Sustav za semantičko unapređenje baza afektivne multimedije korištenjem modela dubokog učenja

Virkes, Katarina

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 480

SUSTAV ZA SEMANTIČKO UNAPREĐENJE BAZA AFEKTIVNE MULTIMEDIJE KORIŠTENJEM MODELA DUBOKOG UČENJA

Katarina Virkes

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 480

SUSTAV ZA SEMANTIČKO UNAPREĐENJE BAZA AFEKTIVNE MULTIMEDIJE KORIŠTENJEM MODELA DUBOKOG UČENJA

Katarina Virkes

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 4. ožujka 2024.

DIPLOMSKI ZADATAK br. 480

Pristupnica:	**Katarina Virkes (0065041508)**
Studij:	Računarstvo
Profil:	Programsko inženjerstvo i informacijski sustavi
Mentor:	doc. dr. sc. Marko Horvat

Zadatak: **Sustav za semantičko unapređenje baza afektivne multimedije korištenjem modela dubokog učenja**

Opis zadatka:

Baze afektivne multimedije su važni alati za istraživanja emocija i pažnje. Pa ipak, modeli ovih baza podržavaju isključivo elementarni semantički opis pohranjenih dokumenata. S obzirom na to da je postupak ručne klasifikacije slika i detekcije objekata dugotrajan, ove zadatke je pogodno izvršiti automatizirano korištenjem učinkovitih postupaka računalnog vida i modela dubokog učenja. U okviru ovog diplomskog rada potrebno je upoznati se s bazama afektivno označene multimedije i modelima emocija u računalnim sustavima. Cilj diplomskog rada je izraditi sustav koji će uspješno prepoznati objekte u afektivno označenim slikama korištenjem modela dubokog učenja i pohraniti rezultate. Koristiti bazu afektivno označenih slika Nencki Affective Picture System (NAPS), definirati prikladan model podataka te migrirati podatke iz baze NAPS u relacijsku bazu podataka. Primijeniti modele dubokog učenja na prepoznavanje i lokalizaciju objekata. Provesti eksperimentalno vrednovanje modela dubokog učenja, uključivo usporedbu s referentnim modelom i statističku obradu rezultata. Prikazati arhitekturu izrađenog sustava i bitne isječke izvornog programskog koda uz potrebna dodatna objašnjenja. Radu priložiti izvorni i izvršni kod razvijenog sustava, semantički označeni skup podataka te potrebnu dokumentaciju.

Rok za predaju rada: 28. lipnja 2024.

Sadržaj

Uvod

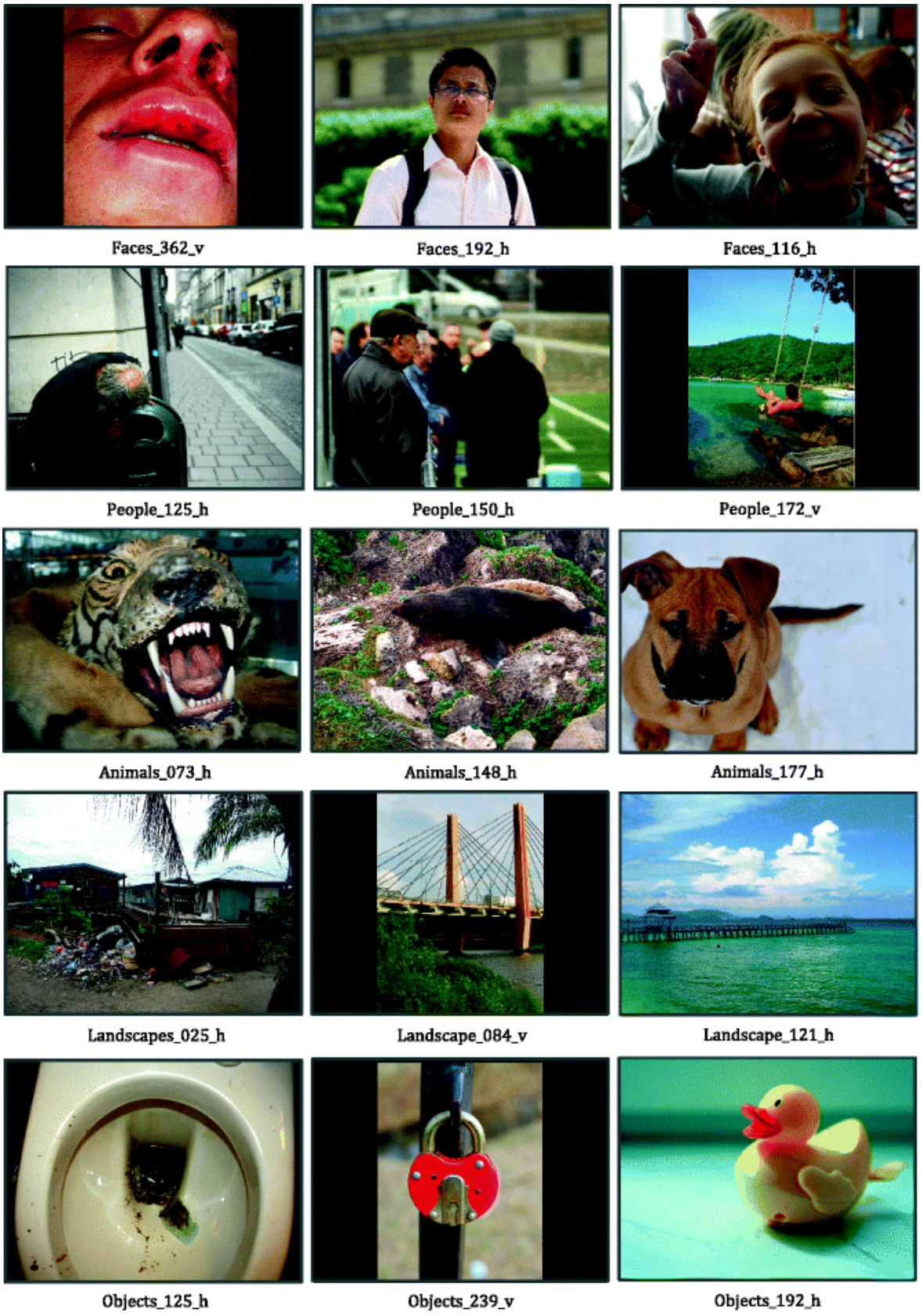
Računala su dio našeg svakodnevnog života te njihovo korištenje utječe na naše emotivno stanje i emocije. Proučavanje tog utjecaja dovodi nas do razvoja baza afektivne medije, kao što je NAPS. Nencki Affective Picture System je baza od 1356 slika, koje su podijeljene u 5 kategorija, ljudi, lica, životinje, objekti te krajolici. Baze afektivne medije sadrže multimediju koja je izabrana baš kako bi izazvala emocije na ljudima koji je konzumiraju. Osim znanosti koje gledaju utjecaj multimedije na ljude, ovakve baze se koriste i u računalnoj znanosti u sustavima za automatizirano proučavanje emocija. U baze su osim same multimedije pohranjene i semantika slike, kao i očekivane emocije. Semantika multimedije, ako i očekivane emocije, se zapisuje ručno, što čini sastavljanje i održavanje ovakvih baza jako zahtjevno. U ovom radu ćemo proučiti moguće metode za automatizirano zapisivanje semantike slika NAPS baze. Koristit ćemo modele dubokog učenja, točnije konvolucijske neuronske mreže te algoritam YOLO. Duboko učenja je grana stojnog učenja koja koristi duboke neuronske mreže, a konvolucijske neuronske mreže su klasa dubokih neuronskih mreža. Naziv su dobile od matematičke operacije konvolucije, linearnog postupka koji se koristi za obradu slika. Umjetne neuronske mreže su mreže umjetnih neurona, a inspirirane su moždanim sustavom, dok su umjetni neuroni inspirirani biološkim neuronima. Veze između neurona u biološkim neuronskim mrežama su predstavljene težinama u umjetnima. Konvolucijske neuronske mreže imaju organizaciju nalik onoj životinjske vizualne kore. YOLO (You Only Look Once) je algoritam za detekciju objekata u stvarnom vremenu. To je jednostupanjski detektor koji koristi konvolucijske neuronske mreže za određivanje graničnih okvira na slici. Odabran je jer, uz svoje odlične rezultate, ima puno manju složenost od nekih drugih poznatih modela kao što su VGG. Proučit ćemo kako na rezultate modela utječe skup podataka na kojima je treniran te vidjeti kolika je razlika u verzijama modela treniranog na istim podacima. Spomenut ćemo uz to gotov alat Azure AI Vision koji ima mogućnost pronalaska objekata na slikama, kao i njihov opis i anotaciju, bez potrebe za ikakvim znanjem programiranja.

1. Afektivna multimedija

Kako su računala te internet postali dio naše svakodnevice, njihov utjecaj na ljude je postala vrlo bitna tema u granama medicine, psihologije te neuroznanosti. Na računalu ljudi mogu biti izloženi raznim multimedijima, od gledanja slika, videa, čitanja teksta do slušanja muzike i ljudi. Afektivna multimedija nam na jednostavan, jeftin te efikasan način omogućava znanstveno proučavanje emotivnog učinka i pažnje (Horvat, 2017). Svrha sadržaja baza afektivne medije jest izazivanje emocija. Takva multimedija je pohranjena zajedno sa semantikom te očekivanim emocijama u baze afektivne multimedije. Zbog toga je ovakve baze teško sastaviti i održavati. Baze afektivne medije se koriste u već spomenutim područjima psihologije i neuroznanosti u svrhu proučavanja medije na stanje čovjeka. Osim toga, sve češće se ovakvi modeli koriste u računalnoj znanosti za analizu sentimenta te automatizirano prepoznavanje emocija. Baze se sastoje od video i audio zapisa, slika te teksta, a ovdje će se promatrat NAPS baza, baza slika.

1.1. NAPS

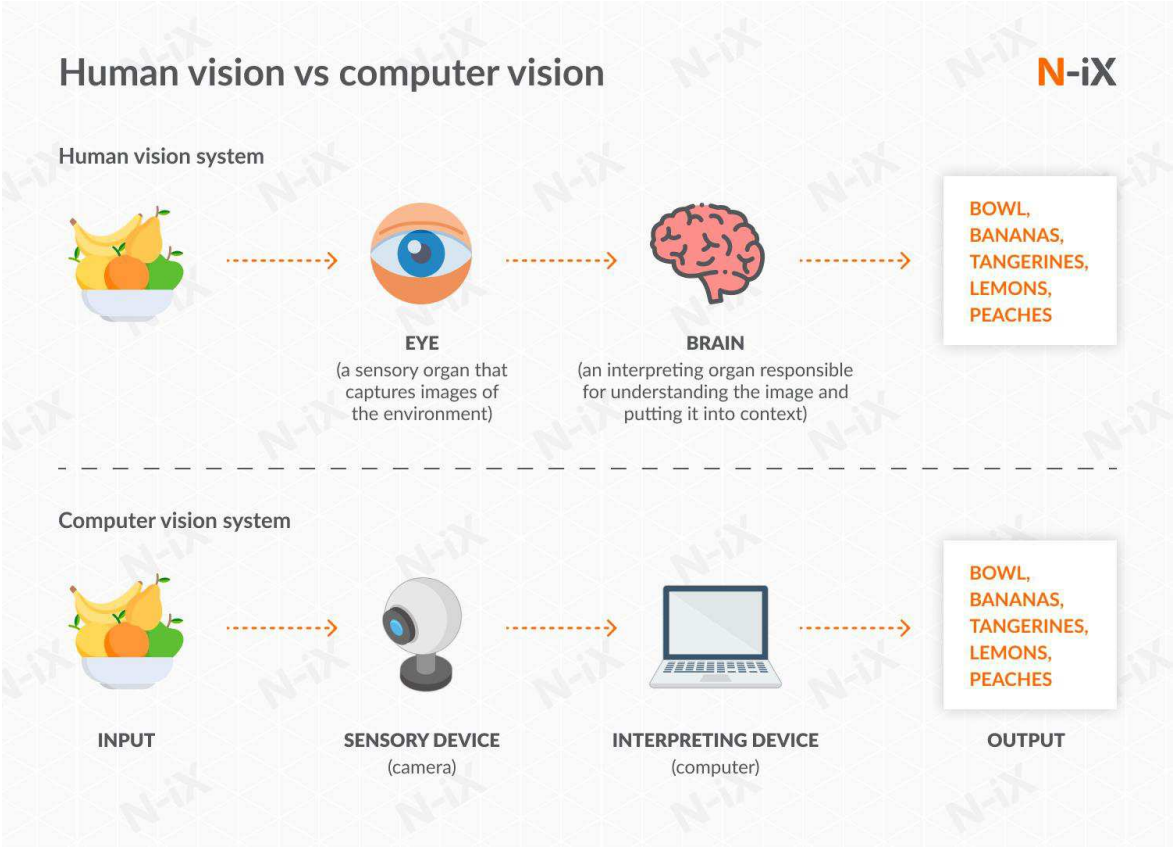
Nencki Affective Picture System (NAPS) je baza podataka sastavljena od 1356 realističnih fotografija. Fotografije su podijeljene u pet kategorija: ljudi, lica, životinje, objekti te krajolici. Primjeri fotografija su prikazani na slici (Sl. 1.1). Za svaku fotografiju napravljene su ocjene nasilja, uzbuđenja te motivacijskog smjera (izbjegavanja ili prilaženje), korištenjem dimenzijske teorije emocija koje su već prije implementirane za prijašnje baze kao što je IAPS. Kao dodatan opis koriste se fizička svojstva fotografije kao što je kontrast. Fotografije su prikupljane 6 godina, od 2006. do 2012. te je iz početnih 5000, izdvojeno 1356 fotografija koje nemaju logotipe ili dobro poznata mjesta. Podaci o bazi te primjer slika preuzeti iz članka (Artur Marchewka, 2013).



Sl. 1.1 Primjeri slika NAPS baze

2. Računalni vid

Računalni vid je grana umjetne inteligencije koja koristi strojno učenje kako bi omogućila računalima da izvedu korisne informacije iz vizualni inputa. Računalni vid je usporediv s ljudskim, dok ljudski koristi retinu, optički živac i vizualni korteks, računala koriste kamere, podatke te algoritme. Ta usporedba je vidljiva na slici (Sl. 2.1). Podaci i slika o sličnosti ljudskog i računalnog vida preuzeti s web stranice (N-iX, 2022). Dok je računalni vid trenutno u velikom zaostatku, njegova brzina procesiranja slika te mogućnost treniranja u specifičnim područjima, nam daje neke prednosti u korištenju u usporedbi s ljudima. Glavne metode koje omogućavaju razvoj računalnog vida su duboko učenje te konvolucijske neuronske mreže.



Sl. 2.1 Usporedba ljudskog i računalnog vida

2.1. Povijest

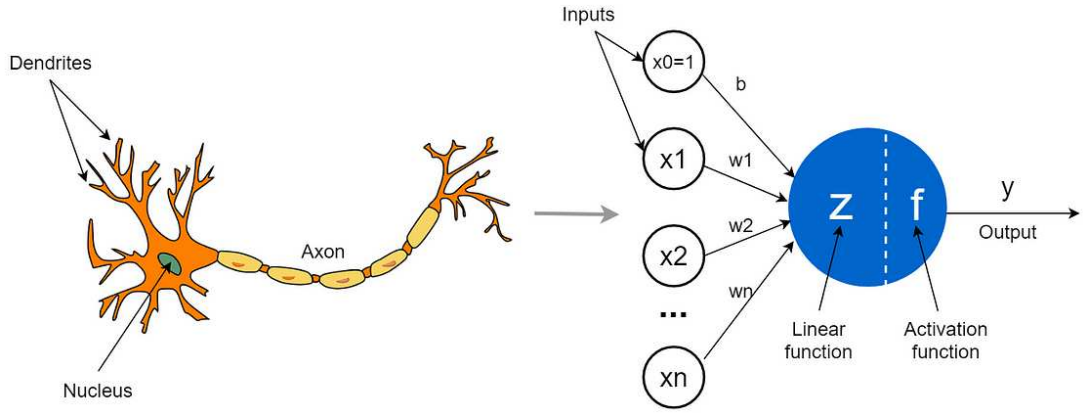
Iako danas sve češće slušamo o računalnom vidu, njegov razvoj je počeo još 50-ih godina prošlog stoljeća. Sve počinje idejom istraživača da nauče računala kako razumjeti i interpretirati vizualne podatke. Frank Rosenblatt je 1957. napravio „Perceptron“, računalnu neuronsku mrežu koja je postavila temelje za kasnije računalne neuronske mreže, koje se koriste u području računalnog vida. 1970-ih znanstvenici su fokusirani na algoritme za detekciju rubova te ekstrakciju značajki iz slika. Napravljen je „Canny edge detector“, detektor rubova koji koristi višefazni algoritam za detekciju širokog ranga rubova na slikama, te „Hough transform“, tehnika za detekciju jednostavnih geometrijskih oblika na slici. 80-ih i 90-ih godina prošlog stoljeća počinju istraživanja prepoznavanja objekata te razumijevanja scene korištenjem tehnika strojnog učenja. Razvija se kaskadna korelacija, algoritam nadziranog učenja za računalne neuronske mreže, te SIFT, Scale-Invariant Feature Transformation, algoritam za transformaciju značajki nepromjenjivog mjerila. GMM, Gaussian Mixture Models, Gaussovi modeli mješavina postaju popularni za grupiranje i modeliranje vizualnih podataka. U 2000-tim godinama dolazi do revolucije računalnog vida značajnim pomacima u području dubokog učenja, a pogotovo kod konvolucijskih neuronskih mreža. Razvijene su arhitekture kao što su AlexNet, VGGNet i ResNet. Od 2010-ih duboke neuronske mreže postaju još korištenije razvojem specijaliziranog hardvera, kao što su GPU, Graphics Processing Unit, i TPU, Tensor Processing Units. Povijest računalnog vida definirana na web stranici (Ambika, 2023).

3. Umjetne neuronske mreže

Umjetne neuronske mreže su mreže ili skupine umjetnih neurona. Težinama su, u umjetnim neuronskim mrežama, predstavljene veze između bioloških neurona. Otkrića u biološkim su nam otkrila da mozak pohranjuje informacije kao uzorke. To dovodi do razvoja umjetnih neuronskih mreža koje korištenjem istih principa razvijaju mogućnost rješavanja problema koje klasično računarstvo nije u stanju riješiti.

3.1. Umjetni neuroni

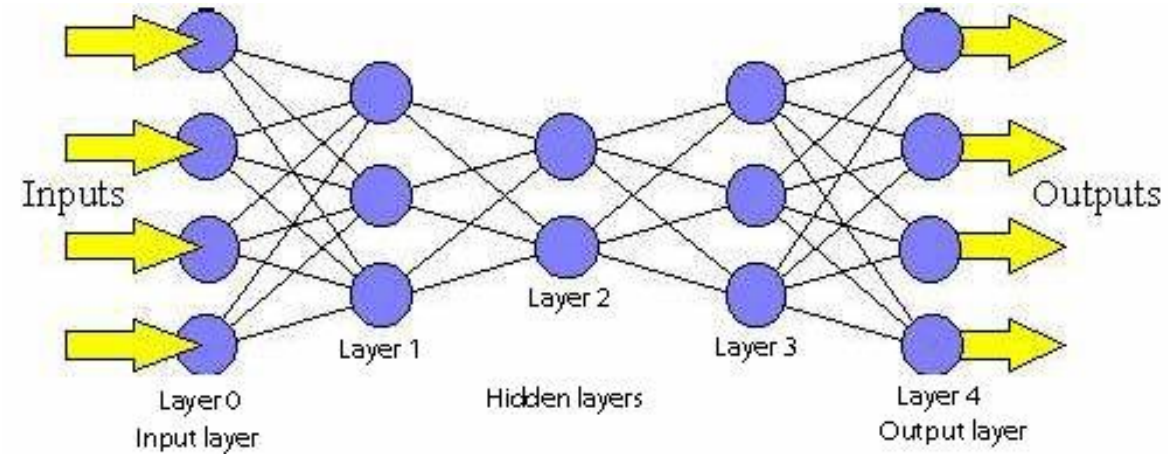
Osnovni element neuronske mreže je neuron. Iako postoji mnogo klasa bioloških neurona, svaki se sastoji od tri osnovna dijela. Dendrita koji dovode živčano uzbuđenje na tijelo stanice. Tijelo stanice koja obrađuje impulse. Aksona koji prenosi živčane impulse s tijela stanice na druge živčane stanice. Umjetni neuroni se najčešće sastoje od ulaznih podataka koji mogu biti samo zbrojeni ili množeni određenom težinom, transformacijske funkcije te izlaza (Ujević Andrijić, 2019). Sličnosti, a i razlike, bioloških i umjetnih neurona su prikazane na slici (Sl. 3.1) koja je preuzeta s web portala Towards Data Science (Pramoditha, 2021). Korištenjem različitih funkcija sumacije i transformacije postižemo različita svojstva umjetnih neuronskih mreža.



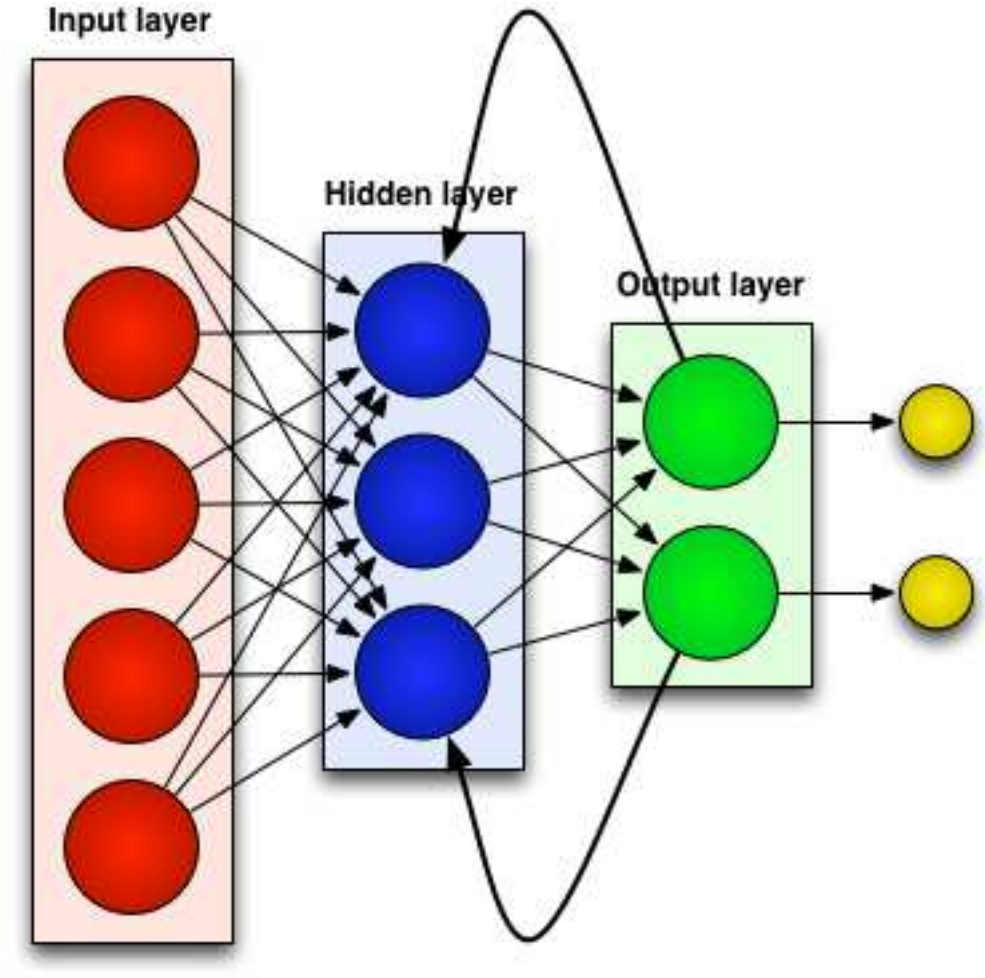
Sl. 3.1 Usporedba bioloških i umjetnih neurona

4. Duboko učenje

Duboko učenje jest grana strojnog učenja koja koristi duboke neuronske mreže za simulaciju kompleksnih mogućnosti ljudskog mozga. Neuronska mreža jest mreža, odnosno skup neurona. Postoje biološke neuronske mreže, kao što je mozak, a ovdje govorimo o umjetnim, ili računalnim, neuronima. Veze među biološkim neuronima su predstavljene težinama u računalnim neuronskim mrežama, a živčani signali realnim brojevima. Dok postoje određene duboke mreže koje nadmašuju ljudski mozak u nekom području, trenutno postojeće neuronske mreže ne uspijevaju modelirati funkcionalnost mozga živih bića. Duboke neuronske mreže, po definiciji, imaju tri ili više slojeva, dok ih u praksi često imaju puno više. Naziv duboko učenje se referencira na algoritme strojnog učenja gdje se korite hijerarhijski slojevi koji pretvaraju ulazne podatke u apstraktniju kompozitnu reprezentaciju. Ako za primjer uzmemo raspoznavanje objekata gdje je ulaz slika, prvi sloj bi pokušao raspoznati samo jednostavne oblike kao što su linije. Drugi sloj bi možda raspoznao u kakvom odnosu su ti oblici na slici, treći bi raspoznao nos i oči, a četvrti bi mogao složiti to u cijelu i raspoznati oči (Wikipedia, 2024). Duboko učenje za treniranje ne zahtijeva strukturirane podatke, već može raditi s kompleksnim podacima kao što su slike i tekst, što smanjuje potrebu za ljudskom intervencijom. Osim toga, duboko učenje, kroz proces gradijentnog silaza i unazadne propagacije, poboljšava preciznost predikcija za svaki novi podatak. Svaki sloj duboke neuronske mreže sastoji se od međusobno povezanih računalnih neurona. Svaki viši sloj nastavlja se na prethodni te optimizira predikciju, taj proces se naziva unaprijedna propagacija. Ulazni i izlazni slojevi neuronskih mreža se nazivaju vidljivim slojevima, dok su ostali skriveni. Razlikujemo acikličku (engl. feedforward) arhitekturu dubokih mreža, prikazanu slikom (Sl. 4.1) preuzetom s web stranice (Standford.edu, n.d.), te mreže s povratnom vezom (engl. recurrent), prikazanu na slici (Sl. 4.2) preuzetoj na portalu Towards Data Science (Roell, 2017). Duljina puta koju signal prođe u acikličkim neuronskim mrežama je jednaka broju slojeva, dok kod mreža s povratnom vezom signal može proći kroz svaki sloj više od jednog puta te mu je duljina puta neograničena. Najveći problemi rada s umjernim neuronskim mrežama jesu pretreniranje i veliko vrijeme izvođenja programa.



Sl. 4.1 Primjer acikličke arhitekture



Sl. 4.2 Primjer arhitekture s povratnom vezom

4.1. Primjene dubokih neuronskih mreža

Računalni vid je samo jedno od područja primjene dubokih neuronskih mreža. One imaju razna područja primjene, a neke od njih koristimo i svakodnevnom životu. Duboke

neuronske mreže koriste se u procesu testiranja novih lijekova. Proces odobravanja novih lijekova je jako dug, a DNN su uspjele malo smanjiti to vrijeme modelima koji predviđaju kako lijek utječe na ostale nutrijente. Neuronske mreže se koriste u sustavima preporuka gdje pamte naše interese te nam predlažu povezane stvari i aktivnosti. Duboke neuronske mreže se koriste u bioinformatici za predviđanje ontologije gena te povezanosti gena. Duboke neuronske mreže pokazuju neusporedive performanse u predviđanju strukture proteina. Svoju primjenu su našle i u stvaranju vremenske prognoze.

4.1.1. Prepoznavanje govornog jezika

Prva široko korištena uspješna upotreba dubokog učenja proizlazi iz automatskog prepoznavanja govornog jezika. Korištenje mreža s povratnom vezom s dugo kratkoročnim pamćenjem, LSTM RNN (engl. Long Short-Term Memory Recurrent Neural Network), omogućava obavljanje ovog zadatka u stvarnom vremenu s latencijom kašnjenja od samo 10 milisekundi. Funkcija aktivacije je pod utjecajem lokalnih aktivacija u blizi, što odgovara kratkoročnom pamćenju, dok na težine sustava utječu izračuni koji se izvode nad cijelim sekvencama, što odgovara dugoročnom pamćenju.

4.1.2. Likovna obrada

Uz razvoj mreža za prepoznavanje slika pojavile su se i metode dubokog učenja raznih likovnih zadataka. Neke od primjena su identifikacija stila u kojim je slika naslikana te generiranje slika na temelju ulaznog opisa.

4.1.3. Obrada prirodnog jezika

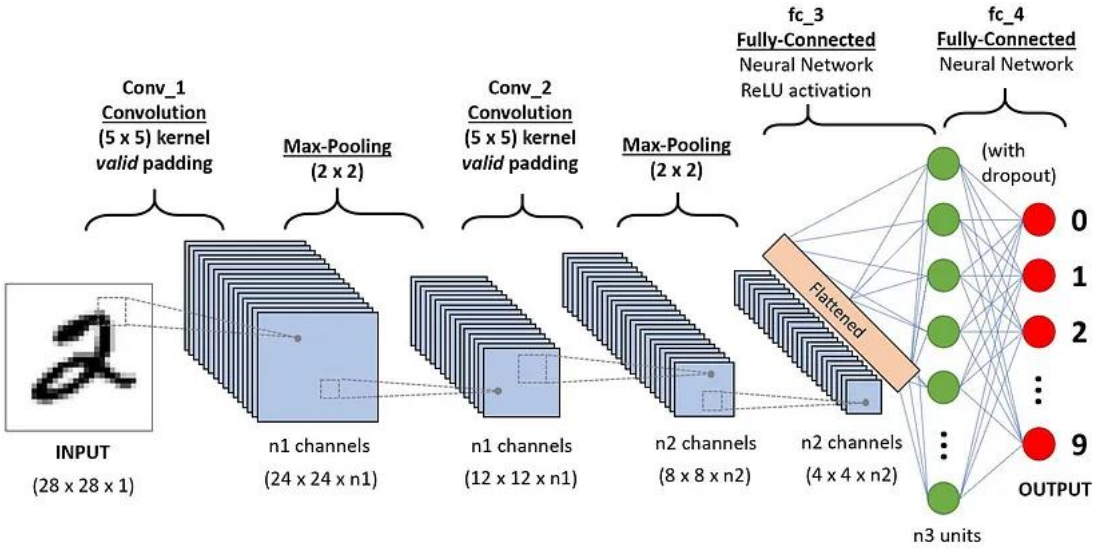
Neuronske mreže se već dugo koriste za obradu prirodnog jezika, a LSTM je donio znatna poboljšanja u prevođenju i jezičnom modeliranju. Duboke neuronske mreže daju najbolje rezultate u područjima razumijevanja govornog jezika, povrata informacija, prijevoda jezika, prepoznavanja stila pisanja klasifikacije teksta kao i mnogih drugih. Googleov prevoditelj koristi veliku LSTM mrežu.

5. Konvolucijske neuronske mreže

Konvolucijske neuronske mreže su klasa duboki neuronskih mreža. Najčešće se koriste za analizu vizualnih slika. Naziv su dobile od matematičke operacije konvolucije, koja označava specijaliziranu vrstu linearnog postupka koji se koristi u konvolucijskim mreža za obradu slika. Područje primjene možemo opisati kao područja u kojima je moguće beskonačno mnogo različitih ulaznih podataka, a izlazni podaci imaju točno određen broj klasa u koje mogu biti svrstani. U ovom radu konkretno, postoji beskonačno mnogo objekata i veličina tih samih objekata koji se mogu nalaziti na ulaznim slikama, međutim izlazi mogu biti samo klase koje su definirane u modelu. Za MS COCO to bi značilo 91 moguć izlaz uz prazan skup. Povezanost konvolucijskih neuronskih mreža inspirirana je organizacijom životinjske vizualne kore. Svaki neuron prima ulaz s određenog broja lokacija u prethodnom sloju te se ulaz u svaki neuron naziva njegovo receptivno polje. Podaci o konvolucijskim neuronskim mrežama preuzeti s web stranice (Wikipedia, 2024).

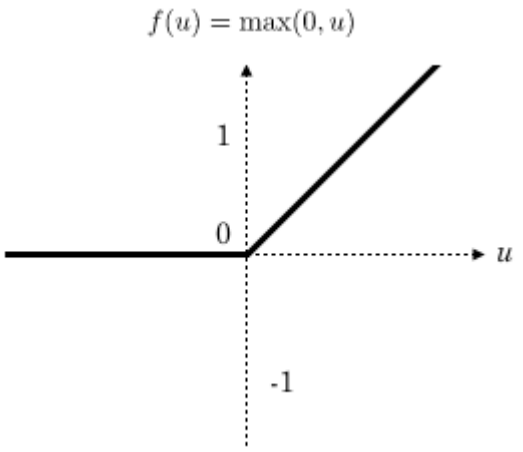
5.1. Arhitektura

Konvolucijske neuronske mreže se sastoje od ulaznog, izlaznog te više skrivenih slojeva među kojima se nalazi niz konvolucijskih slojeva koji se uvijaju s množenjem i drugim točkastim produktima. Primjer arhitekture je vidljiv na slici (Sl. 5.1) preuzetoj s web portala Towards Data Science (Saha, 2018). Točkasti produkt je algebarska operacija koja uzima dvije sekvence brojeva jednake duljine te vraća jedan broj. Prvi od skrivenih slojeva je najčešće aktivacijska funkcija te nakon nje slijede konvolucije, objedinjavanje slojeva, potpuno povezani slojevi i slojevi normalizacije.



Sl. 5.1 Primjer arhitekture konvolucijske neuronske mreže

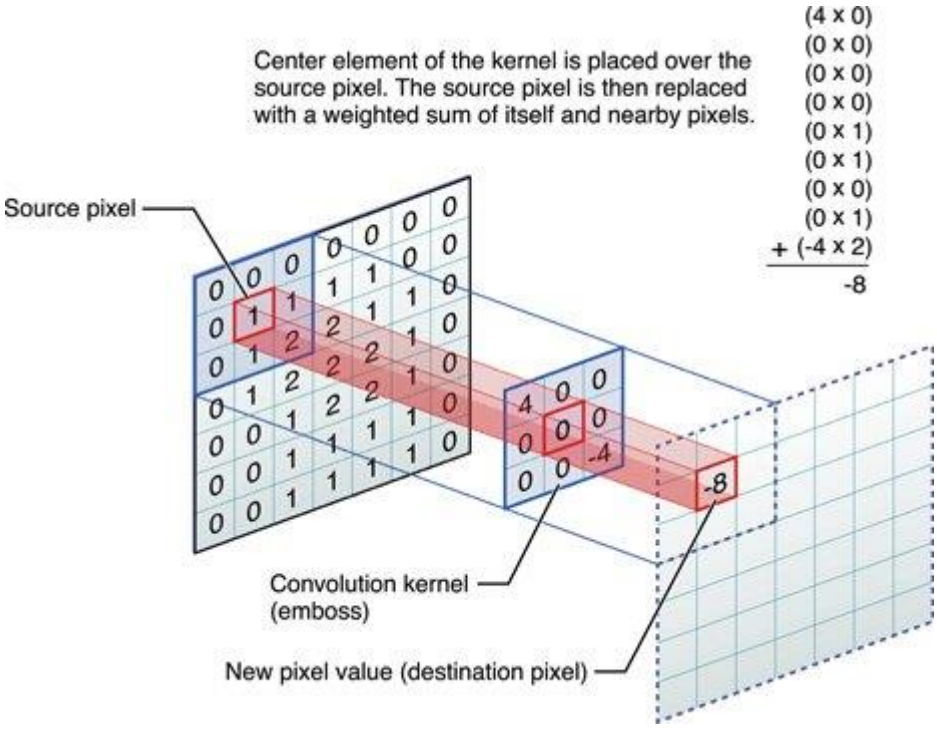
Kao aktivacijska funkcija se najčešće koristi ReLU (engl. rectified linear unit), odnosno ispravljena linearna funkcija koja kao izlaz vraća ulazni podatak ako je veći od 0 ili 0 ako je ulazni podatak manji ili jednak nuli. Postala je popularna jer se ispostavilo da su modeli je koriste lakši za treniranje te da postižu bolje rezultate. Funkcija je linearna za sve ulaze veće od 0, što znači da ima sva željena svojstva linearne aktivacijske funkcije koja nam je potrebna za stohastički gradijentni pad s unazadnom propagacijom grešaka. Funkcija je nelinearna za brojeve manje ili jednake 0 što omogućava učenje složenih odnosa. Graf funkcije se nalazi na slici (Sl. 5.2), preuzetoj s portala Research Gate (Pauly, 2017).



Sl. 5.2 Graf ReLU funkcije

5.1.1. Konvolucijski sloj

Konvolucijski sloj je tenzor, a slika, prolaskom kroz njega, postaje apstrahirana na značajnim kartama. Sam proces konvolucije jest prolazak detektora značajki, koji se još naziva i filter, površinom slike, radi provjere je li značajka prisutna. Detektor značajki je dvodimenzionalno polje težina koji predstavlja dio slike. Detektori se razlikuju u veličini receptivnog polja, ali su najčešće dimenzija 3 * 3. Filter primijenjen na dio slike računa točkasti produkt između piksela ulaznog podatka i filtera, koji se dalje prosljeđuje. Filter se pomiče na drugi dio slike dok ne obiđe cijelu površinu. Krajnji rezultat se naziva karta značajki ili konvolucijska značajka. Kao što je vidljivo na slici (Sl. 5.3) svaki dio ulazne slike se ne preslikava direktno na izlazno polje te se zato ovakvi slojevi nazivaju parcijalno povezani. Slika preuzeta s portala Medium (Singla, 2024). Težine filtera su jednake na svakom dijelu jedne slike, ali se prilagođavaju između obrada slika.

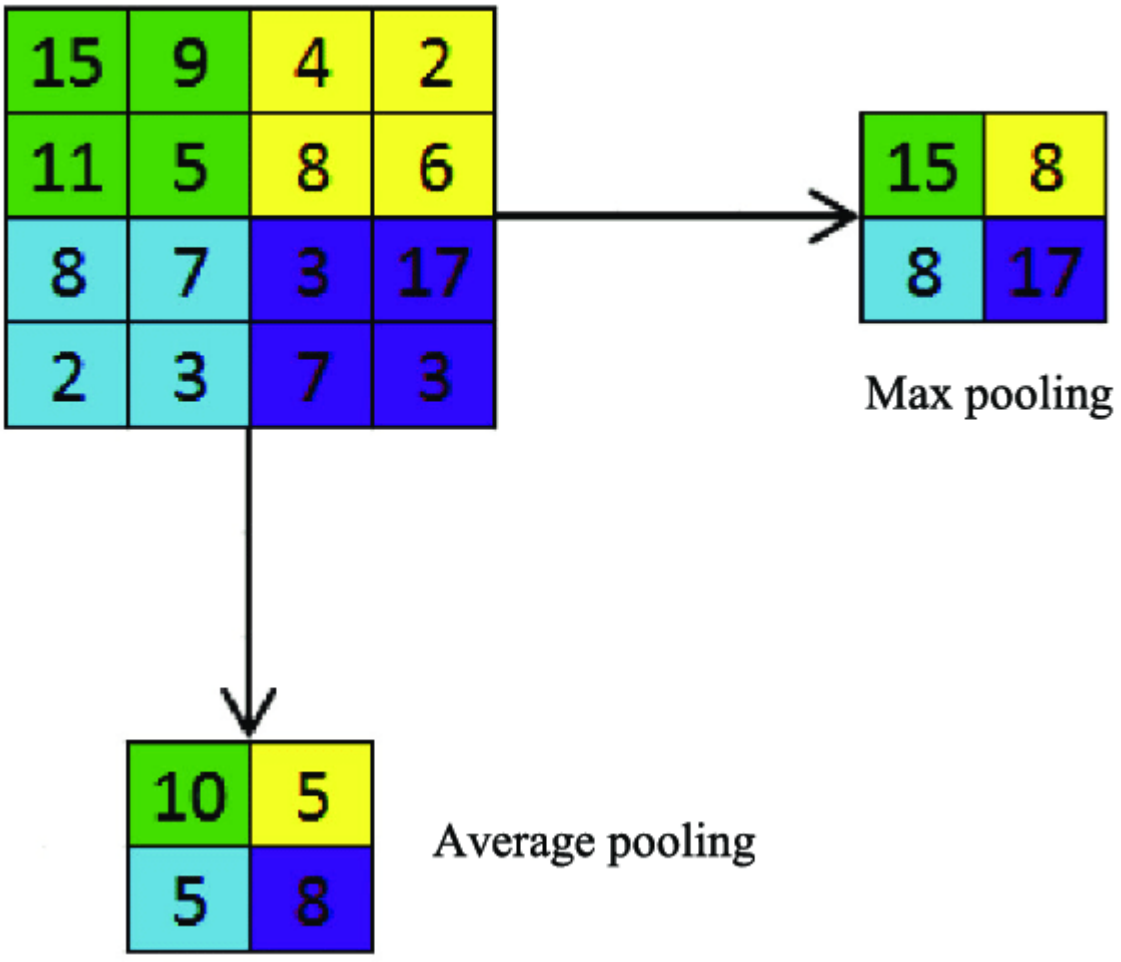


Sl. 5.3 Primjer filtriranja u konvolucijskom sloju

Broj filtera, korak i zero-padding uvijek ostaju isti te se moraju odrediti prije treniranja mreže. Broj filter određuje dubinu izlaza. Korak je udaljenost, odnosno broj piksela, koji filter prelazi na ulaznoj matrici, a zero-padding se koristi ako je filter prevelik za ulaznu sliku. On postavlja sve elemente izvan ulazne matrice na nulu. Nakon svake konvolucijske operacije primjenjuje se ReLU transformacija.

5.1.2. Sloj udruživanja

Sloj udruživanja smanjuje broj parametara u ulazu kroz filter bez težina koji primjenjuje funkciju agregacije na vrijednosti perceptivnog polja. Udruživanja mogu biti lokalna ili globalna. Lokalno kombinira uske nakupine, najčešće 2 * 2, dok globalno djeluje na sve neurone konvolucijskog sloja. Udruživanja mogu biti maksimalna, koristi se maksimalna vrijednost os svake nakupine neurona, ili prosječna, koristi se prosječna vrijednost nakupine. Slikovni prikaz udruživanja nalazi se na (Sl. 5.4) preuzetoj s portala Research Gate (Wang, 2017). Sloj udruživanja smanjuje kompleksnost mreže te povećava efikasnost.

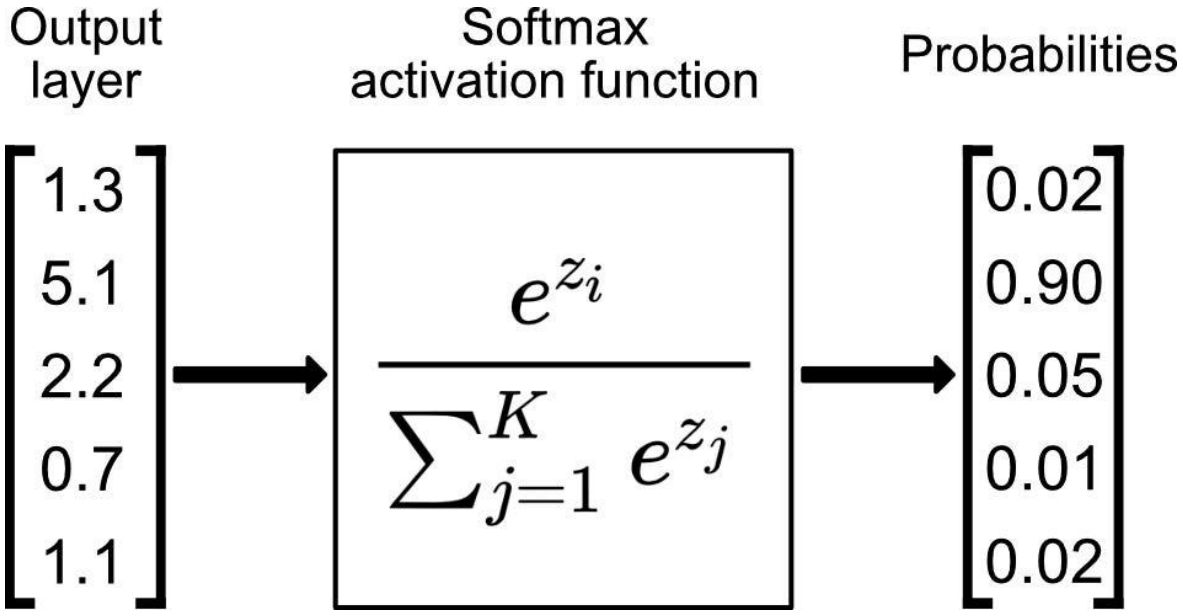


Sl. 5.4 Prikaz razlika maksimalnog i prosječnog udruživanja

5.1.3. Potpuno povezani sloj

Kao što i ime govori u ovom sloju je svaki čvor izlaznog sloja direktno povezan s čvorom prethodnog sloja. Ovaj sloj provodi klasifikaciju temeljenu na rezultatima prethodnih

slojeva, a za to koristi softmax aktivacijsku funkciju. Samo ime nam govori da je to funkcija glatke aproksimacije maksimalnog argumenta funkcije. Softmax funkcija je utemeljena na Luceinom aksiomu izbora, koji na govori da na vjerojatnost odabira jednog predmeta preko drugog predmeta iz izvora ne utječe prisutnost ili odsutnost drugih predmeta u izvoru. Softmax funkcija uzima ulazni vektor sastavljen od realnih brojeva te ga normalizira u vjerojatnosnu distribuciju primjenjujući standardnu eksponencijalnu funkciju vidljivu na slici (Sl. 5.5) preuzetu s portala Towards Data Science (Radečić, 2020).



Sl. 5.5 Prikaz djelovanja softmax funkcije

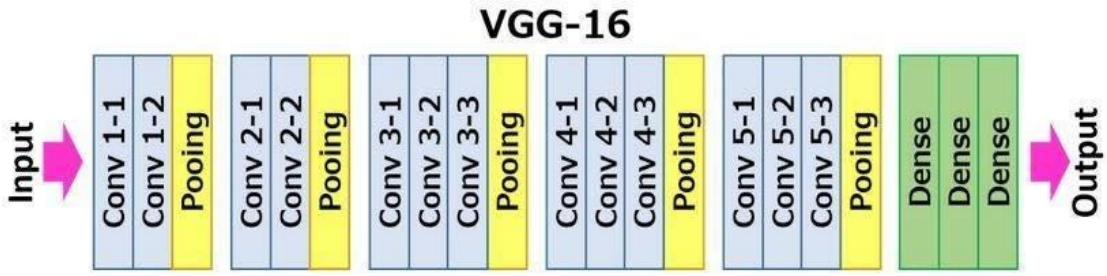
Kao baza funkcije se najčešće koristi *e,* ali baza može biti bilo koji broj veći od 0. Ako je baza u intervalu od 0 do 1, manji ulazi će rezultirati većim izlaznim vrijednostima, a za baze veće od 1, veći ulazi će rezultirati većim izlaznim vjerojatnostima.

6. Detekcija objekata

Jedna od glavnih funkcija računalnog vida jest klasifikacija slika te detekcija objekata. Klasifikacija slika fokusira se na grupiranje slika u predefinirane kategorije. Za detekciju objekata koristi se već istrenirani klasifikator, koji onda na slici nalazi gdje se nalazi neka od predefiniranih klasa. Najčešći modeli koji se koriste za detekciju objekata su YOLO, VGG te Inception. Određene primjene algoritama za detekciju objekata su premašila ljudske mogućnosti, jedan od takvih primjera je algoritam za prepoznavanje prometnih znakova iz 2011. godine.

6.1. VGG

VGG jest model konvolucijske neuronske mreže čija arhitektura je razvijena od strane Visual Geometry Group (VGG) na Oxfordskom Sveučilištu. VGG – 16 je karakteriziran svojom dubinom od 16 slojeva s težinama, od kojih je 13 konvolucijskih slojeva te 3 potpuno povezana sloja, a sadrži još i 5 Max Pooling slojeva. Arhitektura je prikazana na slici (Sl. 6.1).

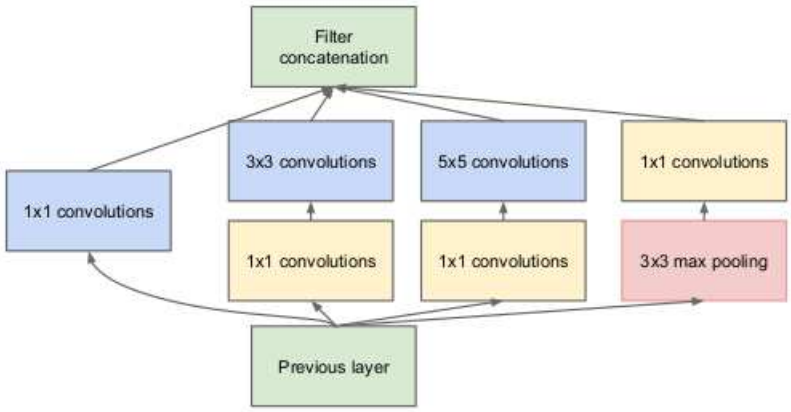


Sl. 6.1 Prikaz arhitekture VGG-16 modela

VGG – 19 se sastoji od 19 slojeva s težinama, odnosno sadrži tri dodatna konvolucijske sloja. VGG se smatra jednim od najboljih modela računalnog vida. Sposoban je klasificirati 1000 slika s 1000 kategorija uz točnost od 92.7 %. Međutim, VGG model je jako spor za treniranje te zauzima puno prostora za pohranu. Podaci o modelima te slika arhitekture preuzeti s portala Medium (Great Learning, 2021).

6.2. Inception

Inception modul je gradivna jedinica za konvolucijske neuronske mreže razvijena od strane Googlea 2014. godine (deepai, n.d.). Predstavljala je veliko napredovanje u području dubokog učenja i računalnog vida. Inception model omogućava konvolucijskim neuronskim mrežama istovremeno izvlačenje više značajki kroz korištenje filtera različitih veličina na istom sloju mrežu. Filteri različitih veličina, isto tako, omogućavaju izvlačenje značajki različitih veličina i kompleksnosti. Primjer modela smanjenih dimenzija je dan na slici (Sl. 6.2) preuzetoj s web stranice Analytics Vidhya (Shaikh, 2023). Uz filtere Inception model koristi pooling grane, koje predstavljaju još jednu formu agregacije.

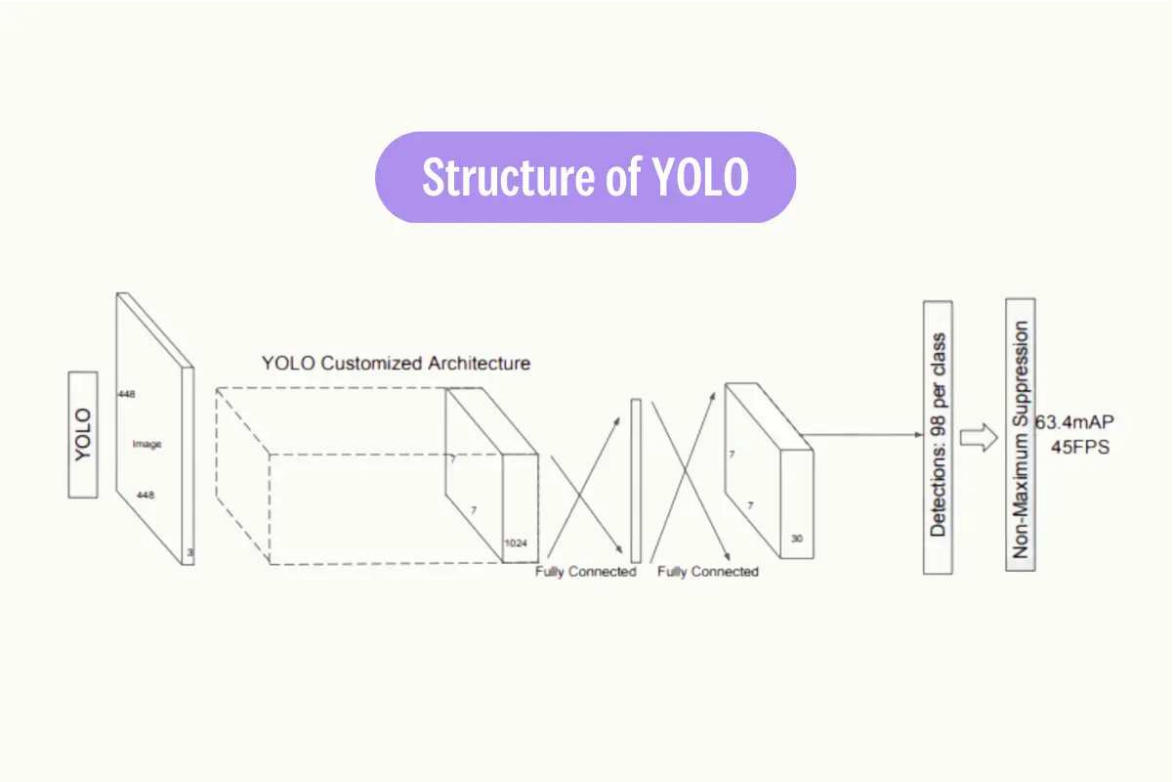


Sl. 6.2 Inception model smanjenih dimenzija

Ovaj model koristi puno manje slojeva neuronskih mreže, ali su sami slojevi kompleksniji. Iako je ovaj model razvijen zbog efikasnosti, on i dalje zahtjeva puno resursa zbog velikog broja operacija koje obavlja.

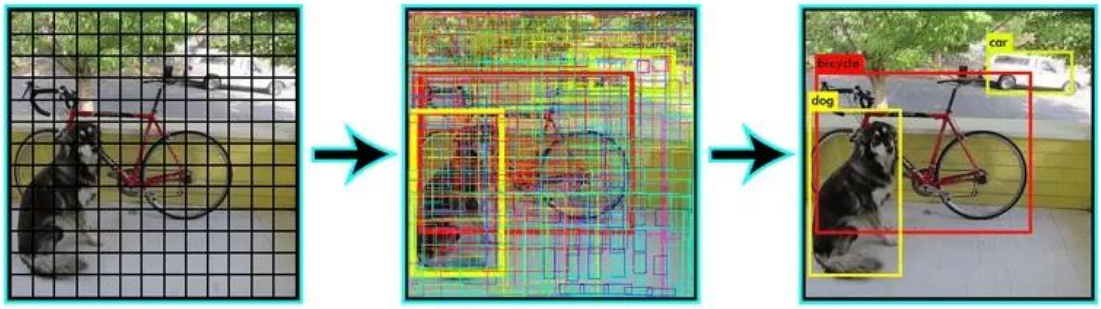
6.3. YOLO

YOLO (You Only Look Once) jest algoritam za detekciju objekata u stvarnom vremenu. Razvili su ga Joseph Redmon i Ali Farhadi 2015. godine. YOLO jest jednostupanjski detektor objekata koji koristi konvolucijske neuronske mreže za predviđanje graničnog okvira (eng. bounding box) te vjerojatnosti klasa objekata na slikama. Arhitektura takvog detektora je vidljiva na slici (Sl. 6.3). Slika i podaci o YOLO algoritmu preuzeti s web stranice Kili (Mirkhan, 2023).



Sl. 6.3 Arhitektura konvolucijske neuronske mreže korištene za YOLO

Prvobitno je implementiran korištenjem Darknet okvira. Darknet jest okvir neuronskih mreža otvorenog izvora, napisan u programskom jeziku C i CUDA. YOLO algoritam razdijeli sliku na ćelije te za svaku ćeliju predvidi vjerojatnost prisutnosti objekta, njegovu klasu te koordinate njegovog graničnog okvira. Faze rada su prvobitan prolaz ulazne slike kroz konvolucijsku neuronsku mrežu kako bi se izvukle glavne strukture slike. Nađene strukture su zatim proslijeđene seriji potpuno povezanih slojeva mreže, koji predviđaju vjerojatnosti klase objekata te njegove granične okvire. Izlaz jest set graničnih okvira te klasa vjerojatnosti za svaku ćeliju. Granični okviri se filtriraju korištenjem post-procesirajućeg algoritma zvanog non-max potiskivanje kako bi se uklonili okviri koji se preklapaju, odnosno kako bi se proslijedili granični okviri s najvećom vjerojatnosti, prikazano djelovanje na slici (Sl. 6.4) preuzetoj s web stranice Labellerr (Singh, 2023). Krajnji rezultat jest set predviđenih graničnih okvira te klase labela za svaki objekt u slici.



Sl. 6.4 Primjer non-max potiskivanja

Trenutno je razvijena deveta verzija algoritma, a svaka verzija je nadogradnja prethodne, što donosi ubrzanje i povećanu preciznost. Algoritam se koristi u mnogim sustavima koji rade u stvarnom vremenu, kao što su samovozeći automobili te sistemi nadzora.

6.3.1. YOLO v2

YOLO v2, odnosno YOLO 9000, je druga verzija YOLO algoritma. Najveća razlika jest postojanje okvira koje YOLO v2 može predvidjeti, dok je u prvoj verziji algoritam mogao predvidjeti samo koordinate granični okvira. To omogućuje drugoj verziji obradu objekata različitih oblika i veličina. Druga glavna razlika jest višerazmjerni pristup. Ulazna slika prolazi kroz CNN na više razmjera, što omogućuje detekciju objekata različitih veličina. YOLO v2 koristi funkciju pogreške kvadratnog zbroja kao funkciju gubitka, što čini algoritam robusnijim te omogućuje brži oporavak modela. YOLO v2 uz to koristi i dublju CNN od YOLO-a, što mu omogućuje prepoznavanje više značajki na slici.

6.3.2. YOLO v3

YOLO v3 je treća verzija YOLO algoritma razvijena od istih autora. YOLO v3 koristi tehniku mreža piramidi značajki ili FPN (engl. feature pyramid network) koja omogućuje raspoznavanje značajki različitih skala. Konvolucijska neuronska mreža u ovoj verziji ima još više slojeva te se koristi nova funkcija gubitka. Koristi se kombinacija funkcija gubitka klasifikacije i lokalizacije što omogućuje modelu učenje vrijednosti klasa te koordinata graničnih okvira. Ova verzija jest jedna od najkorištenijih verzija algoritma čak i danas. Prošla je razna testiranja i validacije na raznim aplikacijama te ja jako stabilna i pouzdana. Uz to modeli su manji, od onih iz novijih verzija, te izvođenje zahtjeva manje računalnih resursa.

6.3.3. YOLO v4

YOLO v4 koristi napredniju arhitekturu neuronske mreže, koja ima parcijalne veze između slojeva. Uz to koristi tehniku zvanu obrada prostorne piramide koja omogućuje nalaženje značajki različitih skala i rezolucija.

6.3.4. YOLO v5

2020. je Ultralytics izdao YOLO v5. YOLO v5 je arhitekturalno jako sličan YOLO v4, ali je temeljen na drugom okviru PyTorch-u umjesto Darkneta te se zbog toga ne smatra službenom verzijom algoritma. PyTorch je biblioteka za strojno učenje temeljena na biblioteci Torch. Najviše se koristi u područjima računalnog vida i obrade prirodnog jezika. Broji softveri za duboko učenje izgrađeni su PyTorch-u, uključujući Tesla Autopilot (Wikipedia, 2024). Uz sličnu arhitekturu i performanse su im slične. YOLO v5 koristi jedan konvolucijski sloj što ga čini fleksibilnijim i primjenjivim za objekte različitih oblika i veličina. Također koristi CmBN (eng. cross mini-batch normalization), tehniku unakrsne normalizacije mini serija, za veću točnost modela. YOLO v5 koristi prijenos učenja, odnosno moguće ga je trenirati na većem skupu podataka, a zatim podesiti na manjem skupu.

6.3.5. YOLO v6

Meituan Technical Team 2022 izdaje YOLO v6 koji je efikasniji te koristi jednostavniju mrežu što ga čini bržim te omogućuje pokretanje uz manje računalne snage.

6.3.6. YOLO v7

YOLO v7 jest jedna od najnovijih stabilnih verzija YOLO algoritma. Najveće poboljšanje u odnosu na prethodnu verziju dolazi od nove implementacije okvira koji se koriste u prvoj fazi detekcije objekata (eng. anchor box). Ovakvi okviri su različiti od graničnih okvira, jer imaju predodređene veličine, dok su granični okviri veličine detektiranog objekta. Ova verzija donosi 9 omjera okvira, što nam omogućuje nalaženje objekata različitih veličina te smanjuje postajanje lažnih pozitiva u rezultatima. Nova verzija, isto tako, donosi novu funkciju zvanu „focal loss“, odnosno žarišni gubitak, koja pomaže u otkrivanju malih objekata namještanjem težina gubitka.

6.3.7. YOLO v8

2023. Ultralytics donosi novu verziju svojeg YOLO algoritma koja ima bolje performanse od prijašnjih verzija. Međutim, neki ga, kao i YOLO v5, ne smatraju YOLO v8 službenom YOLO verzijom. YOLO v8 je zaštićen AGP-3.0 licencom što znači da organizacije moraju platiti Ultralyticsu ako koriste algoritam u komercijalne svrhe.

7. MongoDB

MongoDB je baza podataka s javno dostupnim izvornim kodom koja može raditi na različitim platformama (MongoDB, 2024). Klasificira je kao NoSQL baza, što znači da za naredbe ne koristi SQL jezik. MongoDB koristi dokumente slične JSON objektima, umjesto tradicionalnih tablica s relacijama, uz opcionalne scheme. Razvijen je od strane MongoDB Inc. te su trenutne verzije licencirane pod Server Side Public License. Prva verzija objavljena je 2009. godine. MongoDB je korišten kao baza jer je rezultat pronalaska objekata vraćen kao JSON, odnosno kao lista JSON-a. Takve objekte je puno lakše spremati u nerelacijske baze, a i samo filtriranje nad tim objektima je puno brže.

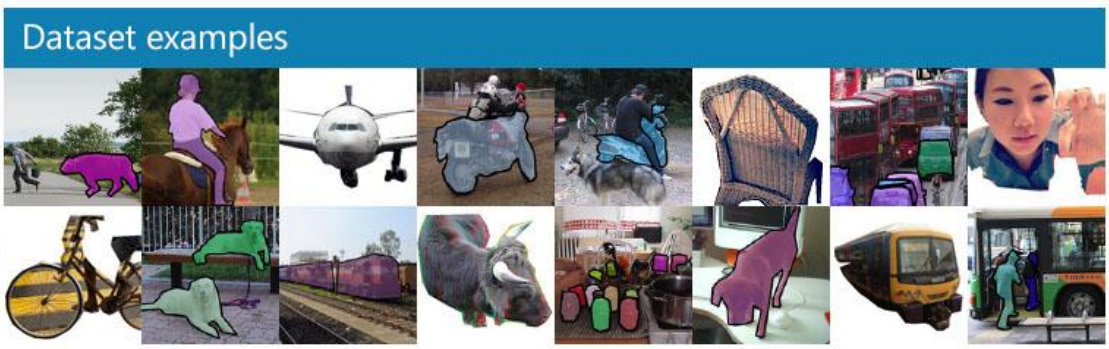
8. Analiza NAPS baze korištenjem YOLO v3 algoritma

8.1. OpenCV

OpenCV je najveća biblioteka za rad s računalnim vidom na svijetu. Otvoren je koda te sadrži preko 2500 algoritama. OpenCV jest pod operativom neprofitne organizacije Open Source Vision Foundation. Nudi mogućnost korištenja C++, Python i Java programskih jezika, te Linux, MacOS, Windows, iOS i android operativnih sustava (OpenCV, 2024).

8.1.1. MS COCO set podataka

MS COCO (Microsoft Common Objects in Context) set podataka se sastoji od 328 tisuća slika, primjeri prikazani na slici (Sl. 8.1) preuzetoj na web stranici (COCO dataset, 2024). Koristi se za detekciju objekata, segmentaciju, detekciju ključnih stvari na slikama te titlovanje. Prva verzija seta je objavljena 2014. godine, a zadnja 2017. godine. Svaka nova verzija donosi nove slike i bolje anotacije. Detekcija objekata se provodi anotacijama koje sadrže granične okvire za 80 klasa objekata. Titlovanje se provodi iz anotacija pisanih prirodnim jezikom koje opisuju slike. Detekcija stvari se ostvaruje segmentacijom slike, a sadrži anotacije kao što su trava, nebo itd. a zajedno s klasama objekata, tvore 91 klasu za detekciju (Papers with Code, 2024).



Sl. 8.1 Primjer slika COCO seta

8.1.2. Open Images set podataka

Open Images jest set podataka koji sadrži približno 9 milijuna anotiranih slika. Sadrži 16 milijuna graničnih okvira na 1.9 milijuna slika, primjeri prikazani slikom (Sl. 8.2). Sadrži 600 klasa objekata koje može detektirati, a svaka slika sadrži prosječno 8.3 objekta na slici. Open Images sadrži i anotacije vizualnih relacija na slici, kao što su „čaša na slici“. Ukupno 3.3 milijuna takvih anotacija tvori 1466 različitih relacija. Segmentacijske maske napravljene su nad 2.8 milijuna objekata, koji su raspoređeni u 350 klasa. Podaci i primjeri slika preuzeti s web stranice (Google Apis, 2022).

Sl. 8.2 Primjer anotacija Open Images seta

8.2. Analiza slika

Početni kod je preuzet s portala Medium (kamal_DS, 2023). Dohvatimo s Darkneta konvolucijsku neutronsku mrežu te učitamo konfiguraciju i težine za odabrani skup slika na kojima želimo da je trenirana (Kôd 8.1).

```
Neural_Network =
cv2.dnn.readNetFromDarknet(os.path.join(main_dir, "darknet-
yolo/darknet/cfg/yolov3-openimages.cfg"),
os.path.join(main_dir, "darknet-yolo/yolov3-
openimages.weights"))
```

Kôd 8.1 Program za konfiguraciju željenog modela

Isto tako pripremimo imena razreda koje mreža može vratiti kao rezultat (Kôd 8.2).

```
k = open(os.path.join(main_dir, "darknet-
yolo/darknet/data/openimages.names"),'r')
```

Kôd 8.2 Program za dohvaćanje imena razreda

Zatim koristimo OpenCV-jevu funkciju `cv2.dnn.blobFromImage()` koja vraća blob ulazne slike nakon oduzimanja srednje vrijednosti n-torke (R, G, B), normalizacije te zamjene R i B komponente (Rosebrock, 2017). Za sve nađene objekte s njihovim okvirima gledamo koji je najvjerojatniji razred objekta te ako je vjerojatnost veća od granične vrijednosti koju smo odredili, u ovom slučaju 50 %, zapisujemo objekt (Kôd 8.3).

```
def bounding_box_prediction(output_data):
    bounding_box = []
    class_labels = []
    confidence_score = []
    for i in output_data:
        for j in i:
            high_label = j[5:]
            classes_ids = np.argmax(high_label)
            confidence = high_label[classes_ids]

            if confidence > Threshold:
                w , h = int(j[2] * image_size) , int(j[3] *
image_size)
                x , y = int(j[0] * image_size - w/2) ,
int(j[1] * image_size - h/2)
                bounding_box.append([x,y,w,h])
                class_labels.append(classes_ids)
                confidence_score.append(confidence)
```

Kôd 8.3 Program za pronalazak razreda objekta

Koristimo non-max potiskivanje kako bi dobili samo jedan granični okvir za svaki pronađeni objekt (Kôd 8.4).

```
prediction_boxes = cv2.dnn.NMSBoxes(bounding_box ,
confidence_score , Threshold , .6)
```

Kôd 8.4 Program za primjenu non-max potiskivanja

Svaki objekt u odgovarajućem obliku zapisujem u bazu podataka.

8.3. Rezultati analize

8.3.1. Model treniran na MS COCO setu

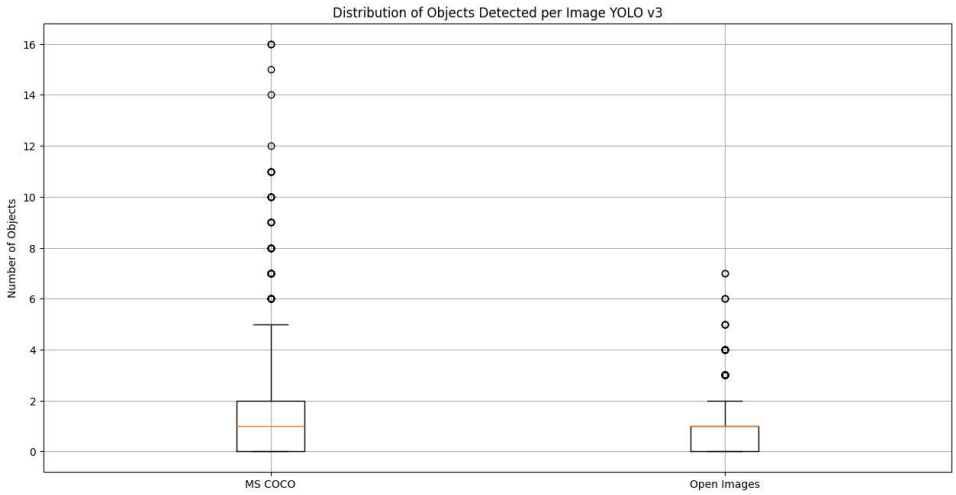
Za 887 slika, odnosno za 65.4 % slika, je nađen barem jedan objekt na slici s odgovarajućom klasom i graničnim okvirom. Prosječna sigurnost klase nađenog objekta je 0.85.

8.3.2. Model treniran na Open Images setu

Za 681 sliku, što čini malo više od 50 %, je nađen barem jedan objekt na slici. Prosječna sigurnost nađenih objekata na slici je 0.711.

8.3.3. Usporedba

Na slici (Sl. 8.3) je prikazana grafička usporedba distribucije nađenih objekata na slikama. Vidljivo je da su rezultati analize puno bolji na modelu treniram na MS COCO setu od onoga treniranog na Open Images setu. To je u kontradikciji s našim očekivanjima da će model treniran na većem skupu podataka dati bolje rezultate. Međutim razlog takvih rezultata je loša anotacija Open Images seta, gdje su određeni objekti označeni graničnim okvirima te je anotirana njihova klasa na određenim slikama, dok na drugim slikama nikako označeni. To nam dokazuje koliko je važna dobra priprema seta za treniranje neuronskih mreža.



Sl. 8.3 Usporedba distribucije nađenih objekta na slikama, lijevo model treniran na COCO setu, desno Open Images

9. Analiza NAPS baze korištenjem YOLO v5 algoritma

9.1. Objects365

Objects365 je set slika razvijen za detekciju različitih objekata u raznim situacijama. Sadrži 365 kategorija, 2 milijuna slika te 30 milijuna graničnih okvira. Primjeri slika dani su na slici (Sl. 9.1), a preuzeti sa web stranice (Ultralytics, 2024). Anotacije se provode kroz pažljivo osmišljeni proces od tri koraka, te je zbog toga ovaj set najveći potpuno anotirani set do sada. Ima 11 superkategorija: čovjek i povezani dodaci, dnevna soba, odjeća, kuhinja, instrument, uredski pribor i životinja (Shuai Shao, 2019).

Sl. 9.1 Primjeri slika Objects365 seta

9.2. Analiza

YOLO v5 je puno lakši za korištenje od prethodne verzije, ali zato i na manje faktora možemo utjecati. Prvo trebamo učitati predtreniran model:

```
model = torch.hub.load(r'yolov5', 'custom',
path=r'yolov5l.pt', source='local')
```

Otvoriti slike koristeći PIL biblioteku:

```
img = Image.open(image_path)
```

Rezultate nam vrati sam model, bez dodatnih predobrada:

```
results = model(img)
```

Dobivene rezultate pohranjujemo u zapisu koji nam odgovara.

9.3. Rezultati analize

Kod dobivanja objekata za ovaj model nije provjeravano je li im sigurnost veća od 50 %.

9.3.1. Model treniran na MS COCO setu

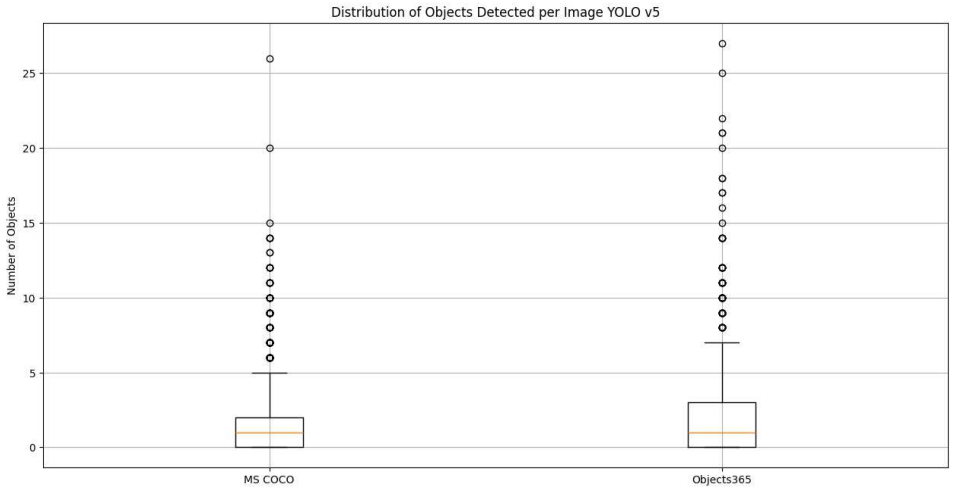
Za 1121 slika, odnosno 82.1 % slika, je nađen barem jedan objekt. Prosječna sigurnost klase nađenog objekta je 0.582. Filtracijom objekata kao u modelu verzije tri dobivamo 899 slika za koje nađen barem jedan objekt te sigurnost od 0.735. Kada ove rezultate usporedimo s modelom YOLO v3 treniranom na istom setu podataka vidimo da je ovaj model našao objekte na još 12 slika. U isto vrijeme sigurnost za klasu objekta na slici mu je pala za 0.115.

9.3.2. Model treniran na Objects365 setu

Za 1060 slika, odnosno 77.7 % slika, je nađen barem jedan objekt na slici. Prosječna vrijednost sigurnosti klase nađenog objekta je 0.558. Filtracijom objekata koji imaju sigurnost manju od 50 %, broj slika s nađenim objektima je 811, a sigurnost je 0.725.

9.4. Usporedba

Na slici (Sl. 9.2) je dana grafička usporedba distribucije nađenih objekata na slikama. Vidimo da je model treniran na MS COCO setu našao objekte na više slika te da ima veću sigurnost u klase nađenih objekata. Međutim, model treniran na Objects365 setu je naša ukupno više objekata na svim slikama. Objects365 set isto tako ima 365 kategorija objekata, za razliku od MS COCO-ih 91, što mu daje prednost na ovakvim bazama podataka kao što je NAPS.



Sl. 9.2 Usporedba distribucije nađenih objekta na slikama, lijevo model treniran na COCO setu, desno Objects365

10. Analiza NAPS baze korištenjem Azure AI Visiona

10.1. Azure AI Vision

Azure AI Vision jest Microsoftov servis koji nudi razne mogućnosti veza za područje računalnog vida. Neke od mogućnosti su analiza slika, čitanje teksta iz slika, detekcija lica te anotaciju slika (Azure, n.d.). Alat nam isto tako može služiti za pripremu slika jer prepoznaje gdje se na slici nalaze praznine te preporučuje podrezivanja slika, ako za time ima potrebe. Korištenje ovog servisa ne zahtijeva nikakvo predznanje strojnog učenja. Analiza slika može vratiti anotacije sadržaja za tisuće raspoznatljivih objekata, živućih bića, krajolika te akcija na slici. Anotacije nisu organizirane u taksonomiju te nemaju hijerarhiju nasljeđivanja. Kolekcija anotacija sadržaja čini bazu za stvaranje opisa slika, koji je formatiran u rečenicu na ljudski čitljivom jeziku (Learn, 2024). Analiza slika nam uz to može vratiti objekte na slikama s njihovim graničnim okvirima i definiranom klasom.

10.2. Analiza slike

Tražimo da nam analizator slike vrati opis slike (eng. caption), objekte na slici (eng. objects) te anotacije sadržaja slike (engl. tags) (Kôd 10.1).

```
result = client.analyze(
        image_data=image_data,
        visual_features=[VisualFeatures.CAPTION,
VisualFeatures.OBJECTS, VisualFeatures.TAGS],
        gender_neutral_caption=True,
        )
```

Kôd 10.1 Konfiguracija analizatora

Pregledavamo jesu li tražene osobine nađene te, ako jesu, spremamo ih u bazu podataka (Kôd 10.2).

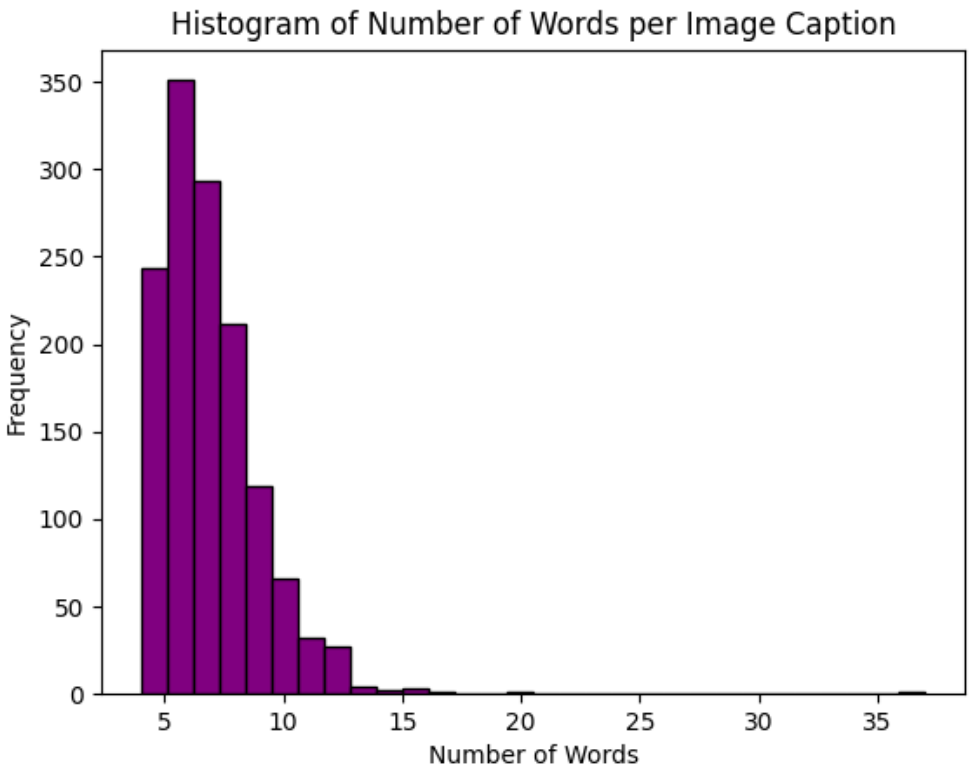
```
if result.tags is not None:
        tag_list = []
        for t in result.tags["values"]:
            tag_dict = {
```

```
                "name": t.name,
                "confidence": t.confidence
            }
            tag_list.append(tag_dict)
mongo_dict["tags"] = tag_list
```

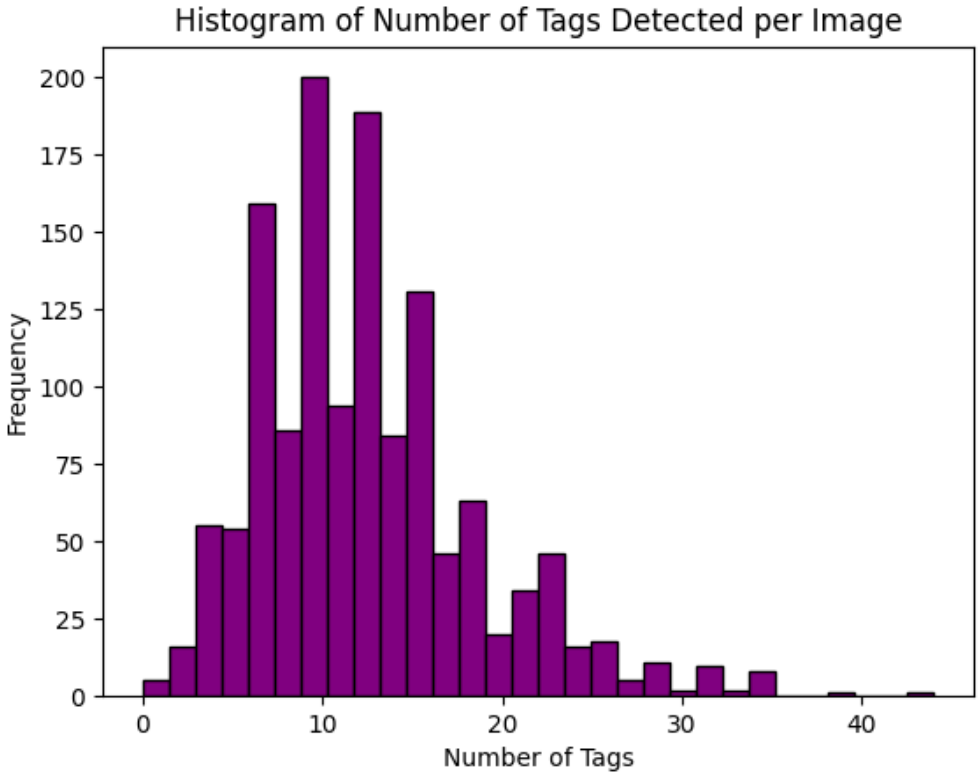
Kôd 10.2 Program za spremanje podataka u bazu

10.3. Rezultati analize

Za svaku sliku iz bazu je dobiveno više anotacija te je za svaku sliku generiran opis slike, dobiven iz najmanje dvije pripadajuće anotacije. Histogram duljina rečenica opisa slika prikazan je na grafu (Sl. 10.1), a broj anotacija na grafu (Sl. 10.2). Za 960 slika, odnosno otprilike 71 % slika, su nađeni objekti sa svojim klasama i graničnim okvirima. Najviše praznih polja objekata je vraćeno za slike iz kategorije krajolika. Pretpostavka je da je kod treniranja modela za takve slike postajala samo anotacija klase, a ne i naznačeni granični okviri. Objekti, za razliku od anotacija, imaju definiranu taksonomiju te hijerarhiju nasljeđivanja te su rezultati i u tome smislu lošiji. Dok dobivena anotacija predstavlja objekt na slici, sama detekcija objekta bi vratila objekt koji je u hijerarhiji nadređen predstavljenom objektu, odnosno kod detekcije objekata nije dobivena precizna semantika. Prosječna sigurnost modela za opise slika je 0.743, za nađene objekte 0.702, a za naznačene anotacije 0.866. Vidljivo je da i sam alat najsigurniji u rezultate anotacija, a najmanje siguran za rezultate pronađenih objekata.



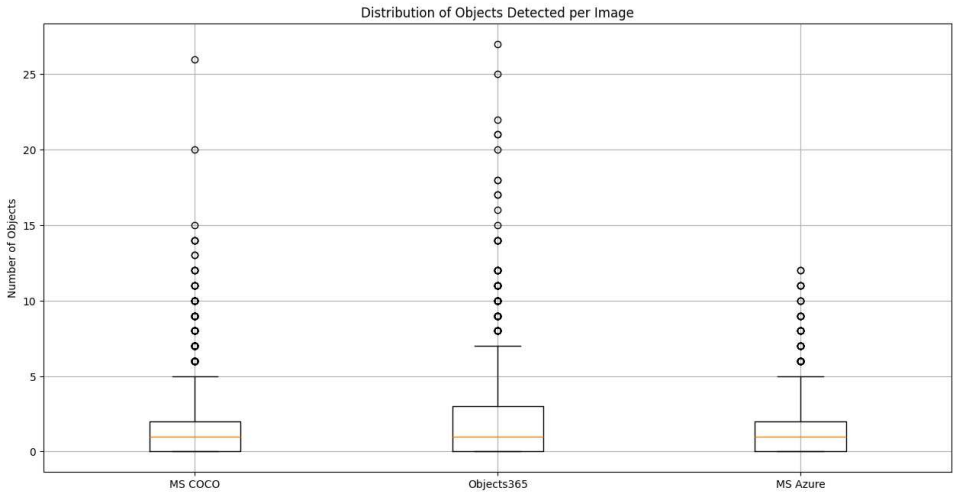
Sl. 10.1 Histogram broja riječi u opisima slika



Sl. 10.2 Histogram broja dobivenih anotacija na slikama

11. Usporedba najboljih modela

Kao što je bilo moguće vidjet Open Images set nema dobru anotaciju te modeli trenirani na njemu ne daju dobre rezultate. Kada gledamo modele trenirane na MS COCO setu podataka, vidimo da je noviji YOLO v5 dao nešto bolje rezultate u odnosu na YOLO v3. Odlične rezultate dobili smo i od modela treniranog na Objects365 setu. Na grafu (Sl. 11.1) je napravljena usporedba dva YOLO v5 modela te MS Azure AI Visiona. Azureov model je našao barem jedan objekt za najviše slika od svih navedenih modela, međutim vidljivo da su YOLO v5 modeli našli više objekata na slikama. Kod odabira koji ćemo od ovih modela koristiti važno je uzeti u obzir da Azure AI Vision osim objekata vraća i opise slike te anotacije, ali se isto tako i plaća.



Sl. 11.1 Usporedba distribucija nađenih objekata na redom model YOLO v5 treniran na COCO set, model YOLO v5 treniran na Objects365 setu te model alata Azure AI Vision

Zaključak

Baze afektivne multimedije su baze čiji sadržaj je okupljen u svrhu izazivanja emocija. Koriste se u područjima medicine, psihologije te neuroznanosti za proučavanje utjecaja multimedija na ljude. Svoju upotrebu su našle i u svijetu računalne znanosti za modele automatiziranog predviđanja emocija. Uz multimediju imaju spremljenu i semantiku te očekivane emocije, stvari koje se zapisuju ručno. Nencki Affective Picture System, NAPS, je baza od 1356 slika podijeljenih u 5 kategorija. Proučili smo neke od modela dubokog učenja, modele koji koriste duboke neuronske, za automatizirano zapisivanje semantika slika u bazu. Točnije proučili smo kako YOLO, You Only Look Once, algoritam pronalazi objekte na slici. Za to smo koristili verzije 3 i 5, te MS COCO, Open Images i Objects365 setove podataka za treniranje. Proučili smo i kako gotov alat Microsoftov Azure AI Vision semantički opsuje slike. Azure alat je našao objekte na najviše slika, dok je model treniran na Objects365 setu našao najviše objekata na slikama. Ovakvi modeli i alati bi mogli unaprijediti proces stvaranja i održavanja baza afektivne medije, ali da bi u potpunosti zamijenili ljudsko opisivanje semantike trebalo bi unaprijediti i setove na kojima se treniraju. Trenutno dostupni setovi za treniranje, od slika koje izazivaju negativne emocije, sadrže samo stvari kojih se određeni ljudi boje, kao što su zmije, ali ne sadrže puno slika ozljeda, nesreća ili katastrofa.

Literatura

[1] M. Horvat, *A brief Overview of Affective Multimedia Databases*, Varaždin, 2017.

[2] A. Marchewka, Ł. Żurawski, K. Jednoróg i A. Grabowska, *The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database*, Springer, 2013.

[3] *Computer vision system: challenges, benefits, use cases*, N-iX, (2022, travanj). Poveznica: https://www.n-ix.com/computer-vision/; pristupljeno 23. ožujka 2024.

[4] Ambika, *What is Computer Vision?(History, Applications, Challenges)*, Medium, (2023, Rujan). Poveznica: https://medium.com/@ambika199820/what-is-computer-vision-history-applications-challenges-13f5759b48a5; pristupljeno 24. ožujka 2024.

[5] R. Pramoditha, *The Concept of Artificial Neurons (Perceptrons) in Neural Networks*, Towards Data Science, (2021, Prosinac). Poveznica: https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc; pristupljeno 25. svibnja 2024.

[6] Ž. Ujević Andrijić i Bolf (ur.), *Osvježimo znanje: Umjetne neuronske mreže*, Kemija u industriji: Časopis kemičara i kemijskih inženjera Hrvatske (2019), str. 219-220

[7] *Deep learning*, Wikipedia, (2024, lipanj). Poveznica: https://en.wikipedia.org/wiki/Deep_learning; pristupljeno 13. travnja 2024.

[8] *Neural Networks,* Standford. Poveznica: https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/Architecture/feedforward.html; pristupljeno 15. travnja 2024.

[9] J. Roell, *Understanding Recurrent Neural Networks: The Preferred Neural Network for Time-Series Data*, Towards Data Science, (2017, lipanj). Poveznica: https://towardsdatascience.com/understanding-recurrent-neural-networks-the-prefered-neural-network-for-time-series-data-7d856c21b759; pristupljeno 3. lipnja 2024.

[10] *Konvolucijske neuronske mreže*, Wikipedia, (2024, Siječanj). Poveznica: https://sr.wikipedia.org/sr-el/Konvolucijske_neuronske_mre%C5%BEe; pristupljeno 18. travnja 2024.

[11] S. Saha, *A Comprehensive Guide to Convolutional Neural Networks – the ELI5 way*, Towards Data Science, (2018. Prosinac). Poveznica: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53; pristupljeno 26. svibnja 2024.

[12] L. Pauly, *Deeper Networks for Pavement Crack Detection*, Research Gate, (2017, Srpanj). Poveznica: https://www.researchgate.net/figure/ReLU-activation-function_fig3_319235847; pristupljeno 8. lipnja 2024.

[13] P. Singla, *Understanding Convolutional Neural Networks (CNNs)*, Medium, (2024, Siječanj). Poveznica: https://medium.com/@prathamsingla4619/understanding-convolutional-neural-networks-cnns-2a2d6d110529; pristupljeno 8. lipnja 2024.

[14] Z. Wang, *Deep Convolutional Neural Networks for Image Classification*, Towards Data Science, (2017, Lipanj). Poveznica: https://www.researchgate.net/figure/Average-versus-max-pooling_fig1_317496930; pristupljeno 9. lipnja 2024.

[15] D. Radečić, *Softmax Activation Function Explained*, Towards Data Science, (2020, lipanj). Poveznica: https://towardsdatascience.com/softmax-activation-function-explained-a7e1bc3ad60; pristupljeno 9. lipnja 2024.

[16] Great Learning, *Everything you need to know about VGG16*, Medium, (2021, rujan). Poveznica: https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918; pristupljeno 16. travnja 2024.

[17] *Understanding the Inception Module in Deep Learning*, DeepAI. Poveznica: https://deepai.org/machine-learning-glossary-and-terms/inception-module; pristupljeno 16. travnja 2024.

[18] J. Shaikh, *Deep Learning in the Trenches: Understanding Inception Network from Scratch*, Analytics Vidhya, (2023, studeni). Poveznica: https://www.analyticsvidhya.com/blog/2018/10/understanding-inception-network-from-scratch/; 16. travnja 2024.

[19] A. Mirkhan, *YOLO Algorithm: Real-Time Object Detection from A to Z*, Kili, (2023)Poveznica: https://kili-technology.com/data-labeling/machine-learning/yolo-algorithm-real-time-object-detection-from-a-to-z#what-is-yolo; pristupljeno 19. travnja 2024.

[20] S. Singh, *Leveraging YOLO Object Detection for Accurate and Efficient Visual Recognition*, Labellerr, (2023, siječanj). Poveznica: https://www.labellerr.com/blog/why-is-the-yolo-algorithm-important/; pristupljeno 21. travnja 2024.

[21] *PyTorch*, Wikipedia, (2024, svibanj).Poveznica: https://en.wikipedia.org/wiki/PyTorch; pristupljeno 29. svibnja 2024.

[22] MongoDB, (2024). Poveznica: https://www.mongodb.com/; pristupljeno 30. ožujka 2024.

[23] OpenCV, (2024). Poveznica: https://opencv.org/; pristupljeno 11. travnja 2024.

[24] COCO dataset, (2024). Poveznica: https://cocodataset.org/#home; pristupljeno 21. svibnja 2024.

[25] *MS COCO (Microsoft Common Objects in Context)*, Papers with Code, (2024). Poveznica: https://paperswithcode.com/dataset/coco; pristupljeno 11. travnja 2024.

[26] *Open Images Dataset V7*, Google Apis, (2022). Poveznica: https://storage.googleapis.com/openimages/web/factsfigures_v7.html; pristupljeno 16. travnja 2024.

[27] kamal_DS, *Mastering Object Detection with YOLOv3 and COCO dataset*, Medium, (2023, Ožujak). Poveznica: https://korlakuntasaikamal10.medium.com/mastering-object-detection-with-yolov3-and-coco-dataset-1a3158fef9ea; pristupljeno 11. travnja 2024.

[28] A. Rosebrock, *Deep learning: How OpenCV's blobFromImage works*, PyImageSearch, (2017, studeni). Poveznica: https://pyimagesearch.com/2017/11/06/deep-learning-opencvs-blobfromimage-works/; pristupljeno 13. lipnja 2024.

[29] S. Shao, Z. Li, T. Zhang, C. Peng, G. Yu, X. Zhang, J. Li i J, Sun, *Objects365: A Large-scale, High-quality Dataset for Object Detection*, IEEE Xplore, 2019.

[30] *Objects365 Dataset*, Ultralytics Docs, (2024, lipanj). Poveznica: https://docs.ultralytics.com/datasets/detect/objects365/; pristupljeno 27. svibnja 2024.

[31] *Azure AI Vision*, Microsoft. Poveznica: https://azure.microsoft.com/en-us/products/ai-services/ai-vision; pristupljeno 10. ožujka 2024.

[32] *Image tagging*, Microsoft, (2024, siječanj). Poveznica: https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/concept-tagging-images; pristupljeno 28. ožujka 2024.

Sažetak

Tema rada je izrada sustava za semantičko unapređenje baza afektivne multimedije korištenjem modela dubokog učenja. Baze afektivne multimedije su napravljene u srhu izazivanja emocija. Jedna od takvih baza je Nencki Affective Picture System, NAPS, koja sadrži 1356 slika raspoređeni u 5 kategorija. Baze uz multimediju imaju zapisanu semantiku te očekivane emocije koji se ručno zapisuju. Proučili smo neke od modela dubokog učenja za automatizirano zapisivanje semantike. Proučili smo kakve rezultate YOLO, You Only Look Once, algoritam daje u ovisnosti o verziji algoritma i setu na kojem je treniran. Vidjeli smo da odabrani set jako utječe na rezultate, dok novije verzije daju nešto bolje rezultate od starijih. Proučili smo i kakve rezultate daje gotov alat Azure AI Vision. Na kraju smo zaključili da bi ovakvi modeli mogli olakšat održavanje baza afektivne medije.

Ključne riječi: afektivna multimedija, baze afektivne multimedije, računalni vid, detekcija objekata, duboko učenje, umjetni neuroni, umjetne neuronske mreže, konvolucijske neuronske mreže, YOLO

Summary

The topic of the paper is the creation of a system for semantic improvement of affective multimedia databases using deep learning models. Bases of affective multimedia are made with the aim of evoking emotions. One such database is the Nencki Affective Picture System, NAPS, which contains 1356 pictures arranged in 5 categories. Databases with multimedia have menually written semantics and expected emotions. We have explored some of the deep learning models for automated semantic writing. We studied what kind of results the YOLO, You Only Look Once, algorithm gives depending on the version of the algorithm and the set on which it was trained. We have seen that the selected set greatly affects the results, while the newer versions give slightly better results than the older ones. We also studied the results of the ready-made tool Azure AI Vision. In the end, we concluded that such models could facilitate the maintenance of databases of affective media.

Keywords: affective multimedia, affective multimedia databases, computer vision, object detection, deep learning, artificial neurons, artificial neural networks, convolutional neural networks, YOLO

Skraćenice

NAPS	*Necki Affective Picture System*	Nencki sustav afektivnih fotografija
IAPS	*Internation Affective Picture System*	međunarodni sustav afektivnih fotografija
SIFT	*Scale-invariant feature transform*	transformacija značajki nepromjenjivog mjerila
GMM	*Gaussian Mixture Model*	Gaussovi modeli mješavina
GPU	*Graphics processing unit*	grafički procesor
TPU	*Tensor processing unit*	tenzorski procesor
ANN	*Artificial Neural Network*	umjetna neuronska mreža
LSTM	*Long Short-Term Memory*	dugo kratkoročno pamćenje
RNN	*Recurrent Neural Network*	mreže s povratnom vezom
CNN	*Convolutional Neural Network*	konvolucijska neuronska mreža
ReLU	*Rectified Linear Unit*	ispravljena linearna funkcija
VGG	*Visual Geometry Group*	
YOLO	*You Only Look Once*	jednostupanjski detektor objekata
MS	*Microsoft*	
COCO	*Common Objects in Context*	
JSON	*JavaScript Object Notation*	JavaScript zapis objekata
AI	*Artificial Intelligence*	umjetna inteligencija