SVEUČILIŠTE U SPLITU

FAKULTET ELEKTROTEHNIKE, STROJARSTVA I BRODOGRADNJE

POSLIJEDIPLOMSKI DOKTORSKI STUDIJ ELEKTROTEHNIKE I INFORMACIJSKIH TEHNOLOGIJA

KVALIFIKACIJSKI ISPIT

DETEKCIJA LJUDI NA SLIKAMA SNIMLJENIH IZ ZRAKA

Danijel Zelenika

Split, siječnja 2020.

Sadržaj:

1 Uvod

Sve veća prisutnost prirodnih katastrofa kao što su požari, potresi i poplave povećava potrebu za pronalaženje učinkovitog načina za spašavanje ljudi koji se nalaze u neposrednoj opasnosti. Prema istraživanju [1], najveću šansu za preživljavanje imaju osobe pronađene unutar prvih 48 sati, od početka akcije spašavanja. Nakon toga, vjerojatnost za spašavanje naglo opada.

Misije, potrage i spašavanja uglavnom su vezane za nepristupačne terene kao što su: šume, more, pustinja i planine, te uključuju prilično veliki broj ljudi. Profil ljudi koji sudjeluje u potragama spašavanja se kreće od posebno obučenih profesionalaca[1], medicinskog osoblja, vatrogasaca i policajaca, pa sve do običnih ljudi koji su usko vezani sa osobama za kojima se traga. Svi oni su vrlo izloženi opasnim situacijama, koje mogu nastati zbog utjecaja različitih prijetnji kao što su klizišta, odroni, ruševine i slično. U ovakvim okolnostima, spasitelj vrlo lako može postati žrtva koju treba spašavati. S ove točke gledišta, postoji vrlo opravdan razlog za pronalazak alternativnog načina, koji će smanjiti rizik i povećati učinkovitost akcije spašavanja.

Zbog toga se u misijama potrage i spašavanja vrlo često koriste potražni psi, jer lako mogu detektirati prisutnost čovjeka. Međutim, u ovakvim situacijama je teško biti potpuno ovisan o njima, jer ne mogu prosuditi situaciju i prenijeti informacije. Zbog toga se potražni psi koriste paralelno kao ispomoć ljudima.

Akcija spašavanja može trajati nekoliko dana, a u nekim slučajevima i po nekoliko mjeseci, što nije rijedak slučaj s obzirom da je potrebno pretražiti relativno veliko područje. Za brže pretraživanje terena obično se koriste helikopteri, pri čemu se vrši vizualna inspekcija. Ovaj način inspekcije je jako skup i ne garantira pronalazak nestale osobe. Nedovoljna fokusiranost promatrača, uzrokovana umorom i iscrpljenošću, neki su od ključnih faktora koji znatno utječu na uspješnost operacije spašavanja. Ovaj nedostatak je moguće djelomično ispraviti, uzimanjem fotografija za vrijeme zračne inspekcije, te na taj način ostaviti mogućnost za dodatnu analizu kako bi sa što većom sigurnošću mogli eliminirati neka područja.

Vizualna inspekcija slika je veoma dugotrajan i iscrpan posao za čovjeka, jer se radi sa slikama visoke rezolucije na kojima ljudi zauzimaju relativno malo prostora u odnosu na veličinu slike.

[1] Osoba koja je uspješno završila SAR obuku.

Prema istraživanju napravljenom u radu [2], za inspekciju jedne slike čovjeku je potrebno od 15 do 90 sekundi. Ovakav način inspekcije nije podoban za opsežne i dugotrajne potrage, jer iziskuje prilično velike računalne i ljudske resurse.

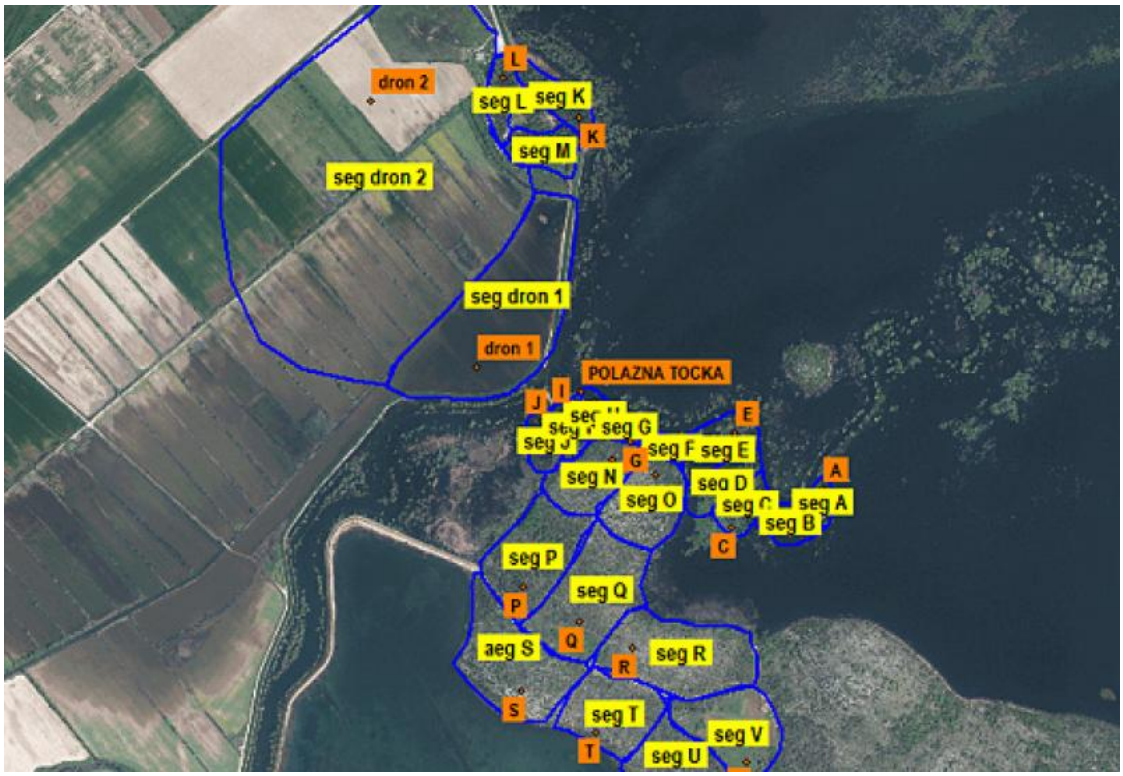
Sustav za vizualno raspoznavanje objekata je jedno od rješenja, koja svoju primjenu mogu pronaći u misijama potrage i spašavanja. Zadaće vizualnog raspoznavanja, kao što su klasifikacija slika i detekcija objekata, dugi niz godina se primjenjuju za rješavanje problema u području medicine, robotike, pametnog upravljanja, razvoju autonomnih automobila i drugim područjima. Zbog toga postoji vrlo opravdan interes za razvojem sustava koji će automatski detektirati ljude na slikama snimljenim iz zraka.

Ovaj rad je podijeljen na sljedeći način. U drugom poglavlju su opisane mogućnosti i problemi primjene bespilotnih letjelica u misijama potrage i spašavanja. U trećem poglavlju dat je pregled dosadašnjih istraživanja, koji razmatraju mogućnosti primjene konvencionalnih algoritama i tehnika za obradu slike, za rješavanje problema detekcije ljudi u slikama snimljenih iz zraka. Opisane su prednosti i nedostatci svakog od pristupa, te klasični problemi s kojima se susreću istraživači u ovom području. U četvrtom poglavlju opisana je arhitektura algoritama temeljenih na dubokom učenju, kao i mogućnosti primjene ovih algoritama za rješavanje već navedenih problema. Opisana je arhitektura klasične konvolucijske neuronske mreže, te najznačajniji algoritmi koji koriste duboko učenje za detekciju objekata. U petom poglavlju opisane su strategije prijenosnog učenja, a u šestom piramide značajki za detekciju objekata. U posljednjem, sedmom poglavlju dat je zaključak.

2 Primjena bespilotnih letjelica za pretraživanje terena

Početak akcije spašavanja započinje pozivom u pomoć, u svrhu prijave nestanka osobe, a završava pronalaskom nestale osobe [3]. Nakon zaprimljenog poziva, potražni tim prikuplja informacije o nestaloj osobi: starosna dob, fizički opis, zdravstveno i psihološko stanje, te vrijeme i lokaciju kada je osoba zadnji put viđena.

Ove informacije su krucijalne, kako bi nestalu osobu svrstali u odgovarajuću kategoriju, na temelju koje je moguće odrediti potražnu mapu. Za svaku kategoriju, postoje točno određeni statistički obrasci ponašanja, koji se koriste za određivanje pravca kretanja nestale osobe. Drugi faktori koji imaju značajnu ulogu u određivanju potražne mape su: tip terena, razina pošumljenosti, meteorološki uvjeti, te vrijeme nestanka osobe. Na Slika 1. je prikazan je primjer potražne mape, koja je izrađena u svrhu izvođenja potražne vježbe[2] na području Hutova Blata.



Slika 1. Potražna mapa, na području Hutova blata

[2] U potražnoj vježbi je sudjelovalo ukupno 72 sudionika iz deset spasilačkih službi saveza gorskih službi spašavanja u Bosni i Hercegovini. U vježbi su sudjelovali i pripadnici policije, vatrogasaca, civilne zaštite i djelatnici Sveučilišta u Mostaru.

Potražna mapa na Slika 1. sastoji se od ukupno 25 segmenata za pretraživanje koji se mogu podijeliti prema tipu terena. Segmenti „dron 1“ i „dron 2“ se nalaze na pretežno travnatom tipu terena, te su predviđeni za inspekciju iz zraka. Ostali segmenti se nalaze na blago pošumljenom terenu i predviđeni su za neposrednu inspekciju s terena, korištenjem potražnih timova i pasa. Manji broj segmenata je pozicioniran na rijeci i na močvarnom tipu terena. Ovo je tipičan primjer hercegovačkog krajolika.

Bespilotne letjelice veoma brzo pokrivaju veliko područje i omogućuju jednostavan pristup udaljenim i teško dostupnim lokacijama [4], zbog čega imaju veoma važnu ulogu u misijama potrage i spašavanja. Bespilotne letjelice pružaju odgovor na novu perspektivu pretraživanja terena, ali su njihove mogućnosti ograničene kada je riječ o sustavnom pretraživanju terena. Relativno veliko područje, ograničeno vrijeme leta, te memorijski i računalni zahtjevi za obradu podataka, neki su od problema koji ograničavaju njihovu primjenu. U nastavku je opisan način planiranja potrage korištenjem bespilotnih letjelica u misijama potrage i spašavanja.

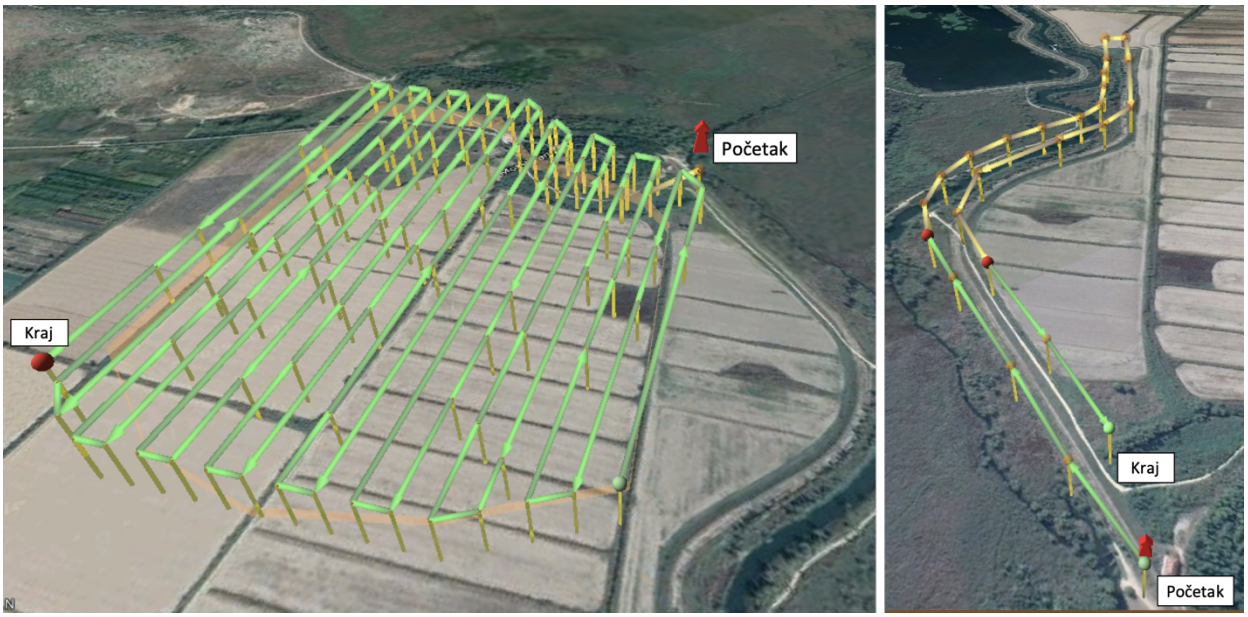
2.1 Planiranje potrage bespilotnim letjelicama

Pretraživanje trase

Prema proceduri za vođenje i planiranje operacije potrage i spašavanja, prva faza uključuje pretraživanje cesta i područja koja se nalaze neposredno uz cestu. U ovoj fazi bespilotna letjelica se programira za pretraživanje terena prema zadanoj trasi koja prati smjer ceste, kao na Slika 2 b). Vrijeme trajanja leta i broj generiranih slika ovisi o parametrima leta i vrsti kamere.

Za planiranje leta za snimanja terena sa Slika 2. b) korišten program *QgroundControl,* koji automatski kreira plan leta na osnovu postavljenih parametara, kako bi na najoptimalniji način pokrio teren iznad kojeg se izvodi let. Snimanje se obavlja s visine od 50 metara, a iznos frontalnog i bočnog preklapanja između slika je 30%. Za snimanje fotografija se koristi *Phantom 4 Pro* s ugrađenom kamerom, čiji je senzor dimenzija 13.2 x 8 mm, a žarišna udaljenost leće 8.8 mm. Pretražuje se područje širine 150 m uz cestu. Ukupna površina terena kojeg je potrebno pokriti iznosi 146.962,00 m^2.

Uz odabrane parametre, potrebno je napraviti ukupno tri preleta kako bi pokrili cijelo područje. Ukupna dužina leta je 3824 m, a vrijeme trajanja leta 12 min i 46 sekundi, pri čemu se generira 85 slika. Vrijeme okidanja između fotografija je 7 sekundi.



a) Pretraživanje rešetkom b) Pretraživanje trasom

Slika 2. Scenarij za pretraživanje terena pomoću bespilotne letjelice

U Tablica 1. je prikazan statistički pregled misija za pretraživanje terena sa Slika 2. b), za nekoliko različitih bespilotnih letjelica. Kao što se vidi iz tablice, postoji značajna razlika u žarišnoj udaljenosti leće između testiranih letjelica, što se reflektira razlikom u veličini vidnog polja i rezoluciji na tlu. Manja rezolucija na tlu znači da kamera može uhvatiti više detalja, što je veoma važno za vizualnu inspekciju slike.

Ljudi na slikama snimljenim iz zraka trebaju imati dovoljan broj detalja, kako bi se mogao napraviti reprezentativan računalni model. S druge strane, premalo vidno polje se odražava na povećan broj slika i ukupno vrijeme leta, što je kontraproduktivno za misije potrage i spašavanja. Stoga je optimalno određivanje ovih parametara ključan faktor za uspjeh misije spašavanja.

Tablica 1. Statistički pregled rezultata pretraživanja terena trasom

UAV	Dimenzije senzora	Fokalna dužina	Dimenzije slike	Broj slika	Interval (s)	Vrijeme (min)	Rezolucija (cm/px)
Phantom 3 Pro	6.17 x 4.55	3.57	4000 x 3000	85	9.1	12:46	2.2
Phantom 4 Pro	13.2 x 8	8.8	5472 x 3684	109	7.1	12:40	1.4
Inspire 1 X5	17.3 x 13	13	4592 x 3448	145	7.0	12:41	1.4

Pretraživanje terena rešetkom

U Tablica 2. su prikazani statistički rezultati pretraživanja terena prikazanog na Slika 2 a), s više različitih bespilotnih letjelica. Pretraživanje se obavlja s visine od 50 m, a ukupna površina terena je 553500 m^2. Izbor kamere znatno utječe na ukupno trajanje misije i broj slika. Broj generiranih slika se kreće u rasponu od 236 do 370, a vrijeme trajanja leta od 33 do 42 minute.

Tablica 2. Statistički pregled rezultata pretraživanja terena rešetkom

UAV	Dimenzije senzora	Fokalna dužina	Dimenzije slike	Broj slika	Interval (s)	Vrijeme (min)	Rezolucija (cm/px)
Phantom 3 Pro	6.17 x 4.55	3.57	4000 x 3000	236	9.1	33	2.2
Phantom 4 Pro	13.2 x 8	8.8	5472 x 3684	324	7.1	38	1.4
Inspire 1 X5	17.3 x 13	13	4592 x 3448	370	7.0	42	1.4

Kao što se vidi iz rezultata analize snimanja terena prikazanih na Slika 2, generira se značajan broj slika koje je potrebno pretražiti. Ručna inspekcija slika je veoma zahtjevan zadatak za čovjeka, stoga je neophodno pronaći način, koji će u najboljem slučaju automatski obraditi slike, ili poluautomatski uz pomoć čovjeka. U okviru sljedećeg poglavlja, opisani su najznačajniji pristupi, koji koriste tradicionalne tehnike za obradu slika i mogućnosti primjene za rješavanje problema u misijama potrage i spašavanja.

3 Tradicionalni sustavi za detekciju objekata

Detekcija ljudi sa slika snimljenih iz zraka je jedan od najzahtjevnijih zadataka računalnog vida. Različiti tipovi terena, varijabilan oblik, različita boja odjeće, te nagle promjene osvjetljenja neki su od glavnih problema s kojima se susreću istraživači u ovom području [5].

U literaturi je dostupan relativno mali broj radova koji se bavi ovom problematikom. U najvećem broju radova, autori se bave razvojem modela za prepoznavanje objekata u pokretu, koji svoju primjenu najčešće nalaze u automobilskoj industriji za prepoznavanje pješaka ili sigurnosnom nadzoru. Najjednostavnija metoda za detekciju objekata u pokretu je tehnika oduzimanja pozadine. Međutim, ova metoda nije prikladna za detekciju objekata u slici, jer se objekt od interesa ne pomjera. Konvencionalne metode za detekciju pješaka koriste informacije o teksturi i obliku objekta, te postižu veoma dobre rezultate. Međutim, na snimkama snimljenim iz zraka nedostaje veliki broj značajki, koje su normalno dostupne za objekte koji zauzimaju veću površinu u slici, što ograničava primjenu ovih metoda u misijama potrage i spašavanja.

U jednom od prvih radova koji se bavi ovom problematikom, autori predlažu detektor za prepoznavanje osoba u pokretu [6]. Detektor se sastoji od dva dijela. U prvom dijelu se biraju najbolje značajke pomoću AdBoost algoritma, a nakon toga se obavlja proces klasifikacije. Sustav za klasifikaciju se sastoji od niza ručno dizajniranih pravila, baziranih na Haarovim značajkama i prostorno-vremenskoj razlici.

Nekoliko autora je predložilo bimodalni sustav, koji koristi bespilotnu letjelicu opremljenu termalnom i optičkom kamerom [7]. Detekcija se obavlja fuzijom informacija dobivenih iz dva izvora podataka. Glavni cilj je razviti sustav, koji u stvarnom vremenu detektira ljude i automobile. Za detekciju automobila se koristi nekoliko Haar klasifikatora spojenih u niz i informacije iz optičke slike, a termalna slika se u ovom slučaju koristi samo za potvrdu detekcije. Za detekciju ljudi koristi se obrnuti postupak. Kaskadni Haarov klasifikator se prvo trenira na slikama iz termalne kamere, a nakon toga se za potvrdnu detekcije na optičkoj slici primjenjuje tehnika multi varijantnog Gaussovog oblika.

U radu [8], autori predlažu histogram orijentiranih gradijenata (eng. HOG, Histogram of Oriented Gradients) za detekciju i identifikaciju ljudi iz zračnih snimaka. Iako su dobiveni vrlo

zadovoljavajući rezultati, ovaj rad nije relevantan za primjenu u misijama potrage i spašavanja, jer se u scenarijima koristi vrlo jednostavno okruženje koje ograničava primjenu ove tehnike.

U radu [9], autori koriste informacije iz termalne i optičke kamere za detekciju ljudi na zračnim slikama. Termalna kamera se koristi za identifikaciju regija s visokom temperaturom, koje se analiziraju u optičkoj slici. Detekcija se obavlja pomoću kaskadnog klasifikatora, koji koristi Haarove značajke.

Glavni nedostatak sustava s termalnom kamerom su izrazito visoke temperature u ljetnom dijelu godine, pogotovo u mediteranskim regijama, gdje se događa najveći dio potraga za izgubljenim turistima. Zbog navedenih problema, u radu [10] autori predlažu metodu, koja se temelji na algoritmu za segmentaciju slike. Predložena metoda, koristi optičku kameru kao glavni izvor informacija te predlaže lokacije na kojima se potencijalno nalazi čovjek, što znatno olakšava proces vizualne inspekcije. Glavni nedostatak metode je velika količina lažnih alarma.

Kako bi prevladali ove nedostatke u radu [11], autori predlažu novi model koji poboljšava postupak vizualne inspekcije, koristeći se tehnikom za otkrivanje ispupčenih objekata u slici i konvolucijskih neuronskih mreža. U radu je napravljena usporedba predloženog modela i trenutno dostupnih algoritama za detekciju malih objekata u slici.

Proces vizualnog prepoznavanja ljudi iz zračnih snimaka se temelji na prepoznavanju odjeće i oblika koji je karakterističan za ljudsko tijelo. Na slikama snimljenim iz zraka, glava i ramena su najistaknutiji dijelovi ljudskog tijela, koji se mogu koristiti za prepoznavanje. Koristeći samo ovaj set podataka, ljudi s velikom preciznošću mogu prepoznati objekte na slici. Na sličan način moguće je podesiti i algoritme za detekciju ljudi na zračnim snimkama. Zbog toga se većina prethodno opisanih algoritama temeljni na deskriptorima značajki, koji na najbolji način opisuju ljude na slici. U nastavku je dat pregled najkorištenijih algoritama za izdvajanje značajki.

3.1 Izdvajanje značajki iz slike

U prethodnih nekoliko desetljeća deskriptori značajki čine sastavni dio bilo kojeg algoritma za detekciju objekata u slici. Proces detekcije se sastoji od pronalaska značajki, povezivanja značajki i klasifikacije. Da bi se postiglo pouzdano prepoznavanje, veoma je važno da značajke budu otporne na promjene u veličini objekta, orijentaciji i osvjetljenju slike. Područja s visokom promjenom intenziteta, kao što su rubovi, krajevi i promjene boje spadaju u detalje koji su najviše

otporni na ove promjene. Zbog toga se većina deskriptora sastoji od detekcije rubova i krajeva objekata.

SIFT (eng. Scale-Invariant Feature Transform)

SIFT je jedan je od najčešće korištenih deskriptora značajki. Glavna prednost algoritma je invarijantnost na skaliranje, rotaciju i promjene osvjetljenja, što smanjuje doprinos pogreškama uzrokovanim lokalnim promjenama u slici. SIFT može veoma dobro identificirati djelomično skrivene objekte u slici. Ovo svojstvo je veoma važno za primjenu u misijama potrage i spašavanja, jer se ljudi obično nalaze u zaklonu od vremenskih nepogoda, kao što su sunce, vjetar i slično. Glavni nedostatak algoritma je sporost, što ograničava primjenu ovog algoritma u sustavima koji rade u stvarnom vremenu.

SURF (eng. Speeded-Up Robust Features)

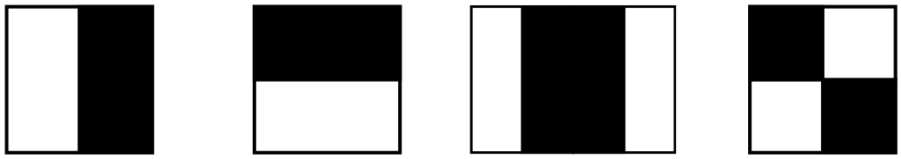
SURF se bazira na sličnim pravilima kao SIFT, ali uz niz matematičkih poboljšanja postiže bolje rezultate u smislu brzine. Algoritam prvo računa Haarove valiće u x i y smjeru, a nakon toga radi vrlo jednostavnu aproksimaciju Hessianove matrice, te na taj način ubrzava proces izdvajanja značajki. Glavna prednost algoritma je dovoljno veliki broj značajki, na temelju kojih se može prepoznati objekt. Ova prednost je ujedno i glavni nedostatak ovog algoritma, pogotovu u situacijama kada objekt nije prisutan u slici. SURF se najčešće koristi za spajanje slika (*eng. stitching*) [12], a za opis značajki se koristi 64 dimenzionalni vektor. U praksi se vrlo često koriste algoritmi za smanjenje dimenzija vektora značajki, radi ubrzanja procesa detekcija. Najčešće korišteni algoritmi su: metoda glavnih komponenti (eng. Principal Component Analysis - PCA) i analiza nezavisnih komponenti (eng. Independent Component Analysis - ICA).

HOG (eng. Histogram of Oriented Gradients)

Histogram orijentiranih gradijenata je jedna od osnovnih tehnika za detekciju i lokalizaciju objekta u slici. Glavna prednost ovih značajki je veoma brz i jednostavan izračun, koji se može brzo obaviti i na običnom procesoru. Informacije o gradijentu se računaju u histogramu orijentacija, koji služi kao vektor značajki, i kao takav se koristi kao ulaz u algoritme strojnog učenja. Na HOG značajkama se obično primjenjuje PCA metoda, što značajno ubrzava proces treniranja. U radu [8] je predstavljen detektor, koji za prepoznavanje objekata iz zraka koristi informacije o obliku objekta.

Haarove značajke

Haarove značajke su prvi put predstavljene u algoritmu za detekciju lica kojeg su 2002. godine predložili Paul Viola i Michael Jones. Iako je primarna svrha detekcija lica, algoritam se u kombinaciji sa strojnim učenjem može iskoristiti za detekciju bilo kojeg objekta. Prednost algoritma dolazi do izražaja zbog činjenice, jer sve operacije na slici radi sa skupinama piksela, što je veoma značajno za obradu u stvarnom vremenu. Haarove značajke se računaju kao razlika sume piksela unutar predefinirani područja u slici. Tako izračunate vrijednosti predstavljaju prisustvo ili odsustvo određenih značajki u slici, kao što su rubovi i linije. U originalnom radu se koriste tri tipa značajki sa dva, tri i četiri pravokutnika. Na Slika 3 je prikazan primjer ovih značajki.



Slika 3. Haarove značajke

Značajke sa dva pravokutnika se obično koriste za detekciju rubova, značajke s tri pravokutnika za detekciju linija, a značajke s četiri pravokutnika za detekciju kose crte.

Haarove značajke jako ovise o kvaliteti slike, što je ujedno i njihov glavni nedostatak. Ukoliko rubovi nisu jasno označeni, detektor će imati lošije rezultate. Klasifikacija se obavlja pomoću metoda strojnog učenja, a u nastavku je dat pregled najznačajnijih algoritama za klasifikaciju objekata.

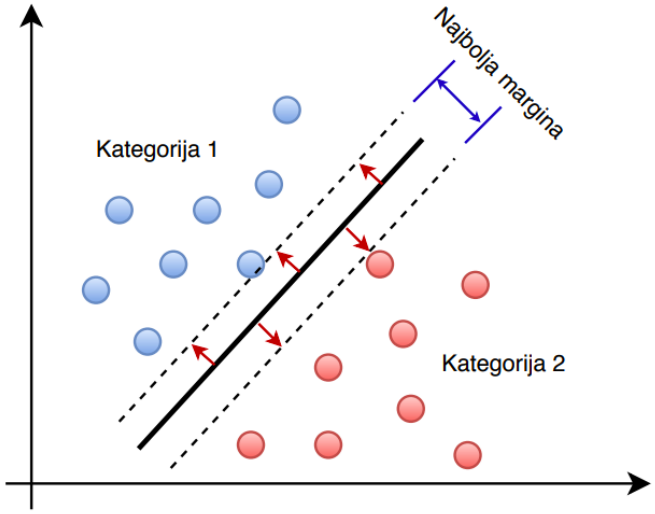
3.2 Klasifikacija objekata

Klasifikacija je operacija koja svaku instancu iz skupa podataka smješta u jednu od unaprijed definiranih kategorija s određenom preciznošću. Algoritam za klasifikaciju generira model na osnovu ulaznog skupa podataka i unaprijed poznate kategorije za svaku instancu. Model se obično naziva klasifikator, a proces klasifikacije se obavlja u tri koraka: treniranje, testiranje, i validacija.

SVM (eng. Support Vector Machine)

SVM je jedna od najpopularnijih metoda za klasifikaciju, koja se bazira na podjeli podataka u dvije kategorije. Na Slika 4. je predstavljena osnovna ideja algoritma. Glavni cilj je pronaći granicu koja

će dati maksimalno odstupanje od ulaznih podataka. Klasifikator je precizniji ukoliko je granica više udaljena od podataka. Ovaj postupak se naziva linearna klasifikacija.



Slika 4. Klasifikacija korištenjem SVM klasifikatora

Funkciju koja dijeli niz ulaznih podataka na dva dijela možemo opisati funkcijom D, prikazanom u izrazu (3.1), gdje je W vektor normale na ravninu, a b slobodni član.

$$D(x) = W_x + b \tag{3.1}$$

Ako je klasifikator pravilno utreniran, onda za ulazni podatak x vrijedi relacija (3.2).

$$x \in \begin{cases} 0, & D(x) < 0 \\ 1, & D(x) \geq 0 \end{cases} \tag{3.2}$$

Postupak treniranja se obavlja dok se svi podaci pravilno klasificiraju, ili dok se ne postigne odgovarajuća preciznost. SVM je prvenstveno predviđen za binarnu klasifikaciju, ali postoje određeni načini na koje se ovaj klasifikator može prilagoditi za rješavanje problema s više kategorija. Jedan od načina je da se za svaku kategoriju koristi zaseban klasifikator, a izlazi se promatraju kao pripadnost ili nepripadnost određenoj kategoriji. SVM je u kombinaciji s HOG značajkama jedan od najrasprostranjenijih pristupa za detekciju objekata, na snimkama snimljenim iz zraka [8] [13].

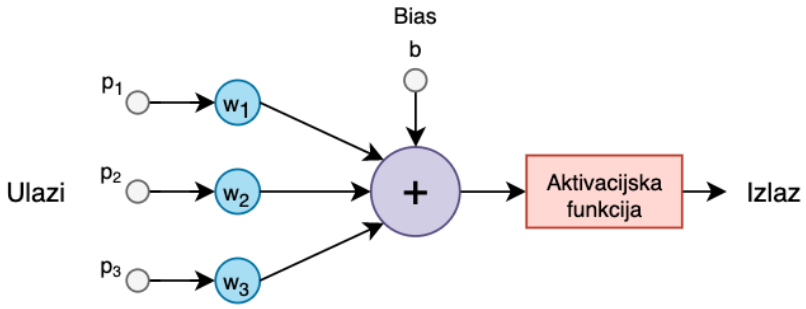
Neuronske mreže

Početak neuronskih mreža se veže za 1943. godinu kada su znanstvenici Warren McCulloch i Walter Pitts objavili rad, kojim su dokazali da se pomoću neuronskih mreža može izračunati bilo

koja aritmetička ili logička funkcija. Ovaj rad je imao jako velik utjecaj na primjenu neuronskih mreža za rješavanje drugih problema. Neuronska mreža se sastoji od velikog broja neurona organiziranih u slojeve. Arhitektura jednog takvog neurona je prikazana na Slika 5. Ulaz u mrežu, označen oznakom p, množi se s težinskim faktorom W i zbraja s internim ulazom pomaka b. Na osnovu ovih koeficijenata proizvodi se aktivacijsko djelovanje mreže, definirano izrazom (3.3).

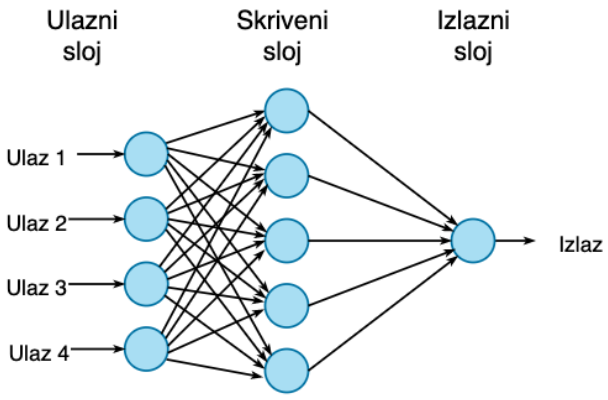
$$h = f(a) = f(W_P + b) \tag{3.3}$$

Ukoliko neuronska mreža ima više ulaza, faktori p i W su predstavljeni s vektorima. Podešavanje neurona se obavlja pomoću težinskih faktora W i pomakom b. Isti ulaz se može dovesti na više pojedinačnih neurona.



Slika 5. Model neurona

Na Slika 6. je prikazan primjer jednoslojne neuronske mreže. U praksi se obično koriste složenije mreže sastavljene od više neurona. Arhitektura mreže se obično sastoji od ulaznog sloja, na čije se ulaze obično dovode podaci sa senzora, skrivenog sloja, te izlaznog sloja. Jedna neuronska mreža obično ima jedan ulazni i izlazni sloj te više skrivenih slojeva.



Slika 6. Arhitektura višeslojne neuronske mreže

3.3 Segmentacija slike

Detekcija malih objekata iz zračnih snimaka može se obaviti klasičnim postupkom za segmentaciju podataka. Glavni cilj bilo kojeg algoritma za segmentaciju, je podijeliti sliku na semantički značajna područja. Glavni motiv za primjenu ovih metoda za detekciju ljudi u zračnim slikama, proizlazi iz činjenice da se misije potrage i spašavanje uglavnom događaju u ne urbanim područjima, na kojima se nalazi vrlo malo ili nimalo neprirodnih objekata. Primjenom klasičnih metoda za segmentaciju moguće je regije na slici podijeliti na prirodne i neprirodne. U jednom od prvih radova koji istražuju mogućnosti primjene ove tehnike za detekciju ljudi na slikama snimljenim iz zraka, autori za segmentaciju koriste algoritam pomaka sredina [10]. Slika se prvo dijeli na nekoliko manjih dijelova nad kojima se obavlja proces segmentacije, a u drugom dijelu se na osnovu centara svih klastera, ponovo primjenjuje ista metoda za segmentaciju. Na kraju se obavlja finalna selekcija, gdje se biraju regije koje imaju najveću vjerojatnost da se unutar njih u ulaznoj slici nalazi neprirodni materijal.

Algoritam pomaka sredina

Algoritam pomaka sredina je jedna od najjednostavnijih metoda za grupiranje podataka. Srž algoritma je postupak za pronalaženje vršnih točaka u višedimenzionalnom prostoru. Jedan od najjednostavnijih postupaka za izglađivanje podataka je postupak konvolucije. Na svaku točku u ulaznom skupu podataka se postavlja jezgra određene širine. Rezultat ove operacije je funkcija gustoće, koja direktno ovisi o propusnosti jezgre. Algoritam pomaka sredina iterativno pomjera sve točke, kako bi došli do najbližeg vrha na funkciji gustoće. Jezgra vrlo uske širine kao rezultat daje funkciju gustoće koja odgovara veličini ulazne slike. Zbog toga će se svaka točka u ulaznom skupu podataka smjestiti u svoj klaster. Suprotan slučaj je primjena jezgre dovoljno velike veličine, na način da se sve točke u ulaznom skupu podataka smjeste u jedan klaster. Sve vrijednosti širine jezgre između ova dva slučaja daju bolju raspodjelu ulaznih podatka.

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x) x_i}{\sum_{x_i \in N(x)} K(x_i - x)} \qquad (3.4)$$

4 Duboko učenje za detekciju objekata

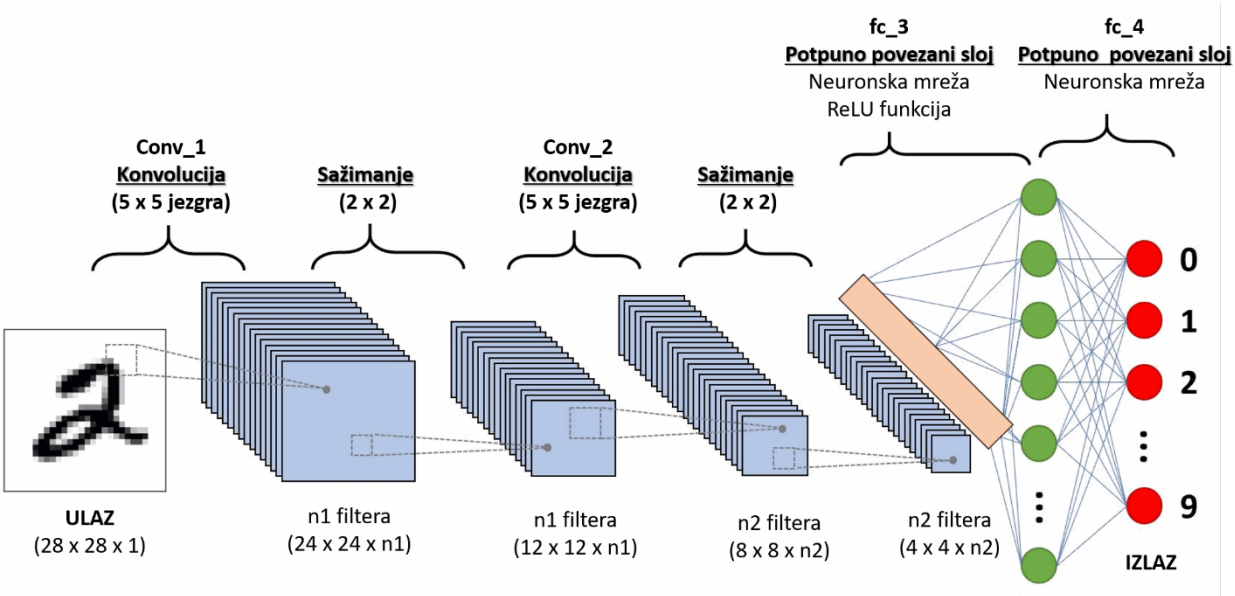
Glavnu komponentu u procesu obavljanja bilo kojeg zadatka čini čovjek. Nakon vrlo kratkog vizualnog pregleda, čovjek može prepoznati sve objekte na slici i odnose između njih. Za ovaj nevjerojatan uspjeh bili su potrebni milijuni godina u ljudskoj evoluciji. Ljudsko oko i mozak rade na savršen način, kako bi razvili veoma snažno vizualno iskustvo. Vizualni korteks unutar mozga i oči su glavni dijelovi sustava koji to ljudima omogućava. Unutar čovjeka postoji sustav koji mu omogućava razumijevanje slike, teksta i drugih zadataka vizualnog raspoznavanja. Čovjek ove zadatke obavlja od djetinjstva. Naučen je kako prepoznati mačku, drvo i druge predmete. Na vrlo sličan način moguće je naučiti i algoritme. Prikazivanjem dovoljne količine slika moguće je naučiti algoritam da prepozna slike koje dosad nije vidio.

Učenje algoritama kako bi razumjeli niz brojeva je veoma izazovan zadatak. Zbog specifične strukture ovi algoritmi se znatno razlikuju od klasičnog programa. Algoritmi bazirani na dubokom učenju sami istražuju podatke, te traže model koji daje najbolje rezultate. Drugim riječima, algoritam ne zahtjeva matematički model, ali je potrebna ogromna količina podataka što je u nekim područjima poseban problem, s obzirom na vrijeme potrebno na označavanje podataka. Razvojem konvolucijskih neuronskih mreža postignut je nevjerojatan uspjeh u prepoznavanju objekata na slikama, konteksta slike, emocija i slično.

4.1 Konvolucijske neuronske mreže

U području računalnog vida, konvolucijske neuronske mreže se koriste za klasifikaciju [14], i grupiranje slika [15], prepoznavanje objekata [16] i slične zadatke. Na Slika 7. je prikazana arhitektura konvolucijske mreže za klasifikaciju brojeva. Slika na ulazu prolazi kroz više filtera, gdje je svaki od njih zadužen za točno određenu vrstu značajki. U prvim slojevima se koriste filteri za horizontalne, vertikalne i okomite linije, a u ostalim slojevima filteri se specijaliziraju za točno određenu vrstu značajki.

Arhitektura mreže je analogna modelu povezanih neurona u ljudskom mozgu. Neuroni reagiraju na podražaje samo u ograničenom području vidnog polja, poznatom kao receptivno polje. Potpuno vizualno polje se sastoji od kolekcije receptivnih polja, koja se međusobno preklapaju. Informacije se prenose u jednom smjeru, od ulaznog do izlaznog sloja, zbog čega se u literaturi vrlo često koristi i naziv unaprijedna mreža.



Slika 7. CNN mreža za klasifikaciju brojeva [17]

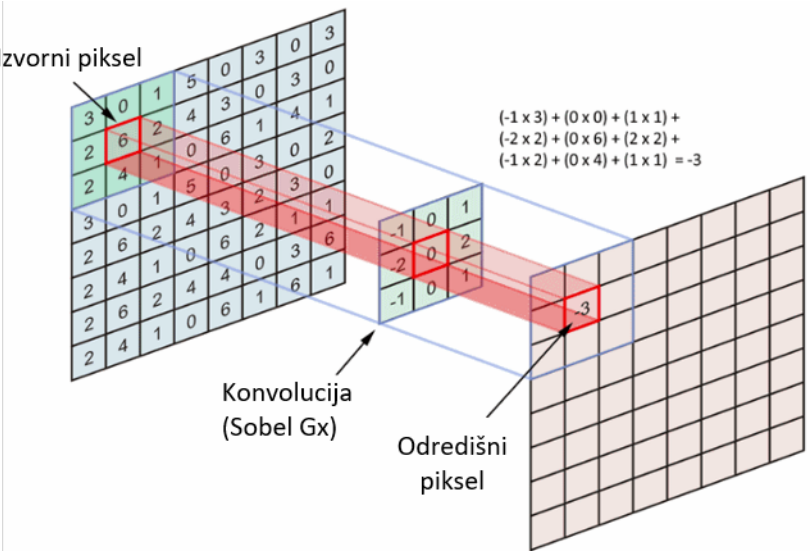
Dva osnovna elementa mreže su: konvolucijski i sloj sažimanja. Na kraju mreže se najčešće koristi jedan ili više potpuno povezanih slojeva, koji služe za klasifikaciju. Promjenom filtera moguće je vrlo uspješno detektirati prostorne ovisnosti na slikama. Smanjenjem broja parametara i ponovnom upotrebom težina, mreža se prilagođava određenom skupu podataka i na taj način bolje razumije detalje na slici.

4.1.1 Konvolucijski sloj

Riječ konvolucija dolazi od latinske riječi „convolvere“ što znači „*valjati se*“. U matematici, konvolucija je mjera preklapanja između dvije funkcije. Možemo je zamisliti kao način spajanja dvaju funkcija. U analizi slike, konvolucija se koristi za traženje značajki. Prva funkcija je ulazna mapa značajki, a druga funkcija je filter s kojim se traže značajke. Funkcije su obično povezane operacijom množenja.

Ulazna mapa značajki je dosta manjih dimenzija od filtera, a konvolucija se obavlja samo nad onim dijelom slike iznad kojeg se nalazi filter. Nakon toga filter se pomjera na sljedeći neuron. Postupak se ponavlja sve dok se ne obradi svaki neuron u ulaznoj mapi značajki. Na Slika 8. je prikazana ilustracija ove operacije.

Prvi konvolucijski sloj je zadužen za izdvajanje značajki niske razine kao što su: bridovi, boje, gradijenti, i sl., a dodatni slojevi se koriste za učenje značajki visoke razine. Ovakav pristup omogućuje bolje razumijevanje detalja u slici.

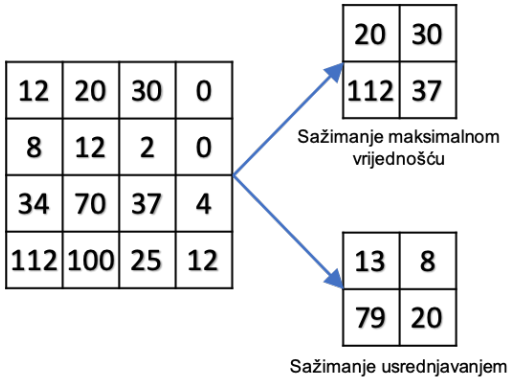


Slika 8. Konvolucija [18]

U većini radova se koristi potpuno povezani konvolucijski sloj, gdje je svaka mapa značajki iz trenutnog sloja povezana sa svim mapama značajki iz prethodnog sloja [19] [20]. Za razliku od klasičnog pristupa, u radu [21] se koristi raspršena povezanost, gdje sve mape značajki u trenutnom sloju, nisu povezane sa svim mapama iz prethodnog sloja.

4.1.2 Sloj sažimanja

Uloga sloja sažimanja je smanjenje dimenzija ulazne mape značajki. Na ovaj način se smanjuju računalni resursi za obradu podataka u mreži. Ovaj sloj se koristi za izdvajanje dominantnih značajki, koje su translacijski i rotacijski neovisne. Postoje dva tipa sažimanja: sažimanje usrednjavanjem i sažimanje maksimalnom vrijednosti. Sažimanje maksimalnom vrijednosti uzima najveću vrijednost iz dijela mape značajki nad kojom se nalazi filter, a u drugom slučaju uzima se aritmetička sredina. Na Slika 9. prikazana je ilustracija ovih operacija.

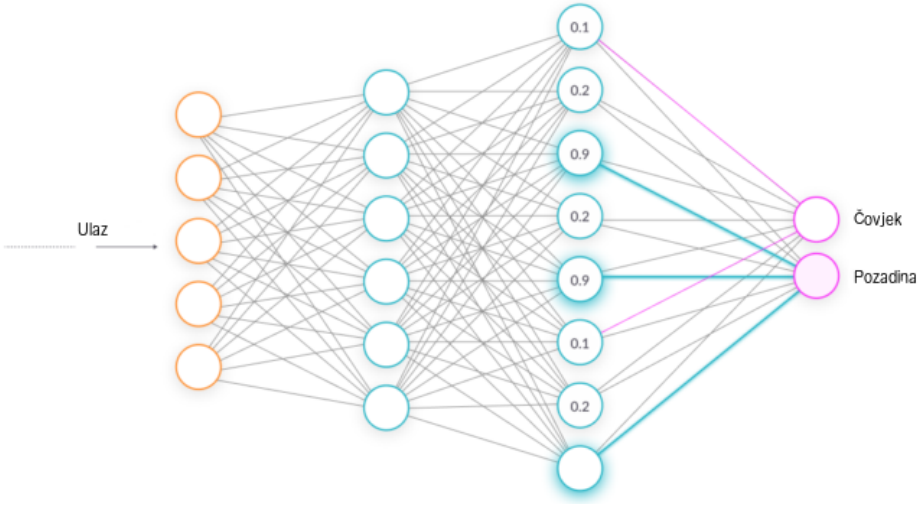


Slika 9. Vrste sažimanja

Sažimanje maksimalnom vrijednosti u potpunosti uklanja šum iz ulazne mape značajki, a sažimanje usrednjavanjem suzbija šum. Konvolucijski sloj i sloja sažimanja zajedno tvore jedan sloj u mreži, a mreža može imati više ovakvih slojeva. Na kraju se najčešće nalazi potpuno povezani sloj, koji obavlja zadatak klasifikacije.

4.1.3 Potpuno povezani sloj

Potpuno povezani sloj je najjednostavniji način za učenje nelinearnih kombinacija značajki. Ovaj sloj na osnovu ulazne mape značajki iz zadnjeg konvolucijskog sloja uči nelinearnu funkciju. Ulazna slika se pomoću konvolucijskih slojeva i slojeva sažimanja pretvara u višedimenzionalni percepton. Potpuno povezani sloj prikazan na Slika 10, u više iteracija uči razlikovati dominantne značajke od značajki niske razine, a zatim ih klasificira Softmax tehnikom.



Slika 10. Potpuno povezani sloj [22]

4.1.4 Parametri slojeva

Da bi mreža pravilno radila potrebno je voditi računa o parametrima za konfiguraciju prethodno opisanih slojeva. U ovom dijelu su opisana uobičajena pravila za konfiguraciju mreže, koja direktno utječu na veličinu.

Veličina ulaznog sloja bi trebala biti nekoliko puta djeljiva s dva, a najčešće dimenzije koje se koriste u većini arhitektura su: 32 (CIFAR-10), 64, 96 (STL-10), ili 224 (ImageNet), 384, i 512.

Najčešća veličina filtera u konvolucijskom sloju je 3 x 3 ili 5 x 5, a veličina koraka i obruba se bira na način da prostorne dimenzije ulazne mape značajki ostaju iste.

Uloga sloja sažimanja je smanjenje prostorne dimenzije ulazne slike, a najčešće se koristi sažimanje maksimalnom vrijednošću. Uobičajena veličina receptivnog polja je 2 x 2, a veličina koraka 2. Na ovaj način se odbacuje 75% aktivacija u ulaznoj mapi značajki. Osim ove konfiguracije koristi se i preklapajuće receptivno polje 3 x 3, s korakom 2. U praksi se vrlo rijetko koristi veće receptivno polje, jer se previše gubi na kvaliteti značajki.

Konfiguracija u kojoj konvolucijski sloj zadržava veličinu ulaza, a sloj sažimanja smanjuje prostornu dimenziju mape značajki se najčešće koristi, jer je ovaj pristup puno ugodniji za dizajn mreže. U suprotnom slučaju bi trebali pažljivo pratiti veličinu ulaza i osigurati simetričnu povezanost arhitekture.

4.1.5 Računalni zahtjevi

Prvih nekoliko konvolucijskih slojeva zahtijeva najviše memorijskih i računalnih zahtijeva, a potpuno povezani sloj ima najviše parametara. U fazi treniranja se moraju čuvati koeficijenti iz svih konvolucijskih slojeva, zbog propagacije greške unatrag, ali pametnom implementacijom faze testiranja moguće je uštediti dosta memorije, jer se koeficijenti iz prethodnih slojeva ne trebaju čuvati u memoriji.

Konvolucijske neuronske mreže imaju nekoliko svojstava, koja im omogućavaju dobru generalizaciju, a to su dijeljenje težina, raspršena povezanost i translacijska invarijantnost. U literaturi su predložene različite arhitekture, koje postižu veoma dobre rezultate. U nastavku su opisane najčešće korištene.

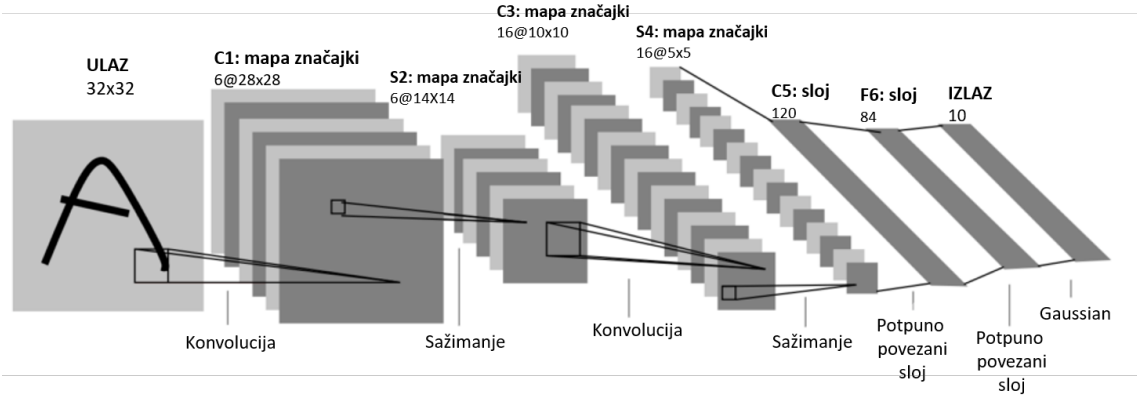
4.2 Standardne arhitekture

Iako se mreža sastoji od vrlo jednostavnih slojeva, postoji neograničen broj načina na koje je moguće rasporediti slojeve, te na taj način prilagoditi mrežu konkretnom problemu. Upravo to je i najzanimljiviji dio korištenja konvolucijskih neuronskih mreža.

4.2.1 LeNet-5

LeNet-5 je prva uspješna primjena konvolucijskih neuronskih mreža, za klasifikaciju rukom pisanih brojeva [21]. Mreža je testirana na MNIST bazi, a sastoji se od ukupno sedam slojeva. U prvom dijelu se dva i po puta ponavljaju konvolucijski sloj i sloj sažimanja, a na kraju se nalaze dva potpuno povezana sloja. Ulaz u mrežu je slika dimenzija 32 x 32. Na izlazu prvog sloja dobije se 6, a nakon drugog 16 mapa značajki. Zadnji sloj je sastavljen od 10 neurona, koji predstavljaju znamenke od 0 do 9.

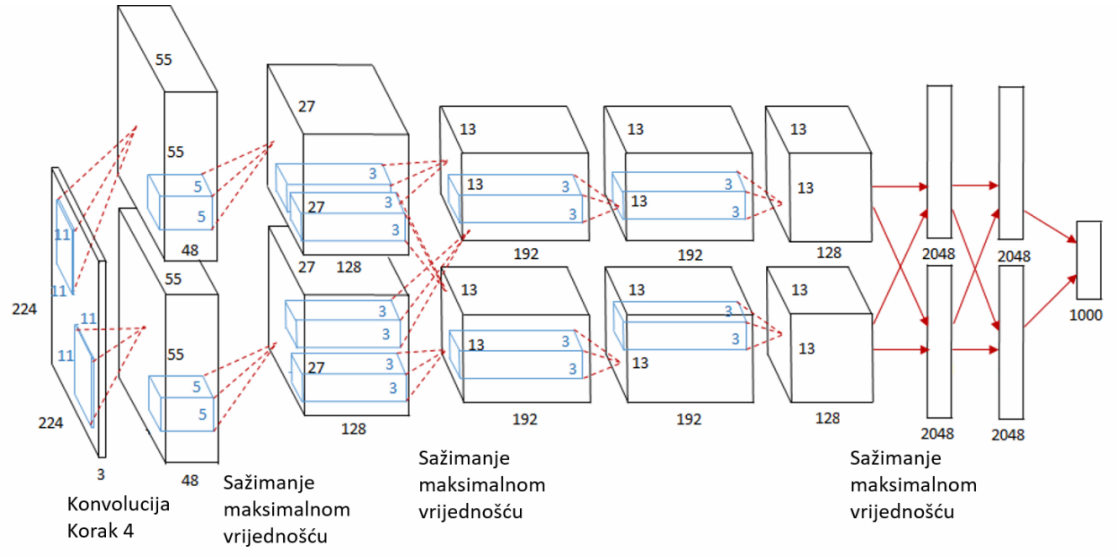
Ponavljajući parovi konvolucijskog i sloja sažimanja danas se koriste u svim modernim arhitekturama. Zanimljivo je da mreža u prvom sloju koristi veoma malo filtera, dok drugi sloj ima mnogo više filtera, ali su manjih dimenzija u odnosu na prethodni sloj. Broj filtera se povećava s dubinom mreže, a klasifikacija se obavlja s potpuno povezanim slojevima kao i kod svih modernih arhitektura.



Slika 11. Arhitektura LeNet-5 mreže za prepoznavanje znamenki [21]

4.2.2 AlexNet arhitektura

AlexNet [23] je pobjednik ImageNet ILSVRC natjecanja 2012. godine, za klasifikaciju objekata. Mreža je dizajnirana od metoda koje nisu bile široko korištene i prihvaćene, ali su danas postali standard u korištenju konvolucijskih neuronskih mreža.

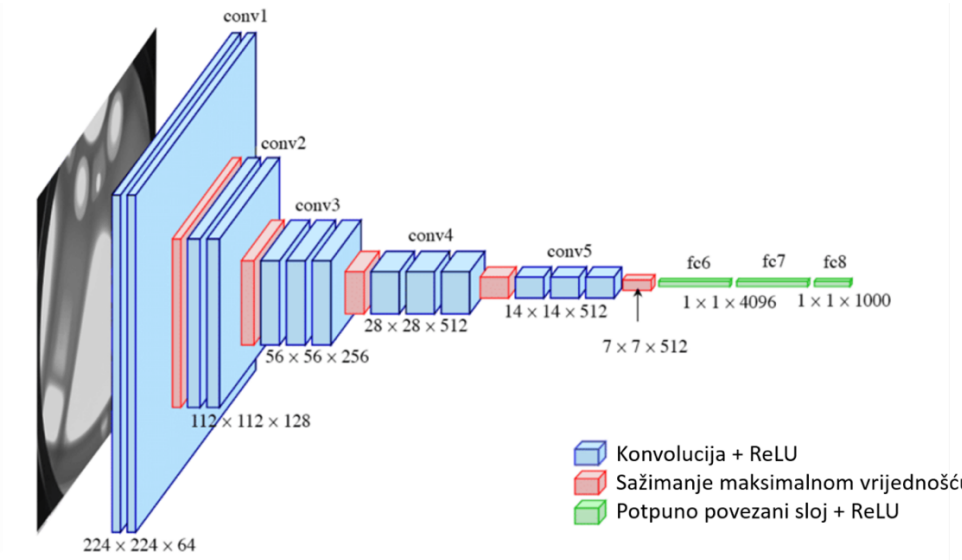


Slika 12. Arhitektura AlexNet mreže [23]

Nakon svakog konvolucijskog sloja koristi se ReLU sloj, umjesto S-funkcije koja je do tada bila uobičajena. Za klasifikaciju se koristi softmax funkcija, a umjesto sažimanja usrednjavanjem koje se koristilo u LeNET-5 arhitekturi, koristi se sažimanje maksimalnom vrijednošću. AlexNet se sastoji od 5 konvolucijskih slojeva, a podijeljen je na dva cjevovoda kako bi se treniranje moglo obaviti na GPU. Mreža koristi i dobro definirane komponente iz LeNet arhitekture, kao što je ponavljajući niz konvolucijskog sloja i sloja sažimanja.

4.2.3 VGG-16

Sljedeći značajan napredak je mreža poznata pod nazivom VGG-16 [24]. Model je prvi put predstavljen na ILSVRC natjecanju 2014 godine. Najvažnija razlika u odnosu ostale arhitekture je uporaba velikog broja malih filtera. U većini konvolucijskih slojeva se koristi sažimanje sa maksimalnom vrijednosti.



Slika 13. Arhitektura VGG-16 mreže [25]

Mreža koristi dva, tri ili čak četiri konvolucijska sloja u nizu prije sloja sažimanja. Na taj način se postiže sličan učinak kao i kod jednog konvolucijskog sloja većom veličinom filtera. Druga značajna razlika je veliki broj filtera, čiji se broj znatno povećava s dubinom mreže. Na Slika 13. je prikazana arhitektura VGG-16 mreže. Ulaz u mrežu je RGB slika dimenzija 224 x 224. Slika prolazi kroz niz konvolucijskih slojeva, koji koriste veoma malo receptivno polje veličine 3 x 3. U jednoj od konfiguracija koristi se i filter veličine 1 x 1. Ovaj filter možemo zamisliti kao linearnu transformaciju ulaznih kanala. U konvolucijskim slojevima filteri se pomjeraju s korakom veličine 1, a nakon svakog konvolucijskog sloja sačuvana je prostorna rezolucija. Mreža koristi pet slojeva

sažimanja s maksimalnom vrijednošću, a usrednjavanje se obavlja s prozorom veličine 2 x 2 i korakom 2. Na kraju mreže se nalaze tri potpuno povezana sloja. Prva dva sloja imaju 4096 kanala, a treći 1000. Za klasifikaciju se koristi softmax funkcija.

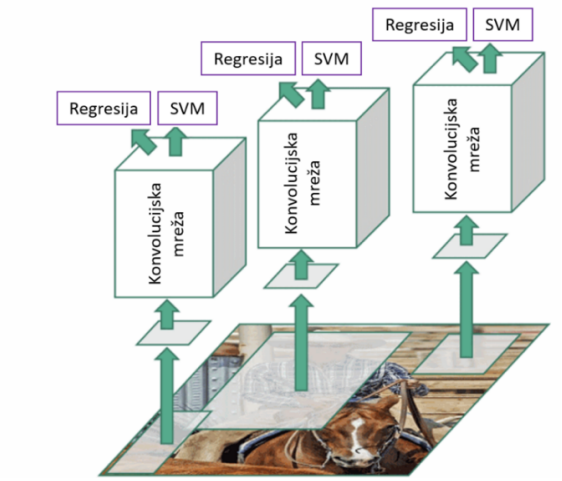
4.3 Algoritmi dubokog učenja za detekciju objekata

Glavni cilj bilo kojeg algoritma za detekciju objekata je ucrtati granični okvir oko objekta u slici. Jedna slika može sadržavati i više objekata, što u praksi znači da je potrebno ucrtati granični okvir oko svakog objekta u slici. Drugim riječima, veličina izlaza je varijabilna, te ovisi o broju pojavljivanja objekata od interesa, što ograničava primjenu standardnih konvolucijskih neuronskih mreža za detekciju objekata u slici.

Najjednostavniji pristup za rješavanje problema je uzeti različite regije od interesa, te ih zasebno klasificirati nekom od standardnih konvolucijskih neuronskih mreža. S obzirom da se objekti mogu nalaziti na različitim lokacijama i omjerima u slici, ovakav pristup je računalno prezahtjevan. Međutim, u literaturi su predloženi različiti pristupi koji su posebno prilagođeni da veoma brzo pronađu objekte na slici. U narednom dijelu opisan je jedan od prvih algoritama koji koristi duboke konvolucijske mreže za detekciju objekata.

4.3.1 R-CNN

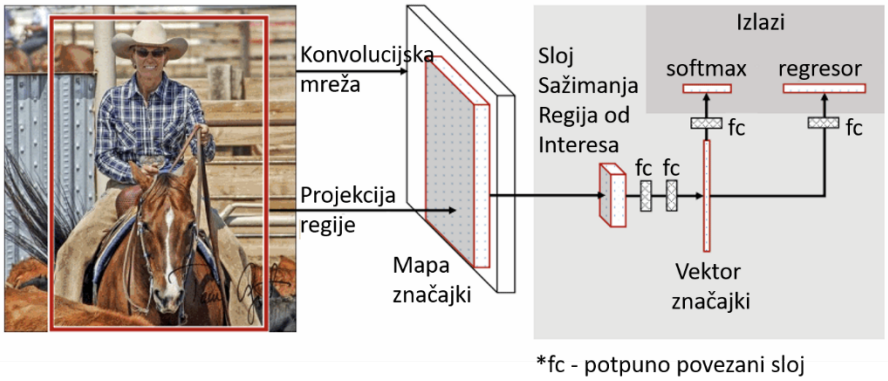
Kako bi riješio problem velikog broja regija, R-CNN [26] koristi algoritam selektivnog pretraživanja, koji izdvaja 2000 regija po slici. Regije se propuštaju kroz klasičnu konvolucijsku neuronsku mrežu, koja se u ovom slučaju koristi za izdvajanje značajki. Za klasifikaciju prisutnosti objekta u regiji koristi se SVM klasifikator. Na Slika 14. prikazan je princip rada R-CNN algoritma.



Slika 14. Princip rada RCN algoritma [27]

4.3.2 Fast R-CNN

Glavni nedostatak R-CNN-a je taj što svaku regiju od interesa mora propustiti kroz unaprijednu konvolucijsku mrežu. Fast R-CNN [28] problem rješava na način da cijelu sliku propušta kroz mrežu kako bi izračunao mapu značajki. Regije od interesa se uzorkuju iz mape značajki, a nakon toga se sažimaju na dimenzije prilagođene potpuno povezanom sloju. Operaciju sažimanja obavlja dodatni sloj u arhitekturi kojeg, autori nazivaju sloj sažimanja regija od interesa (*eng. ROI pooling layer*). Ovaj sloj je neophodan jer potpuno povezani sloj zahtijeva značajke točno određenih dimenzija. Na Slika 15. je prikazana arhitektura ovog algoritma. Slika se propušta samo jednom kroz unaprijednu konvolucijsku mrežu, umjesto 2000 puta kao što je to kod R-CCN-a, što je glavni razlog zašto je ovaj algoritam mnogo brži od svog prethodnika.

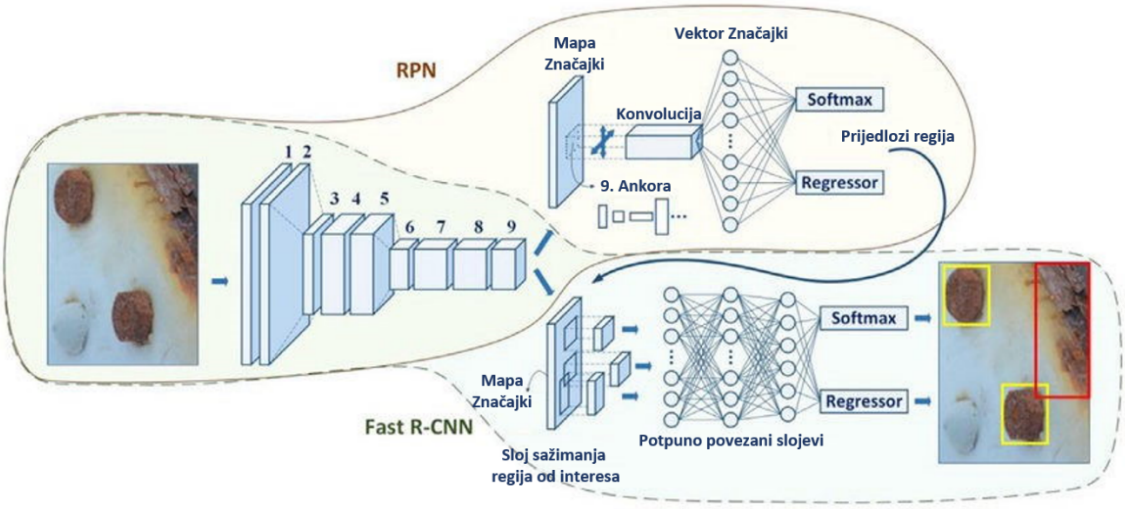


Slika 15. Arhitektura Fast R-CNN-a [28]

4.3.3 Faster R-CNN

Oba prethodno opisana algoritma koriste algoritam selektivne pretrage, za pronalazak regija od interesa. Analizom performansi se može veoma lako uočiti da je algoritam selektivnog pretraživanja spor u odnosu na performanse mreže, zbog čega Faster R-CNN izbacuje algoritam selektivne pretrage i uvodi RPN mrežu. RPN mreža, kao ulaz uzima mapu značajki bilo koje veličine, a na izlazu daje listu regija od interesa s priloženom vjerojatnosti da se unutar regije nalazi objekt. Regije od interesa se nakon toga sažimaju pomoću sloja sažimanje regija od interesa (*eng. ROI pooling layer*) na dimenzije prilagođene potpuno povezanom sloju.

Arhitektura Faster R-CNN-a je prikazana na Slika 16, a sastoji se od konvolucijske neuronske mreže za izdvajanje značajki, RPN algoritma za predlaganje regija od interesa i Fast R-CNN detektora za klasifikaciju.



Slika 16. Arhitektura Faster R-CNN algoritma

Veličina izlazne mape značajki direktno ovisi o arhitekturi mreže. Mreže, VGG-16 i ZF-Net imaju korak 16, što znači da dvije uzastopne točke u izlaznoj mapi značajki odgovaraju dvjema točkama međusobno udaljenim 16 piksela na ulaznoj slici. Za svaku točku u izlaznoj mapi značajki, mreža treba naučiti da li je na odgovarajućoj lokaciji u ulaznoj slici prisutan objekt, te procijeniti njegovu veličinu. U tu svrhu koristi se niz okvira (eng. Anchor box), različitih veličina i omjera.

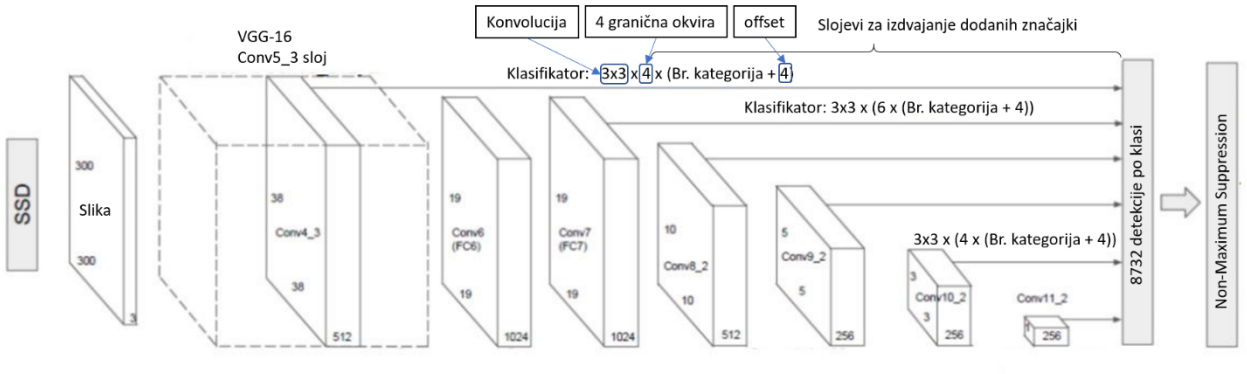
RPN algoritam prolazi kroz sve lokacije u izlaznoj mapi značajki, te provjerava da li se na ulaznoj slici unutar svakog okvira nalazi objekt. Za ulaznu sliku čije su dimenzije 800 x 640, izlazna mapa značajki je veličine 50 x 40. Budući da se za svaku lokaciju u izlaznoj mapi značajki provjerava 9 različitih okvira, ukupan broj okvira je 18000. Algoritam, odbacuje one okvire koji prelaze dimenzije slike, što značajno smanjuje njihov broj.

4.3.4 SSD Single Shot Detector

SSD algoritam detekciju obavlja u samo jednom koraku, za razliku od algoritama baziranih na prijedlozima regija. Ulazna slika prvo prolazi kroz CNN mrežu radi izdvajanja mape značajki, čija je veličina (M x N x P). Za svaku lokaciju se dobije K graničnih okvira različitih veličina i omjera. Svaki granični okvir je definiran s C kategorija i četiri vrijednosti, koje su vezane za dimenzije okvira. Ukupan broj izlaznih vrijednosti je (C + 4) x K x M x N.

Kako bi detekcija bila preciznija, određene mape značajki se propuštaju kroz 3 x 3 konvolucijski sloj. Negativni uzorci se biraju na način da se prvo uzlazno sortiraju prema grešci klasifikacije

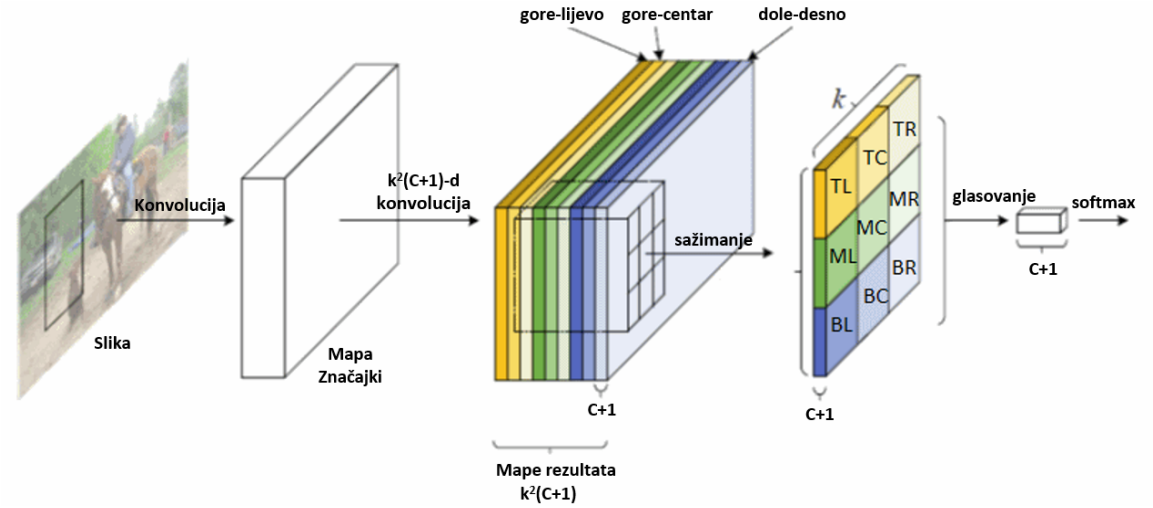
objekata za granični okvir, a uzima se samo onaj na vrhu. Omjer pozitivnih i negativnih uzoraka treba biti najviše 3:1.



Slika 17. Arhitektura SSD mreže

4.3.5 R-FCN

Prethodno opisani algoritmi za klasifikaciju koriste potpuno povezani sloj, s čime se znatno povećava broj parametara i usporava proces detekcije, zbog čega F-RCN [29] (eng. Region-based Fully Convolutional Network) izbacuje potpuno povezani sloj, a klasifikaciju obavlja na temelju mapa rezultata (eng. score maps) koje generira posljednji konvolucijski sloj. Za predlaganje regija se koristi RPN mreža, a svaka regija se propušta kroz sloj sažimanja regija od interesa. Algoritam koristi samo jednu mapu rezultata za sve regije od interesa.



Slika 18. Princip rada F-RCN-a

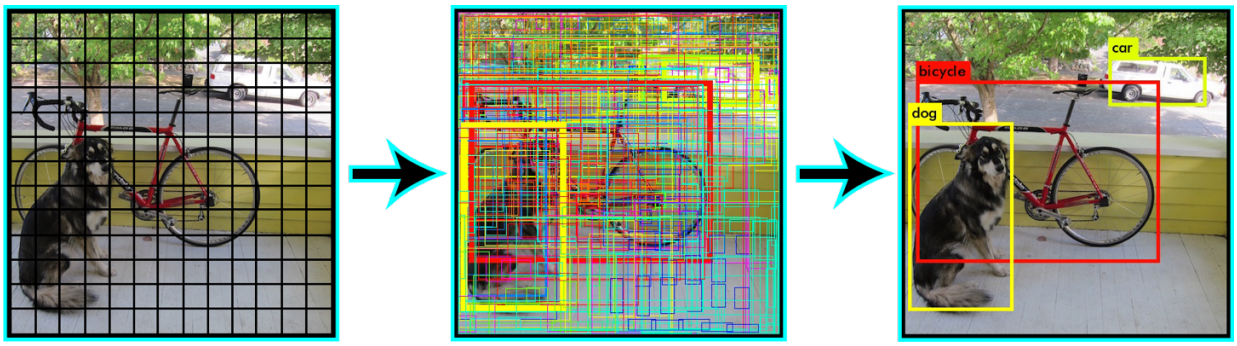
Klasifikacija se obavlja jednostavnim matematičkim operacijama i ne oduzima puno vremena, pa je algoritma čak i brži od Faster R-CNN-a. Kao što se vidi sa Slika 18, posljednji konvolucijski sloj generira mapu rezultata veličine $k^2(C+1)$. Parametar C označava broj kategorija koje je potrebno prepoznati, a dodatna kategorija se koristi za detekciju pozadine.

Za svaku kategoriju se koristi k^2 mapa značajki, gdje je svaka mapa zadužena samo za samo jedan dio objekta koji se želi prepoznati. Svaki objekt se sastoji od ukupno 9 dijelova (gore-lijevo, gore-centar, … , dole-desno). Na kraju se obavlja glasovanje prosječnom vrijednosti (eng. average voting), kako bi dobili vektor veličine C+1, nad kojim se obavlja proces klasifikacija korištenjem softmax funkcije.

4.3.6 YOLO

YOLO algoritam ujedinjuje komponente za detekciju objekata u jednu konvolucijsku mrežu, te na taj način detekciju svodi na regresijski problem. Za predviđanje graničnog okvira koriste se značajke iz cijele slike. Drugim riječima, mreža globalno razmatra cijelu sliku i sve objekte na slici.

Ulazna slika se dijeli na S x S dijelova i propušta kroz konvolucijsku neuronsku mrežu za računanje značajki. Nakon toga se obavlja proces linearne regresije, pomoću dva potpuno povezana sloja, koja za svaku lokaciju generiraju B graničnih okvira. Primjer predloženih graničnih okvira je prikazan na Slika 19 b).



a) *b)* *c)*

Slika 19. Princip rada YOLO algoritma

Svaki granični okvir je definiran s 5 vrijednosti: x, y, w, h i rezultatom povjerenja (*eng. confidence score*) definiranom u izrazu (4.1). Koordinate (x, y) predstavljaju centar okvira u odnosu na ćeliju kojoj pripada, a (w, h) su širina i visina okvira. Rezultat povjerenja, održava prisustvo ili odsustvo

objekta, a ovisi o vjerojatnosti da predloženi okvir sadrži objekt i IOU (*eng. Intersection over union)* mjeri između predviđenog i stvarnog graničnog okvira.

$$Rezultat\ povjerenja\ =\ Vjerojatnsot\ *\ IOU \qquad (4.1)$$

Rezultat povjerenja je jednak nuli, ako granični okvir ne sadrži objekt. U suprotnom se nastoji postići da rezultat povjerenja bude jednak IOU mjeri. Ako se centar objekta ne nalazi unutar ćelije, onda ta ćelija nije zadužena za predviđanje kategorije. Za predviđanje kategorije je odgovorna samo ona ćelija u kojoj se nalazi centar objekta. Na Slika 19. c) prikazan je primjer finalne predikcije, gdje se koriste samo oni okviri čiji je rezultat povjerenja veći od 0.25.

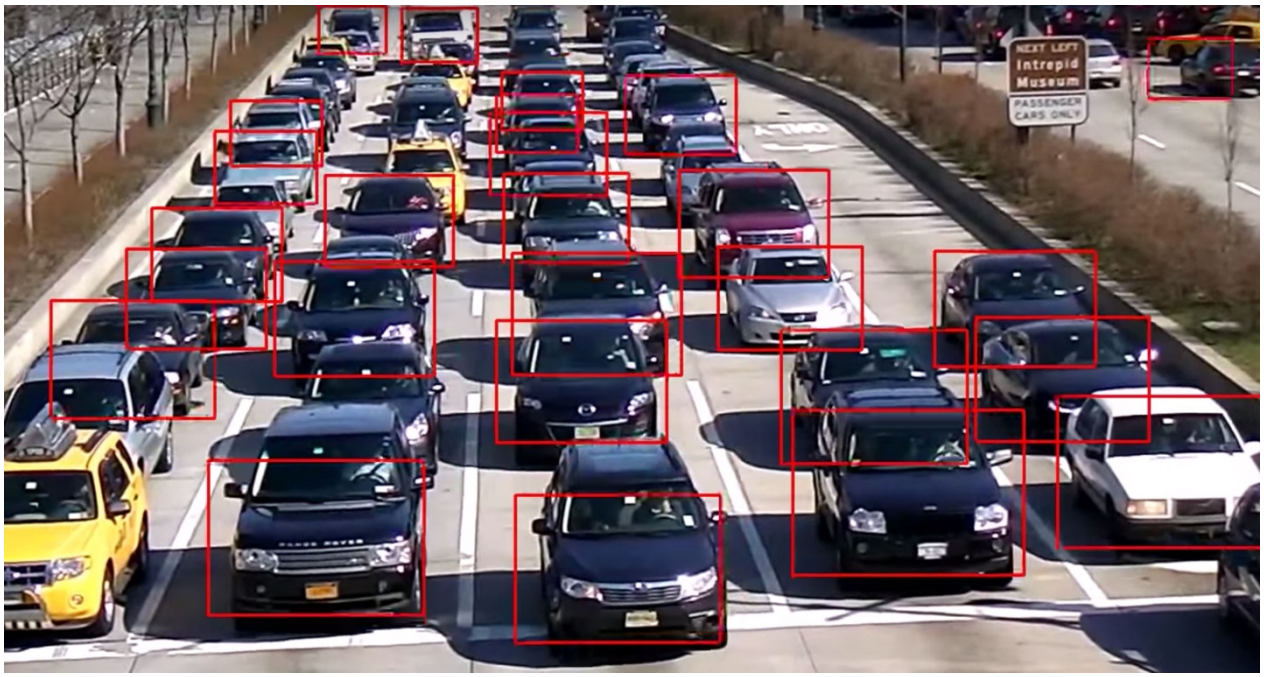
YOLO na izlazu daje ukupno S x S x B x 5 + C vrijednosti, koje su vezane za predikciju graničnog okvira i kategorije. Za svaku ćeliju se koristi samo jedan vektor C, bez obzira na broj graničnih okvira. Vektor C se množi sa svakim graničnim okvirom unutar ćelije, a rezultat označava koliko je dobro obuhvaćen objekt i kategorija kojoj pripada.

Budući da cijelu sliku razmatra odjednom algoritam postiže jako dobre rezultate i na nepoznatom skupu podataka. Ovo svojstvo pridonosi dosta manjem broju lažnih alarma u odnosu na druge algoritme.

4.3.7 YOLOv2

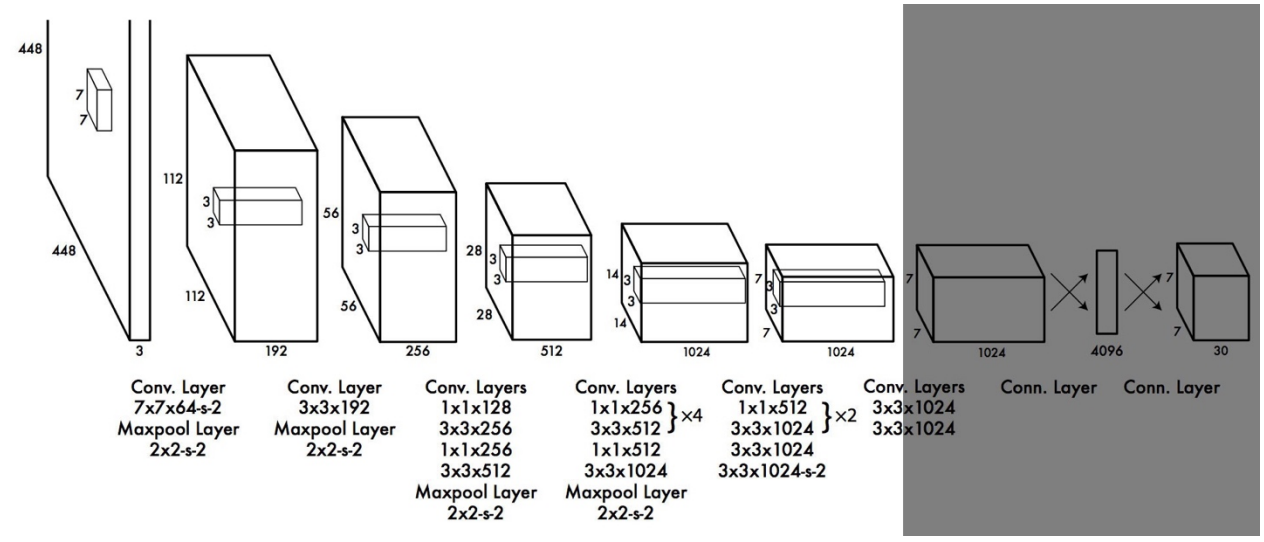
Proces treniranja YOLO algoritma obavlja se u dva dijela. U prvom se dijelu trenira mreža za klasifikaciju slika, u drugom dijelu se umjesto potpuno povezanog sloja stavlja konvolucijski sloj, te se mreža trenira ispočetka. Mreža za klasifikaciju slika se trenira na slikama dimenzija 224 x 224, a mreža za detekciju objekata koristi slike dimenzija 448 x 448. Za razliku od svog prethodnika, YOLOv2 mrežu za klasifikaciju slika trenira u dva koraka. Mreža se prvo trenira sa slikama dimenzija 224 x 224, a u drugom sa slikama dimenzija 448 x 448.

U ranim fazama treninga YOLO je podložan naglim promjenama gradijenta, jer predviđa granični okvir proizvoljne veličine i omjera. Drugim riječima, predikcije se međusobno bore oko veličine i omjera graničnog okvira, za koji će se specijalizirati. Ovo mu daje prednost kod detekcije objekata koji mijenjaju položaj i rotaciju. Međutim, kao što se može vidjeti na Slika 20., granični okvir za stvarne objekte nije uvijek proizvoljnih vrijednosti. Svi automobili imaju sličnu veličinu i omjer. Zbog toga YOLOv2, na početku treninga definira prosječni omjer i veličinu graničnog, a predviđa se odstupanje od predefinirani vrijednosti.



Slika 20. Granični okvir objekt automobil [30]

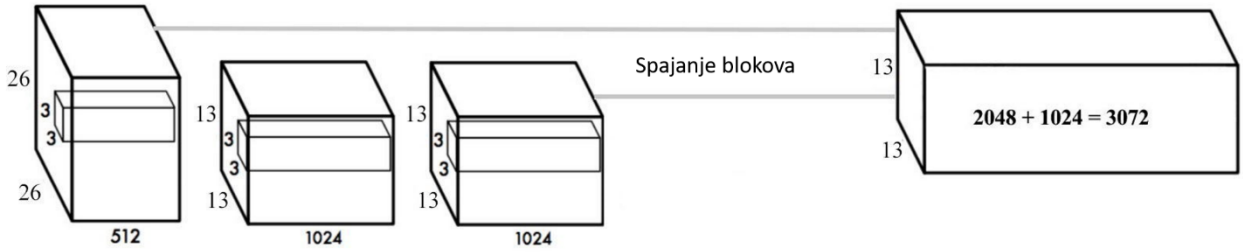
Na Slika 21. prikazana je arhitektura YOLOv2 algoritma. Potpuno povezani sloj zadužen za predikciju graničnog okvira je izbačen, a kategorija se predviđa na razini graničnog okvira, umjesto na razini ćelije. Granični okvir je definiran s 25 parametara: 4 za dimenzije graničnog okvira, 1 za rezultat povjerenja (eng. confidence score) i 20 za predviđanje kategorije. Za svaku ćeliju se predviđa ukupno 5 graničnih okvira.



Slika 21. Arhitektura YOLOv2 mreže [31]

Umjesto zadnjeg konvolucijskog sloja ubačen je (3 x 3) sloj, koji na izlazu daje mapu značajki dubine 1024. Nakon toga se pomoću (1 x 1) konvolucije, mapa značajki smanjuje na 7 x 7 x 125. Ulazna slika je dimenzija 416 x 416, umjesto 448 x 448, na taj način se na zadnjem sloju dobije neparna mapa značajki, koja dolazi do izražaja u detekciju velikih objekata, jer se može preciznije odrediti ćelija koja sadrži najkvalitetnije značajke o objektu.

Konvolucijski slojevi postupno smanjuju prostornu dimenziju ulazne slike, što otežava detekciju malih objekata. Da bi riješio ovaj problem SSD algoritam, detekciju obavlja na više mapa značajki. Na taj način se svaki sloj specijalizira za detekciju objekata različite veličine. Sličan pristup koristi se i u YOLOv2 algoritmu. Zadnji sloj se dijeli na četiri dijela, a dobiveni dijelovi se nadovezuju u jedan blok. Konačni blok na kojem se obavlja predikcija dobije se spajanjem preoblikovanog bloka i originalnog bloka. Ovaj postupak je ilustriran na Slika 22.



Slika 22. Povezivanje blokova [31]

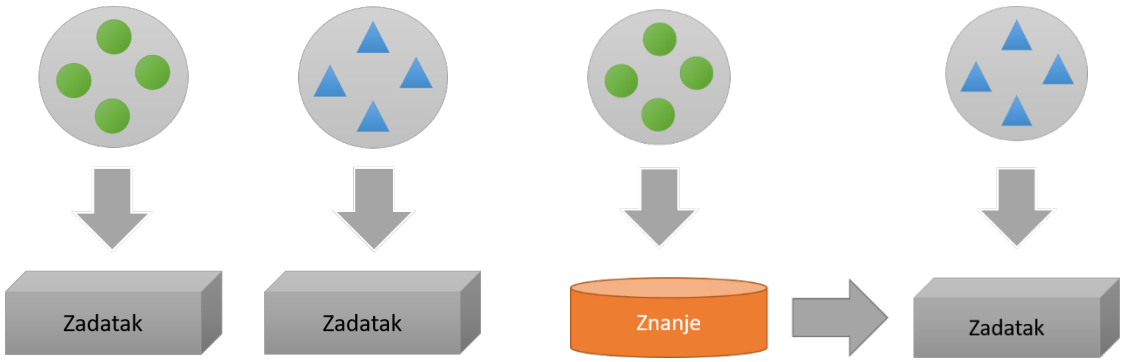
Algoritme za detekciju objekata opisane u ovom poglavlju možemo podijeliti u dvije skupine. Jednu skupinu čine algoritmi koji detekciju obavljaju u dva koraka. Primjer takvih algoritama su R-CNN, Fast R-CNN, Faster R-CNN, R-FCN. Sličnu arhitekturu koriste najpoznatije duboke neuronske arhitekture kao što su AlexNet [23] i VGG-16 [24] namijenjene za rješavanje problema klasifikacije slika, pa možemo zaključiti da je osnovna ideja ovih algoritama svesti problem detekcije objekata na problem klasifikacije slika.

Drugu skupini čine YOLO, SSD, YOLOv2, koji detekciju obavljaju u jednom koraku. U usporedbi s detektorima koji se baziraju na prijedlozima regija, ovi algoritmi daju bolji odziv (eng. Recall).

5 Prijenosno učenje

Ljudi imaju svojstvenu sposobnost prenošenja znanja kroz zadatke. Znanje stečeno rješavanjem jednog zadatka, lako se može primijeniti na rješavanje sličnih zadataka. Proces prijenosa znanja je efikasniji, što su zadaci sličniji. Tradicionalni algoritmi dubokog učenja osposobljeni su za rješavanje vrlo specifičnih zadataka u izoliranom okruženju. Svaki put kada se distribucija značajki promijeni, model se treba trenirati ispočetka. Ideja prijenosnog učenja nadilazi tradicionalnu paradigmu dubokog učenja, te omogućava ponovnu upotrebu naučenog znanja. Prvi put se pojavljuje 1995. godine, kada je predstavljen rad na temu „*Prijenos znanja u induktivnim sustavima*“. Glavna motivacija proizlazi iz činjenice da je za izradu složenijih modela potrebna velika količina podataka, što u nekim područjima nije nimalo jednostavan zadatak, jer je potrebno uložiti dosta truda i vremena na označavanje podataka.

Modeli dubokog učenja su usko specijalizirani na točno određeni zadatak. Čak i najsuvremeniji modeli mogu pokazati vrlo loše rezultate pri rješavanju novog zadatka, koji ima dosta sličnosti sa zadatkom za koji je model obučen. Prijenosno učenje nadilazi specifičnost zadatka i pokušava pronaći način za ponovno iskorištavanje znanja.



Slika 23. Usporedba tradicionalnog pristupa učenju i prijenosnog učenja

Prijenosno učenje nije novi koncept dubokog učenja, ali je značajna razlika između tradicionalnog pristupa u učenju dubokih modela i korištenjem metodologije koja slijedi principe prijenosnog učenja. Na Slika 23. je prikazana usporedba tradicionalnog pristupa učenju i prijenosnog učenja, koje je posebno zanimljivo u situacijama kada nemamo veliki skup podataka.

5.1 Matematički model prijenosnog učenja

Okvir za prijenosno učenje Pan i Yang [32] definiraju prema izrazu (5.1), koristeći domenu, zadatak i marginalnu vjerojatnost. Domena D se sastoji do prostora značajki X, i marginalne vjerojatnosti P(X), gdje je X niz ulaznih podataka.

$$D = \{ X, P(X) \} \tag{5.2}$$

Zadatak T možemo definirati prema izrazu (5.1), a sastoji se do prostora oznaka Y, i funkcije cilja η.

$$T = \{Y, \mathrm{P}(\mathrm{Y} \mid \mathrm{X})\} = \{Y, \eta\} \tag{5.2}$$

Funkcija cilja uči iz ulaznih podataka, gdje je Y vektor značajki, a X vektor oznaka.

5.2 Strategije prijenosnog učenja

U procesu prijenosnog učenja potrebno je pronaći odgovore na neka od ključnih pitanja. Prvi i najvažniji korak je odrediti koji dio znanja se prenosi na novi zadatak. Da bi riješili ovaj problem potrebno je pronaći dijelove znanja koji su specifični za izvorni zadatak, te zajedničke dijelove između novog i izvornog zadatka.

Prijenosno učenje ne garantira poboljšanje rezultata, već postoji i scenarij u kojem se rezultati mogu pogoršati, zbog čega treba biti vrlo oprezan s primjenom ove tehnike. Ovaj problem je u literaturi poznat pod nazivom negativno učenje (eng. negative transfer).

Nakon što je poznato koji dio znanja se prenosi, potrebno je odrediti način prijenosa znanja. Ovaj proces uključuje promjenu postojećih algoritama i primjenu različitih strategija i tehnika prijenosnog učenja. Izbor strategije prijenosnog učenja ovisi o domeni, zadatku i dostupnosti podataka. Prema radu [32] prijenosno učenje možemo podijeliti na: induktivno, nenadzirano i transduktivno prijenosno učenje.

Kod induktivne strategije prijenosnog učenja oba zadatka pripadaju istoj domeni, ali se međusobno razlikuju. Za poboljšanje preciznosti rezultata novog zadatka algoritam iskorištava znanje iz izvorne domene. Ova strategija se može dodatno podijeliti, ovisno o tome da li izvorna domena sadrži označene podatke.

Kod nenadzirane strategije prijenosnog učenja domena izvornog zadatka je slična domeni novog zadatka, ali se zadaci međusobno razlikuju. Ova strategija dijeli neke zajedničke karakteristike kao

i induktivna strategija prijenosnog učenja. Glavna razlika je u dostupnosti označenih podataka, jer u nenadziranom učenju podaci nisu dostupni niti u jednoj domeni.

Kod transduktivne strategije prijenosnog učenja izvorni zadatak je sličan novom zadatku, ali se domene međusobno razlikuju. Znanje se prenosi iz domene koja ima puno označenih podataka, u domenu koja ima malo ili nimalo označenih podataka.

U Tablica 3 je prikazana usporedba prethodno opisanih strategija prijenosnog učenja, u odnosu na komponente koje se prenose na novi zadatak.

Tablica 3. Vrste komponenti koje se prenose u prijenosnom učenju

	Induktivno prijenosno učenje	Transduktivno prijenosno učenje	Ne nadzirano prijenosno učenja
Prijenos instanci	X	X	
Prijenos reprezentacije značajki	X	X	X
Parametarski prijenos	X		
Relacijski prijenos znanja	X		

Direktna primjena naučenog znanja je idealan primjer prijenosnog učenja, ali se u većini slučajeva znanje ne može direktno primijeniti na drugi zadatak. Međutim, postoje određeni **dijelovi instanci,** koje se mogu iskoristiti za učenje novog zadatka. Odabrane instance se koriste zajedno s novim podacima, kako bi poboljšali preciznost. Većina algoritama koji spadaju u induktivnu strategiju prijenosnog učenja koriste modificiranu verziju AdBoost algoritma.

Prijenos reprezentacije značajki ima za cilj pronaći dobre značajke iz izvornog zadatka, kako bi minimizirao odstupanje od domene i smanjili pogrešku klasifikacije. Ovaj pristup se može primijeniti na nadzirane i nenadzirane strategije prijenosa učenja ovisno o tome da li su dostupni označeni podaci.

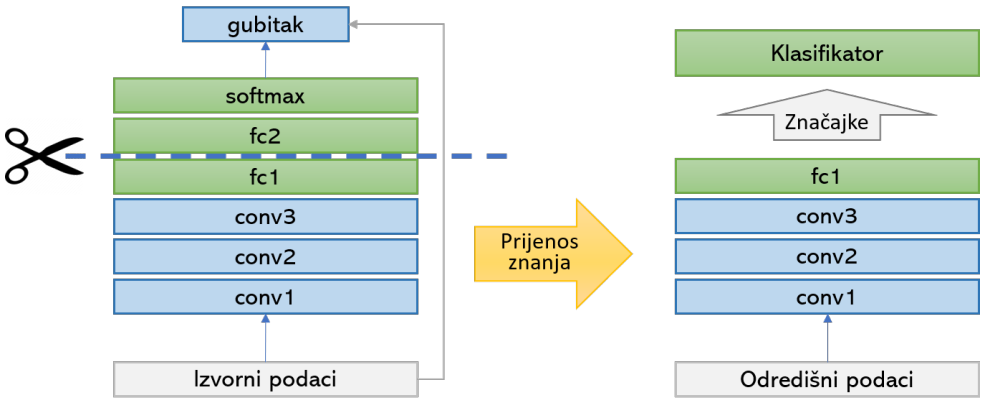
Parametarski pristup radi na pretpostavci da se modeli za zadatke koji su povezani trebaju podijeliti na zajedničke parametre. Većina algoritama prijenosnog učenja koji koriste ovaj pristup su napravljeni za rad s više zadataka, ali se mogu lako primijeniti za prijenosno učenje.

Relacijski prijenos znanja obrađuje podatke koju su jednako raspoređeni, ali nisu neovisni. Drugim riječima, skupove podataka kod kojih svaki podatak ima vezu s drugim podacima u skupu. Ovaj prijenos znanja može se iskoristiti nad podacima iz društvenih mreža.

5.3 Prijenos znanja u dubokom učenju

Modeli dubokog učenja su reprezentativni za ono što je poznato kao induktivno učenje. Glavni cilj induktivnog učenja je naučiti pravila za donošenje odluka iz seta trening podataka. U primjeru klasifikacije, model uči pravila preslikavanja između ulaznih značajki i označenih podataka. Algoritam za učenje koristi niz pretpostavki vezanih uz raspodjelu trening podataka, kako bi model davao dobre rezultate i na novom skupu podataka. Neprekidni razvoj u ovom području je omogućio rješavanje vrlo kompleksnih problema sa zavidnim rezultatima.

Duboke neuronske mreže koriste slojevitu arhitekturu koja uči značajke na različitim slojevima. Na kraju se obično nalazi potpuno povezani sloj koji služi za konačnu detekciju. Slojevita arhitektura omogućava ponovno iskorištavanje unaprijed obučene mreže za rješavanje drugih zadataka. Na Slika 24. je prikazan primjer prijenosnog učenja kod dubokih neuronskih mreža. Prethodno naučeni model se koristi isključivo za izdvajanje značajki, a parametri modela se ne ažuriraju za vrijeme treniranja novog zadataka. Na ovaj način dovoljno je uzeti naučeni model, kao što je AlexNet bez zadnjeg klasifikacijskog sloja te ga iskoristiti za treniranje novog zadatka.

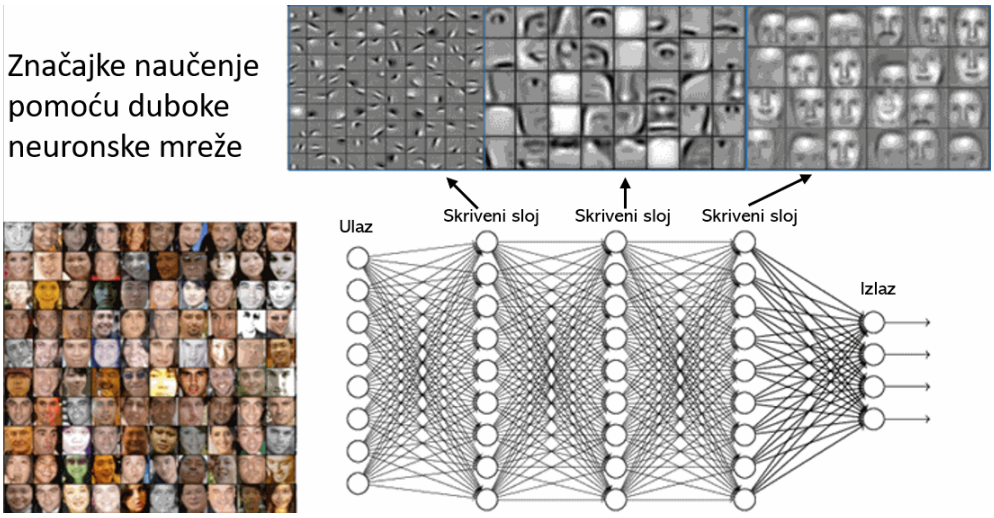


Slika 24. Prijenosno učenje kao ekstraktor značajki

Na osnovu skrivenih slojeva ulazna slika se transformira u vektor značajki veličine 4096. Mreža na ovaj način omogućuje izdvajanje značajki u novoj domeni. Ovo je jedan od najčešćih načina korištenja prijenosnog učenja u dubokim neuronskim mrežama, međutim u praksi je potrebno i selektivno istrenirati i neke od prethodnih slojeva.

Duboke neuronske mreže su konfigurabilne arhitekture s velikim brojem parametara. Početni slojevi vide samo generičke značajke, a dublji slojevi su više fokusirani na specifičan zadatak. Na

Slika 25. je prikazan problem prepoznavanja lica, gdje se jasno vidi da početni slojevi u mreži uče generičke značajke, a viši slojevi u mreži su zaduženi za vrlo specifične zadatke.

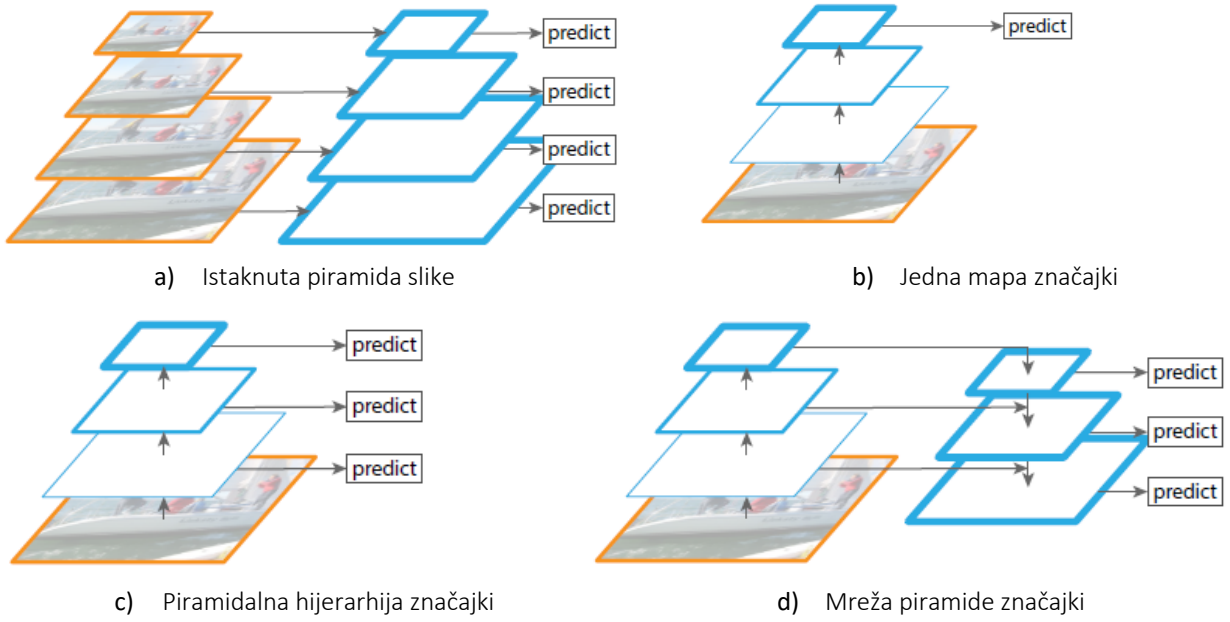


Slika 25. Duboka neuronska mreža za prepoznavanje lica

Na ovaj način moguće je zamrznuti određene slojeve tijekom treniranja, a ostatak mreže prilagoditi potrebama novog zadatka. U ovom slučaju koristimo arhitekturu cijele mreže kako bi fino podesili mrežu i postigli bolje rezultate.

6 Piramida značajki za detekciju objekata

Piramide značajki čine osnovnu komponentu bilo kojeg tradicionalnog sustava za detekciju objekata, ali većina prethodno opisanih algoritama baziranih na dubokom učenju, izbjegava piramidalnu reprezentaciju značajki, najvećim dijelom zbog toga što su piramide značajki računalno i memorijski vrlo zahtjevne. Međutim, uvođenjem jednostavnog okvira za izdvajanje piramide značajki, postignuta su značajna poboljšanja u detekciji. Proces koji je predložen u radu [33], ne troši previše računalnih, niti memorijskih resursa. Na Slika 26. su prikazane različite arhitekture za detekciju objekata koje koriste piramidalnu reprezentaciju značajki.



Slika 26. Arhitekture piramide značajki [34]

Istaknuta piramida slike prikazana na Slika 26 a) je standardno rješenje koje se veoma često koristilo u doba kada su značajke bile ručno dizajnirane. Ove značajke su invarijantne na promjenu veličine objekta, u smislu da se problem može nadoknaditi promjenom razine u piramidi. Pretraživanjem svih pozicija i razina unutar piramide, moguće je detektirati objekte bilo koje veličine.

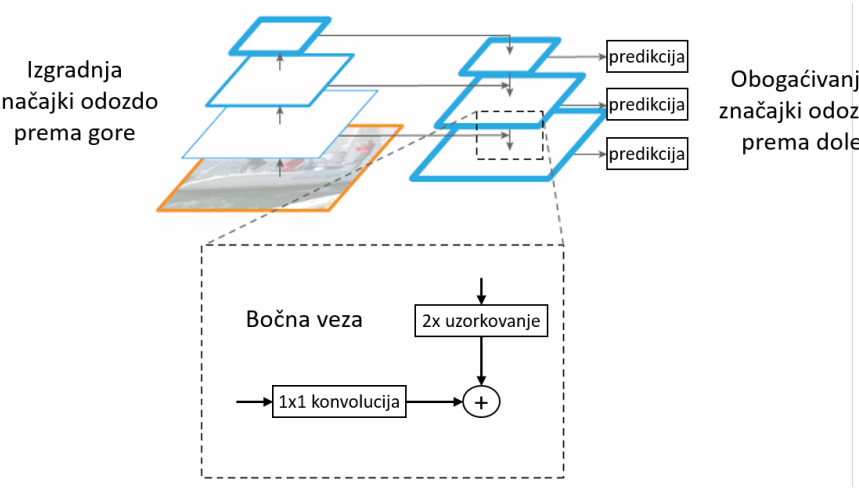
U novije vrijeme za detekciju objekata se najčešće koriste duboke konvolucijske mreže. Sustavi bazirani na dubokom učenju, robusniji su na promjenu veličine objekta, jer koriste značajke visoke razine. Značajke u nižim slojevima imaju lošiju sposobnost prepoznavanja, pa se detekcija obavlja samo na najvišem sloju u piramidi, kao što je prikazano na Slika 26 b).

SSD [35] algoritam je prvi primjer korištenja predikcije na više razina, unutar dubokih konvolucijskih mreža. Primjer ove predikcije je ilustriran na Slika 26 c). Kako bi izbjegao značajke niske razine, SSD piramidu počinje graditi tek od viših slojeva u mreži. Za VGG-16 mrežu, piramida se počinje graditi tek od četvrtog sloja. Na taj način se propušta mogućnosti ponovnog iskorištavanja mapa visoke razlučivosti, što je veoma važno za prepoznavanje malih objekata [33].

Zbog toga, autori u radu [33] predlažu proces za spajanje značajki visoke i niske razine, kako bi na svim slojevima u hijerarhiji dobili semantički jače značajke. Predloženi proces ne zahtijeva previše računalnih niti memorijskih resursa. Ovaj primjer je ilustriran na Slika 26 d). Slična arhitektura se koristi i u radovima [36] [37] [38], ali se predikcija obavlja na samo jednoj mapi značajki.

Proces izgradnje piramide značajki je ilustriran na Slika 27. Prvo se unaprijednim prolazom kroz mrežu računaju značajke na svim razinama, a nakon toga spajaju i postaju semantički jače. Spajanje se obavlja na način da se prostorna rezolucija sloja iznad uzrokuje s faktorom 2, a nakon toga se pomoću lateralne veze spajaju mape značajki iste veličine. U zadnjem koraku se smanjuju dimenzije mape značajki i ublažava efekt uzrokovanja. Dimenzije se smanjuju primjenom 1 x 1 konvolucijskog sloja, a efekt uzrokovanja primjenom 3 x 3 konvolucijskog sloja. Budući da konvolucijske neuronske mreže mogu proizvesti više mapa značajki iste veličine, kao referentni set se uzima samo ona mapa značajki koju proizvodi zadnji sloj.

Mreža piramide značajki se može veoma efikasno primijeniti za obogaćivanje mape značajki u RPN (eng. region proposal network) koja se koristi u Faster R-CNN arhitekturi. Na ovaj način je moguće svaku razinu piramide prilagoditi za točno određenu veličinu objekta.



Slika 27. Proces izgradnje piramide značajki [33]

7 Zaključak

Kroz ovaj rad opisani su najznačajniji pristupni pristupi za detekciju objekata u slici, te mogućnosti njihove primjene za detekciju ljudi u slikama snimljenih iz zraka. Za snimanje terena danas se najčešće koriste bespilotne letjelice, stoga rad započinje primjenom bespilotnih letjelica za snimanje terena u misijama potrage i spašavanja. Bespilotne letjelice pružaju novu perspektivu pretraživanja terena, ali postoje određena ograničenja koja smanjuju mogućnost njihove primjene u misijama potrage i spašavanja. Vizualna inspekcija slika je prezahtjevan zadatak za čovjeka, stoga je potrebno pronaći rješenje koje će u najmanju ruku čovjeku olakšati posao, ili u najboljem slučaju riješiti problem bez intervencije čovjeka.

Iako postoji relativno mali broj radova koji se bavi ovim problemom, pronađeno dosta zanimljivih rješenja. Najistaknutiji dijelovi ljudskog tijela, na snimkama snimljenim iz zraka su glava i ramena, zbog čega se većina predloženih rješenja bazira na detekciji ovih detalja u slikama. Ljudi na zračnim slikama zauzimaju relativnom malo piksela u odnosu na cijelu sliku, što dodatno otežava problem detekcije. Većina pristupa koji se koriste za detekciju pješaka koristi informacije o teksturi i obliku objekta, koji na slikama snimljenim iz zraka nisu dovoljno reprezentativni što ograničava primjenu ovih algoritama u misijama potrage i spašavanja. Najveći dio radova bazira na metodama za segmentaciju slike i izdvajanje značajki.

Kroz ovaj rad su također opisani i najznačajniji algoritmi koji za detekciju objekata koriste duboko učenje. Većina predloženih radova se bazira na različitim arhitekturama dubokih neuronskih mreža. Iako se mreža sastoji od dosta jednostavnih slojeva, postoji neograničen broj kombinacija za njihovo slaganje. Ulazna slika bi trebala biti djeljiva s dva, a parametri konvolucijskog sloja se podešavaju na način da se prostorne dimenzije ulazne mape značajki ne mijenjaju. Zanimljivo je da sve arhitekture koriste ponavljajući uzorak konvolucijskog sloja i sloja sažimanja. Ovaj uzorak se koristi i u prvoj uspješnoj promjeni konvolucijske neuronske mreže za klasifikaciju rukom pisanih brojeva, koju je predložio Yann LeCun.

Prvi pokušaj primjene konvolucijskih neuronskih mreža za detekciju objekata je R-CNN algoritam. Najveći nedostatak ovog algoritma u smislu vremena za detekciju je algoritam selektivnog pretraživanja, kojemu treba oko 40 sekundi za obradu jedne slike. Nakon pojave ovog algoritma u literaturi je predloženo dosta prijedloga za njegovo poboljšanje u smislu, smanjenja memorijskih i računalnih zahtijeva, vremena treniranja i preciznosti detekcije. Prvo značajno poboljšanje je

smanjenje vremena za obradu slike u fazi treniranja, gdje se cijela slika samo jednom propušta kroz mrežu radi izdvajanja značajki, umjesto za svaku regiju posebno. Nakon toga Faster R-CNN, izbacuje algoritam selektivnog pretraživanja te za predlaganje regija koristi RPN mrežu. Ostala skupina algoritama kao što su SSD, YOLO, F-RCN detekciju objekata obavljaju u jednom koraku.

Duboke neuronske mreže su usko specijalizirane na točno određeni zadatak. Svaki put kada se distribucija značajki promijeni, mrežu je potrebno ispočetka trenirati. Za treniranje dubokih modela potrebno je dosta vremena i podataka, koje je za neke zadatke vrlo teško dobiti s obzirom na vrijeme potrebno za označavanje podataka. Zbog toga se u literaturi pojavljuje ideja prijenosnog učenja. Cilj prijenosnog učenja je pronaći najbolji način i metode za prijenos znanja između modela.

Detekcija objekata se obično obavlja na zadnjoj mapi, koja sadrži najkvalitetnije značajke. Značajke na niži razinama nisu pogodne za detekciju objekata, te daju vrlo loše rezultate. Prvi pokušaj detekcije objekata, na više razina je SSD algoritam. Nakon toga je napravljeno nekoliko poboljšanja koje iskorištavaju već izračunate mape značajki, te grade kvalitetnije značajke na svim slojevima u hijerarhiji značajki unutar konvolucijske neuronske mreže.

8 Literatura

[1] B. Shah and H. Choset, "Survey on Urban Search and Rescue Robots," *J. Robot. Soc. Japan*, 2004.

[2] Ž. Marušić, D. Zelenika, T. Marušić, and S. Gotovac, "Visual Search on Aerial Imagery as Support for Finding Lost Persons," in *2019 8th Mediterranean Conference on Embedded Computing, MECO 2019 - Proceedings*, 2019.

[3] R. J. Koester, Lost Person Behavior: A search and rescue guide on where to look - for land, air and water. dbs productions llc, 2008.

[4] S. P. Yeong, M. King, and S. S. Dol, "A Review on Marine Search and Rescue Operations Using Unmanned Aerial Vehicles," *World Acad. Sci. Eng. Technol. Int. J. Mar. Environ. Sci.*, 2015.

[5] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: Survey and experiments," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

[6] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Proceedings of the IEEE International Conference on Computer Vision*, 2003.

[7] A. Gaszczak, T. P. Breckon, and J. Han, "Real-time people and vehicle detection from UAV imagery," in *Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques*, 2011.

[8] O. Oreifej, R. Mehran, and M. Shah, "Human identity recognition in aerial images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010.

[9] P. Rudol and P. Doherty, "Human body detection and geolocalization for UAV search and rescue missions using color and thermal imagery," in *IEEE Aerospace Conference Proceedings*, 2008.

[10] H. Turić, H. Dujmić, and V. Papić, "Two-stage Segmentation of Aerial Images for Search and Rescue," *Inf. Technol. Control*, vol. 39, 2010.

[11] D. Božić-Štulić, Ž. Marušić, and S. Gotovac, "Deep Learning Approach in Aerial Imagery for Supporting Land Search and Rescue Missions," *Int. J. Comput. Vis.*, 2019.

[12] J. Senthilnath, A. Dokania, M. Kandukuri, R. K.N., G. Anand, and S. N. Omkar, "Detection

of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV," *Biosyst. Eng.*, 2016.

[13] X. Dong, J. Dong, and L. Qu, "Enteromorpha detection in aerial images using support vector machines," in *Proceedings - 2009 IEEE Youth Conference on Information, Computing and Telecommunication, YC-ICT2009*, 2009.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, 2017.

[15] C. Hsu and C. Lin, "Unsupervised convolutional neural networks for large-scale image clustering," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 390–394.

[16] S. Hayat, S. Kun, Z. Tengtao, Y. Yu, T. Tu, and Y. Du, "A Deep Learning Framework Using Convolutional Neural Network for Multi-Class Object Recognition," in *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, 2018, pp. 194–198.

[17] "A Comprehensive Guide to Convolutional Neural Networks." [Online]. Available: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

[18] "An intuitive guide to Convolutional Neural Networks." [Online]. Available: https://medium.com/free-code-camp/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050.

[19] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2003.

[20] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012.

[21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, 1998.

[22] "Fully Connected Layers in Convolutional Neural Networks: The Complete Guide." [Online]. Available: https://missinglink.ai/guides/convolutional-neural-networks/fully-connected-layers-convolutional-neural-networks-complete-guide/.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012.

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[25] M. Ferguson, R. Ak, Y.-T. T. Lee, and K. H. Law, "Automatic localization of casting defects with convolutional neural networks," 2018.

[26] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014.

[27] "R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms." [Online]. Available: https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e.

[28] R. Girshick, "Fast R-CNN," in Proceedings of the IEEE International Conference on Computer Vision, 2015.

[29] 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.

[30] "YouTube." [Online]. Available: https://www.youtube.com/watch?v=xVwsr9p3irA.

[31] "Real-time Object Detection with YOLO, YOLOv2 and now YOLOv3." [Online]. Available: https://medium.com/@jonathan_hui/real-time-object-detection-with-yolo-yolov2-28b1b93e2088.

[32] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*. 2010.

[33] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017.

[34] "Review: FPN — Feature Pyramid Network (Object Detection)." [Online]. Available: https://towardsdatascience.com/review-fpn-feature-pyramid-network-object-detection-262fc7482610.

[35] W. Liu et al., "SSD: Single shot multibox detector," in Lecture Notes in Computer Science

(including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016.

[36] A. Shrivastava, R. Sukthankar, J. Malik, and A. Gupta, “Beyond Skip Connections: Top-Down Modulation for Object Detection,” *ArXiv*, vol. abs/1612.0, 2016.

[37] P. O. Pinheiro, T. Y. Lin, R. Collobert, and P. Dollár, “Learning to refine object segments,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016.

[38] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2015.