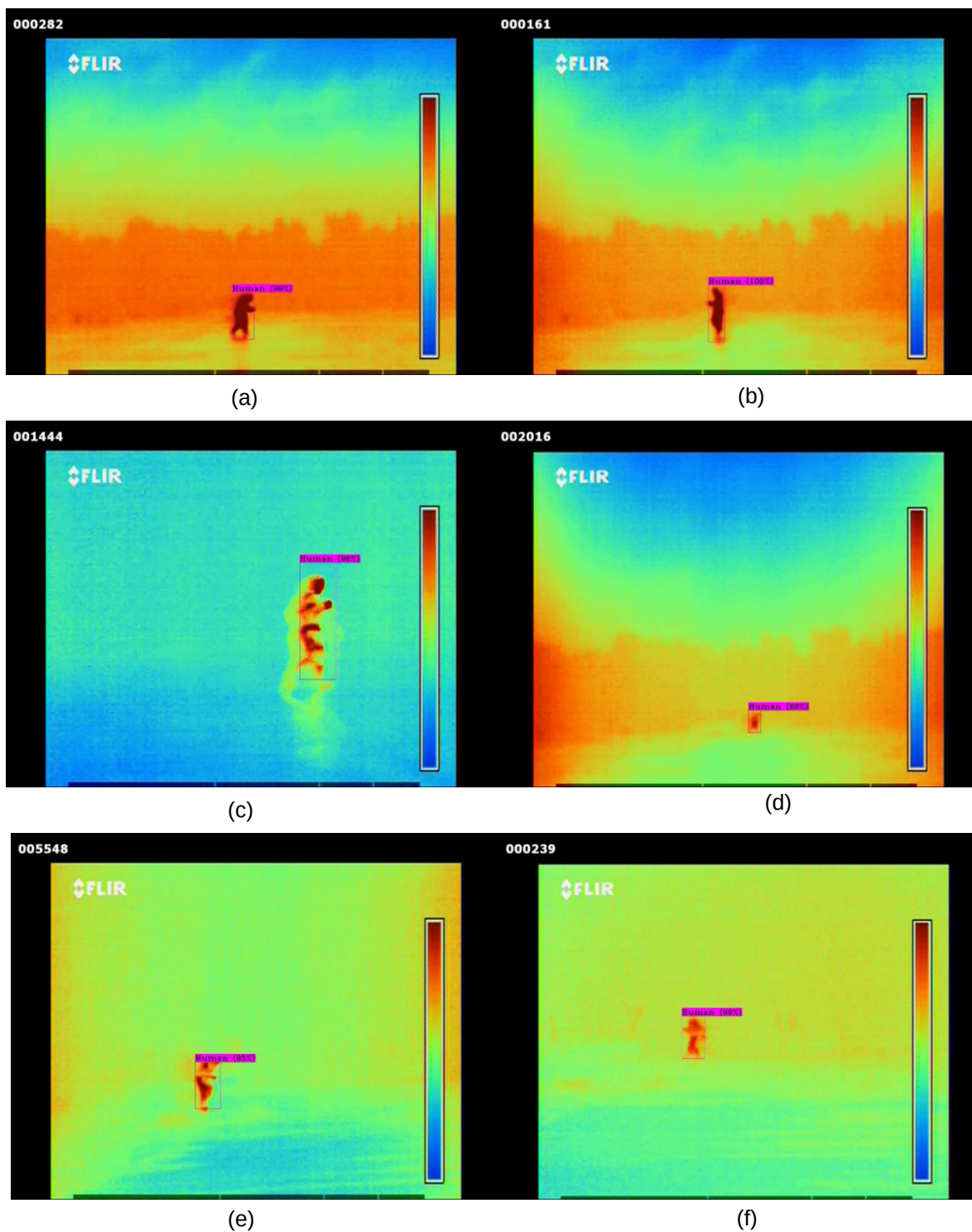


Slika 56. Primjeri detekcije osoba po gustom magli (snimano korištenjem teleobjektiva): (a) udaljenost manje od 30 m, normalan hod; (b) udaljenost 30 m, normalan hod; (c) udaljenost 50 m, normalan hod; (d) udaljenost 50 m, pogureno hodanje

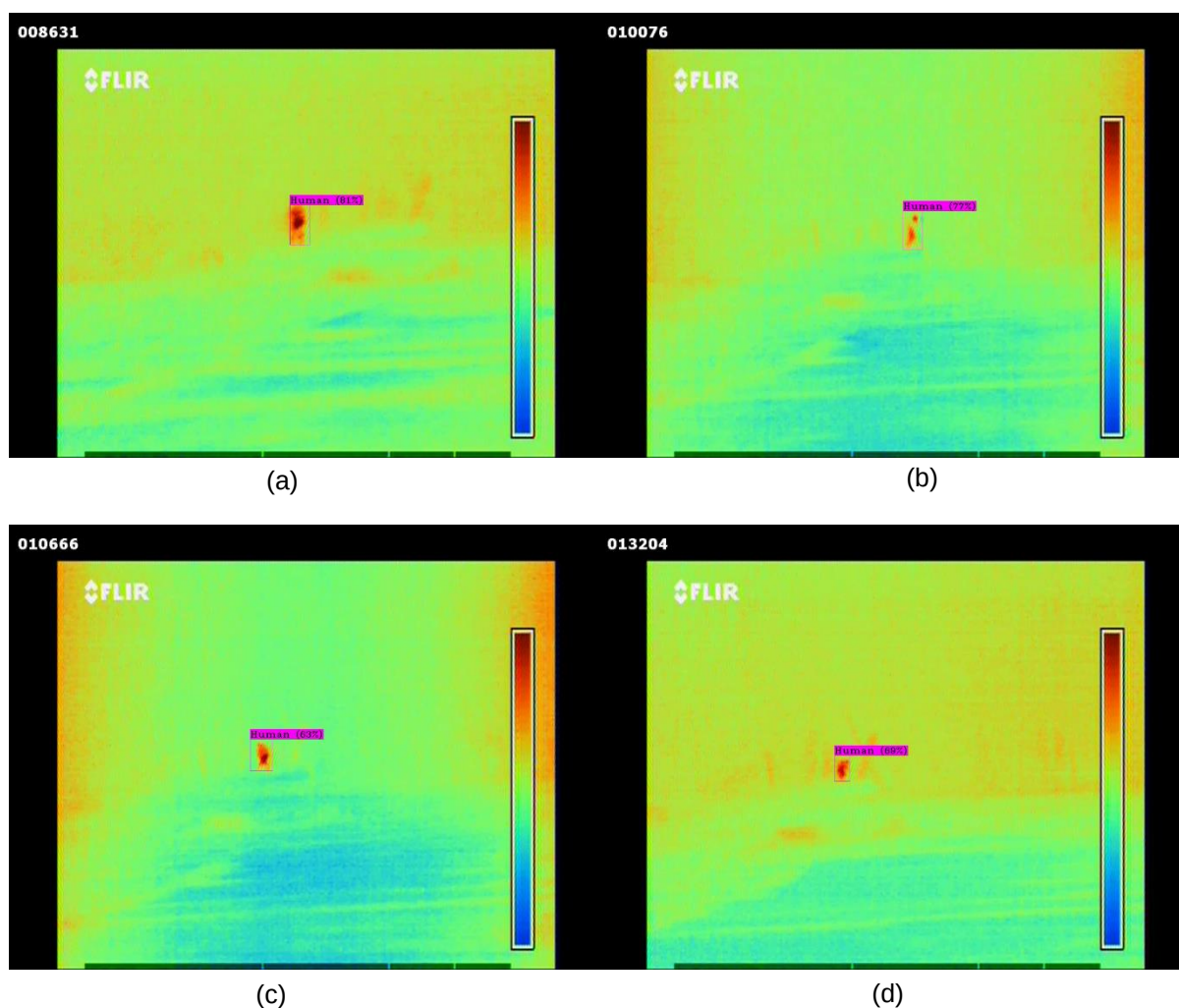
Nadalje, slika 56(a) je primjer pozitivne detekcije osobe u normalnom hodu po gustom magli, primjer (b) također prikazuje pozitivnu detekciju dvije osobe na udaljenosti od oko 30 metara, kao i da se na slici, osim osoba, nalazi i pas čija prisutnost nije zbunila detektor, te nije detektiran kao objekt na slici. Treba naglasiti i da u ovom eksperimentu detektor nije treniran na *Non-Human* klasi, nego je treniran isključivo na klasi *Human* (osoba) s ciljem detekcije i prepoznavanja osoba, neovisno o prisutnosti drugih objekata na slici koje po konturama i silueti liče, odnosno podsjećaju na osobu. Primjeri na (c) i (d) također prikazuju pozitivnu detekciju na slikama snimljenima po gustom magli na udaljenosti od oko 50 metara, s tim da je na slici (c) prikazana osoba pri normalnom hodu, dok je na slici (d) prikazana osoba pri pogurenom hodu.

Po jakoj kiši, slika 57., vidljiva je velika temperaturna razlika između osobe i okoline, pogotovo pri korištenju teleobjektiva (c). Model *tY* je uspješno detektirao osobe

neovisno o položaju tijela ili udaljenosti od kamere. Također je razvidno da temperaturni kontrast između osobe i okoline značajno oscilira, čak i u istim vremenskim uvjetima (a, b, d, e, f). Slika 57. (a – f) snimljene su isti dan po istim vremenskim vjetima, odnosno jakoj kiši, ali je na (c) vidljivo da je okolina hladnija od osobe i prikazana je u plavim i zelenim tonovima, dok na ostalim slikama neposredna okolina oko osobe izgleda toplijom i prikazana je sličnim tonovima kao i osoba (crveni tonovi). Nadalje, na slici 58 (a) do (d) prikazana je detekcija osoba na udaljenostima od 140, 170, 180 i 215 metara, a na istima je također vidljiva promjena u temperaturi okoliša.



Slika 57. Primjeri detekcije osoba po jakoj kiši: (a) 30 m, pogureno hodanje, standardni objektiv; (b) 30 m, normalan hod; (c) 30 m, trčanje, teleobjektiv; (d) 70 m, pogureno hodanje, standardni objektiv; (e) 70 m, trčanje, teleobjektiv; (f) 110 m, normalno hodanje, teleobjektiv

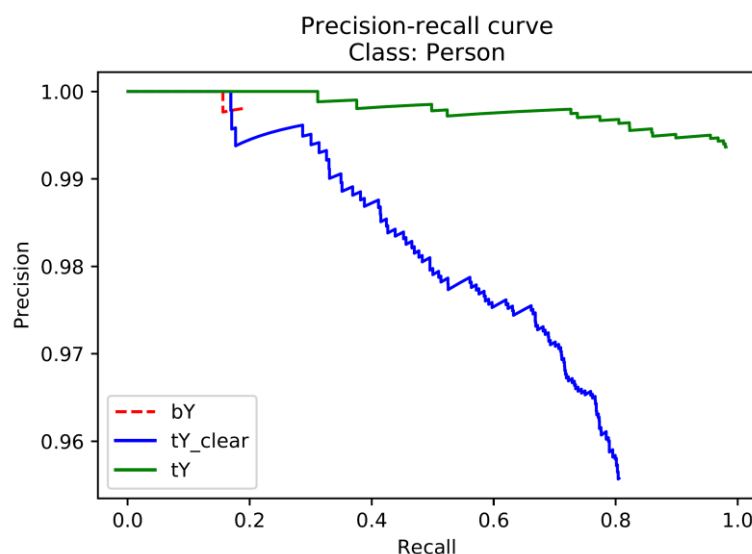


Slika 58. Primjeri detekcije osoba po jakoj kiši: (a) 140 m, normalno hodaње, teleobjektiv; (b) 170 m, normalan hod, teleobjektiv; (c) 180 m, trčanje, teleobjektiv; (d) 215 m, normalno hodaње, teleobjektiv

Rezultati detekcije prikazani na slici 58. (a) – (d), prikazuju rezultat treniranja originalnog YOLO modela (*bY*) za detekciju osoba, odnosno za detekciju i raspoznavanje klase Osoba (*Human/Person*) na termalnim slikama, a dobiveni novi model *tY* postiže izvrsne rezultate detekcije i raspoznavanja osoba u različitim vremenskim uvjetima. Detekcija i raspoznavanje su uspješni čak i kada su osobe na većim udaljenostima ili kada pokušavaju izbjeći detekciju promjenom položaja i brzine hoda (pogureno, trčanje).

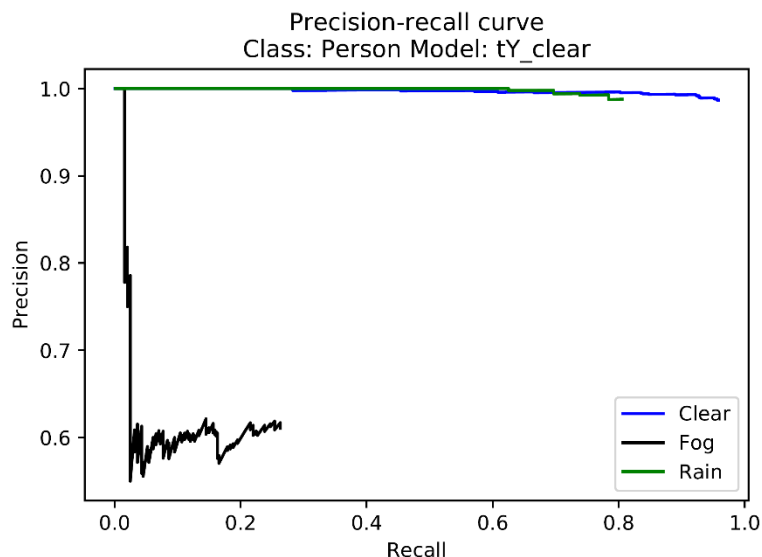
6.4.1.2 Rezultati evaluacije detekcija osoba primjenom bY i tY modela na podskupu slika *clear_ALL*

Nadalje, YOLO model *tY_clear* treniran je samo na slikama snimljenima po vedrom vremenu koji sadrži 2.700 slika, a generalizacijske mogućnosti prikazane su na slici 59., te uspoređene s modelima *bY* i *tY* prema vrijednostima PR krivulje za klasu Human (Person) za cijeli testni skup slika. Već je ranije istaknuto da model *bY* nije treniran na termalnim slikama, kao i da je model *tY_clear* treniran samo na termalnim slikama snimljenima po vedrom vremenu te postiže znatno bolje rezultate nego model *bY*, pri čemu postiže preciznost 79,39% za klasu Human/Person prema AP vrijednosti od 19,63% za model *bY*. S druge strane, a s obzirom da je korišteno manje podataka za treniranje modela *tY_clear*, kao i da su slike snimljene po kiši i magli isključene iz skupa za učenje, model *tY_clear* ne postiže tako dobar rezultat od 97,93%, kao što je to slučaj kod modela *tY* koji je treniran na slikama iz skupa *All In One*.



Slika 59. PR krivulja za bazni YOLO model *bY*, trenirani YOLO model, *tY* i model treniran samo na slikama snimljenima po vedrom vremenu *tY_clear* [106]

Također, ovaj eksperiment je pokazao da dodatno treniranje modela na proširenom skupu slika značajno poboljšava rezultate detekcije, ali i da slike snimljene samo po vedrom vremenu mogu biti korištene za treniranje modela koji može postići dobre rezultate po kiši. Nadalje, na slici 60, koja usporedno prikazuje vrijednost PR krivulje za testiranje *tY_clear* modela na skupovima *tY_fog* i *tY_rain*, a razvidno je da, kada je riječ o detekciji osoba primjenom IR termovizije, magla ima negativniji utjecaj nego kiša.



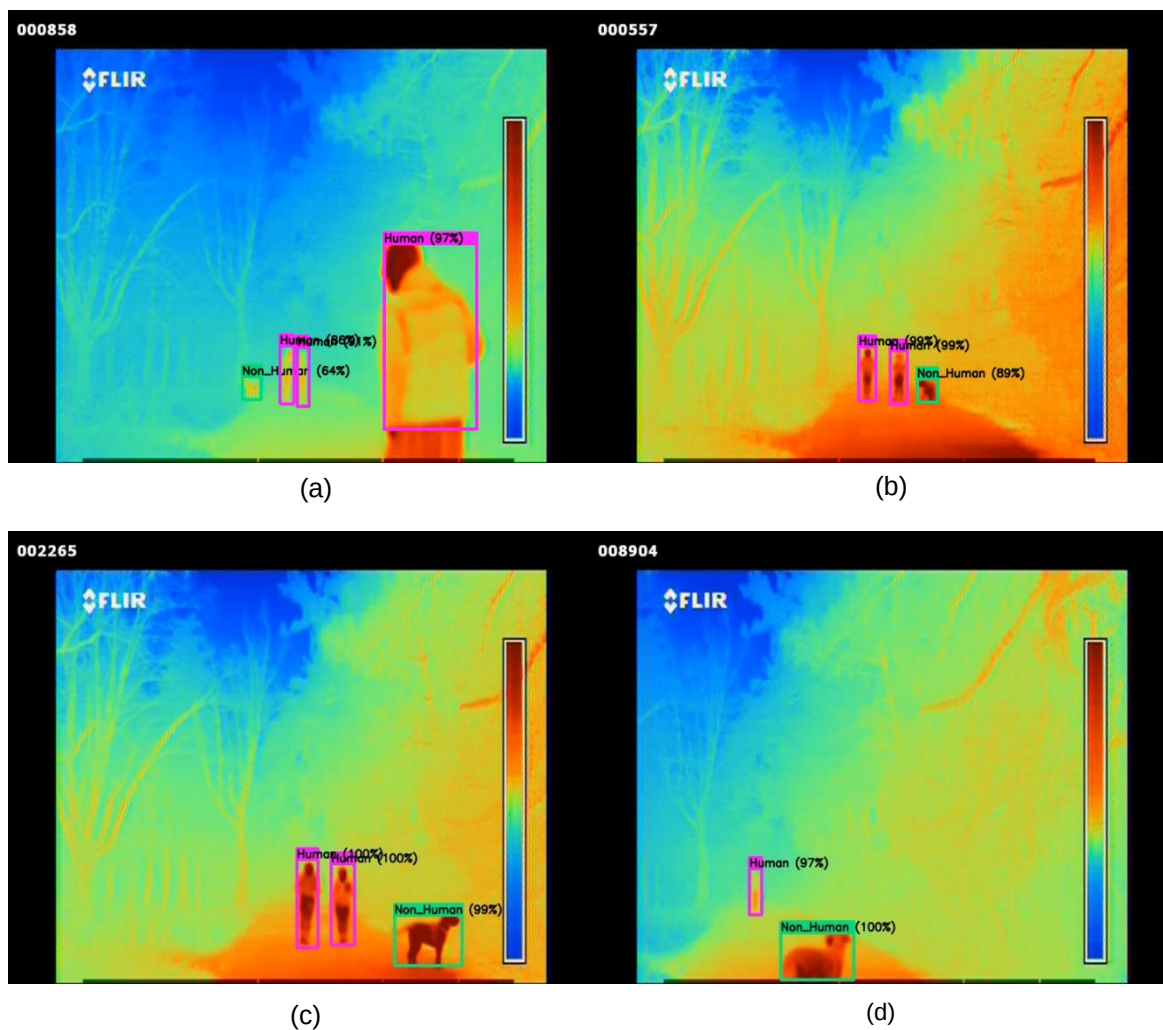
Slika 60. PR krivulja za model tY_clear testiran na skupovima slika snimljenih u različitim vremenskim uvjetima tY_fog (magla) i tY_rain (kiša) [106]

6.5 Detekcija i raspoznavanje osoba u odnosu na druge objekte na slici (Human/Non-Human)

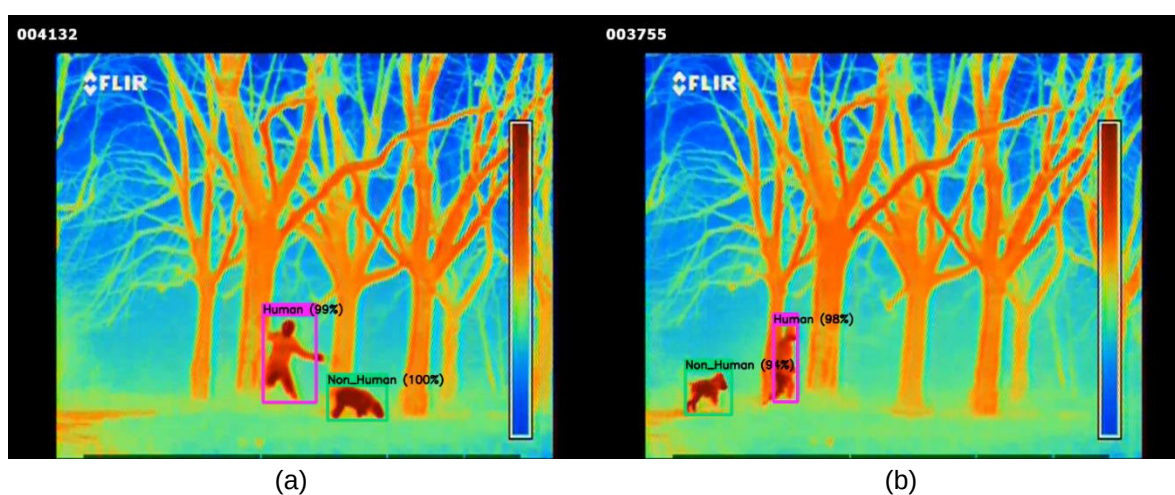
Nakon uspješnog testiranja modela treniranog na skupu slika UNIRI-TID s ciljem detekcije osoba pri čemu je model naučen samo na jednoj klasi – *Human* (osoba), sljedeći eksperiment je bio učenje modela za raspoznavanje osobe od drugih objekata na slici, što je u ovom slučaju bio pas. Nakon treniranja modela, isti je testiran na skupu slika koji je sadržavao osobe i psa, a slike su snimljene po gustom magli korištenjem teleobjektiva, na udaljenostima 30 metara i 50 metara, s tim da je učenje modela provedeno na skupu slika koje su izabrane po maksimalnom razlikovanju među položajima tijela osobe te položaju i brzini kretanja psa, a testiran je na video snimci. Rezultat testiranja, za raspoznavanje osobe od životinje, odnosno psa, AP je iznosio 97,98%. Za napomenuti je da je za testiranje na samo jednoj klasi, *Human*, u prethodnom eksperimentu AP iznosio 97,86%, dok je za binarnu klasifikaciju Human/Non-Human, AP iznosio 98,10%. U ovom slučaju, Odziv (engl. *Recall*) je iznosio 98%, a F1 rezultat 97%. Dobiveni rezultat pokazuje da YOLOv3 detektor s dodatnim treniranjem na dvije klase u potpunosti zadovoljava osnovne ciljeve eksperimenta, a to je uspješno razlikovanje osobe od životinje na termovizijskim slikama i video zapisima, što je ključno za primjenu ovakvog detektora u različitim nadziranima područjima, posebno ukoliko je riječ o teško prohodnim terenima na kojima je velika zahtjevnost angažmana ljudskih posada odnosno policijskih, vojnih ili drugih sličnih sigurnosnih ekipa. Također, jedna od ključnih stvari kod ovakvog tipa detektora

osoba je implementacija i alarmnog modula (kao što će kasnije biti opisano u poglavlju 7, nap.a.), dok visoki rezultat prosječne preciznosti, kao i visoke vrijednosti drugih metrika ukazuju da ovakav sustav može biti pouzdan u smislu da će broj lažno pozitivnih detekcija biti minimalan, što će dovesti do minimizacije potrebe za angažmanom ljudske posade u cilju dodatne provjere i kontrole točnosti detekcije koja je rezultat detektora temeljenog na konvolucijskoj neuronskoj mreži ili konkretno na YOLO detektoru.

Nekoliko primjera detekcije i raspoznavanja prikazano je na slikama 60. i 61. Na prikazanim primjerima vidljivo je da različite udaljenosti, položaji tijela i vidljivost nisu utjecali na pouzdanost detekcije i raspoznavanja osobe od životinje na termalnom video zapisu. Na testnom skupu slika broj lažno pozitivnih detekcija je bio minimalan, što ne utječe na eventualnu primjenjivost ovakvog sustava u realnim uvjetima s obzirom da detektor ovakvog tipa može biti kofiguriran da prije aktivacije alarma određeno vrijeme (nekoliko sekundi) ili određen broj sličica prati osobu za vrijeme kretanja prema ili kroz štićeno područje. Pored toga, kao dodatni kontrolni aktivator alarma može biti postavljen i prag pouzdanosti (engl. *threshold*) s određenom vrijednosti kako bi se izbjegle lažno pozitivne detekcije i preuranjeno ili lažno aktiviranje alarma.



Slika 61. Rezultat detekcije i raspoznavanja osobe od životinje na termalnoj snimci snimljenoj po gustom magli na udaljenosti od oko 30 i 50 metara korištenjem teleobjektiva pri normalnom hodu. Također je razvidno da osobe i pas nisu na istoj udaljenosti, odnosno da je pas bliže kameri nego osobe [106]



Slika 62. Rezultat detekcije i raspoznavanja osobe od životinje na termalnoj snimci po gustom magli na udaljenosti od oko 30 m, korištenjem teleobjektiva dok je snimana osoba mijenjala položaj tijela: normalan hod (a) i trčanje (b) [106]

6.6. Raspoznavanje aktivnosti na termovizijskim slikama i video snimkama

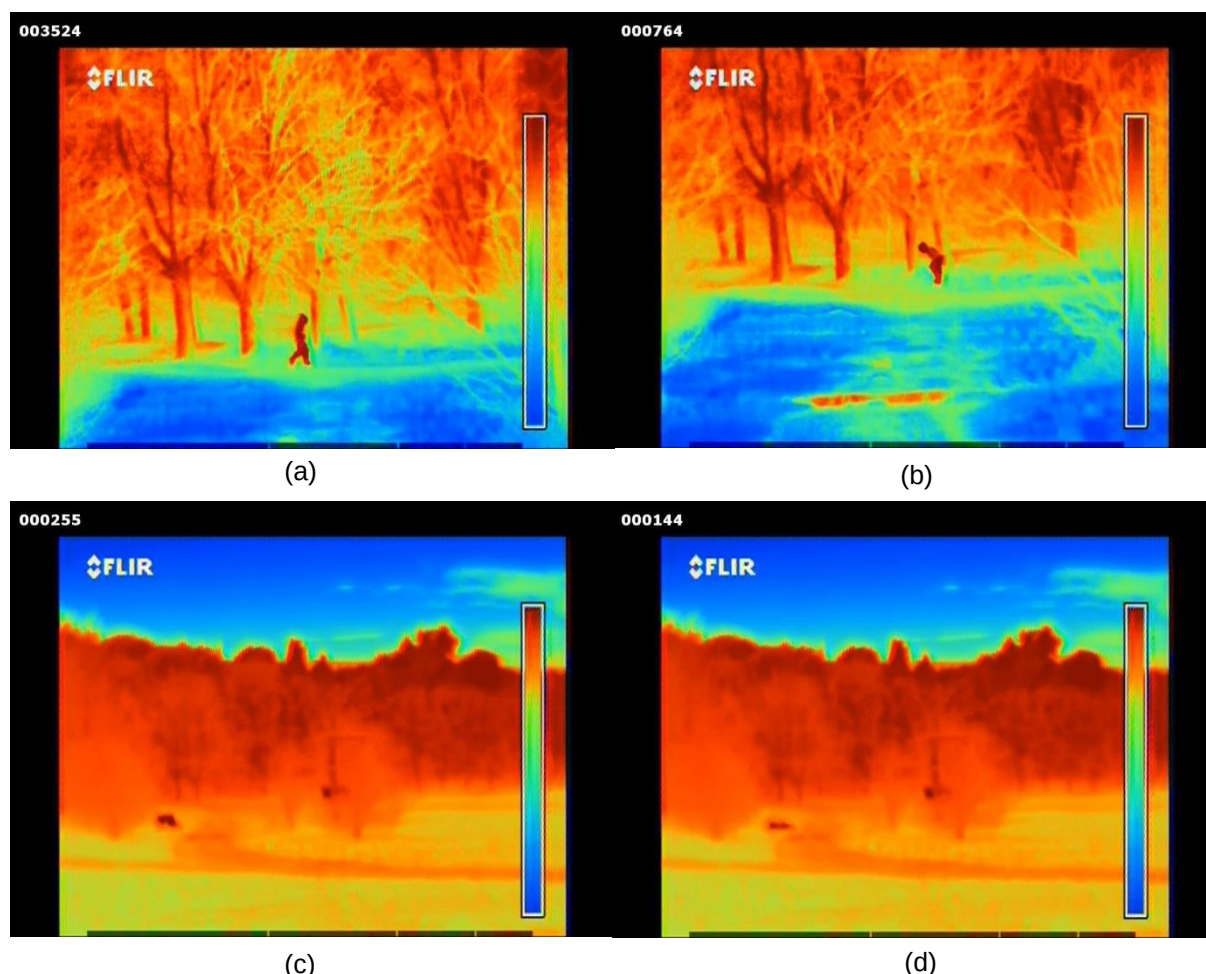
U ovom eksperimentu cilj je bio utvrditi pouzdanost detektora za raspoznavanje različitih aktivnosti osoba snimljenih termalnom kamerom, na različitim udaljenostima i u različitim vremenskim uvjetima (*activityYOLO* model ili *act_Y*). Učenje modela i testiranje je izvedeno u dvije faze. U prvoj fazi model je naučen na slikama po vedrom vremenu, a testiran zasebno na slikama po magli i kiši. U drugoj fazi, kreiran je sveobuhvatni skup slika snimljenih po svim vremenskim uvjetima u kojem su slike anotirane za različite aktivnosti, odnosno brzine kretanja i položaje tijela, s tim da je izmijenjen omjer slika za učenje i testiranje, te je u ovom slučaju korišten omjer 10:90, odnosno 10 % slika je korišteno za učenje, a 90 % za testiranje. Za treniranje u prvoj fazi eksperimenta kreiran je novi skup slika temeljen na prethodno opisanom koji je kreiran za potrebe izrade ovog doktorskog rada, s tim da je zahtijevao dodatnu obradu odnosno pripremu. Najprije je bilo potrebno iznova ručno anotirati slike u pet klasa – normalan hod, trčanje, poguren hod, četveronoški hod i puzanje, odnosno ležanje na tlu. Nakon dodatne obrade, kreiran je skup koji je za svaki tip kretanja sadržavao određen broj slika koje maksimalno različito prikazuju položaje tijela pri određenom načinu kretanja, a broj slika iz skupa za učenje za pojedini način kretanja prikazan je u tablici 6. Za kreiranje skupa slika za učenje modela za prvu fazu korištene su isključivo slike snimljene po vedrom vremenu na različitim udaljenostima korištenjem standardnog i teleobjektiva, a testiranje je porevedno na slikama i video zapisima koje su snimljene po gustoj magli (na udaljenostima od 30 i 50 m korištenjem teleobjektiva), i jakoj kiši (snimljene na udaljenostima od 30 do 215 m).

Tablica 6. Broj slika za pojedini način kretanja u trening skupu slika

Način kretanja	Broj slika	Vremenski uvjeti	Udaljenost
Normalan hod	1.876	Vedro	165 m
Trčanje	549	Vedro	165 m
Poguren hod	818	Vedro	165 m
Četveronoški hod	156	Vedro	110 m
Puzanje, ležanje na zemlji	491	Vedro	110 m

Ukupan broj slika u skupu za učenje je 3.890 slika, a provedeno je s omjerom 80:20, kao i za prethodne eksperimente. Slika 63, prikazuje navedene položaje tijela i načine, odnosno brzine kretanja za svaku pojedinu udaljenost i različite vremenske uvjete, dok

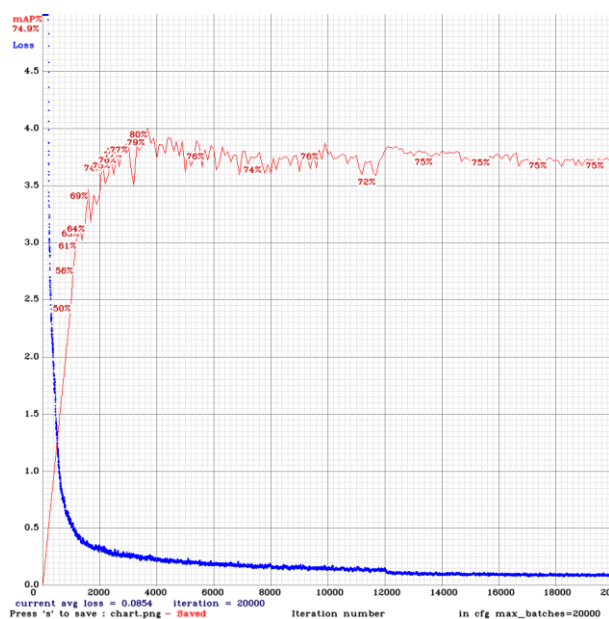
slika 65, prikazuje jednu i više osoba snimljenih po vedrom vremenu pri normalnom hodu, ali na različitim udaljenostima.



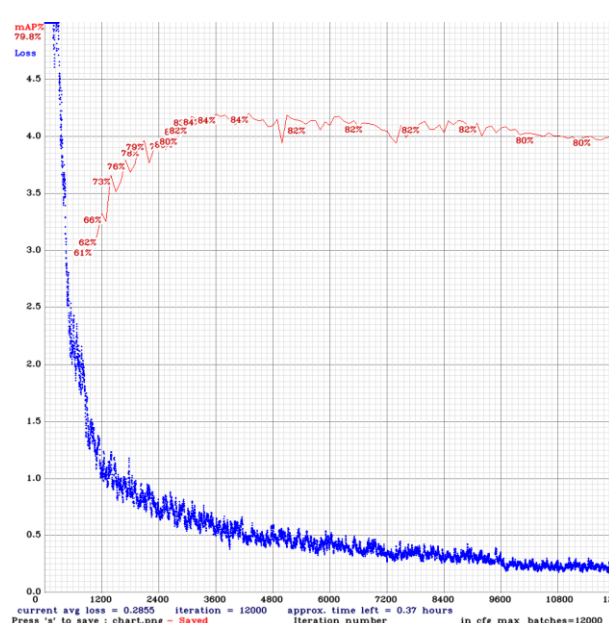
Slika 63. Prikaz osoba po vedrom vremenu (bez klasifikacije aktivnosti): prilikom trčanja, 165 m, teleobjektiv (a); pogureni hod, 165 m, teleobjektiv (b); četveronoški hod, 110 m, standardni objektiv (c); puzanje, 110 m, standardni objektiv (d)

Nadalje, na slici 64, prikazane krivulje za vrijednost mAP i funkcije gubitka (engl. *Loss Function*) za proces učenja, odnosno testiranja naučenog modela za YOLOv3 i YOLOv4 detektore. S tim u svezi, treba napomenuti da je učenje YOLOv3 modela provedeno kroz 20.000, dok je učenje YOLOv4 provedeno na 12.000 iteracija, s obzirom da su prethodni eksperimenti pokazali da je dovoljan minimalan broj iteracija ovisno o broju klasa, odnosno 2.000 iteracija po klasi. Navedeno također pokazuje da proces učenja YOLOv4 modela brže konvergira. Pored toga, prikaz rezultata treniranja/testiranja raspoznavanja aktivnosti primjenom YOLOv3 i YOLOv4 detektora prikazan je u tablicama 7. i 8. Nakon učenja modela i njegova testiranja na cijelom

skupu slika za raspoznavanje aktivnosti, testiranje je provedeno i na video zapisima te pojedinačnim slikama, a rezultati detekcije prikazani su na slikama 66. i 67.



(a)



(b)

Slika 64. Prikaz rezultata učenja/testiranja YOLOv3 (a) i YOLOv4 (b) detektora na skupu t_Act, odnosno na skupu termalnih slika za raspoznavanje aktivnosti

Tablica 7. Rezultat treniranja/testiranja YOLOv3 za raspoznavanje aktivnosti

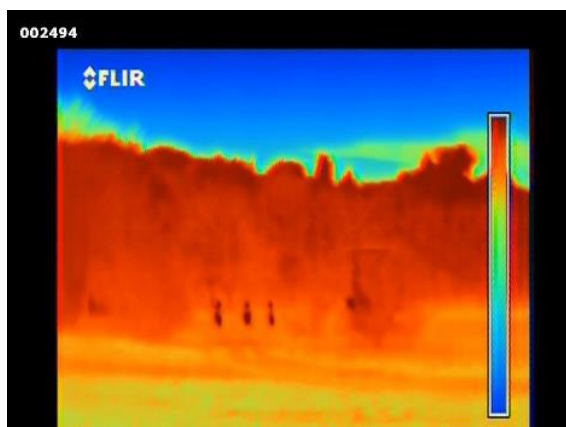
YOLOv3			
mAP	Recall	F1	Average IoU
74,89%	80%	84%	66,76%
Tip aktivnosti			
AP	TP	FP	
Normalan hod (Human_Normal_Walk)	84,33%	9.467	669
Trčanje (Human_Running)	56,73%	728	330
Poguren hod (Human_Hunched)	48,71%	593	591
Ležanje na tlu, puzanje (Human_Lying_On_The_Ground)	90,59%	334	91
Četveronoški hod (Human_Fourleg_Walking)	79,31%	286	64
Objekt koji nije osoba (Ne-Čovjek) (Non_Human)	89,66%	2.140	83

Tablica 8. Rezultat treniranja/testiranja YOLOv4 za raspoznavanje aktivnosti

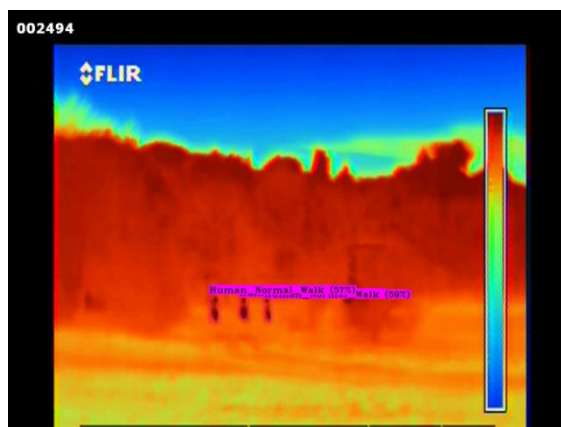
YOLOv4			
mAP	Recall	F1	Average IoU
79,83%	92%	91%	72,87%
Tip aktivnosti	AP	TP	FP
Normalan hod (Human_Normal_Walk)	94,80%	10.894	684
Trčanje (Human_Running)	66,61%	988	433
Poguren hod (Human_Hunched)	49,88%	566	411
Ležanje na tlu, puzanje (Human_Lying_On_The_Ground)	86,70%	316	48
Četveronoški hod (Human_Fourleg_Walking)	82,89%	318	74
Objekt koji nije osoba (Ne-Čovjek) (Non_Human)	98,08%	2.511	78

Glede postignutih rezultata, razvidno je da već i YOLOv3 postiže i više nego dobre rezultate što se tiče raspoznavanja aktivnosti s obzirom da mAP vrijednost iznosi gotovo 75%. Međutim, također je vidljivo da postoji određen broj lažno pozitivnih detekcija, a odnosi se na pogrešno klasificiranje među pojedinim klasama. Najčešće pogreške se odnose na zamjenu između pogurenog hoda i trčanja, s obzirom da je položaj tijela prilikom pogurenog hoda i trčanja gotovo identičan, s tim da je glavna razlika u brzini kretanja, što u ovom eksperimentu nije uzeto u obzir.

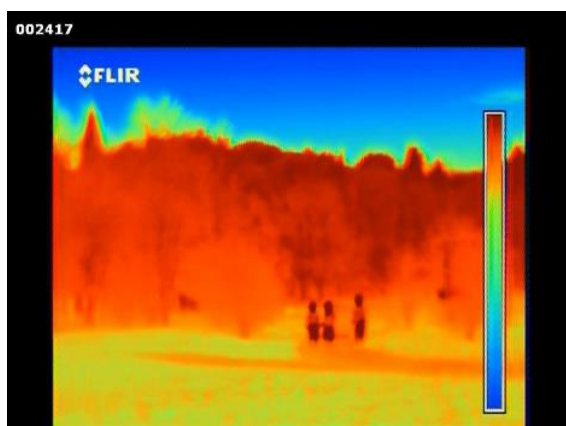
Naknadno je provedeno treniranje i testiranje korištenjem YOLOv4 detektorka koji postiže bolji rezultat u odnosu na prethodnu verziju, što je i vidljivo s obzirom na mAP vrijednost koja iznosi gotovo 80%, što je i više nego dobar rezultat s obzirom na zahtjevnost raspoznavanja među pojedinim klasama, odnosno s obzirom na sličnost pojedinih klasa, odnosno položaja tijela pri testiranim aktivnostima.



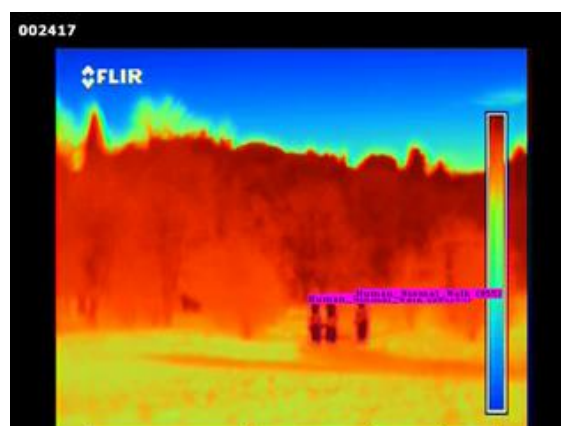
(a)



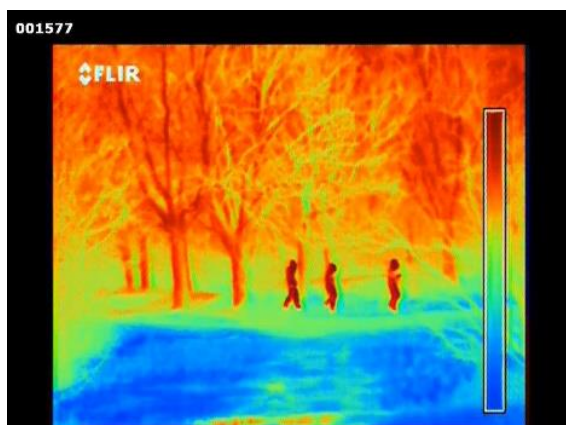
(a)



(b)

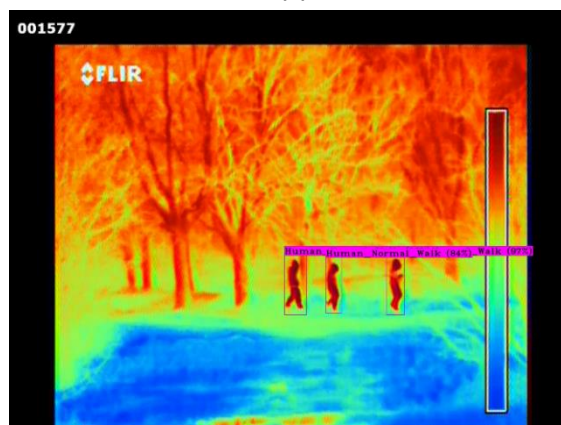


(b)



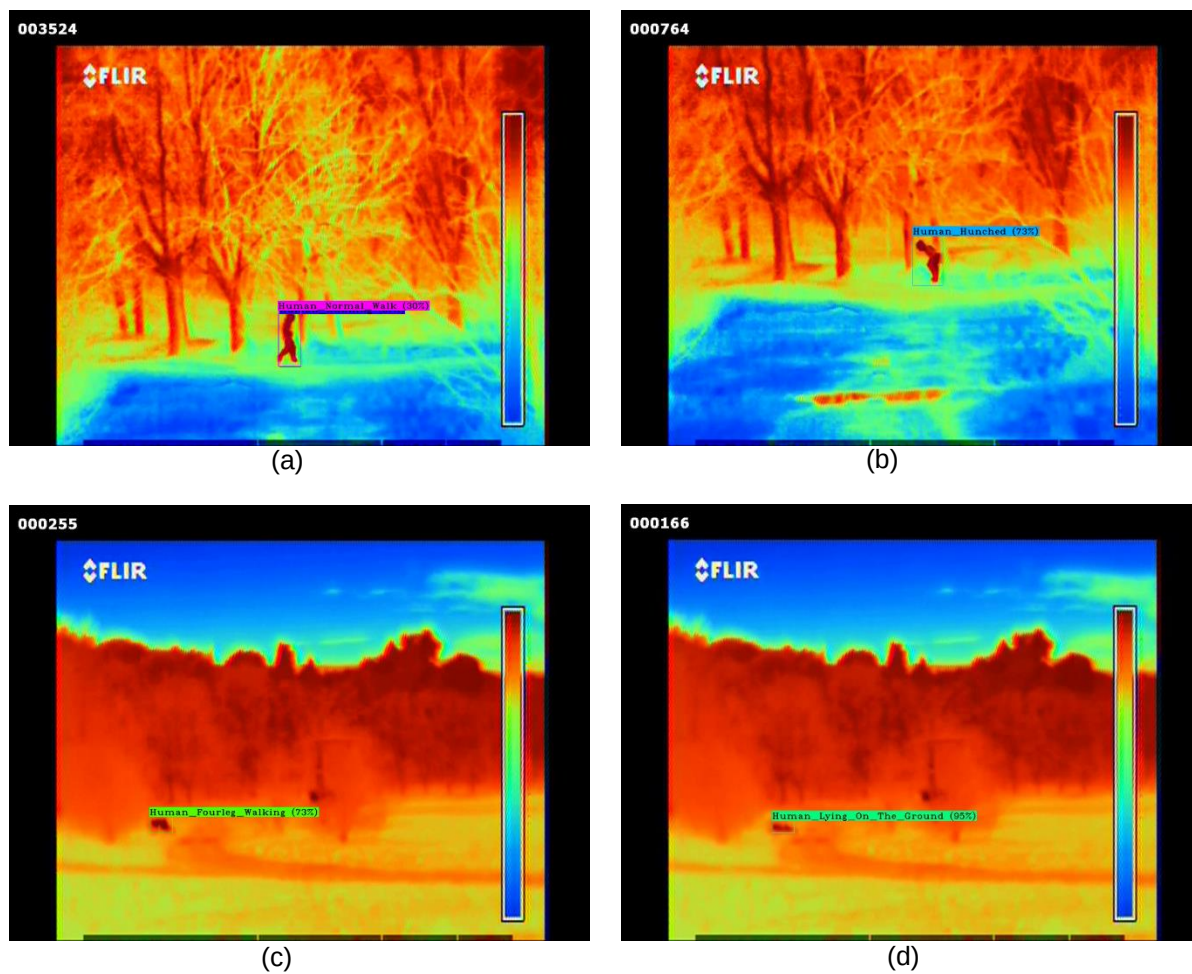
(c)

Slika 65. Primjeri osoba u normalnom hodu (bez klasifikacije načina hoda) koje su snimljene po vedrom vremenu na udaljenostima od: 110 m, standardni objektiv (a); 110 m, teleobjektiv (b); 165 m, teleobjektiv (c)



(c)

Slika 66. Vedro vrijeme - Primjeri osoba u normalnom hodu (s klasifikacijom aktivnosti) koje su snimljene po vedrom vremenu na udaljenostima od: 110 m, standardni objektiv (a); 110 m, teleobjektiv (b); 165 m, teleobjektiv (c)



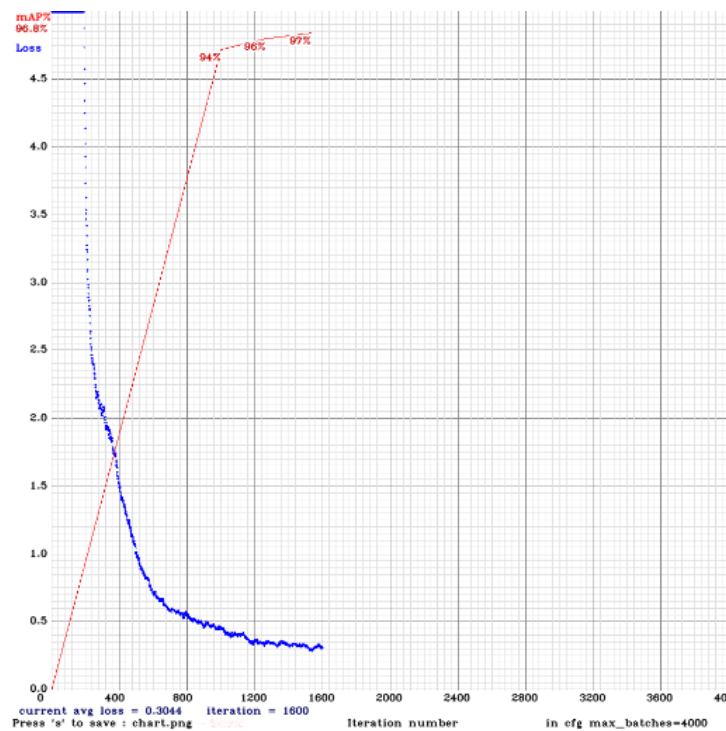
Slika 67. Prikaz osoba po vedrom vremenu (s klasifikacijom aktivnosti): prilikom trčanja, 165 m, teleobjektiv (a); pogureni hod, 165 m, teleobjektiv (b); četveronoški hod, 110 m, standardni objektiv (c); puzanje, 110 m, standardni objektiv (d)

6.7. Treniranje YOLO modela na malom skupu termovizijskih slika

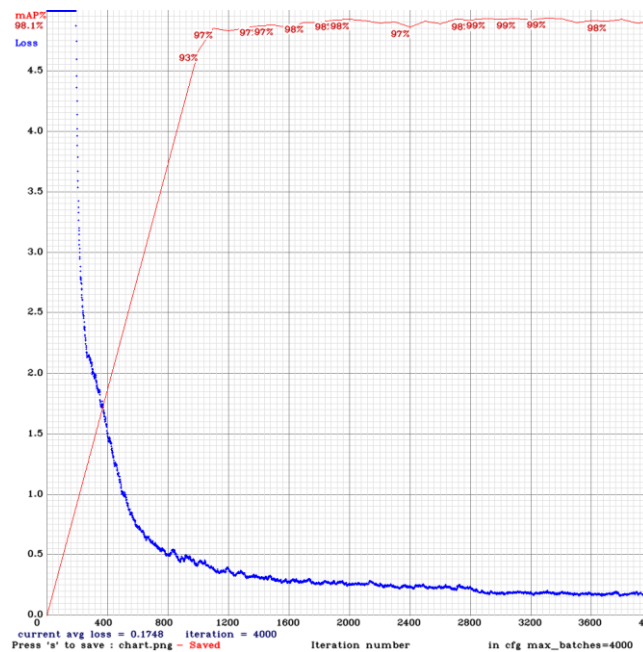
Dodatno na prethodne eksperimente u cilju detekcije i raspoznavanja osoba u otežanim vremenskim uvjetima, a za koji je učenje provedeno prema uobičajenoj raspodjeli 80:20, gdje je 80% skupa slika korišteno za učenje, a 20% za testiranje (slika 68), ovaj eksperiment je proveden u cilju testiranja uspješnosti modela za detekciju osoba koji je treniran na malom broju slika. Prvo je korišten *Clear_All* skup koji obuhvaća samo slike snimljene po vedrom vremenu, ali uključuje sve udaljenosti, brzine i načine kretanja, kao i slike snimljene korištenjem standardnog i teleobjektiva. Drugi korišteni skup je *All_In_One* koji obuhvaća slike snimljene po svim vremenskim uvjetima (vedro, kiša, magla), te uključuje slike svih načina i brzina kretanja, kao i slike snimljene korištenjem standardnog i teleobjektiva.

U prvom slučaju, model je treniran na slikama snimljenima po vedrom vremenu (*Clear_All* skup) i to tako da je on nasumično podijeljen na skup za učenje i testni skup u omjeru 10:90, gdje je 10% slika korišteno za učenje, a 90% za testiranje. Kao i ranije, učenje i testiranje modela provedeno je na Darknet platformi koja je korištena unutar Linux Ubuntu 18.04 operativnog sustava. Slika 69, prikazuje vrijednosti mAP i funkcije gubitka (engl. *Loss function*), a vrijednosti su dobivene izravno tijekom faze treniranja modela tijekom koje se naučeni model testira na zadanom skupu nakon određenog broja iteracija (koji se u konfiguracijskoj datoteci može postaviti sukladno vlastitim procjenama i potrebama). U ovom slučaju je maksimalni broj iteracija postavljen na 4.000, za razliku od prethodnih eksperimenata kod kojih je maksimalni broj iteracija bio 40.000. Dodatno, u ovom eksperimentu, osim temeljnog cilja ispitivanja pouzdanosti modela naučenog na malom broju slika, također je ispitana pouzdanost modela naučenog kroz relativno mali, odnosno minimalni broj iteracija (ukupan broj iteracija jednak je umnošku broja klasa s 2.000). Rezultati učenja/testiranja su pokazali visoku vrijednost mAP metrike u smislu vagane prosječne preciznosti, kao i visoki odziv. Kao što je već navedeno, ponovljeno je treniranje korištenjem standardne podjele 80:20, koje je namjerno zaustavljeno nakon 1.600 iteracija s obzirom da se vrijednost funkcije gubitka prestala značajnije smanjivati, a što je jedan od pokazatelja (prekid treninga je moguć nakon što se vrijednost funkcije gubitka smanji ispod 0,5), jer daljnje učenje modela može dovesti do prenaučenosti modela. Nakon toga, treniranje je provedeno koristeći nasumičnu raspodjelu 20:80 i 10:90 (slika 69). Rezultati vidljivi na prikazanim grafovima potvrđuju da je YOLO model moguće naučiti i na vrlo malom broju slika i

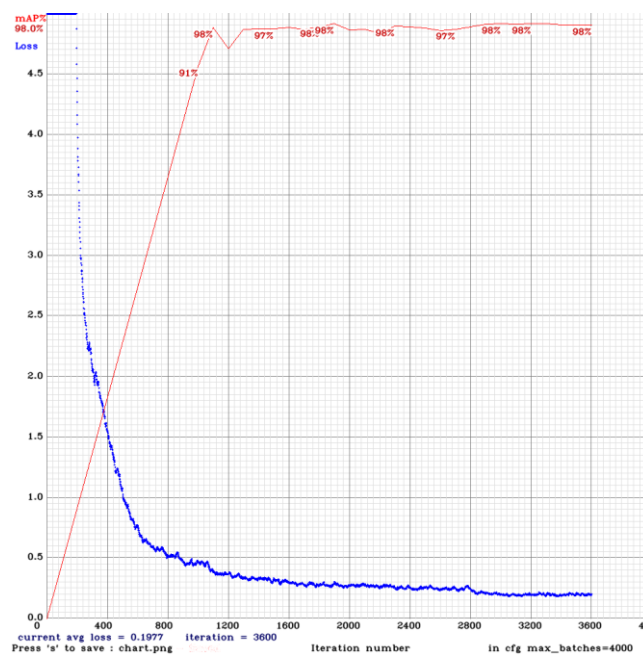
s relativno malim brojem iteracija što uvelike skraćuje vrijeme potrebno za učenje modela bez rizika od gubitka pouzdanosti modela.



Slika 68. Vrijednost mAP i funkcije gubitka za skup Clear_ALL: treniranje/testiranje 80:20 [106]



(a)



(b)

Slika 69. Vrijednost mAP i funkcije gubitka za skup Clear_All: treniranje/testiranje (a) 10:90; (b) 20:80 [106]

6.8 Testiranje naučenog YOLO modela na drugim skupovima termalnih slika i video zapisa

U ovom eksperimentu testiran je prvi YOLO model (tY), koji je naučen na skupu slika *All_In_One* koji obuhvaća slike iz svih podskupova *Clear_All*, *Fog_All* i *Rain_All*, a model je testiran na vanjskim skupovima slika koji se učestalo koriste za testiranje pouzdanosti modela za detekciju osoba na termalnim slikama (tablica 9). Treba napomenuti da je ovaj *All_In_One* skup slika dodatno povećan na način da su originalne RGB slike konvertirane u crno-bijele i to u dvije verzije, prva na kojoj su osobe, odnosno toplija područja, prikazane svijetlim nijansama prema bijeloj boji, a druga na kojima su osobe prikazane tamnim nijansama, odnosno prema crnoj boji. Na ovaj način ukupan broj slika povećan je na 19.008. Ovaj skup je korišten s obzirom da su usporedni skupovi za testiranje naučenog modela različito koncipirani, te je cilj bio povećati mogućnosti generalizacije treniranog modela, posebice u kontekstu njegove aplikacije za realne uvjete nadzora, odnosno detekcije i raspoznavanja osoba, jer je na ovaj način pokriven i veći spektar različitih tipova termalnih kamera koje se koriste u takvim uvjetima. Naučeni model označen je sa, a naučeni model označen je sa $tY_transform$, a za njegovo učenje korišteno je ukupno 11.437 slika iz skupa *All_In_One*.

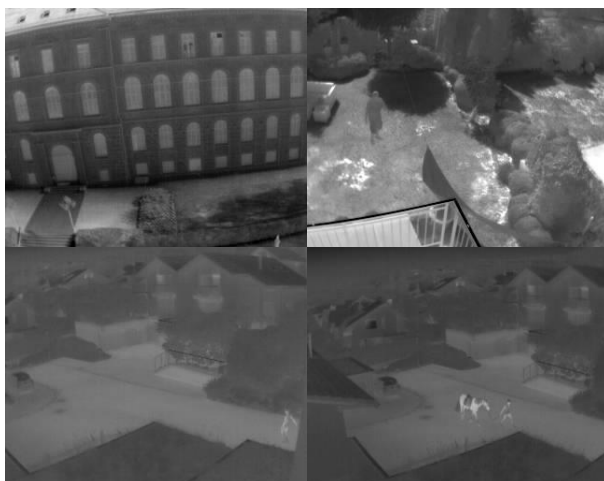
Za usporedno testiranje korišteni su dobro poznati i učestalo korišteni skupovi: ASL_ETH_FLIR [300, 311], LITIV2012 Dataset [301, 302], KAIST Multispectral Pedestrian Detection Benchmark [89], OSU Thermal Pedestrian Database from OTCBVS Benchmark Dataset Collection [303, 304, 312], Terravic Motion IR Database [304, 305], CVC-09: FIR Sequence Pedestrian Dataset [306 – 308], and VOT-TIR2015 Dataset [309, 310], a detalji svakog pojedinog skupa prikazani su u tablici 9.

Tablica 9. Broj slika u svakom pojedinom skupu slika koji je korišten za usporedno testiranje naučenog YOLO modela [106]

Skup	Broj slika	Termovizijska kamera/rezolucija
ASL ETH FLIR [300, 311]	4.381	FLIR Tau 320/324×256 pixels
LITIV2012 Dataset [301, 302]	6.325	320x240 pixels
KAIST Multispectral Pedestrian Detection Benchmark [89]	3.500	FLIR-A35/640x480 pixels
OSU Thermal Pedestrian Database from OTCBVS Benchmark Dataset Collection [303, 304, 312]	6.799	Raytheon 300D/360 x 240 pixels
Terravic Motion IR Database [304, 305]	20.255	320x240 pixels

CVC-09: FIR Sequence Pedestrian Dataset [306 – 308]	10.006	FLIR Tau 2/ 640 × 512 pixels; IDS UI-3240CP/ 1280 × 1024 pixels
VOT-TIR2015 Dataset [309]	7.279	320×240 to 1920×480 pixels

Slike od 70. do 73., prikazuju primjere iz skupova slika korištenih za usporedno testiranje naučenog YOLO modela.



Slika 70. ASL ETH FLIR [300, 311]



Slika 71. CVC IR 09 (gore lijevo) [306 – 308], KAIST dataset (gore desno) [89], LITIV 2012 (dolje) [301, 302]



Slika 72. VOT-TIR2015 [309]



Slika 73. OSU thermal dataset (gore) [303, 304, 312], Terravic motion dataset (dolje) [304, 305]

Tablica 10. Rezultati YOLO metrike za detekciju osoba u svakom pojedinom usporednom skupu slika – *tY_transform* naučeni model [106]

Skup slika	mAP	Avg. IOU	Recall	F1 score
ASL ETH FLIR [300, 311]	0.36	0.38	0.27	0.35
LITIV2012 Dataset [301, 302]	0.71	0.44	0.75	0.64
KAIST [89]	0.35	0.19	0.69	0.36
OSU Thermal Pedestrian [303, 304, 312]	0.84	0.50	0.87	0.72
Terravic Motion IR Database [304, 305]	0.97	0.66	0.98	0.92
CVC-09: FIR Pedestrian [306 – 308]	0.49	0.23	0.66	0.42
VOT-TIR2015 Dataset [309]	0.83	0.54	0.86	0.77

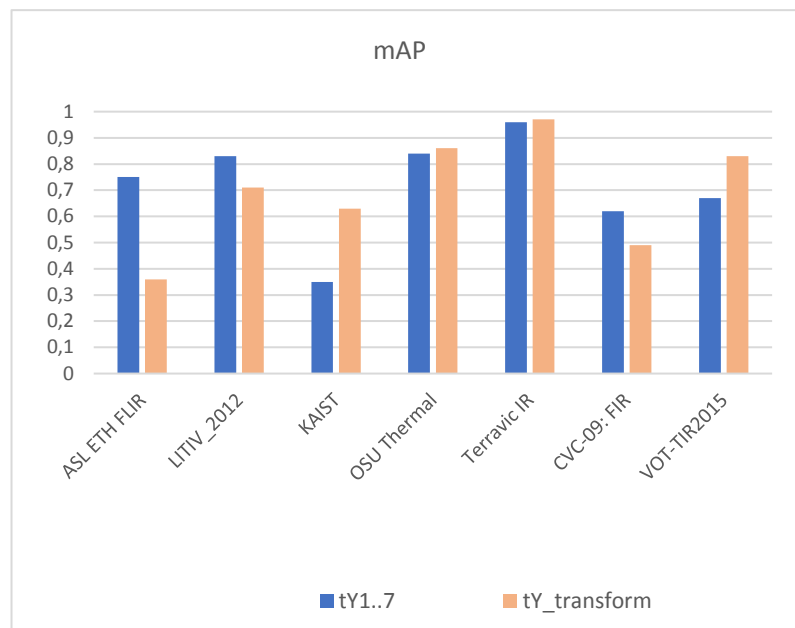
Razvidno je da su najbolji rezultati od 97% mAP i 92% F1 korištenjem *tY_transform* naučenog modela postignuti na skupu podataka Terravic motion [304, 305] jer su termalne siluete iz tog skupa najsličnije slikama iz našeg izvornog skupa. Nešto lošiji, ali i dalje vrlo dobri rezultati detekcije, od oko 83% mAP i 77% F1, model postiže na VOT-TIR2015 [309] i OSU Thermal [303, 304, 312] skupovima podataka, što ukazuje na činjenicu da model dobro generalizira, posebno uzimajući u obzir specifičnost termalne slike, odnosno postojanje značajnih razlika u termalnim siluetama osoba ovisno o udaljenosti i uvjetima snimanja. Najlošiji rezultati, od oko 35% mAP, postignuti su na skupovima podataka AST ETH [300, 311] i KAIST [89] koji imaju značajno različite položaje, udaljenosti i uvjete snimanja [106].

Uz to, kako bismo mogli vjerodostojnije potvrditi izvedbu modela *tY_transform* na referentnim skupovima podataka termalnih slika, izveden je dodatni usporedni eksperiment. Naičeni su i testirani YOLO modeli, *tY1...tY7*, na svakom od referentnih termalnih sklopova iz tablice 10 u omjeru učenje/testiranje 10:90 kroz 20.000 iteracija. Postignuti rezultati kada je svaki pojedini model učen i testiran na istom referentnom skupu prikazani su u tablici 11. i na slici 74. [106]

Tablica 11. Rezultati YOLO metrike za detekciju osoba za učenje/testiranje na svakom pojedinom referentnom skupu slika [106]

Skup slika	mAP	Avg. IOU	Recall	F1 score
ASL ETH FLIR [300, 311]	0.75	0.76	0.79	0.58
LITIV2012 Dataset [301, 302]	0.83	0.84	0.87	0.69
KAIST [89]	0.63	0.69	0.70	0.49
OSU Thermal Pedestrian [303, 304, 312]	0.86	0.67	0.89	0.89
Terravic Motion IR Database [304, 305]	0.96	0.95	0.95	0.75
CVC-09: FIR Pedestrian [306 – 308]	0.62	0.66	0.69	0.52
VOT-TIR2015 Dataset [309]	0.67	0.64	0.75	0.65

U svezi postignutih rezultata potrebno je napomenuti da je očekivano da naučeni modeli postižu bolje rezultate kada se uče na slikama iz izvornog skupa iz kojeg su izdvojene slike za učenje i testiranje, nego kad se uče na slikama iz potpuno drugog skupa kao u slučaju modela *tY_transform*. Navedeno se pokazalo točnim za one skupove koji se uvelike razlikuju od kreiranog skupa UNIRI-TID, a u ovom slučaju riječ je o lošim rezultatima koji su postignuti na skupovima ASL ETH i KAIST. Međutim, zanimljivo je da je na skupovima gdje je model *tY_transform* postigao dobre rezultate, ti rezultati su slični ili još bolji od rezultata koje je postigao model naučen na originalnom skupu. [106]

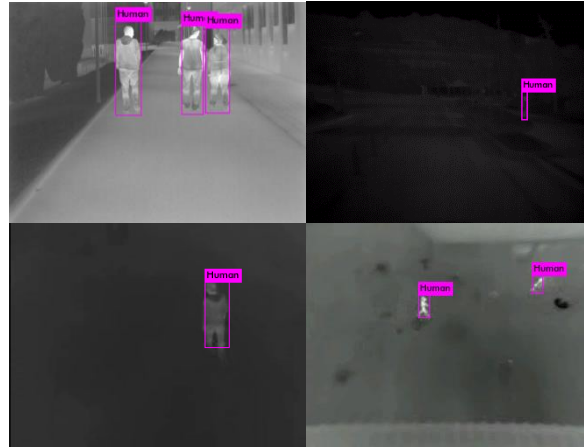


Slika 74. mAP rezultati na različitim skupovima slika za model naučen na skupu UNIRI_TID (*tY_transform*) i svakom skupu podataka (*tY1...tY7*) [106]

Nadalje, slike od 75. do 78., prikazuju rezultate detekcije na pojedinim slikama iz skupova za usporedno testiranje naučenog YOLO modela.



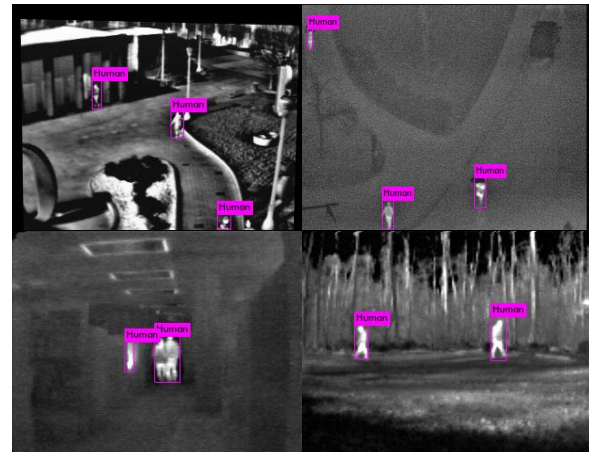
Slika 75. Detekcije na slikama iz skupa ASL ETH FLIR [300, 311]



Slika 76. Detekcije na slikama iz skupova: CVC IR 09 (gore lijevo) [306 – 308], KAIST dataset (gore desno) [89], LITIV 2012 (dolje) [301, 302]

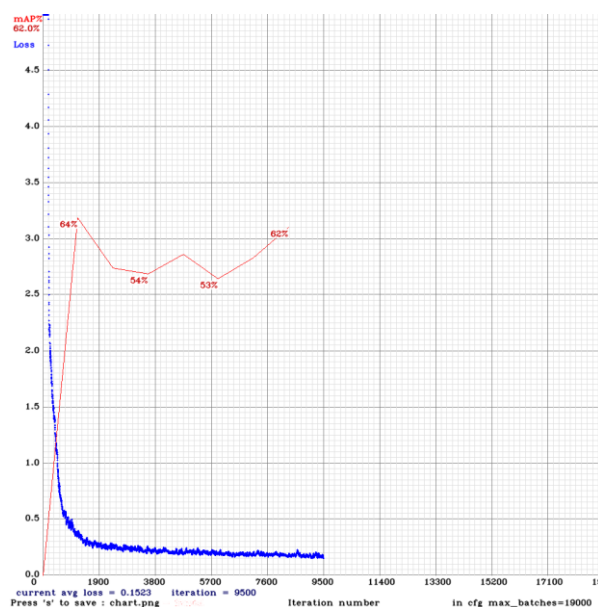


Slika 77. Detekcije na slikama iz skupa VOT-TIR2015 [309]



Slika 78. Detekcije na slikama iz skupova: OSU thermal dataset (gore) [303, 304, 312],

Posljednji eksperiment je također izveden u cilju testiranja mogućnosti generalizacije naučenog YOLO modela, a u ovom slučaju riječ je o slikama koje su ekstrahirane iz video snimke snimljene u realnim uvjetima i okolnostima. Riječ je o termovizijskoj snimci koja je snimljena tijekom vojne vježbe "Swift Response 2019" [313], a snimljena je termovizijskom kamerom GUIDE IR 516B (LWIR) rezolucije 800x600 piksela, po vedrom vremenu s različitih udaljenosti snimanja. Prvi test je proveden kao dio procesa učenja YOLOv3 modela, tijekom kojeg su slike iz video zapisa korištene kao testni skup. Skup sadrži 4.732 slike, a slika 79, prikazuje vrijednost mAP i funkcije gubitka za ovaj slučaj. Pored toga, testirana je i pouzdanost dobivenog modela korištenjem standardne metrike za detekciju objekata, a postignuti su rezultati mAP = 63,71%, recall = 0,59, F1 = 0,60 te prosječni IoU = 44,67%. Dodatno, testiranje je provedeno i izravno na video zapisu, a pojedini rezultati detekcije su prikazani na slici 80.



Slika 79. mAP i Loss Function krivulja za učenje/testiranje YOLOv3 modela na termalnim slikama iz video snimke vojne vježbe "Swift Response 2019"



Slika 80. "Swift Response 2019", primjeri detekcije osoba korištenjem naučenog YOLOv3 modela

7 Sustav za detekciju i raspoznavanje osoba u otežanim uvjetima primjenom IR termovizije

Sustav za detekciju i raspoznavanje osoba u otežanim vremenskim uvjetima primjenom IR termovizije zamišljen je kao sustav temeljen na detektoru objekta razvijenom na bazi konvolucijske neuronske mreže (npr. YOLO, SSD,...) s mogućnošću neprekidnog rada (24/7) u svim vremenskim uvjetima i s minimalno potrebnim angažmanom ljudske posade. Jedan od ciljeva je također i maksimalno pojednostavljivanje sustava, kako u kontekstu modula koji ga čine, tako i u kontekstu budućih potreba i mogućnosti za buduća unaprjeđenja i povezivanje s novim tipovima senzora. Također će biti predviđena mogućnost implementacije već postojećih detekcijskih sustava u cilju maksimizacije pouzdanosti detekcijskog sustava, s obzirom da se zamišljena primjena odnosi na nadzor državnih granica u teško pristupačnim područjima, nadzor područja zračnih luka i drugih značajnih infrastrukturnih objekata, te drugih štićenih prostora i objekata. Svakako uvodno treba naglasiti da sustav ovog tipa mora imati visoku razinu pouzdanosti u radu, jednostavnosti u izgradnji istoga, kao i za konačni cilj visoku točnost raspoznavanja osoba u odnosu na druge pokretne objekte (u prvom redu životinje koje ne predstavljaju prijetnju štićenim objektima i prostorima).

Glede visoke pouzdanosti u radu, sustav treba zadovoljavati kriterije u kontekstu robusnosti i fizičke neosjetljivosti na sve vremenske uvjete i temperature u okolišu koje mogu dovesti do pogrešaka u radu, stoga fizički dijelovi sustava (termovizijska kamera, kablovske veze, radio odašiljači i sl.) moraju zadovoljavati visoke kriterije u smislu otpornosti na ekstremne vremenske uvjete. Nadalje, pouzdanost svakako uključuje i stabilnost električnog napajanja u kontekstu spoja na postojeću električnu mrežu ali zahtijeva postojanje i pomoćnih autonomnih sustava napajanja električnom energijom kako bi sustav mogao neometano funkcionirati i u slučajevima ispadanja iz električne mreže. Kako je već spomenuto, sustav pomoćnog napajanja mora biti neovisan u smislu samostalne proizvodnje dovoljne količine električne energije za napajanje svake pojedine termalne kamere koja je dio detekcijskog sustava. To mogu biti sustavi napajanja temeljeni na iskorištavanju solarne ili vjetroenergije. Pored navedenog, pouzdanost sustava svakako podrazumijeva i implementaciju adekvatne razine krypto zaštite na komunikacijskim kanalima između termalnih kamera kao isturenih ulaznih senzora koji čine prvi korak u nadzoru štićenog prostora. Osim implementacije krypto

zaštite između kamera i centralnog dijela detekcijskog sustava u kojem se vrši obrada video snimki prikupljenih s termalnih kamera, ista razina zaštite nužna je i na komunikacijskim kanalima prema zapovjednim mjestima te prema definiranim postajama sa ljudskom posadom, a sve u cilju maksimizacije zaštite komunikacije i osiguravanja od neovlaštenih pristupa u komunikacijske kanale sustava te eventualno njegovo gašenje ili promjena u cilju onemogućavanja njegovog rada. Pojedine segmente ovakvog sustava moguće je osigurati kroz razvoj odgovarajuće aplikativne i krypto zaštite, dok je drugi dio na ljudskom faktoru, odnosno osobama koje su izravno uključene u rad i održavanje sustava. Svakako treba napomenuti da ljudski faktor predstavlja veći rizik od proboja u zaštićene sustave, nego što je to neovlašteni pristup putem digitalnih kanala, stoga je svakako nužna implementacija sigurnosnih mjera u kontekstu minimizacije rizika koji mogu proisteći od strane ljudskog faktora. Pouzdanost sustava treba promatrati i u kontekstu brzine obrade podataka pri čemu je nužna obrada u realnom vremenu, jer u kriznim situacijama jedino takva obrada omogućava ljudskoj posadi (policajcima, vojnicima, zaštitarima...) pravovremenu reakciju i sprečavanje ulaska neovlaštenih osoba u štićeni prostor ili objekt, odnosno hvatanje počinitelja (npr. pri pokušajima nezakonitog prelaska državne granice, krijumčarenja i sl.)

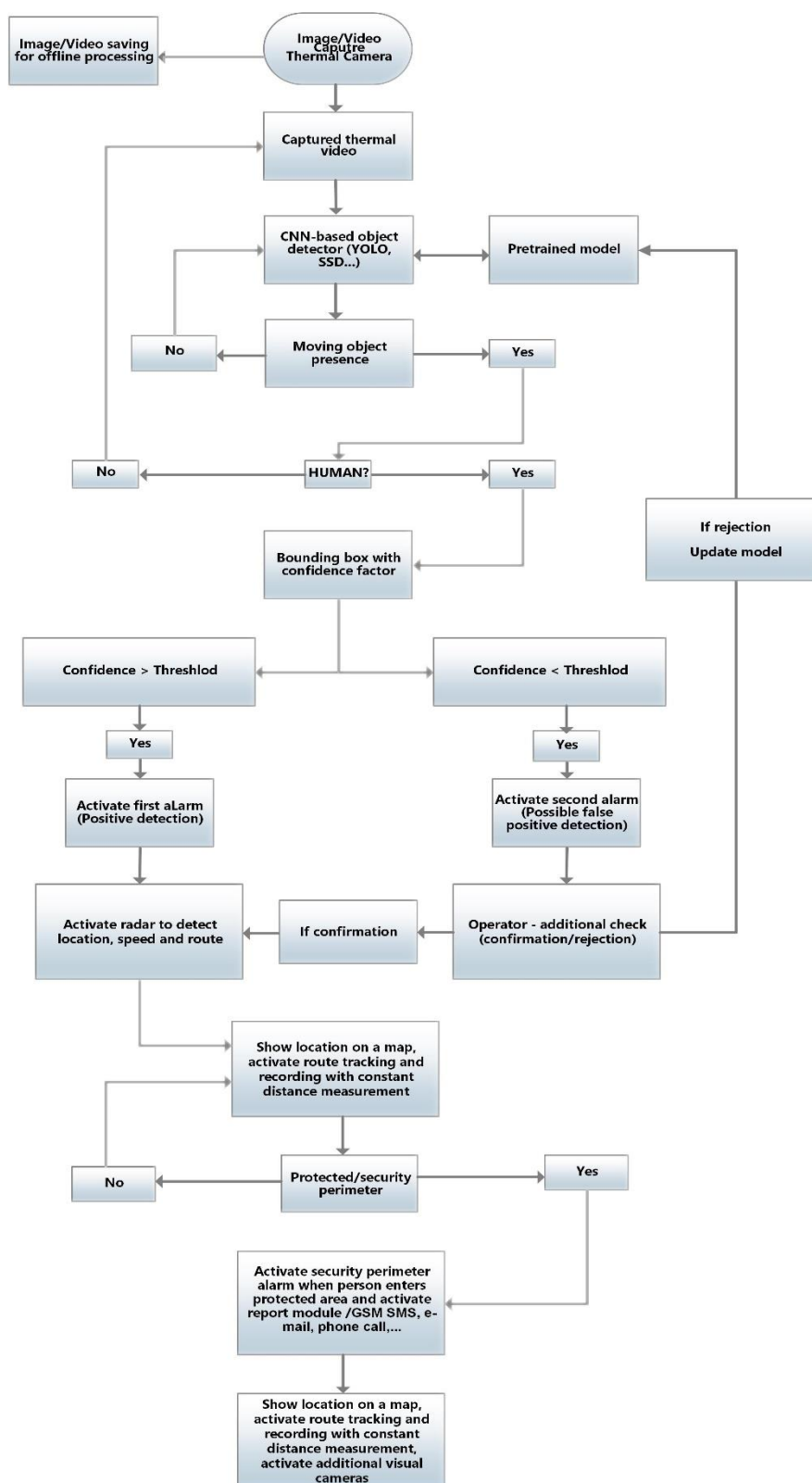
Druga značajna karakteristika sustava je postizanje visoke točnosti u radu istog. Ova karakteristika ovisi o izboru najboljih aplikativnih rješenja i implementacije metoda koje će postići najbolji rezultat u kontekstu detekcije i raspoznavanja, odnosno utvrđivanja da je riječ o osobi, a ne o životinji ili drugom tipu pokretnog objekta koji ne predstavlja prijetnju štićenom objektu ili prostoru. Idealna mjera je postizanje točnosti od 100%, bez lažno pozitivnih detekcija jer je jedino u tom slučaju potrebna minimalan, odnosno gotove je nepotreban angažman ljudske posade u cilju potvrde da su detektirani objekti doista neovlaštene osobe. U drugim slučajevima, kada eventualno postoji sumnja u točnost detekcije, isto zahtijeva angažman ljudske posade za potvrdu ili odbacivanje rezultata da je riječ o osobi, a kako je sustav predviđen i za primjenu na teško pristupačnim područjima, takav angažman je zahtjevan, kako vremenski tako i u smislu mobilizacije tehničke oprema (vozila, besposadne letjelice ili sl.). Dodatna otegotna okolnost je predviđena primjena ovakvog sustava u otežanim vremenskim uvjetima (kiša, snijeg, magla, i sl.), što onda zahtijeva i dodatni angažman ljudske posade u cilju samo potvrde ili odbacivanja da je detektirani objekt osoba, a ne životinja. Stoga je

postizanje visoke točnosti nužnost, jer i najmanji postotak lažno pozitivnih detekcija smanjuje pouzdanost sustava. Slučajeve lažno pozitivnih detekcija moguće je riješiti i kroz angažman ljudskog operatora koji će vizualno pregledati snimku te utvrditi ili odbaciti pretpostavku da je riječ o osobi, međutim sustav koji se predstavlja tip nenadziranog sustava, potpuno autonomnog s visokim stupnjem umjetne inteligencije, jer samo takav sustav smanjuje potrebu za angažmanom ljudske posade koja je zapravo najskuplji element svakog sigurnosnog sustava, pa tako i sustava za nadzor državne granice ili drugih štićenih objekata i prostora, te značajnih infrastrukturnih objekata i sl.

Također treba napomenuti da postoje komercijalna rješenja dostupna na tržištu koja se temelje na termalnoj kameri ili sustavu termalnih kamera za nadzor štićenih prostora i područja. Pored navedenog, postoje i sustavi koji kombiniraju radar za detekciju udaljenosti i smjer kretanja detektiranog objekta sa termalnom kamerom za raspoznavanje detektiranog objekta. Međutim, kako je riječ o komercijalnim sustavima, u promotivnim materijalima, osim kod jednog, uglavnom ne spominje da li se sustav detekcije temelji na neuronskoj mreži ili klasičnim klasifikatorima. Osim toga, također se ne spominje da li su takvi sustavi, osim jednog, imaju mogućnost raspoznavanja između osobe i drugih detektiranih objekata. Nadalje, u dostupnim promotivnim video zapisima nije vidljivo da takvi sustavi uz detekciju pokazuju i vjerojatnost da li je riječ o osobi niti pokazuju naziv detektiranog objekta. Nekoliko primjera postojećih sustava: PTZ sustavi izraelske tvrtke OPGAL temeljeni na više kamera (engl. *Multi-Camera PTZ Systems, PTZ – Pan-Tilt-Zoom*)) [URL: <https://www.opgal.com/ptz-cameras/>], sustavi termalnog nadzora tvrtke Stanley [URL: <https://www.stanleysecuritysolutions.com/thermaldetection>], glede kojih je navedeno da imaju mogućnost raspoznavanja osoba od drugih objekata na slikama i videozapisima, međutim na navedenoj internetskoj stranici nije javno dostupan video iz kojeg bi to bilo vidljivo. Nadalje, kineska tvrtka HIKVISION u svojoj ponudi također ima sustave za video nadzor temeljene na termalnoj kameri, a u dostupnom promotivnom video zapisu istaknuto je da prateći softverski paket ima mogućnost raspoznavanja osoba i vozila od drugih pokretnih objekata kao što životinje, lišće i slično. [URL: <https://www.hikvision.com/en/products/Thermal-Products/Security-thermal-cameras/>]. Francuska tvrtka HGH Infrared Systems također razvija tržišne sustave nadzora temeljene na termalnoj kameri naziva Spynel koji su upareni s

popratnim aplikativnim rješenjem naziva Cyclope. Glede njihova sustava, iz promotivnih video zapisa razvidno je da Cyclope nema mogućnosti automatskog raspoznavanja, odnosno klasifikacije detektiranih objekata nego taj dio ovisi isključivo o operateru, odnosno ljudskom faktoru [URL: <https://www.hgh-infrared.com/Products/Optronics-for-Security/Spynel-S-Panoramic-Detection-System>]. Pored navedenih, sustave nadzora temeljene na termalnoj kameri također razvija i američka tvrtka FLIR, jedan od najpoznatijih svjetskih proizvođača termalnih kamera, ali i zemaljskih radara [URL: <https://www.flir.com/security/perimeter-protection/>]. Ovakve sustave razvija i kanadska tvrtka Infiniti Electro-Optics (dio Ascendent Technology Group, Inc.), međutim na njihovim internetskim stranicama nema navoda o popratnim aplikativnim rješenjima za detekciju i raspoznavanje osoba. [URL: <https://www.infiniioptics.com/>]

7.1 Arhitektura sustava za detekciju i raspoznavanje osoba



Slika 81. Moguća arhitektura sustava za detekciju i raspoznavanje osoba temeljem na termalnoj kameri

7.2 Modul za snimanje

Ulazni senzor računalnog vida je ranije spomenuta termalna kamera. Ranije je također navedeno postojanje različitih dijelova IR spektra (NIR, SWIR, MWIR, LWIR), međutim ovdje će biti naglasak na snimanju unutar LWIR područja. Navedeno područje predstavlja pasivno snimanje topline emitirane od strane promatranog objekta. Nadalje, LWIR područje pokriva valnu duljinu od 8 do 12 μm , (po nekim autorima i do 14 μm), a EM valovi ovih valnih duljina imaju najmanju frekvenciju od svih dijelova IR spektra, ali zbog toga ima veću energiju te njihova detekcija i snimanje moguće i sa većih udaljenosti. Svrha modula za snimanje je upravljanje termalnom kamerom, snimanje i prosljeđivanje snimljenog videa u dva smjera. Jedan smjer je snimanje video zapisa prije obrade, odnosno video zapisa na kojem nisu detektirane osobe, odnosno drugi detektirani objekti. Drugi smjer je prosljeđivanje snimljenog video zapisa prema detektoru objekata temeljenom na metodama dubokog učenja za provođenje *online* obrade i detekcije, odnosno raspoznavanja detektiranih objekata.

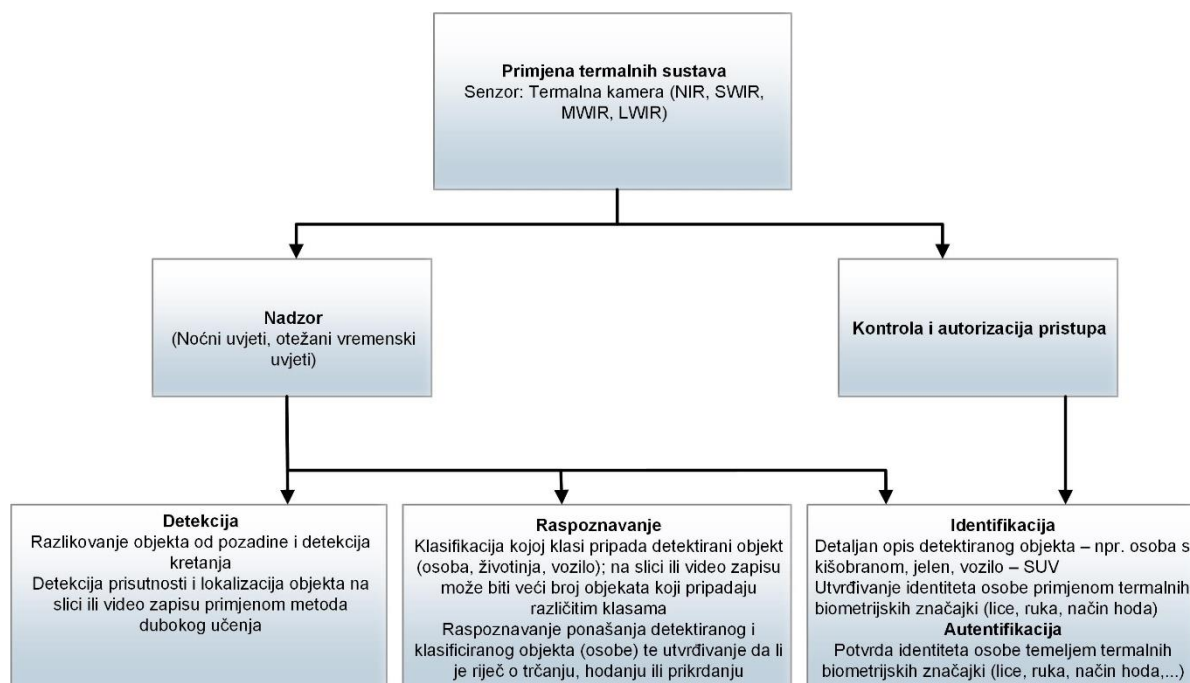
7.3 Modul za detekciju i klasifikaciju temeljen na dubokom učenju

Detektor objekata je centralni modul sustava za detekciju i raspoznavanje osoba na termalnim slikama odnosno video zapisima, bilo da je riječ o *offline* ili *online* obradi. Naglasak je na primjeni *online* obrade i detekcije odnosno raspoznavanja, jer je riječ o obradi u stvarnom vremenu, te je moguća pravovremena reakcija najbliže ljudske posade, u cilju terenske potvrde ili u cilju hvatanja osobe koja se neovlašteno kreće unutar štićenog područja. Iz ranije prikazanih rezultata razvidno je da je moguće postići obradu u realnom vremenu s obzirom na mogućnost obrade velikog broja sličica u sekundi.

Pored navedenog, u sustav se može implementirati podmodul koji će tijekom samog procesa obrade video signala, dodatno spremati i zasebno, drugim tipom alarma upozoriti operatora o čemu je riječ te zatražiti provjeru te potvrdu ili odbacivanje detekcije. Nadalje, nakon operatorova pregleda izdvojene slike, on će detekciju potvrditi ili odbaciti. Ukoliko detekciju potvrdi, proces se dalje nastavlja, te se aktivira radar za detekciju lokacije, brzine i smjera kretanja. Ukoliko operator odbaci sliku, ona se prosljeđuje modulu koji sadrži model te ažurira model.

Mogućnosti detekcije i raspoznavanja, kao i rezultati provedenih eksperimenata koji prikazuju uspješnost detekcije osoba te raspoznavanja od životinja prethodno je

detaljno opisano u poglavlju 6. Međutim, ovdje treba dodatno pojasniti nekoliko osnovnih pojmova koji se odnose na detekciju, raspoznavanje i identifikaciju detektiranih objekata. Međuodnos detekcije, raspoznavanja i identifikacije prikazan je na slici 80.



Slika 82. Primjena termalnih sustava za nadzor u kontekstu nacionalne sigurnosti

Ranije je spomenuta potreba za implementacijom sustava za praćenje osoba kako bi se utvrdio smjer i brzina kretanja detektirane osobe te procijenila stvarna namjera o prolasku granične crte na državnoj granici ili druge definirane granice štićenog područja. Za ovaj modul je potrebna implementacija dodatnih algoritama koji će izračunati smjer i brzinu kretanja uz naznaku vjerojatnosti da će detektirana osoba ili skupina doista i prijeći granicu nadziranog odnosno štićenog područja. Smjer i brzinu kretanja moguće je izračunati iz podataka sa ulazne video snimke ili podatke o smjeru i brzini kretanja može dati radar za detekciju objekata na zemlji.

7.4 Izvještajni modul

S obzirom na uvijek moguća odstupanja i lažno pozitivne detekcije, nužnost je implementacija označavanja detektirane osobe s izraženim postotkom pouzdanosti da je doista riječ o osobi, a ne npr. životinji, s obzirom da će osobe prilikom prolaska nadziranih i ulaska u štićene prostore, osim uspravnog hoda, biti sklone i promjeni položaja tijela te će hodati pogureno ili četveronoške, što može dovesti do lažno negativne detekcije. Također je moguće da neka divlja životinja bude zamijenjena za

osobu koja se kreće četveronoške ili puzi po tlu, što će biti prikazano kao lažno pozitivna detekcija, što može aktivirati ugrađene alarmne sustave i aktivirati najbližu ljudsku posadu te odvući pažnju od neke stvarno pozitivne detekcije na drugom dijelu nadziranog ili štićenog prostora. Pored navedenog, implementacija sustava za označavanje detektirane osobe pomaže u daljnjem koraku a to je praćenje kretanja u nadziranom području. Naime, sustav je zamišljen da nadzire određeno štićeno područje koje je omeđeno granicom koja u stvarnosti može (ograđeno žicom ili sl.), ali i ne mora postojati, kao što je najčešći slučaj kod državnih granica. U takvim uvjetima najčešće ne postoji fizička prepreka koja bi onemogućavala ili otežala prolazak granične crte, stoga je uz pomoć praćenja kretanja detektiranih osoba znatno olakšano praćenje smjera kretanja detektirane osobe ili skupine osoba. Navedeno dodatno pomaže ljudskoj posadi da procijeni da li detektirane osoba imaju stvarnu namjeru prelaska granične crte.

Za pouzdano utvrđivanje smjera kretanja kao i detekcije da je detektirana osoba prešla granicu nadziranog odnosno štićenog prostora, nužna je implementacija odgovarajuće kartografske ili GIS podloge. Temeljem podataka dobivenih iz video snimke ili s radara za detekciju objekata na zemlji te utvrđivanja stvarnog položaja na kartografskoj podlozi sustav može utvrditi da li je osoba prešla graničnu crtu, odnosno koliko blizu iste se nalazi, te aktivirati odgovarajući alarmni modul za uzbunjivanje ljudske posade koja se nalazi najbliže području u kojem je detektirano kretanje, te je utvrđeno da je kretanje rezultat hoda osobe ili skupine osoba.

Posljednji modul sustava za detekciju i raspoznavanje osoba u otežanim vremenskim uvjetima je također alarmni modul koji će se aktivirati tek kada su ispunjeni svi prethodni uvjeti. Naime, ovaj alarmni modul bi se trebao aktivirati tek kada je visoki postotak pouzdanosti da je doista riječ o osobi ili skupini osoba, odnosno kada postotak pouzdanosti prelazi postavljenu vrijednost praga, zatim kada utvrđeni smjer kretanja također s visokim postotkom pouzdanosti ukazuje da će detektirana osoba ili skupina doista prijeći granicu nadziranog odnosno štićenog područja, te kada, temeljem podataka o brzini kretanja, bude izračunato vrijeme prelaska granične crte, što ostavlja dovoljno vremena za pravodobnu i adekvatnu reakciju najbliže ljudske posade.

7.5. Mogućnosti skrivanja od detekcije termalnom kamerom

Koliko se radi na razvoju i istraživanju sustava detekciju i raspoznavanje osoba, a u ovom radu su opisani rezultati i arhitektura sustava temeljenog na termalnoj kameri, tako se paralelno radi na istraživanju i razvoju materijala i metoda koje će omogućiti skrivanje od detekcije ovakvih sustava. Iako termalna kamera ima mogućnost snimanja objekata u svim vremenskim uvjetima, a uz potporu metoda dubokog učenja i mogućnost detekcije i raspoznavanja, postoje mogućnosti kako se uspješno sakriti od takvih sustava.

Jedan od najjednostavnijih i najjeftinijih načina je korištenje pokrivača koji se uobičajeno koristi za očuvanje tjelesne topline odnosno tzv. svemirskog pokrivača (engl. *space blanket*) [320, 321], koji u potpunosti blokira širenje toplinskog zračenja, te osoba koje je njime pokrivena nije vidljiva na termalnoj kameri. Međutim, problem korištenja ovog tipa kamuflaže za izbjegavanje detekcije termalnom kamerom je to što pojedini dijelovi tijela postaju vidljivi onaj trenutak kada više nisu prekriveni, a pokrivač je izuzetno tanak i lagan zbog čega je osjetljiv na pomake prilikom kretanja ili pri vjetrovitom vremenu, stoga je i njegova stvarna primjena upitna. Unatoč tomu, za objekte koji miruju, pruža savršenu zaštitu od detekcije termalnom kamerom.

Nadalje, na tržištu postoje i određena komercijalna rješenja kao što tekstili tvrtke UF-PRO [322], za koje proizvođač navodi da omogućuje uspješno skrivanje od termalnih IR kamera, te iako na internetskim stranicama proizvođača nisu navedeni detaljniji detalji glede same tehnologije, istaknuto je da njihov tekstil od kojeg je moguće izraditi uniforme i drugu odjeću, nosioca uspješno skriva od NIR, SWIR i MWIR IR kamera, dok glede LWIR spektra nema nikakvih navoda. Nadalje, na internetskim stranicama proizvođača nema izvornih snimki, nego je riječ o promotivnom videu koji sadrži značajan broj specijalnih efekata, te nije razvidno koliko je materijal, od kojeg je moguće izraditi uniforme koje će pružati značajno bolju zaštitu od termalne detekcije nego ranije spomenuti "svemirski pokrivač". Na navedeno je na forumu [323] komentare ostavilo nekoliko korisnika, na koje je proizvođač odgovorio da je tkanina još, kao i sama tehnologija još uvijek u fazi testiranja. Navode o sličnim svojstvima prezentirala je i izraelska tvrtka Fibrotex [324], tvrtka koja također proizvodi materijal od kojeg je načinjeno odijelo naziva "*Nightwalker*", odnosno "noćni šetač" za kojeg je navedeno da uspješno skriva od termalnih kamera, a priloženi video pokazuje osobe koje nose navedeni tekstil i jedna koja ne nosi taj materijal, na udaljenosti od 900

metara, te je prikazan termalni otisak tih osoba. Na slici je vidljivo da je termalni otisak osobe koja nosi ovaj materijal doista smanjen, međutim i dalje je vidljivo kretanje osobe, iako su nijanse prikaza gotovo crne, ali je kretanje zamjetno. Niti jedna niti druga tvrtka u svojim promotivnim materijalima nisu priložili detaljne specifikacije ispitivanja niti snimke termalne kamere na manjim udaljenostima kako bi se mogla donijeti utemeljena ocjena da li ovakvi materijali doista mogu u potpunosti prikriti termalnu siluetu osobe. Također niti jedna tvrtka nije navela podatke da sam materijal preuzima dio tjelesne topline i nakon koliko vremena i to odijelo samo po sebi počinje kreirati termalnu siluetu.

Iako predstavljeni materijali blokiraju distribuciju tjelesne topline, na promotivnim video zapisima se, unatoč tomu, može zamijetiti kretanje osoba, što ostavlja mogućnost detekcije zemaljskim radarom ili eventualno pasivnim IR sustavom.

8 Zaključak

Temeljni cilj doktorskog rada je ispitati mogućnosti, uspješnost i pouzdanost konvolucijskih neuronskih mreža u kombinaciji s termovizijskom kamerom za detekciju i raspoznavanje osoba u otežanim vremenskim uvjetima (kiša, magla). U tu svrhu korišten je YOLOv3, model neuronske mreže koji je postigao izvrsne rezultate prepoznavanja objekata na RGB slikama, s tim da je u ovom slučaju korišten na termovizijskim slikama i videozapisima. Osnovni eksperiment, a to je detekcija i raspoznavanje osoba u otežanim vremenskim uvjetima, proveden je na prilagođenom skupu podataka snimljenom tijekom zime u različitim vremenskim uvjetima (vedro vrijeme, kiša, magla), po noći i na različitim udaljenostima od kamere, u rasponu od 30 do 215 m. Kretanje osoba variralo je od uobičajenog hodanja, trčanja do pokušaja skrivanja puzeći ili hodajući pogureno kako bi simulirali ilegalna kretanja oko državne granice ili u drugim nadziranim područjima.

Performanse izvorne mreže YOLOv3 trenirane na COCO RGB skupu podataka korištene su kao osnovni model (s oznakom bY) i uspoređene su s modelom iste arhitekture koja je dodatno trenirana na skupu termovizijskih slika snimljenih po vedrom vremenu, a testiranje uspješnosti detekcije provedeno je na slikama snimljenima po jakoj kiši i gustoj magli. Testiranje korištenjem baznog modela treniranog na COCO RGB skupu slika, provedeno je uz pretpostavku postojanja određenih sličnosti između RGB i termovizijskih slika, međutim originalni YOLO model (bY), za detekciju osoba (klasa Human ili Person) postigao je prosječnu preciznost (AP) od samo 19,63%, uz odziv od 15,5% sa 100% preciznošću, što je znatno niži rezultat od prijavljenog AP-a od oko 90% za klasu Person u RGB slikama. Iako je postotak detekcije i prepoznavanja modela treniranog na RGB slikama mali, ipak je uspijevaao prepoznati osobe na određenom broju termovizijskih slika te je poslužio kao dobra polazna točka za treniranje novog, zasebnog modela treniranog na termovizijskim slikama.

Model (s oznakom tY) treniran je na skupu od oko 3.000 termovizijskih slika iz skupa kreiranog za potrebe ovog doktorskog rada te je postigao znatno bolje rezultate na testnom setu, s vrijednošću AP od 97,93% za sve vremenske uvjete. Treniranje modela na malom uzorku slika iz originalnog trening skupa također se pokazao dovoljnim za postizanje izvrsnih rezultata detekcije osoba u svim testiranim scenarijima, odnosno po različitim vremenskim uvjetima, položajima tijela i varijacijama

te udaljenosti od kamere. Također je pokazano da se podaci prikupljeni po vedrom vremenu mogu uspješno upotrijebiti za treniranje modela (s oznakom tY_clear) koji će biti primijenjen na slike i video zapise snimljene i po jakoj kiši i gustoj magli, u kojima postiže AP od gotovo 100%.

Pored toga, treniran je model za raspoznavanje, odnosno razlikovanje osoba od drugih objekata na snimkama (u ovom slučaju razlikovanje čovjeka od psa) na termalnim slikama pri čemu je postignut mAP od 97,98%, što ukazuje na mogućnost korištenja ovog ili sličnih CNN modela za razvoj samostalnog sustava za automatski nadzor zaštićenih objekata i područja, uz minimalan angažman ljudske posade.

Nadalje, dodatno treniranje YOLOv3 modela pokazalo je da je moguće dobiti pouzdan model korištenjem relativno malog broja slika i s malim brojem iteracija, što znatno skraćuje potrebno vrijeme treninga. Uz to, tako trenirani YOLOv3 model pokazuje dobra generalizacijska svojstva u odnosu na rezultate postignute treniranjem na eksternim skupovima slika. Još bolji model može se dobiti kombiniranjem svih skupova korištenih u ovom radu, što će skup dodatno proširiti različitim prikazima termalnih silueta, što je u slučaju treniranja modela za detekciju i raspoznavanje osoba na termalnim slikama ključno za postizanje izuzetno pouzdanih rezultata za implementaciju modela u stvarnim uvjetima.

Iako su postignuti izvrsni rezultati u području detekcije i raspoznavanja osoba, dodatna generalizacijska svojstva modela, a samim tim i povećanje pouzdanosti modela moguće je postići proširenjem osnovnog skupa slika dodavanjem novih slika snimljenih u drugim uvjetima (primarno po snijegu), kao i novih klasa objekata, nakon čega bi model trebao imati mogućnost raspoznavanja i šireg spektra objekata, te svakako pouzdanije detekcije i raspoznavanja osoba.

Pored navedenog, eksperimenti iz ovog rada su, osim navedenih doprinosa, pokazali i da kombinacija termovizije i neuronskih mreža otvara cijeli spektar novih mogućnosti implementacije u području sigurnosti, posebice zbog činjenice da su moduli termalnih kamera dostupni na širokom tržištu po relativno povoljnim cijenama, zbog čega i postoji ekonomičnost razvoja te implementacije ovakvih sustava u svakodnevne aktivnosti (od kontrole pristupa vlastitom objektu – npr. uz pomoć termovizijskog biometrijskog sustava temeljenog ili na cijelom tijelu, odnosno načinu hoda ili samo temeljem snimke

lica do podizanja razine sigurnosti na bankomatima bez potrebe za korištenjem bankovnih kartica i PIN-ova).

9 Znanstveni doprinos i perspektiva budućih istraživanja

Ishodište doktorskog rada bilo je u očekivanim znanstvenim doprinosima koji su navedeni u poglavlju 1.3, ali ćemo ih ovdje ponoviti uz obrazloženje za svaki pojedinačno te naglasak da li je isti ostvaren.

Sukladno navedenom, očekivani znanstveni doprinosi su bili:

- baza slika osoba snimljenih termovizijskom kamerom u otežanim vremenskim uvjetima uz anotacije i slike pogodne za usporedbu i raspoznavanje – kao što je razvidno iz prethodnih poglavlja u kojima je predstavljena baza slika sa pripadajućim podskupovima slika, te je ista korištena u eksperimentalnom dijelu doktorskog rada. Svakako treba istaknuti da kreirana baza slika ima znatno širi potencijal od početne namjere kreiranja iste za detekciju i raspoznavanje osoba u otežanim vremenskim uvjetima. Naime, razvidno je da baza, osim termalnih slika snimljenih u različitim vremenskim uvjetima sadrži i termalne slike osoba snimljene na različitim udaljenostima, pri različitim položajima tijela, brzinama kretanja, kao i termalne slike osoba zajedno s drugim objektima na slikama (pas), što je iskorišteno za dokazivanje mogućnosti neuronskih mreža za uspješno raspoznavanje osoba od životinja, kao i za uspješno raspoznavanje različitih položaja i brzina kretanja osoba. Upravo ovakva struktura kreiranu bazu uvelike čini jedinstvenom u odnosu na druge slične skupove termovizijskih slika koji se već naširoko koriste u domeni detekcije i raspoznavanja osoba na termovizijskim slikama, Kreiranu bazu podataka također je moguće koristiti u daljnjim istraživanjima iz područja računalnog vida u kontekstu detekcija osoba u različitim situacijama, npr. za učenje modela koji mogu biti korišteni za traganje i spašavanje osoba primjenom IR termovizije i sl. S tim u svezi potrebno je napomenuti da je ovo vjerojatno prvi ovakav skup javno dostupan znanstvenoj zajednici za daljnja istraživanja.
- postupak detekcije osoba na termovizijskim slikama korištenjem temperaturnih pragova i biometrijskih značajki – vezano za ovaj doprinos potrebno je istaknuti da je korištenje temperaturnih pragova kao jednog od klasifikacijskih parametara bilo početna ideja, međutim tijekom razrade teme, posebice tijekom

izvođenja eksperimenata korištenjem metoda dubokog učenja, odnosno YOLO detektora, ovaj parametar se pokazao nepotrebnim za ostvarivanje dobrih rezultata odnosno visoke stope pouzdanosti detekcije osoba na termovizijskim slikama. Tako je, umjesto temperaturnih pragova u eksperimentalnom dijelu korišten detektor objekata temeljen na konvolucijskoj neuronskoj mreži, što je aktualno standard u području računalnog vida i detekcije, odnosno raspoznavanja objekata. Glede drugog dijela ovog doprinosa, razvidno je da su u radu korištene tzv. meke biometrijske značajke odnosno termalna silueta s obzirom da termovizijska kamera, pogotovo na većim udaljenostima i u otežanim vremenskim uvjetima može snimati samo termalnu siluetu osobe, što se kroz postignute rezultate pokazalo kao dovoljan parametar za pouzdanu detekciju i raspoznavanje osoba kako u otežanim vremenskim uvjetima tako i u odnosu na druge objekte iz okoline.

- arhitektura sustava za detekciju i klasifikaciju osoba u realnom vremenu, koja se temelji na sustavu infracrvenih termovizijskih kamera – navedeni sustav predstavljen je u poglavlju 7, a razvidno je da ovakav sustav pogotovo kada se temelji na konvolucijskim neuronskim mrežama donosi cijeli niz novih mogućnosti koje je moguće realizirati u realnom vremenu. Isto se očituje u tome, da osim detekcije kretanja osoba, ovakav sustav može imati i mogućnost raspoznavanja osobe od drugih objekata na slikama odnosno video zapisu, prvenstveno u odnosu na životinje i druge objekte iz okoline, zatim da ovakav sustav uparen s dodatnim senzorima i opremom može osigurati visoku razinu autonomne zaštite štićenog objekta ili područja, a isto se može prilagoditi i različitim razinama i potrebama.
- postupak ocjenjivanja uspješnosti primjene autonomnih infracrvenih termovizijskih sustava u području sigurnosti - za postizanje ovog doprinosa također nije bilo potrebe za razvojem novog postupka ocjenjivanja s obzirom da je početna ideja o korištenju temperaturnih pragova napuštena nakon postizanja vrhunskih rezultata primjenom metoda dubokog učenja. Sukladno tomu, korištena je standardna metrika za mjerenje uspješnosti odnosno pouzdanosti detekcije objekata u području računalnog vida primarno razvijena za slike snimljene standardnim optičkim kamerama, pri čemu je dokazano da se ova metrika može uspješno primijeniti i na područje detekcije osoba primjenom termovizijskih kamera u otežanim vremenskim uvjetima.

Pored navedenog, kroz izvedene eksperimente dokazana je velika mogućnost primjene konvolucijskih neuronskih mreža treniranih na termovizijskim slikama za detekciju i raspoznavanje osoba u otežanim vremenskim uvjetima, posebice po noći, kada standardne optičke kamere nemaju nikakve mogućnosti detektirati osobe, posebice na većim udaljenostima i u potpunom mraku. Nadalje, dodatni doprinos također se očituje i u dokazivanju mogućnosti neuronskih mreža za pouzdano razlikovanje osoba od drugih objekata u nadziranom području. Ovakva mogućnost, posebice kada je povezana s alarmnim modulom uvelike smanjuje potrebu za angažmanom ljudskih posada u nadziranim područjima (primjerice zona državne granice gdje je uglavnom riječ o teško pristupačnim područjima) što svakako pridonosi i ekonomičnosti nadzora državne granice i drugih štićenih prostora i područja. Također je dokazano da je, kada je riječ o detekciji i raspoznavanju osoba, moguće učenje modela na malom skupu termalnih slika koji je moguće uspješno implementirati za detekciju i raspoznavanje osoba na termalnim slikama što također dovodi do značajnog skraćivanja vremena potrebnog za pripremu materijala za skup za učenje, zbog čega se i vrijeme učenja modela također značajno skraćuje, te se ubrzava cjelokupni proces učenja i implementacije naučenog modela u realne sustave.

Daljnji rad i znanstveni doprinosi temeljeni na ovom doktorskom radu kao svojevrsnoj polaznoj točki usmjeren je na eksperimentiranje i ispitivanje mogućnosti detektora objekata temeljenih na neuronskim mrežama i kao i učenje modela dubokog učenja za identifikaciju osoba na termovizijskim slikama prema načinu hoda koji je jedinstvena meka biometrijska značajka svake osobe. S obzirom na specifičnost termovizijskih slika, provedena su preliminarne testiranja mogućnosti ljudskog mozga, odnosno ljudske sposobnosti za identifikaciju osoba na termovizijskim slikama. Naime, provedeni eksperiment je uključivao osobu – promatrača koja dio snimanih osoba poznaje jako dobro i duži niz godina, a dio poznaje kraće vrijeme i površno, te je unatoč tomu uspješno, sa 100 % pouzdanosti identificirala dobro poznate osobe na termovizijskim slikama i video zapisima oslanjajući se isključivo na termalnu siluetu i način hoda tih osoba. S druge strane, osobe koje promatrač poznaje kraće vrijeme, prepoznala je s 50 % pouzdanosti, odnosno identificirala je po imenu jednu od dvije takve osobe. Sukladno tomu, postavlja se i hipoteza da je moguće naučiti model temeljen na neuronskoj mreži koji će pouzdano i kontinuirano imati mogućnost

identifikacije osoba na termovizijskim slikama koristeći hod kao meku biometrijsku značajku. U tom području javlja se cijela lepeza novih eksperimenata, a samim tim i mogućnosti implementacije u realnom svijetu za primjenu ovakvih sustava. Pored toga, buduća istraživanja moguće je usmjeriti i u područje identifikacije osoba temeljem termovizijskih slika lica, odnosno implementacija takvih sustava u područje sigurnosti i svakodnevne aktivnosti kao što je eventualna ugradnja takvih sustava primjerice u bankomate pri čemu se gotovo u potpunosti anulira potreba za korištenjem bankovnih kartica i PIN brojeva, jer vlastito lice, odnosno struktura krvnih žila lica jedinstvena je za svaku osobu, pri čemu osoba mora biti živa da bi termovizijska kamera uspješno snimila lice, a kroz kasniju obradu moguće je izlučiti uzorak krvnih žila lica te kroz sustav ga usporediti s uzorcima iz baze korisnika i sl.

Literatura

- [1] URL: <https://www.kbvresearch.com/video-surveillance-market/>
- [2] URL: <https://ourworldindata.org/terrorism#how-many-people-are-killed-by-terrorists-worldwide>
- [3] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." *CVPR* (1) 1 (2001): 511-518.
- [4] Eveland, Christopher K., Diego A. Socolinsky, and Lawrence B. Wolff. "Tracking human faces in infrared video." *Image and Vision Computing* 21.7 (2003): 579-590.
- [5] Krišto, Mate, and Marina Ivasic-Kos. "An overview of thermal face recognition methods." *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2018.
- [6] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." *international Conference on computer vision & Pattern Recognition (CVPR'05)*. Vol. 1. IEEE Computer Society, 2005.
- [7] Zhao, Zhong-Qiu, et al. "Object detection with deep learning: A review." *IEEE transactions on neural networks and learning systems* 30.11 (2019): 3212-3232.
- [8] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.
- [9] Girshick, Ross. "Fast R-CNN." *Proceedings of the IEEE international conference on computer vision*. 2015.
- [10] W. Liu et al., "SSD: Single shot multi-box detector," at a European conference on computer vision, 2016, pp. 21–37.
- [11] K. He, G. Gkioxari, and Dollar, "Mask r-CNN," in 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988.
- [12] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Advances in neural information processing systems*, 2016, pp. 379–387.
- [13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [14] Redmon, Joseph, and Ali Farhadi. "YOLO9000: better, faster, stronger." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- [15] Redmon, Joseph, and Ali Farhadi. "YOLOv3: An incremental improvement." *arXiv preprint arXiv:1804.02767* (2018).
- [16] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "YOLOv4: Optimal Speed and Accuracy of Object Detection." *arXiv preprint arXiv:2004.10934* (2020).

- [17] M. Buric, M. Pobar, and M. Ivasic-Kos, "Adapting YOLO network for Ball and Player Detection," in 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2019), 2019, pp. 845–851.
- [18] M. Buric, M. Pobar, and M. Ivasic-Kos, "Ball detection using YOLO and Mask R-CNN," in 2018 International Conference on Computational Science and Computational Intelligence (CSCI), 2018.
- [19] M. Ivasic-Kos and M. Pobar, "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), 2018, pp. 1–6.;
- [20] Laroca, R., Severo, E., Zanlorensi, L. A., Oliveira, L. S., Gonçalves, G. R., Schwartz, W. R., & Menotti, D. (2018, July). A robust real-time automatic license plate recognition based on the YOLO detector. In 2018 International Joint Conference on Neural Networks (IJCNN) (pp. 1-10). IEEE
- [21] Shen, D., Chen, X., Nguyen, M., & Yan, W. Q. (2018, April). Flame detection using deep learning. In 2018 4th International Conference on Control, Automation and Robotics (ICCAR) (pp. 416-420). IEEE
- [22] Zhang, Xiang, et al. "A Fast Learning Method for Accurate and Robust Lane Detection Using Two-Stage Feature Extraction with YOLOv3." *Sensors* 18.12 (2018): 4308.
- [23] George, J., Skaria, S., & Varun, V. V. (2018, February). Using YOLO based deep learning network for real-time detection and localization of lung nodules from low dose CT scans. In *Medical Imaging 2018: Computer-Aided Diagnosis* (Vol. 10575, p. 105751I). International Society for Optics and Photonics.
- [24] Kharchenko, V., & Chyrka, I. (2018, July). Detection of Airplanes on the Ground Using YOLO Neural Network. In 2018 IEEE 17th International Conference on Mathematical Methods in Electromagnetic Theory (MMET) (pp. 294-297). IEEE.
- [25] Krišto, Mate, and Marina Ivasic-Kos. "Thermal Imaging Dataset for Human Detection." 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, 2019
- [26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-CNN: Towards real time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.
- [27] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into High-Quality Object Detection, arXiv:1712.00726 [cs], Dec. 2017, Accessed: Jun. 08, 2020. [Online]. Available: <http://arxiv.org/abs/1712.00726>.
- [28] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully Convolutional One-Stage Object Detection," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), Oct. 2019, pp. 9626–9635, DOI: 10.1109/ICCV.2019.00972.
- [29] Lasanen, Roope. Infrared thermography in the evaluation of skin temperature. Diss. Thesis, University of Eastern Finland Dissertations in Forestry and Natural Sciences No186, Kuopio, 2015.
- [30] Vollmer, Michael, and M. Ã. Klaus-Peter. Infrared thermal imaging: fundamentals, research and applications. John Wiley & Sons, 2017., str. 6

- [31] [URL:
http://www.optris.com/software?file=tl_files/pdf/Downloads/Brochures%20US/optris-basic-brochure.pdf
- [32] Kaplan, Herbert. Practical applications of infrared thermal sensing and imaging equipment. Vol. 75. SPIE press, 2007. god., str. 5
- [33] URL: <https://www.micro-epsilon.com/download/products/dat--infrared-basics--en-us.pdf>
- [34] URL: <http://argosfp7project.blogspot.hr/2015/11/short-wave-infrared-swir-imaging.html>
- [35] Chrzanowski, Krzysztof. *Non-contact thermometry: measurement errors*. SPIE Polish Chapter, 2001., str. 18 - 20
- [36] Bhowmik, Mrinal Kanti, et al. "Thermal infrared face recognition—a biometric identification technique for robust security system." Reviews, refinements and new ideas in face recognition. InTech, 2011.
- [37] URL:
http://www.vworld.fr/images/banrotation/Bandeau_vieWTerra_sensors_01.png
- [38] Vollmer, Michael, and M. Å. Klaus-Peter. Infrared thermal imaging: fundamentals, research and applications. John Wiley & Sons, 2017., str. 352 – 359.
- [39] M. Riedl, "Optical Design Fundamentals for Infrared Systems", *Second Edition, SPIE Press, Bellingham, WA*, 2001., str. 5
- [40] Chrzanowski, Krzysztof. *Non-contact thermometry: measurement errors*. SPIE Polish Chapter, 2001., str. 22 – 27
- [41] M. Riedl, "Optical Design Fundamentals for Infrared Systems", *Second Edition, SPIE Press, Bellingham, WA*, 2001., str 7
- [42] Vollmer, Michael, and M. Å. Klaus-Peter. Infrared thermal imaging: fundamentals, research and applications. John Wiley & Sons, 2017., str. 32
- [43] A. Rogalski, K. Chrzanowski, "Infrared devices and techniques", *Opto-Electronics Review*, Vol. 10, No. 2, pp. 111 – 136, 2002.
- [44] C. Solomon, T. Breckon, "Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab", *Wiley Publishing*, 2011., str. 49
- [45] Rogalski, A. "History of infrared detectors." *Opto-Electronics Review* 20.3 (2012): 279-308.
- [46] URL: <https://www.britannica.com/technology/bolometer>, uvid 24.2.2018. godine
- [47] Rogalski, Antoni. "Infrared detectors: status and trends." *Progress in quantum electronics* 27.2-3 (2003): 59-210.
- [48] Kaplan, Herbert. Practical applications of infrared thermal sensing and imaging equipment. Vol. 75. SPIE press, 2007. god., str. 49-50
- [49] Vollmer, Michael, and M. Å. Klaus-Peter. Infrared thermal imaging: fundamentals, research and applications. John Wiley & Sons, 2017., str. 327
- [50] Bernard, Erwan and Rivi re, Nicolas and Renaudat, Mathieu and Guiset, Pierrick and P alat, Michel and Zenou, Emmanuel *Experiments and Models of Active and*

- Thermal Imaging Under Bad Weather Conditions*. (2013) In: SPIE Defence + Security, 23 September 2013 - 26 September 2013 (Dresden, Germany).
- [51] Chris Solomon and Toby Breckon; *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*; John Wiley & Sons, Ltd. 2011
 - [52] Williams, Thomas. *Thermal imaging cameras: characteristics and performance*. CRC Press, 2009.
 - [53] Vollmer, Michael, and M. Å. Klaus-Peter. *Infrared thermal imaging: fundamentals, research and applications*. John Wiley & Sons, 2017., str. 173
 - [54] Kaplan, Herbert. Practical applications of infrared thermal sensing and imaging equipment. Vol. 75. SPIE press, 2007., str. 28
 - [55] Vollmer, Michael, and M. Å. Klaus-Peter. *Infrared thermal imaging: fundamentals, research and applications*. John Wiley & Sons, 2017., str. 186
 - [56] Bhan, R. K., et al. "Uncooled infrared microbolometer arrays and their characterisation techniques." *Defence Science Journal* 59.6 (2009): 580.
 - [57] FLIR: Uncooled detectors for thermal imaging cameras. Technical note (2011). FLIR Commercial Vision Systems B.V.
 - [58] Gade, Rikke, and Thomas B. Moeslund. "Thermal cameras and applications: a survey." *Machine vision and applications* 25.1 (2014): 245-262.
 - [59] Ostrower, Daniel. "Optical thermal imaging—replacing microbolometer technology and achieving universal deployment." *III-Vs Review* 19.6 (2006): 24-27.
 - [60] Chrzanowski, Krzysztof. "Testing thermal imagers." Practical guide, MUT, Warszawa (2010).
 - [61] Rai, Mritunjay, Tanmoy Maity, and R. K. Yadav. "Thermal imaging system and its real time applications: a survey." *Journal of Engineering Technology* (ISSN: 0747-9964) 6.2 (2017): 290-303.
 - [62] Battalwar, Pooja, Janhvi Gokhale, and Utkarsha Bansod. "Infrared thermography and IR camera." *History* 1.2 (2015).
 - [63] Wardlaw, Jessica, et al. "A new approach to thermal imaging visualisation." EngD Group Project, University College London (2010).
 - [64] Havens, Kirk J., and Edward J. Sharp. *Thermal Imaging Techniques to Survey and Monitor Animals in the Wild: A Methodology*. Academic Press, 2015.
 - [65] Gim, LEE Cheow, EE Kok Tiong, and HENG Yinghui Elizabeth. "Performance challenges for high resolution imaging sensors for surveillance in tropical environment.", *DSTA HORIZONS*, 2015, pp. 80 – 88.
 - [66] Holst, Gerald C. *Common sense approach to thermal imaging*. Washington: SPIE Optical Engineering Press, 2000.
 - [67] Swinehart, D. F. "The beer-lambert law." *Journal of chemical education* 39.7 (1962): 333.
 - [68] Zhang, Huaizhong, et al. "A novel infrared video surveillance system using deep learning based techniques." *Multimedia Tools and Applications* 77.20 (2018): 26657-26676.

- [69] H. S. Hadi, M. Rosbi, and U. U. Sheikh, "A Review of Infrared Spectrum in Human Detection for Surveillance Systems," *Int. J. Interact. Digit. Media*, vol. 1, no. 3, pp. 13–20, 2013.
- [70] W. K. Wong, P. N. Tan, C. K. Loo, and W. S. Lim, "An Effective Surveillance System Using Thermal Camera," *J. Comput. Sci. Eng.*, vol. 3, no. 2, pp. 42–51, 2010.
- [71] Negied, Nermin K. "Human Biometrics: Moving Towards Thermal Imaging." *International Journal of Recent Technology and Engineering (IJRTE)*, ISSN (2014): 2277-3878.
- [72] E. T. Gilmore, C. Ugbome, C. Kim, "An IR-based Pedestrian Detection System Implemented with Matlab-Equipped Laptop and Low-Cost Microcontroller," *International Journal of Computer Science and Information Technology*, Vol. 3, No. 5, pp. 79 - 87, 2011.
- [73] R. Mieziako, D. Pokrajac, "People detection in low-resolution infrared videos," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Honeywell Labs, Minneapolis, MN, pp. 1 - 6, 2008.
- [74] E. Jeon, et al., "Human Detection Based on the Generation of a Background Image by Using a Far-Infrared Light Camera," *Sensors*, Vol. 15, No. 3, pp. 6763 - 6788, 2015.
- [75] M. Everingham, S. M. A. Eslami, L. Van Gool, C.K.I. Williams, J. Winn and A. Zisserman, "Challenges of Visual Classes in Pascal: Retrospective," *Int. J. Comput. Vis.*, Vol. 111, no. 1, p. 98-136, 2014
- [76] W. Wong, et al., "Omnidirectional Thermal Imaging Surveillance System Featuring Trespasser and Faint Detection," *International Journal of Image Processing (IJIP)*, Vol. 6, No. 6, pp. 518 - 538, 2011.
- [77] I. Riaz, J. Piao, H. Shin, "Human detection by using centrist features for thermal images," *IADIS International Journal on Computer Science and Information Systems*, Vol. 8, No. 2, pp. 1 - 11, 2013.
- [78] Wu, J. et al., 2011. Real-Time Human Detection Using Contour Cues. In *Proc. ICRA*, Shanghai, China, pp. 860-867.
- [79] Theodoridis, Sergios, and Konstantinos Koutroumbas. "Pattern recognition. 2003." *Google Scholar Google Scholar Digital Library Digital Library* (2009).
- [80] E. Chen, O. Haik, Y. Yitzhaky, "Classification of moving objects in the atmospherically degraded video," *Optical Engineering*, vol. 51, no. 10, p. 101710, 2012.
- [81] F. Coutts, S. Marshall, P. Murray, "Human detection and tracking through temporal feature recognition," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European, Lisbon*, pp. 2180 - 2184, 2014.
- [82] J. Castillo, et al., "Segmenting Humans from Mobile Thermal Infrared Imagery", in *3rd International Work-Conference on The Interplay Between Natural and Artificial Computation: Part II: Bioinspired Applications in Artificial and Natural Computation (IWINAC '09)*, Springer-Verlag, Berlin, Heidelberg, pp. 334 - 343, 2009.

- [83] Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (1981)
- [84] Akram, Saba, and Q. U. Ann. "Newton raphson method." *International Journal of Scientific & Engineering Research* 6.7 (2015): 1748-1752.
- [85] B. Besbes, et al., "Pedestrian Detection in Far-Infrared Daytime Images Using a Hierarchical Codebook of SURF", *Sensors*, Vol. 15, No. 4, pp. 8570 - 8594, 2015.
- [86] Bertozzi, M.; Broggi, A.; Felisa, M.; Vezzoni, G. Low-level Pedestrian Detection by means of Visible And Far Infra-red Tetra-vision. In Proceedings of the 2006 IEEE Intelligent Vehicles Symposium, Tokyo, Japan, 13–15 June 2006; pp. 231–236
- [87] J. Li, W. Gong, "Real Time Pedestrian Tracking using Thermal Infrared Imagery," *JCP*, Vol. 5, No. 10, pp. 1606 - 1613, 2010.
- [88] Baek, Jeonghyun, et al. "Efficient pedestrian detection at nighttime using a thermal camera." *Sensors* 17.8 (2017): 1850.
- [89] Hwang, S., Park, J., Kim, N., Choi, Y., & Kweon, I. S. (2015). Multispectral pedestrian detection: Benchmark dataset and baseline. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07–12–June, 1037–1045.
- [90] F. Lee, F. Chen, J. Liu, "Infrared Thermal Imaging System on a Mobile Phone," *Sensors*, Vol. 15, No. 5, pp. 10166-10179, 2015.
- [91] Ma, Yalong, et al. "Pedestrian detection and tracking from low-resolution unmanned aerial vehicle thermal imagery." *Sensors* 16.4 (2016): 446.
- [92] Rajkumar, S., and PVSSR Chandra Mouli. "Target detection in infrared images using block-based approach." *Informatics and Communication Technologies for Societal Development*. Springer, New Delhi, 2015. 9-16.
- [93] Li, Jianfu, et al. "Robust pedestrian detection in thermal infrared imagery using the wavelet transform." *Infrared Physics & Technology* 53.4 (2010): 267-273.
- [94] Soundrapandiyan, Rajkumar, and PVSSR Chandra Mouli. "Robust pedestrian detection in infrared images using rotation and scale invariant-based structure element descriptor." *International Journal of Signal and Imaging Systems Engineering* 10.3 (2017): 157-167.
- [95] Li, Wei, et al. "An effective approach to pedestrian detection in thermal imagery." 2012 8th International Conference on Natural Computation. IEEE, 2012.
- [96] Zhao, X., He, Z., Zhang, S., & Liang, D. (2015). Robust pedestrian detection in thermal infrared imagery using a shape distribution histogram feature and modified sparse representation classification. *Pattern Recognition*, vol. 48, no. 6, pp. 1947-1960.
- [97] Askari, M., Asadi, M., Asilian, B. A., & Ebrahimpour, H. (2016). Isolated Persian/Arabic handwriting characters: Derivative projection profile features, implemented on GPUs. *Journal of AI and Data Mining*, vol. 4, pp. 9-17.
- [98] Gonzalez, R. C. (2002). Richard E. woods. Digital image processing.

- [99] Olmeda, Daniel, Arturo de la Escalera, and José María Armingol. "Far infrared pedestrian detection and tracking for night driving." *Robotica* 29.4 (2011): 495-505.
- [100] Kim, Jong Hyun, Hyung Gil Hong, and Kang Ryoung Park. "Convolutional neural network-based human detection in nighttime images using visible light camera sensors." *Sensors* 17.5 (2017): 1065.
- [101] Yang, C., Liu, H., Liao, S., & Wang, S. (2015). Pedestrian Detection in Thermal Infrared Image Using Extreme Learning Machine. In *Proceedings of ELM*, Springer, Cham. vol. 2, pp. 31-40.
- [102] Cheng, M. M., Zhang, Z., Lin, W. Y., & Torr, P. (2014). BING: Binarized normed gradients for objectness estimation at 300fps. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3286-3293.
- [103] Cao, Jiuwen, Kezhi Mao, Jonathan Wu, and Amaury Lendasse. (2015). *Proceedings of ELM-2015*, Springer, Vol. 2.
- [104] Marina Ivašić-Kos, Mate Krišto, and Miran Pobar. 2019. Human Detection in Thermal Imaging Using YOLO. In *Proceedings of the 2019 5th International Conference on Computer and Technology Applications (ICCTA 2019)*. ACM, New York, NY, USA, 20-24. DOI: <https://doi.org/10.1145/3323933.3324076>
- [105] Ivasic-Kos, Marina, Mate Kristo, and Miran Pobar. "Person Detection in thermal videos using YOLO." *Proceedings of SAI Intelligent Systems Conference*. Springer, Cham, 2019.
- [106] Krišto, Mate, Marina Ivasic-Kos, and Miran Pobar. "Thermal Object Detection in Difficult Weather Conditions Using YOLO." *IEEE Access* 8 (2020): 125459-125476.
- [107] Krišto, Mate, and Marina Ivasic-Kos. "Thermal Imaging Dataset for Person Detection" *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. 2019.
- [108] Ghazali, Seyed M., and Y. Baleghi. "Pedestrian Detection in Infrared Outdoor Images Based on Atmospheric Situation Estimation." *Journal of AI and Data Mining* 7.1 (2019): 1-16.
- [109] Heo, Duyoung, Eunju Lee, and Byoung Chul Ko. "Pedestrian detection at night using deep neural networks and saliency maps." *Electronic Imaging* 2018.17 (2018): 1-9.
- [110] M. Jeong, B. C. Ko, and J. Y. Nam, "Early detection of sudden pedestrian crossing for safe driving during summer nights," *IEEE Trans. Circuits. Syst. Video Technol.* 27, 1368–1380 (2017)
- [111] Herrmann, Christian, Miriam Ruf, and Jürgen Beyerer. "CNN-based thermal infrared person detection by domain adaptation." *Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything*. Vol. 10643. International Society for Optics and Photonics, 2018.
- [112] Rodger, Iain, Barry Connor, and Neil M. Robertson. "Classifying objects in LWIR imagery via CNNs." *Electro-Optical and Infrared Systems: Technology and Applications XIII*. Vol. 9987. International Society for Optics and Photonics, 2016.

- [113] Rodin, Christopher Dahlin et al. "Object Classification in Thermal Images using Convolutional Neural Networks for Search and Rescue Missions with Unmanned Aerial Systems." *2018 International Joint Conference on Neural Networks (IJCNN)* (2018): 1-8.
- [114] Shahid, Nazeer, et al. "Real-Time Implementation of Human Detection in Thermal Imagery Based on CNN." *한국정보기술학회논문지* 17.1 (2019): 107-121.
- [115] J. Li, W. Gong, "Real Time Pedestrian Tracking using Thermal Infrared Imagery", *JCP*, Vol. 5, No. 10, pp. 1606 - 1613, 2010.
- [116] T. Bourlai, N. Kalka, B. Čukić, et al., "Ascertaining human identity in night environments", Book Chapter, *Distributed Video Sensor Networks*, London, England, Springer, pp. 451-467, 2011.
- [117] D. Tan, et al., "Efficient Night Gait Recognition Based on Template Matching", in *Pattern Recognition, ICPR 2006, 18th International Conference on*, Hong Kong, pp. 1000 - 1003, 2006.
- [118] URL: <http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp>
- [119] Bourlai, Thirimachos, and Bojan Cukic. "Multi-spectral face recognition: identification of people in difficult environments." *Intelligence and Security Informatics (ISI)*, 2012 IEEE International Conference on. IEEE, 2012
- [120] Seal, Ayan, et al. "Automated thermal face recognition based on minutiae extraction." *International Journal of Computational Intelligence Studies* 2.2 (2013): 133-156.
- [121] Xie, Zhihua, and Guodong Liu. "Blood perfusion construction for infrared face recognition based on bio-heat transfer." *Bio-medical materials and engineering* 24.6 (2014): 2733-2742
- [122] E. Chen, O. Haik, Y. Yitzhaky, "Classification of moving objects in the atmospherically degraded video", *Optical Engineering*, vol. 51, no. 10, p. 101710, 2012.
- [123] E. T. Gilmore, C. Ugbome, C. Kim, "An IR-based Pedestrian Detection System Implemented with Matlab-Equipped Laptop and Low-Cost Microcontroller", *International Journal of Computer Science and Information Technology*, Vol. 3, No. 5, pp. 79 - 87, 2011.
- [124] E. Jeon et al., "Human Detection Based on the Generation of a Background Image by Using a Far-Infrared Light Camera", *Sensors*, Vol. 15, No. 3, pp. 6763 - 6788, 2015.
- [125] F. Coutts, S. Marshall, P. Murray, "Human detection and tracking through temporal feature recognition", in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European, Lisbon*, pp. 2180 - 2184, 2014.
- [126] U. Dias, M. Rane, "Motion Based Object Detection And Classification For Night Surveillance", *ICTACT Journal on Image and Video Processing*, Vol. 3, No. 2, pp. 518 - 521, 2012
- [127] Berg, Amanda, Jörgen Ahlberg, and Michael Felsberg. "A thermal infrared dataset for evaluation of short-term tracking methods." *Swedish Symposium on Image Analysis*. 2015.

- [128] URL: <https://www.flir.eu/oem/adas/adas-dataset-form/>
- [129] W. Wong, et al., "Omnidirectional Thermal Imaging Surveillance System Featuring Trespasser and Faint Detection", *International Journal of Image Processing (IJIP)*, Vol. 6, No. 6, pp. 518 - 538, 2011.
- [130] I. Riaz, J. Piao, H. Shin, "Human detection by using centrist features for thermal images", *IADIS International Journal on Computer Science and Information Systems*, Vol. 8, No. 2, pp. 1 - 11, 2013
- [131] F. Lee, F. Chen, J. Liu, "Infrared Thermal Imaging System on a Mobile Phone", *Sensors*, Vol. 15, No. 5, pp. 10166-10179, 2015.
- [132] B. DeCann, A. Ross, J. Dawson, "Investigating gait recognition in the short-wave infrared (SWIR) spectrum: dataset and challenges", in *SPIE 8712, Biometric and Surveillance Technology for Human and Activity Identification X*, 2013.
- [133] J. Yun, S. Lee, "Human Movement Detection and Identification Using Pyroelectric Infrared Sensors", *Sensors*, Vol. 14, No. 5, pp. 8057-8081, 2014.
- [134] Gurney, Kevin. *An introduction to neural networks*. CRC press, 2014.
- [135] Kohonen, Teuvo. "Self-organized formation of topologically correct feature maps." *Biological cybernetics* 43.1 (1982): 59-69.
- [136] URL: <https://sergioskar.github.io/assets/img/posts/neuron.jpg>
- [137] Andrijić, Ž. Ujević. "Umjetne neuronske mreže." *Kem. Ind* 68.5-6 (2019): 219-220.
- [138] Sher, Gene I. *Handbook of neuroevolution through Erlang*. Springer Science & Business Media, 2012.
- [139] Sze, Vivienne, et al. "Efficient processing of deep neural networks: A tutorial and survey." *Proceedings of the IEEE* 105.12 (2017): 2295-2329.
- [140] McCulloch, Warren S., and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity." *The bulletin of mathematical biophysics* 5.4 (1943): 115-133.
- [141] URL: <https://towardsdatascience.com/rosenblatts-perceptron-the-very-first-neural-network-37a3ec09038a>;
- [142] URL: https://miro.medium.com/max/369/1*fDHIg9iNo0LLK4czQqqO9A.png
- [143] TURING, INTELLIGENCE BY AM. "Computing machinery and intelligence-AM Turing." *Mind* 59.236 (1950): 433.
- [144] URL: <http://vision.stanford.edu/Science-2015-Turing.pdf>
- [145] URL: <https://towardsdatascience.com/did-google-duplex-beat-the-turing-test-yes-and-no-a2b87d1c9f58>
- [146] URL: <https://isturingtestpassed.github.io/>
- [147] M. Gilberto, "The Turing test and android science," *J Robotics Autom*, vol. 2, no. 1, 2018.
- [148] Samuel, A. L. "Some studies in machine learning using the game of checkers." *IBM Journal of Research and Development* 3.3 (1959): 210-229.
- [149] Rosenbaltt, F. "The perceptron—a perciving and recognizing automation." *Report 85-460-1 Cornell Aeronautical Laboratory, Ithaca, Tech. Rep.* (1957).

- [150] Hubel, David H., and Torsten N. Wiesel. "Receptive fields of single neurones in the cat's striate cortex." *The Journal of physiology* 148.3 (1959): 574-591.
- [151] Kelley, Henry J. "Gradient theory of optimal flight paths." *Ars Journal* 30.10 (1960): 947-954.
- [152] Farlow, Stanley J. "The GMDH algorithm of Ivakhnenko." *The American Statistician* 35.4 (1981): 210-215.
- [153] Werbos, Paul. "Beyond Regression:" New Tools for Prediction and Analysis in the Behavioral Sciences." *Ph. D. dissertation, Harvard University* (1974).
- [154] Fukushima, Kunihiko. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position." *Biological cybernetics* 36.4 (1980): 193-202.
- [155] Hopfield, John J. "Neural networks and physical systems with emergent collective computational abilities." *Proceedings of the national academy of sciences* 79.8 (1982): 2554-2558.
- [156] Du, Ke-Lin, and M. N. S. Swamy. "Recurrent neural networks." *Neural networks and statistical learning*. Springer, London, 2019. 351-371.
- [157] Sejnowski, Terrence J., and Charles R. Rosenberg. "Parallel networks that learn to pronounce English text." *Complex systems* 1.1 (1987): 145-168.
- [158] Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors." *Cognitive modeling* 5.3 (1988): 1.
- [159] LeCun, Yann, et al. "Backpropagation applied to handwritten zip code recognition." *Neural computation* 1.4 (1989): 541-551.
- [160] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86.11 (1998): 2278-2324.
- [161] Watkins, Christopher John Cornish Hellaby. "Learning from delayed rewards." (1989).
- [162] Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." *Neural networks* 61 (2015): 85-117.
- [163] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.
- [164] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009.
- [165] Miller, George A., et al. "Introduction to WordNet: An on-line lexical database." *International journal of lexicography* 3.4 (1990): 235-244.
- [166] URL: <https://www.mturk.com/>
- [167] Russakovsky, Olga, et al. "Imagenet large scale visual recognition challenge." *International journal of computer vision* 115.3 (2015): 211-252.
- [168] Cireşan, Dan Claudiu, et al. "Deep, big, simple neural nets for handwritten digit recognition." *Neural computation* 22.12 (2010): 3207-3220.
- [169] Nair, Vinod, and Geoffrey E. Hinton. "Rectified linear units improve restricted boltzmann machines." *Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010.

- [170] Le, Quoc V., et al. "Building high-level features using large scale unsupervised learning." *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 2013.
- [171] Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.
- [172] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).
- [173] Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." *European conference on computer vision*. Springer, Cham, 2014.
- [174] Lin, Min, Qiang Chen, and Shuicheng Yan. "Network in network." *arXiv preprint arXiv:1312.4400* (2013).
- [175] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
- [176] Szegedy, Christian, et al. "Going deeper with convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.]
- [177] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [178] Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- [179] Larsson, Gustav, Michael Maire, and Gregory Shakhnarovich. "Fractalnet: Ultra-deep neural networks without residuals." *arXiv preprint arXiv:1605.07648* (2016).
- [180] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
- [181] Lee, Chen-Yu, et al. "Deeply-supervised nets." *Artificial intelligence and statistics*. 2015.
- [182] Pezeshki, Mohammad, et al. "Deconstructing the ladder network architecture." *International Conference on Machine Learning*. 2016.
- [183] Huang, Gao, et al. "Deep networks with stochastic depth." *European conference on computer vision*. Springer, Cham, 2016.
- [184] Chollet, François. "Xception: Deep learning with depthwise separable convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- [185] Alom, Md Zahangir, et al. "Inception recurrent convolutional neural network for object recognition." *arXiv preprint arXiv:1704.07709* (2017).
- [186] Li, Yikang, et al. "Vip-cnn: Visual phrase guided convolutional neural network." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.
- [187] URL: <https://devblogs.nvidia.com/deep-learning-object-detection-digits/>

- [188] Haykin, Simon. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.
- [189] Rosenbaltt, F. "The perceptron – A perciving and recognizing automation." *Report 85-460-1 Cornell Aeronautical Laboratory, Ithaca, Tech. Rep.* (1957).
- [190] Aggarwal, Charu C. "Neural networks and deep learning." *Cham: Springer International Publishing* (2018).
- [191] K. Dhana Sree, Data Analytics: Role of Activation function In Neural Net, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-5 March, 2019
- [192] Nwankpa, Chigozie, et al. "Activation functions: Comparison of trends in practice and research for deep learning." *arXiv preprint arXiv:1811.03378* (2018).
- [193] URL: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>
- [194] Duch, Włodzisław, and Norbert Jankowski. "Survey of neural transfer functions." *Neural computing surveys* 2.1 (1999): 163-212.
- [195] URL: <https://www.analyticsvidhya.com/blog/2017/10/fundamentals-deep-learning-activation-functions-when-to-use-them/>
- [196] Neal, Radford M. "Connectionist learning of belief networks." *Artificial intelligence* 56.1 (1992): 71-113.
- [197] Glorot, Xavier, and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks." *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. 2010.
- [198] Hochreiter, Sepp. "The vanishing gradient problem during learning recurrent neural nets and problem solutions." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6.02 (1998): 107-116.
- [199] Karlik, Bekir, and A. Vehbi Olgac. "Performance analysis of various activation functions in generalized MLP architectures of neural networks." *International Journal of Artificial Intelligence and Expert Systems* 1.4 (2011): 111-122.
- [200] URL: <https://www.oreilly.com/library/view/intelligent-projects-using/9781788996921/assets/58f98f0c-6a81-4acc-bafb-9b0f683ad9c9.png>
- [201] URL: <https://dmitry.ai/uploads/default/original/1X/cd999cf88ef1284eedf1584304a4525dc3921bfc.png>
- [202] Gao, Bolin, and Lacra Pavel. "On the properties of the softmax function with application in game theory and reinforcement learning." *arXiv preprint arXiv:1704.00805* (2017).
- [203] URL: <https://medium.com/@prateekvishnu/activation-functions-in-neural-networks-bf5c542d5fec>
- [204] Ramachandran, Prajit, Barret Zoph, and Quoc V. Le. "Swish: a self-gated activation function." *arXiv preprint arXiv:1710.05941* 7 (2017).
- [205] Ramachandran, Prajit, Barret Zoph, and Quoc V. Le. "Searching for activation functions." *arXiv preprint arXiv:1710.05941* (2017).

- [206] Eger, Steffen, Paul Youssef, and Iryna Gurevych. "Is it time to swish? comparing deep learning activation functions across NLP tasks." *arXiv preprint arXiv:1901.02671* (2019).
- [207] URL: <https://medium.com/@jaiyamsharma/experiments-with-swish-activation-function-on-mnist-dataset-fc89a8c79ff7>
- [208] URL: <https://medium.com/@sh.tsang/review-nasnet-neural-architecture-search-network-image-classification-23139ea0425d>
- [209] URL: <https://towardsdatascience.com/review-inception-v4-evolved-from-googlenet-merged-with-resnet-idea-image-classification-5e8c339d18bc>
- [210] URL: <https://thefork.com/wp-content/uploads/2019/02/final.png>
- [211] Mendel, Jerry M. "Adaptive learning and pattern recognition systems." *Theory and applications* (1970).
- [212] Weiss, Karl & Khoshgoftaar, Taghi & Wang, DingDing. (2016). A survey of transfer learning. *Journal of Big Data*. 3. 10.1186/s40537-016-0043-6.
- [213] Saha, Rohan. (2018). Transfer Learning - A Comparative Analysis. 10.13140/RG.2.2.31127.39848.
- [214] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [215] Alom, Md Zahangir, et al. "The history began from AlexNet: A comprehensive survey on deep learning approaches." *arXiv preprint arXiv:1803.01164* (2018).
- [216] Alom, Md Zahangir, et al. "A state-of-the-art survey on deep learning theory and architectures." *Electronics* 8.3 (2019): 292.
- [217] Muhammad, Iqbal, and Zhu Yan. "SUPERVISED MACHINE LEARNING APPROACHES: A SURVEY." *ICTACT Journal on Soft Computing* 5.3 (2015).
- [218] Alloghani, Mohamed, et al. "A systematic review on supervised and unsupervised machine learning algorithms for data science." *Supervised and Unsupervised Learning for Data Science*. Springer, Cham, 2020. 3-21.
- [219] Cios, Krzysztof J. "Deep neural networks—a brief history." *Advances in Data Analysis with Computational Intelligence Methods*. Springer, Cham, 2018. 183-200.
- [220] Albawi, Saad, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network." *2017 International Conference on Engineering and Technology (ICET)*. IEEE, 2017.
- [221] Rumelhart, David E., et al. "Sequential thought processes in PDP models." *Parallel distributed processing: explorations in the microstructures of cognition* 2 (1986): 3-57.
- [222] Zhang, Aston, et al. "Dive into Deep Learning." *May 19* (2019): 2019.
- [223] Sutton, Richard S., and Andrew G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [224] François-Lavet, Vincent, et al. "An introduction to deep reinforcement learning." *Foundations and Trends® in Machine Learning* 11.3-4 (2018): 219-354.

- [225] Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems*. 2014.
- [226] Ballard, Dana H. "Modular Learning in Neural Networks." *AAAI*. 1987.
- [227] Ackley, David H., Geoffrey E. Hinton, and Terrence J. Sejnowski. "A learning algorithm for Boltzmann machines." *Cognitive science* 9.1 (1985): 147-169.
- [228] Fischer, Asja, and Christian Igel. "An introduction to restricted Boltzmann machines." *Iberoamerican Congress on Pattern Recognition*. Springer, Berlin, Heidelberg, 2012.
- [229] Widrow, Bernard, and Marcian E. Hoff. *Adaptive switching circuits*. No. TR-1553-1. Stanford Univ Ca Stanford Electronics Labs, 1960
- [230] Du, Ke-Lin, and Madisetti NS Swamy. *Neural networks and statistical learning*. Springer Science & Business Media, 2013.
- [231] Mehrotra, Kishan, Chilukuri K. Mohan, and Sanjay Ranka. *Elements of artificial neural networks*. MIT press, 1997.
- [232] Bolf, N., and I. Jerbić. "Primjena umjetnih neuronskih mreža pri identificiranju i vođenju procesa." *Kemija u industriji: Časopis kemičara i kemijskih inženjera Hrvatske* 55.11 (2006): 457-468.
- [233] David, Kriesel. "A Brief Introduction to Neural Networks." (2005).
- [234] Abiodun, Oludare Isaac, et al. "State-of-the-art in artificial neural network applications: A survey." *Heliyon* 4.11 (2018): e00938.
- [235] URL: <https://www.digitalvidya.com/blog/types-of-neural-networks/>
- [236] Hagan, Martin T., and Mohammad B. Menhaj. "Training feedforward networks with the Marquardt algorithm." *IEEE transactions on Neural Networks* 5.6 (1994): 989-993.
- [237] Hertz, John A. *Introduction to the theory of neural computation*. CRC Press, 2018.
- [238] Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review* 65.6 (1958): 386.
- [239] Elman, Jeffrey L. "Finding structure in time." *Cognitive science* 14.2 (1990): 179-211.
- [240] Cho, Kyunghyun, et al. "On the properties of neural machine translation: Encoder-decoder approaches." *arXiv preprint arXiv:1409.1259* (2014).
- [241] Chung, Junyoung, et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling." *arXiv preprint arXiv:1412.3555* (2014).
- [242] Šíma, Jiří, Pekka Orponen, and Teemu Antti-Poika. "On the computational complexity of binary and analog symmetric Hopfield nets." *Neural computation* 12.12 (2000): 2965-2989.
- [243] Šíma, Jiří, and Pekka Orponen. "Continuous-time symmetric Hopfield nets are computationally universal." *Neural Computation* 15.3 (2003): 693-733.
- [244] Hinton, Geoffrey. "Boltzmann machines." *Encyclopedia of Machine Learning and Data Mining* (2014): 1-7.
- [245] Salakhutdinov, Ruslan, and Geoffrey Hinton. "Deep boltzmann machines." *Artificial intelligence and statistics*. 2009.

- [246] Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets." *Neural computation* 18.7 (2006): 1527-1554.
- [247] Bourlard, Hervé, and Yves Kamp. "Auto-association by multilayer perceptrons and singular value decomposition." *Biological cybernetics* 59.4-5 (1988): 291-294.
- [248] Jolliffe, Ian. *Principal component analysis*. Springer Berlin Heidelberg, 2011.
- [249] URL: <https://www.tensorflow.org/>
- [250] Abadi, Martín, et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems." *arXiv preprint arXiv:1603.04467* (2016).
- [251] URL: <https://translate.google.hr/>
- [252] URL: <https://keras.io/>
- [253] URL: <https://caffe.berkeleyvision.org/>
- [254] URL: <https://deepdreamgenerator.com/>
- [255] URL: <https://pytorch.org/>
- [256] URL: <https://docs.microsoft.com/en-us/cognitive-toolkit/>
- [257] Nguyen, Giang, et al. "Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey." *Artificial Intelligence Review* 52.1 (2019): 77-124.
- [258] URL: <https://spark.apache.org/>
- [259] URL: <https://scikit-learn.org/stable/>
- [260] URL: <https://mahout.apache.org/>
- [261] URL: <https://singa.apache.org/>
- [262] URL: <https://aws.amazon.com/machine-learning/>
- [263] URL: <http://accord-framework.net/>
- [264] Wu, Jianxin. "Introduction to convolutional neural networks." *National Key Lab for Novel Software Technology. Nanjing University. China* 5 (2017): 23.
- [265] O'Shea, Keiron, and Ryan Nash. "An introduction to convolutional neural networks." *arXiv preprint arXiv:1511.08458* (2015).
- [266] Nielsen, Michael A. *Neural networks and deep learning*. Vol. 25. San Francisco, CA, USA:: Determination press, 2015.
- [267] URL: <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>
- [268] Ruder, Sebastian. "An overview of gradient descent optimization algorithms." *arXiv preprint arXiv:1609.04747* (2016).
- [269] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.
- [270] URL: <https://stackoverflow.com/questions/57753640/convert-yolov3-output-to-coordinates-of-bounding-box-label-and-confidence>

- [271] Misra, Diganta. "Mish: A self regularized non-monotonic neural activation function." *arXiv preprint arXiv:1908.08681* (2019).
- [272] URL: <https://github.com/lutzroeder/netron>
- [273] URL: <http://cs231n.github.io/convolutional-networks/#conv>
- [274] Du, Juan. "Understanding of Object Detection Based on CNN Family and YOLO." *Journal of Physics: Conference Series*. Vol. 1004. No. 1. IOP Publishing, 2018.
- [275] Michelucci, Umberto. *Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks*. Apress, 2018.
- [276] URL: https://www.termogram.com/pdf/therma_cam_p_series/thermacam_p10/P10_datasheet.pdf
- [277] URL: <https://www.pass-thermal.co.uk/flir-131-mm-7-degree-telephoto-p-b-series-lens>
- [278] URL: <https://play.google.com/store/apps/details?id=com.augmentra.viewranger.android&hl=hr>
- [279] URL: https://www.gsmarena.com/cat_s60-7928.php
- [280] URL: <https://earth.google.com/web/>
- [281] Gim, LEE Cheow, EE Kok Tiong, and HENG Yinghui Elizabeth. "Performance challenges for high-resolution imaging sensors for surveillance in tropical environment.", DSTA HORIZONS, 2015, pp. 80 – 88.
- [282] URL: <http://www.videosoftdev.com/free-video-editor>
- [283] URL: <https://www.mathworks.com/products/matlab.html>
- [284] URL: <http://www.robots.ox.ac.uk/~vgg/software/via/>
- [285] URL: <https://www.labelbox.com/>
- [286] URL: <http://labelme.csail.mit.edu/Release3.0/>
- [287] URL: <https://annotorious.github.io/>
- [288] URL: <https://github.com/drainingsun/ybat>
- [289] URL: <https://pjreddie.com/darknet/yolo/>
- [290] URL: <http://host.robots.ox.ac.uk/pascal/VOC/>
- [291] URL: <http://cocodataset.org/#home>
- [292] URL: <https://timebutt.github.io/static/how-to-train-yolov2-to-detect-custom-objects/>
- [293] URL: <https://github.com/AlexeyAB/darknet>
- [294] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.
- [295] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on*

- Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826, DOI: 10.1109/CVPR.2016.308.
- [296] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” arXiv:1512.03385 [cs], Dec. 2015, Accessed: Jun. 08, 2020. [Online]. Available: <http://arxiv.org/abs/1512.03385>.
- [297] Z. Cai and N. Vasconcelos, “Cascade R-CNN: Delving into High-Quality Object Detection,” arXiv:1712.00726 [cs], Dec. 2017, Accessed: Jun. 08, 2020. [Online]. Available: <http://arxiv.org/abs/1712.00726>.
- [298] T.-Y. Lin et al., “Microsoft COCO: Common objects in context,” in European conference on computer vision, 2014, pp. 740–755.
- [299] URL: <https://www.youtube.com/watch?v=7INYPMIAYCY>
- [300] URL: <https://projects.asl.ethz.ch/datasets/doku.php?id=ir:iricra2014>
- [301] Bilodeau, Guillaume-Alexandre, et al. "Thermal–visible registration of human silhouettes: A similarity measure performance evaluation." *Infrared Physics & Technology* 64 (2014): 79-86.
- [302] URL: <https://www.polymtl.ca/litiv/en/codes-and-datasets>
- [303] Davis, James W., and Vinay Sharma. "Background-subtraction using contour-based fusion of thermal and visible imagery." *Computer vision and image understanding* 106.2-3 (2007): 162-182.;
- [304] URL: <http://vcip1-okstate.org/pbvs/bench/>
- [305] Mieziako, Roland. "Terravic research infrared database." *IEEE OTCBVS WS Series Bench* (2005).
- [306] González, Alejandro, et al. "Pedestrian detection at day/night time with visible and FIR cameras: A comparison." *Sensors* 16.6 (2016): 820,
- [307] URL: <http://adas.cvc.uab.es/elektra/enigma-portfolio/item-1/>
- [308] Socarrás, Yainuvis, et al. "Adapting pedestrian detection from synthetic to far infrared images." *ICCV–Workshop Visual Domain Adaptation and Dataset Bias*. 2013.
- [309] Felsberg, Michael, et al. "The thermal infrared visual object tracking VOT-TIR2015 challenge results." *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2015.
- [310] URL: <https://www.votchallenge.net/vot2015/dataset.html>
- [311] J. Portmann, S. Lynen, M. Chli, and R. Siegwart, “People detection and tracking from aerial thermal views,” in 2014 IEEE International Conference on Robotics and Automation (ICRA), May 2014, pp. 1794–1800, DOI: 10.1109/ICRA.2014.6907094.
- [312] J. W. Davis and M. A. Keck, “A two-stage template approach to person detection in thermal imagery,” in 2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION’05)-Volume 1, 2005, vol. 1, pp. 364–369.
- [313] MORH – Ministarstvo obrane Republike Hrvatske (MORH RH), Swift Response 2019, URL: <https://www.youtube.com/watch?v=7INYPMIAYCY>, upotreba video zapisa u znanstvene svrhe odobrena od strane vlasnika – MORH RH.

- [314] URL: <https://www.opgal.com/ptz-cameras/>
- [315] URL: <https://www.stanleysecuritysolutions.com/thermaldetection>
- [316] URL: <https://www.hikvision.com/en/products/Thermal-Products/Security-thermal-cameras/>
- [317] URL: <https://www.hgh-infrared.com/Products/Optronics-for-Security/Spynel-S-Panoramic-Detection-System>
- [318] URL: <https://www.flir.com/security/perimeter-protection/>
- [319] URL: <https://www.infinitioptics.com/>
- [320] URL: https://en.wikipedia.org/wiki/Space_blanket;
- [321] URL: <https://en.wikipedia.org/wiki/BoPET>
- [322] URL: <https://ufpro.com/camoshield>
- [323] URL: <https://krposts.com/level/camoshield-thermal-infrared-protection-for-professionals/l9yLILXS16ahx6k>
- [324] URL: <http://www.fibrotex-tech.com/>
- [325] URL: <https://opencv.org/>
- [326] URL: <https://developer.nvidia.com/cuda-zone>
- [327] URL: <https://www.khronos.org/opencv/>
- [328] URL: <https://github.com/AlexeyAB/darknet#yolo-v4-in-other-frameworks>

Popis slika

Slika 1.	Elektromagnetski spektar [31].....	9
Slika 2.	Podjela IR segmenta elektromagnetskog spectra [34].....	10
Slika 3.	Prikaz područja snimljenog u vizualnom i različitim dijelovima IR spectra [37].....	10
Slika 4.	Pregled povijesnog razvoja IR detektora i sustava [45].....	13
Slika 5.	Primjer snimanja istog objekta primjenom aktivnog IR snimanja (lijevo) i termalnog snimanja (desno). Fotografija prikazuje katedralu Saint Etienne u Bourgesu (Francuska) [50].....	16
Slika 6.	Blok dijagram glavnih komponenti IR termovizijske kamere [53].....	17
Slika 7.	Atomsferski "prozori" za snimanje u vizualnom i IR spektru [65].....	24
Slika 8.	Tipičan model (a) klasične arhitekture sustava za nadzor koji se temelji na termoviziji i (b) sustava temeljenog na neuronskoj mreži [69].....	31
Slika 9.	Primjeri slika iz nekoliko skupova: a) OTCBVS (fusion of thermal and color imagery), b) KAIST pedestrian dataset c) CASIA infrared night gait dataset.....	45
Slika 10.	Usporedba biološkog i umjetnog neurona [136].....	48
Slika 11.	Međudodnosi umjetne inteligencije (AI), strojnog učenja (engl. ML – Machine Learning), umjetnih neuronskih mreža, dubokog učenja (engl. DL – Deep Learning) i Spiking Neural Networks (SNN) [139]....	49
Slika 12.	Warren McCulloch (lijevo), Walter Pitts (sredina) i njihov model umjetnog neurona (desno) [141, 142].....	50
Slika 13.	Mark I Perceptron Machine [141].....	51
Slika 14.	Prikaz osnovne građe perceptrona – lijevo bez prirasta (dodatni prag ili unutarnji pomak neurona), desno – s prirastom [190].....	61
Slika 15.	Osnovna arhitektura unaprijedne neuronske mreže sa dva skrivena i jednim izlaznim slojem. Iako svaka jedinica sadrži jednu skalarnu varijablu, jedna često predstavlja sve neurone u jednom sloju kao jedan vector. Vektorske jedinice se često prikazuju kao pravokutnici i spojeni su matricama između njih [190].....	63
Slika 16.	Funkcija s pragom osjetljivosti (engl. <i>step function</i>) [195].....	67
Slika 17.	Sigmoidna funkcija [193].....	69
Slika 18.	Hiperbolna tangens – Tanh funkcija [200].....	71
Slika 19.	Usporedba ReLU, LReLU i RLReLU aktivacijskih funkcija [201].....	72
Slika 20.	Prikaz računanja Softmax funkcije [203].....	73
Slika 21.	Usporedba ReLU i SWISH aktivacijskih funkcija [207].....	73
Slika 22.	Pregled aktivacijskih funkcija [210].....	74

Slika 23. Vrste i međudodnos vrsta dubokog učenja [215].....	80
Slika 24. Klasifikacija neuronskih mreža [234].....	87
Slika 25. Tipovi neuronskih mreža [235].....	88
Slika 26. Arhitektura konvolucijske neuronske mreže [267].....	92
Slika 27. Vizualni pozitiv (lijevo) i negativ (desno) koji prikazuje IoU kriterij.....	98
Slika 28. Arhitektura YOLOv3 [270].....	104
Slika 29. Korištena termalna kamera i teleobjektiv.....	112
Slika 30. Prikaz lokacije i kretanja prilikom snimanja po vedrom [106].....	114
Slika 31. Usporedni prikaz osoba snimljenih termalnom kamerom po vedrom vremenu na udaljenosti od 110 m: (a) standardni objektiv, (b) teleobjektiv.....	114
Slika 32. Usporedni prikaz osoba snimljenih termalnom kamerom po vedrom vremenu na udaljenosti od 165 m: (a) standardni objektiv (b) teleobjektiv.....	114
Slika 33. Prikaz lokacije i putanje kretanja za snimanje po gustoj magli.....	115
Slika 34. Termalna snimka osobe u gustoj magli: (a) teleobjektiv na 30 m, (b) teleobjektiv na 50 m.....	116
Slika 35. Prikaz lokacije za snimanje po jakoj kiši.....	116
Slika 36. Termalna snimka osobe po jakoj kiši: (a) teleobjektiv na 30 m, (b) teleobjektiv na 70 m (pogureno); (c) teleobjektiv na 140 m; (d) teleobjektiv na 215 m.....	117
Slika 37. Ciklus koraka prilikom trčanja, vedro vrijeme, udaljenost 165 m, snimljeno korištenjem teleobjektiva.....	118
Slika 38. Termalna silueta: (a) udaljenost 110 m, vedro, hodanje, standardni objektiv; (b) 110 m, vedro, standardni objektiv, četveronoški hod; (c) 110 m, vedro, teleobjektiv, trčanje; (d) 165 m, vedro, hodanje, teleobjektiv; (e) 165 m, vedro, teleobjektiv, trčanje; (f) 165 m, pogureno trčanje, teleobjektiv; (g) 165 m, pogureno hodanje, teleobjektiv.....	120
Slika 39. Termalna silueta: (a) udaljenost 30 m, magla, hodanje, teleobjektiv; (b) 50 m, magla, teleobjektiv, hodanje; (c) 50 m, magla, teleobjektiv, pogureno hodanje; (d) 50 m, magla, pogureno trčanje, teleobjektiv (magla se dodatno pojačala).....	120
Slika 40. Termalna silueta: (a) udaljenost 30 m, kiša, hodanje, standardni objektiv; (b) 30 m, kiša, standardni objektiv, pogureno hodanje; (c) 30 m, kiša, teleobjektiv, trčanje; (d) 70 m, kiša, hodanje, standardni objektiv; (e) 70 m, kiša, teleobjektiv, hodanje; (f) 70 m, kiša, teleobjektiv, pogureno hodanje; (g) 110 m, kiša, teleobjektiv, hodanje; (h) 110 m, kiša, teleobjektiv, pogureno hodanje.....	121

Slika 41. Termalna silueta: (a) udaljenost 140 m, kiša, pogureno hodanje, teleobjektiv; (b) 140 m, kiša, teleobjektiv, hodanje; (c) 170 m, kiša, teleobjektiv, hodanje; (d) 170 m, kiša, pogureno hodanje, teleobjektiv; (e) 180 m, kiša, teleobjektiv, hodanje; (f) 180 m, kiša, teleobjektiv, trčanje; (g) 180 m, kiša, teleobjektiv, pogureno hodanje; (h) 180 m, kiša, teleobjektiv, pogureno trčanje; (i) 215 m, kiša, teleobjektiv, hodanje; (j) 215 m, kiša, teleobjektiv, trčanje; (k) 215 m, kiša, teleobjektiv, pogureno hodanje.....	..121
Slika 42. Korisničko suč enje anotacijskog alata Yolo BBox Annotation Tool.....	..124
Slika 43. Korisničko suč enje anotacijskog alata Yolo BBox Annotation Tool s uč itanom slikom i označ enim objektima (osobama).....	..124
Slika 44. YOLO format anotacija za tri označ ena objekta na slici koji pripadaju istoj klasi.....	..125
Slika 45. YOLO format anotacija za tri označ ena objekta na slici koji pripadaju dvjema različ itim klasama (Human – Non-Human).....	..125
Slika 46. AP rezultat i krivulja preciznosti/odziva za bazni YOLO model, <i>bY</i> (a), i za YOLO model, <i>tY</i> (b), nauč en na pripremljenom termalnom skupu slika [105].....	..130
Slika 47. Krivulja preciznosti/odziva za vedro vrijeme [105].....	..131
Slika 48. Krivulja preciznosti/odziva za jaku kišu [105].....	..131
Slika 49. Krivulja preciznosti/odziva za gustu maglu [105].....	..132
Slika 50. Primjer detekcije osoba primjenom <i>bY</i> modela (a) i <i>tY</i> modela (b), slike su snimljene korištenjem standardnog objektiva termalne kamere po vedrom vremenu na udaljenosti od oko 150 m.....	..132
Slika 51. Rezultati detekcije osoba (pogureni hod) primjenom <i>bY</i> modela (a) i <i>tY</i> modela (b). Slike su snimljene korištenjem standardnog objektiva kamere po jakoj kiši na udaljenosti od oko 70 m.....	..133
Slika 52. Rezultat detekcije osoba (pogureni hod) primjenom <i>bY</i> modela (a) i <i>tY</i> modela (b). Slike su snimljene po jakoj kiši korištenjem teleobjektiva na udaljenosti od oko 100 m.....	..133
Slika 53. Rezultat detekcije osobe (trč anje) primjenom <i>bY</i> modela (a) i <i>tY</i> modela (b). Slike su snimljene po jakoj kiši korištenjem teleobjektiva na udaljenosti od oko 215 m.....	..134
Slika 54. Rezultat detekcije osoba (normalan hod) primjenom <i>bY</i> modela (a) i <i>tY</i> modela (b). Slike su snimljene po gustom magli korištenjem teleobjektiva na udaljenosti od oko 50 m.....	..134
Slika 55. Rezultat detekcije osoba na slikama koje su snimljene teleobjektivom po vedrom vremenu na udaljenosti od 165 m; normalan hod (a), trč anje (b).....	..135

Slika 56. Primjeri detekcije osoba po gustoj magli (snimano korištenjem teleobjektiva): (a) udaljenost manje od 30 m, normalan hod; (b) udaljenost 30 m, normalan hod; (c) udaljenost 50 m, normalan hod; (d) udaljenost 50 m, pogureno hodaње.....	136
Slika 57. Primjeri detekcije osoba po jakoj kiši: (a) 30 m, pogureno hodaње, standardni objektiv; (b) 30 m, normalan hod; (c) 30 m, trčanje, teleobjektiv; (d) 70 m, pogureno hodaње, standardni objektiv; (e) 70 m, trčanje, teleobjektiv; (f) 110 m, normalno hodaње, teleobjektiv.....	138
Slika 58. Primjeri detekcije osoba po jakoj kiši: (a) 140 m, normalno hodaње, teleobjektiv; (b) 170 m, normalan hod, teleobjektiv; (c) 180 m, trčanje, teleobjektiv; (d) 215 m, normalno hodaње, teleobjektiv.....	139
Slika 59. PR krivulja za bazni YOLO model bY, trenirani YOLO model, tY i model treniran samo na slikama snimljenima po vedrom vremenu tY_clear [106].....	140
Slika 60. PR krivulja za model tY_clear testiran na skupovima slika snimljenih u različitim vremenskim uvjetima tY_fog (magla) i tY_rain (kiša) [106].....	141
Slika 61. Rezultat detekcije i raspoznavanja osobe od životinje na termalnoj snimci snimljenoj po gustoj magli na udaljenosti od oko 30 i 50 metara korištenjem teleobjektiva pri normalnom hodu. Također je razvidno da osobe i pas nisu na istoj udaljenosti, odnosno da je pas bliže kameri nego osobe [106].....	143
Slika 62. Rezultat detekcije i raspoznavanja osobe od životinje na termalnoj snimci po gustoj magli na udaljenosti od oko 30 m, korištenjem teleobjektiva dok je snimana osoba mijenjala položaj tijela: normalan hod (a) i trčanje (b) [106].....	143
Slika 63. Prikaz osoba po vedrom vremenu (bez klasifikacije aktivnosti): prilikom trčanja, 165 m, teleobjektiv (a); pogureni hod, 165 m, teleobjektiv (b); četveronoški hod, 110 m, standardni objektiv (c); puzanje, 110 m, standardni objektiv (d).....	145
Slika 64. Prikaz rezultata učenja/testiranja YOLOv3 (a) i YOLOv4 (b) detektora na skupu t_Act, odnosno na skupu termalnih slika za raspoznavanje aktivnosti.....	146
Slika 65. Primjeri osoba u normalnom hodu (bez klasifikacije načina hoda) koje su snimljene po vedrom vremenu na udaljenostima od: 110 m, standardni objektiv (a); 110 m, teleobjektiv (b); 165 m, teleobjektiv (c).....	148
Slika 66. Vedro vrijeme - Primjeri osoba u normalnom hodu (s klasifikacijom aktivnosti) koje su snimljene po vedrom vremenu na udaljenostima od: 110 m, standardni objektiv (a); 110 m, teleobjektiv (b); 165 m, teleobjektiv (c).....	148

Slika 67. Prikaz osoba po vedrom vremenu (s klasifikacijom aktivnosti): prilikom trčanja, 165 m, teleobjektiv (a); pogureni hod, 165 m, teleobjektiv (b); četveronoški hod, 110 m, standardni objektiv (c); puzanje, 110 m, standardni objektiv (d).....	149
Slika 68. Vrijednost mAP i funkcije gubitka za skup Clear_ALL: treniranje/testiranje 80:20 [106].....	151
Slika 69. Vrijednost mAP i funkcije gubitka za skup Clear_ALL: treniranje/testiranje (a) 10:90; (b) 20:80 [106].....	152
Slika 70. ASL ETH FLIR [300, 311].....	154
Slika 71. CVC IR 09 (gore lijevo) [306 – 308], KAIST dataset (gore desno) [89], LITIV 2012 (dolje) [301, 302].....	154
Slika 72. VOT-TIR2015 [309].....	154
Slika 73. OSU thermal dataset (gore) [303, 304, 312], Terravic motion dataset (dolje) [304, 305].....	154
Slika 74. mAP rezultati na različitim skupovima slika za model naučen na skupu UNIRI_TID (<i>tY_transform</i>) i svakom skupu podataka (<i>tY1...tY7</i>) [106].....	156
Slika 75. Detekcije na slikama iz skupa ASL ETH FLIR [300, 311].....	157
Slika 76. Detekcije na slikama iz skupova: CVC IR 09 (gore lijevo) [306 – 308], KAIST dataset (gore desno) [89], LITIV 2012 (dolje) [301, 302].....	157
Slika 77. Detekcije na slikama iz skupa VOT-TIR2015 [309].....	157
Slika 78. Detekcije na slikama iz skupova: OSU thermal dataset (gore) [303, 304, 312], Terravic motion dataset (dolje) [304, 305].....	157
Slika 79. mAP i Loss Function krivulja za učenje/testiranje YOLOv3 modela na termalnim slikama iz video snimke vojne vježbe "Swift Response 2019".....	158
Slika 80. "Swift Response 2019", primjeri detekcije osoba korištenjem naučenog YOLOv3 modela.....	159
Slika 81. Moguća arhitektura sustava za detekciju i raspoznavanje osoba temeljem na termalnoj kameri.....	164
Slika 82. Primjena termalnih sustava za nadzor u kontekstu nacionalne sigurnosti.....	166
Slika 83. Sadržaj datoteke .data unutar Darknet platforme.....	207

Popis tablica

Tablica 1.	Pregled baza termovizijskih slika za detekciju osoba [107].....	46
Tablica 2.	Broj slika u kreiranoj bazi slika za treniranje modela dubokog učenja..	119
Tablica 3.	Dimenzije označavajućih okvira na slikama ovisno o udaljenosti snimanja i korištenom objektivu kamere.....	122
Tablica 4.	Usporedni rezultati za detekciju osoba na termalnim slikama [106]....	126
Tablica 5.	Pregled naučenih modela tijekom izvođenja [106].....	129
Tablica 6.	Broj slika za pojedini način kretanja u trening skupu slika.....	144
Tablica 7.	Rezultat treniranja/testiranja YOLOv3 za raspoznavanje aktivnosti.....	146
Tablica 8.	Rezultat treniranja/testiranja YOLOv4 za raspoznavanje aktivnosti.....	147
Tablica 9.	Broj slika u svakom pojedinom skupu slika koji je korišten za usporedno testiranje naučenog YOLO modela [106].....	153
Tablica 10.	Rezultati YOLO metrike za detekciju osoba u svakom pojedinom usporednom skupu slika – tY_transform naučeni model [106].....	155
Tablica 11.	Rezultati YOLO metrike za detekciju osoba za učenje/testiranje na svakom pojedinom referentnom skupu slika [106].....	156

Popis kratica

Kratica	Značenje (engleski jezik)	Značenje (hrvatski jezik)
COVID-19	<i>Coronavirus Disease</i>	Bolest uzrokovana SARS-CoV-2 virusom otkrivenim 2019. godine
RGB	<i>Red-Green-Blue</i>	RGB je aditivni model boja kod kojeg se zbrajanjem osnovnih boja dobiva bijela boja. Jedna boja se opisuje kroz tri vrijednosti: dio crvene, dio zelene i dio plave boje. Svaki dio boje varira između 0% i 100%.
Engl.	<i>English</i>	Engleski jezik
CNN ili ConvNet	<i>Convolutional Neural Networks</i>	Konvolucijska neuronska mreža
Fast R-CNN	<i>Fast Region-CNN ili Fast Region with Convolutional Neural Network</i>	Algoritam za brzu detekciju objekata temeljen na konvolucijskoj neuronskoj mreži temeljen na detekciji regija u kojima se moguće nalaze objekti
SSD	<i>Single Shot Detector</i>	Algoritam za detekciju objekata s mogućnošću detekcije više objekata u jednom pregledu slike
Mask R-CNN	<i>Mask Region-CNN ili Mask Region with Convolutional Neural Network</i>	Algoritam za detekciju objekata temeljen na konvolucijskoj neuronskoj mreži temeljen na kreiranju maski detektiranih objekata
R-FCN	<i>Region-based Fully Convolutional Networks</i>	Algoritam za detekciju objekata temeljen na potpuno povezanoj konvolucijskoj neuronskoj mreži koja objekte detektira skeniranjem regija slike u kojima se moguće nalaze objekti
YOLO	<i>You Only Look Once</i>	"Ti Gledaš Samo Jedanput" – algoritam za detekciju objekata temeljen na konvolucijskoj neuronskoj mreži koja objekte detektira samo jednim pregledom slike
LWIR	<i>Long Wave Infra-red</i>	Dugovalni infracrveni segment elektromagnetskog spektra
EM	<i>Electromagnetic spectrum</i>	Elektromagnetski spektar
YOLOv2, YOLOv3, YOLOv4	<i>You Only Look Once, versions 2 – 4</i>	YOLO algoritam, verzije 2 – 4
Faster R-CNN	<i>Faster Region-CNN ili Faster Region with Convolutional Neural Network</i>	Novija i brža verzija algoritma Fast R-CNN
Cascade R-CNN	<i>Cascade Region-CNN or Cascade Region with Convolutional Neural Network</i>	Višefazni algoritam za detekciju objekata temeljen na konvolucijskoj neuronskoj mreži
FCOS	<i>Fully Convolutional One-Stage Object Detection</i>	Jednofazni algoritam za detekciju objekata temeljen na potpuno povezanoj konvolucijskoj neuronskoj mreži
IR	<i>Infrared</i>	Infracrveni segment elektromagnetskog spektra
μm	<i>Micrometre</i>	Mikrometar
THz	<i>Terahertz</i>	Terahertz
NIR	<i>Near-Infrared</i>	Područje blizu infracrvenog dijela EM spektra (od 0,7 to 1 μm)

SWIR	<i>Short-Wave infrared</i>	Kratkovalno infracrveno zračenje (od 1 do 3 μm)
MWIR	<i>Mid-Wave infrared</i>	Srednjevalno infracrveno područje (od 3 do 5 μm)
LWIR	<i>Long-Wave infrared</i>	Dugovalno infracrveno područje (od 8 do 14 μm)
VLWIR	<i>Very Long-Wave infrared</i>	Dugo-Dugovalno infracrveno područje (više od 14 μm)
QWIP	<i>Quantum Well Infrared Photodetector</i>	Kvantni infracrveni fotodetektor
FPA	<i>Focal Plane Array</i>	Detektor termalnog zračenja koji ima mogućnost "promatranja" cijelog vidnog polja kamere odjednom, bez potrebe za skeniranjem
Hz	<i>Hertz</i>	Hertz – osnovna jedinica za mjerenje frekvencije
FOV	<i>Field of View</i>	Područje koje je vidljivo kroz objektiv određenog uređaja za snimanje ili promatranje (u ovom slučaju riječ je o području koje je vidljivo kroz objektiv termalne kamere)
IRFPA	<i>Infrared Focal Plane Array</i>	Infracrvena žarišna ravnina
K	<i>Kelvin</i>	Kelvin – jedinica za temperaturu u SI sustavu jedinica
NETD	<i>Noise Equivalent Temperature Difference</i>	Šum ekvivalentan razlici temperature
F-broj	<i>F-number</i>	Omjer fokalne duljine leća i promjera irisa ili otvora blende.
USB	<i>Universal Serial Bus</i>	Univerzalna serijska sabirnica, međunarodni standard za spajanje uređaja
JPEG	<i>Joint Photographic Experts Group</i>	komprimirani slikovni format s gubicima
H264 ili AVC	<i>Advanced Video Coding</i>	Međunarodni standard za kompresiju video zapisa
MPEG	<i>Moving Picture Experts Group</i>	Međunarodni standard za kompresiju video zapisa
mK	<i>miliKelvin</i>	miliKelvin, jedinica za mjerenje temperature u SI sustavu jedinica
EM	<i>Electromagnetic radiation</i>	Elektromagentsko zračenje
°C	<i>Degree Celsius</i>	Celzijev stupanj - mjerna jedinica temperature u izvedenom SI sustavu
ROI	<i>Region of Interest</i>	Regija interesa – područje na slici ili videu koje moguće sadrži objekt
FP	<i>False Positives</i>	Lažno pozitivne detekcije pri detekciji osoba na slikama ili video zapisu
HOG	<i>Histogram of Oriented Gradients</i>	Histogram gradijenta
SVM	<i>Support Vector Machine</i>	Stroj s potpornim vektorima
FPS	<i>Frame per Second</i>	Sličica u sekundi - brzina snimanja ili prikazivanja video zapisa

KNN	<i>K-nearest neighbor</i>	Algoritam K-najbližih susjeda
mRTCT	<i>modified Real-Time Tracking Compressive</i>	Modificirana verzija algoritma RTTC
FIR	<i>Far Infra Red</i>	Pojam koji se u literaturi povremeno koristi za LWIR segment IR dijela EM spektra
SURF	<i>Speeded Up Robust Features</i>	Algoritam za detekciju objekta na slikama
RBF-SVM	<i>Radial Basis Function Support Vector Machine</i>	Stroj s potpornim vektorima koji koristi RBF funkciju kao kernel
TPIHOG ili TπHOG	<i>Thermal-Position-Intensity-Histogram Of Oriented Gradient</i>	Termalna pozicija intenziteta histograma orijentiranih gradijenata
AKSVM	<i>Additive Kernel SVM</i>	Stroj s potpornim vektorima s additive kernelom
TIM	<i>Thermal Imaging Module</i>	Modul za termalno snimanje
UAV	<i>Unmanned Aerial Vehicle</i>	Bespilotna letjelica ili dron
DCT	<i>Discrete Cosine Transformation</i>	Diskretna kosinusna transformacija
2D DD DT CWT	<i>2-Dimensional Double Density Dual Tree Continuous Wavelet Transform</i>	Dvodimenzionalna kontinuirana valična transformacija dvostruke gustoće dvostrukog stabla
SDH	<i>Shape Distribution Histogram</i>	Značajke histograma distribucije oblika
CSM	<i>Contour Saliency Map</i>	Mape kontura
UKF	<i>Unscented Kalman filter</i>	Nepromijenjeni Kalmanov filter
ADAS	<i>Advanced Driver Assistance Systems</i>	Napredni sustav za pomoć vozaču
BING	<i>Binarized Normed Gradients</i>	
FFNN	<i>Feedforward Neural Network</i>	Unaprijedna neuronska mreža
ELM	<i>Extreme Learning Machine</i>	
ABMS	<i>Adaptive Boolean-map-based saliency</i>	
ATD/R	<i>Automatic Target Detection/Recognition</i>	Sustav za automatsku detekciju i raspoznavanje meta (objekata)
ANN ili NN	<i>Artificial Neural Networks; Neural Networks</i>	Umjetne neuronske mreže ili kraće neuronske mreže
AI	<i>Artificial Intelligence</i>	Umjetna inteligencija
ML	<i>Machine Learning</i>	Strojno učenje
DL	<i>Deep Learning</i>	Duboko učenje
SNN	<i>Spiking Neural Networks</i>	
MCP neuron	<i>McCulloch-Pitts neuron</i>	Model umjetnog neurona koji su predstavili McCulloch i Pitts
TLU	<i>Threshold Logic Unit</i>	Poboljšana verzija MCP umjetnog neurona
GMDH	<i>Group Method of Data Handling</i>	Grupna metoda za obradu podataka
RNN	<i>Recurrent Neural Network</i>	Ponavljajuća neuronska mreža
LSTM	<i>Long Short-Term Memory</i>	

ReLU	<i>Rectified Linear Unit</i>	Tip aktivacijske funkcije u neuronskoj mreži
Asynchronous SGD ili Async SGD	<i>Asynchronous Stochastic Gradient Descent</i>	Stohastički gradijentni spust
GAN	<i>Generative Adversarial Networks</i>	Generativne suparničke neuronske mreže
GAP	<i>Global Average Pooling</i>	
VGG net	<i>The Visual Geometry Group neural network</i>	Neuronska mreža koju je razvila The Visual Geometry Group
GoogLeNet	<i>Google neural network</i>	Googleova neuronska mreža
ResNet	<i>Residual Network</i>	Neuronska mreža koju je razvio Microsoft
DensNet	<i>Densely connected convolutional networks</i>	
FCN	<i>Fully Convolutional Networks</i>	Potpuno povezana neuronska mreža
FFD	<i>Feed-Forward Network</i>	Unaprijedna neuronska mreža
LReLU)	<i>leaky ReLU</i>	Propusna ReLU aktivacijska funkcija
RReLU	<i>Randomized ReLU</i>	Randomizirana ReLU aktivacijska funkcija
DNN	<i>Deep Neural Networks</i>	Duboke neuronske mreže
GRU	<i>Gated Recurrent Units</i>	Propusne povratne jedinice
AE	<i>Auto-Encoders</i>	Autoenkoderi
RBM	<i>Restricted Boltzmann Machines</i>	Ograničeni Boltzmanov stroj
BPVS	<i>Backpropagation with Variable Step Size</i>	Algoritam propagacije unazad s promjenjivom veličinom koraka
OABP	<i>Online Adaptive Backpropagation</i>	online adaptivni algoritam propagacije unazad
DBD	<i>delta-bar-delta algorithm</i>	delta-bar-delta algoritam
ALECO	<i>Algorithm for Learning Efficiently with Constrained Optimisation</i>	Algoritam za učinkovito učenje uz ograničenu optimizaciju
RLS BP		
FNN	<i>fuzzy neural networks</i>	Neizrazite neuronske mreže
CMAC	<i>Cerebellar Model Articulation Controller</i>	
TDNN	<i>Time Delay Neural Networks</i>	Višeslojna neuronska mreža s elementima zadržke
SCN	<i>Symmetrically Connected Networks</i>	Simetrično povezane neuronske mreže
DBN	<i>Deep Belief Network</i>	Duboka probabilistička mreža
SBN	<i>Sigmoid Belief Nets</i>	Sigmoidne probabilističke mreže
PCA	<i>Principal Component Analysis</i>	Metoda analize svojstvenih komponenti
TF	<i>TensorFlow</i>	Cjelovita platforma za strojno učenje otvorenog koda koju su razvili istraživači i inženjeri koji rade na Googleu
CPU	<i>Central Processing Unit</i>	Računalni procesor

GPU	<i>Graphical Processing Unit</i>	Grafički procesor
CLI	<i>Command Line Interface</i>	Sučelje naredbenog retka
IoU	<i>Intesection over Union</i>	Presjek preko unije
WRC	<i>Weighted-Residual-Connections</i>	Ponderirane rezidualne veze
CSP	<i>Cross-Stage-Partial-Connections</i>	
CmBN	<i>Cross mini-Batch Normalization</i>	
SAT	<i>Self-adversarial-training</i>	
CSPDarknet53	<i>Cross Stage Partial Network Darknet53</i>	Poboljšana verzija Daknet53 neuronske mreže korištene u YOLOv3
SPP	<i>Spatial pyramid pooling</i>	
PAN	<i>Path Aggregation Network</i>	
mAP	<i>Mean Average Pecision</i>	Prosječna srednja preciznost
m	<i>meter</i>	metar
VOC ili Pascal VOC	<i>Pascal Visual Object Classes</i>	Skup slika za učenje modela dubokog učenja za detekciju objekata, segmentaciju slika i sl.
MS COCO	<i>Microsoft Common Objects in Context</i>	Skup slika za učenje modela dubokog učenja za detekciju objekata, segmentaciju slika i sl.
AP	<i>Average Precision</i>	Prosječna peciznost, jedna od standardnih metrika koja se koristi u detekciji objekata
CONV	<i>Convolutional layer</i>	Često korištena kratica za konvullucijski sloj u konvolucijskoj neuronskoj mreži
TP	<i>True Positive</i>	Broj točno prepoznatih primjera pojedine klase – stvarno pozitivni rezultati
TN	<i>True Negatives</i>	Broj ispravno prepoznatih primjera koji ne pripadaju klasi – stvarno negativni primjeri
FP	<i>False Positives</i>	Broj primjera koji su pogrešno dodijeljeni klasi – lažno pozitivni primjeri
FN	<i>False Negatives</i>	Broj primjera koji nisu prepoznati kao primjeri klase – lažno negativni primjeri
PR Curve	<i>Precision/Recall curve</i>	Krivulja Preciznosti/Odziva
UNIRI-TID	<i>University of Rijeka – Thermal Imaging Dataset</i>	Skup termalnih slika za detekciju i raspoznavanje osoba u otežanim vremenskim uvjetima nastao u sklopu Sveučilišta u Rijeci.
PTZ	<i>Pan-Tilt-Zoom</i>	Panoramska kamera koja ima mogućnost pomicanja po nagibu i uvećavanja (zumiranja)

Dodatak: Programska implementacija

Za izvođenje eksperimenata korišten je YOLO detektor, koji je prethodno detaljnije opisan, implementiran u Darknet koji je izvorno razvio i predstavio Joseph Redmon, a prva verzija je bila predstavljena nedugo nakon objave radova o YOLO detektoru [289]. Darknet je platforma neuronske mreže otvorenog koda napisana u programskim jezicima C i CUDA. Riječ je platformi koja je jednostavna za instalaciju, omogućuje izvođenje na CPU i GPU, moguće ju je koristiti i u Windows i Linux okruženju. S tim u svezi treba napomenuti da je za izvedene eksperimente korišten Darknet u Linuxu (Ubuntu 18.04), koji je instaliran, odnosno radnoj stanici HP Z820 sa dva CPU-a Intel Xeon E5 2690v2 @3.00 Ghz (10 fizičkih jezgri po CPU), 32 GB DDR3 ECC RAM, GPU Nvidia GTX 1080 Ti 11 GB, SSD 2 TB Samsung SCSI disk. Nadalje, vezano za programsku podršku, Darknet zahtijeva instalaciju OpenCV (engl. *Open Source Computer Vision Library*) biblioteka [325], a riječ je o bibliotekama softvera za računalni vid i strojno učenje otvorenog koda. OpenCV je izgrađen kako bi pružio zajedničku infrastrukturu za programe računalnog vida i ubrzao upotrebu percepcije strojeva u komercijalnim proizvodima. Budući da je proizvod s licencom BSD, OpenCV olakšava tvrtkama upotrebu i izmjenu koda. Pored toga, Darknet zahtijeva i CUDA (engl. *Compute Unified Device Architecture*) koju je razvila tvrtka Nvidia [326], a riječ je o računalnoj platformi za paralelnu obradu podataka i modelu programiranja koji je namijenjen izvođenju operacija na Nvidia GPU-ima. CUDA programerima omogućava ubrzavanje računalno zahtjevnih aplikacija iskorištavanjem snage GPU-a za paralelnu obradu. Iako su predloženi i drugi API-ji za GPU-ove, kao što je OpenCL [327], a postoje i konkurentni GPU-ovi drugih tvrtki, poput AMD-a, kombinacija CUDA platforme i Nvidia GPU-a dominira u nekoliko područja primjene, uključujući duboko učenje i temelj je za neke od njih. Nakon instalacije navedenih zahtjeva i potrebnih biblioteka, instalacija Darknet platforme je vrlo jednostavna, posebice u Linux Ubuntu okruženju. Sam proces instalacije detaljno je opisan na internetskim stranicama Darkneta [289]. Osim izvorne verzije Darkneta koju je razvio J. Redmon, dostupno je nekoliko drugih verzija od kojih svaka ima svoje određene specifičnosti, odnosno prednosti i nedostatke u odnosu na druge verzije. Sukladno tomu, za izvođenje eksperimenata u okviru ovog doktorskog rada odabrana je verzija Darkneta koju je razvio Alexey Bochkovskiy, koji je i jedan od autora četvrte verzije YOLO detektora (YOLOv4). Ova verzija Darkneta dostupna je na GitHub repozitoriju [293], te posljednja verzija dostupna za download obuhvaća sve četiri verzije YOLO detektora (YOLO v1 –

v4), a u repozitoriju se nalaze i detaljna uputstva za instalaciju i korištenje Darkneta, odnosno upute za učenje i testiranje vlastitih modela za detekciju, raspoznavanje i klasifikaciju objekta. Darknet se izvodi upisivanjem naredbi u naredbeni redak u Linux terminalu (engl. CLI – *Command Line*), a u svakoj naredbi je potrebno specificirati određene parametre. Nadalje, za početno učenje vlastitog modela, osim definiranja skupa za učenje i skupa testiranje, potrebno je koristiti jedan od postojećih modela na temelju kojeg se uči novi model.

Prije početka učenja novog modela, osim skupa za učenje i testnog skupa, potrebno je prilagoditi osnovne parametre unutar konfiguracijske datoteke (.cfg), te kreirati odgovarajuću .data i .names datoteku. Unutar .cfg datoteke nužno je prilagoditi broj filtera broju klasa u skupu za učenje nakon svakog YOLO sloja, pri čemu se broj filtera određuje prema formuli:

$$\text{filters} = (\text{classes} + 5) \times 3 \quad (32)$$

gdje parametar *classes* označava broj klasa u skupu za učenje.

Nadalje, u .cfg datoteci moguće konfiguriranje i drugih parametara značajnih za točnost učenja modela, koji su prethodno opisani u 6.3.1. Osim nužnih izmjena, potrebno je pripremiti .data datoteku čiji je sadržaj prikazan na slici 80., a u ovom slučaju riječ je .data datoteci koja je krištenja za učenje modela na skupu podatak All_in_One

```
classes= 1
train = ALL_IN_ONE_RGB-GREY_train_10.txt
valid = ALL_IN_ONE_RGB-GREY_test_90.txt
names = ALL_IN_ONE_RGB-GREY.names
backup = backup/
```

Slika 83. Sadržaj datoteke .data unutar Darknet platforme

U navedenom primjeru *classes* označava broja klasa u skupu za učenje, *train* označava putanju do skupa za učenje, *valid* označava putanju do skupa za testiranje, dok *backup* označava putanju do mape u kojoj će biti spremljen naučeni model.

Treća datoteka koju treba pripremiti prije početka učenja novog modela je `.names`, koja sadrži nazive svih klasa u skupu za učenje, pri čemu se naziv svake klase upisuje u novi red.

Učenje novog modela pokreće se naredbom:

```
./darknet detector train data/ALL_IN_ONE_RGB-GREY_20-80.data cfg/ALL_IN_ONE_RGB-GREY.cfg darknet53.conv.74 -map,
```

gdje argument `-map` dodatno označava da će tijekom procesa učenja biti izračunavana mAP vrijednost na zadanom skupu za testiranje, te istovremeno prikazana na grafu koji prikazuje vrijednost mAP i vrijednost funkcije gubitka, kao što je prikazano na slici 64.

Trajanje učenja novog modela se može ograničiti prema broju iteracija, ali ne manje od dvostrukog broja klasa u skupu za učenje, odnosno ukoliko skup sadrži dvije klase, tada učenje traje minimalno 4.000 iteracija, a drugi parametar za raniji prekid učenja je kada vrijednost funkcije gubitka padne ispod 0,05 jer nakon toga može doći do pretreniranosti modela.

Osim u okviru opisane platforme, YOLO detektor je moguće koristiti u okviru drugih platformi, pa tako primjerice YOLOv4 je dostupan za korištenje u sklopu Tensorflow, OpenCV.dnn, Intel OpenVINO 2020 R4, Tencent/ncnn, PyTorch>ONNX, TensorRT, TVM, OpenDataCam [328].

Životopis

Mate Krišto, rođen 3.4.1980. godine u Tomislavgradu (BiH), gdje je završio Opću gimnaziju Marka Marulića (1999. god), a nakon toga je diplomirao na Ekonomskom fakultetu Sveučilišta u Rijeci – smjer management (2006. god). Doktorski studij - modul inteligentni računalni sustavi na Odjelu za informatiku Sveučilišta u Rijeci upisao je 2012. godine. Zaposlen, oženjen, otac dvoje djece.

Popis objavljenih radova:

1. Krišto, Mate. "Review Of Methods For The Surveillance And Access Control Using The Thermal Imaging System." Review of Innovation and Competitiveness: A Journal of Economic and Social Research 2.4 (2016): 71-91.
2. Krišto, Mate, and Marina Ivasic-Kos. "An overview of thermal face recognition methods." 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, 2018.
3. Ivašić-Kos, Marina, Mate Krišto, and Miran Pobar. "Human detection in thermal imaging using YOLO." Proceedings of the 2019 5th International Conference on Computer and Technology Applications. 2019.
4. Krišto, M., and Marina Ivašić-Kos. "Thermal imaging dataset for person detection." 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, 2019.
5. Ivasic-Kos, Marina, Mate Kristo, and Miran Pobar. "Person Detection in thermal videos using YOLO." Proceedings of SAI Intelligent Systems Conference. Springer, Cham, 2019.
6. Krišto, Mate, Marina Ivasic-Kos, and Miran Pobar. "Thermal Object Detection in Difficult Weather Conditions Using YOLO." IEEE Access 8 (2020): 125459-125476.

Sudjelovanja u projektima:

Research Class: Aktivosti u sklopu projekta HRZZ "Automatsko raspoznavanje akcija i aktivnosti u multimedijalnom sadržaju iz domene sporta - RAASS"

Popis priloga i dodataka

Prilog 1. Odobrenje Ministarstva obrane Republike Hrvatske za korištenje termnog video zapisa s vojne vježbe Swift Response 2019

Prilog 2. USB memorija koja sadrži UNIRI Thermal imaging dataset for human detection and recognition