

MODELI DUBOKOG UČENJA ZA KLASIFIKACIJU SLIKA PRIMENOM KONVOLUCIONIH NEURONSKIH MREŽA

Andreja Samčović, Nikola Matijašević, Marko Đogatović
Univerzitet u Beogradu - Saobraćajni fakultet
andrej@sf.bg.ac.rs, m.djogatovic@sf.bg.ac.rs, nikola.matijasevic@sf.bg.ac.rs

Rezime: *Obuka računarskog vida se realizuje korišćenjem dubokog učenja. Duboko učenje se može primeniti za rešavanje raznih problema klasifikacije. Za ovaj tip problema, kao najbolje rešenje pokazalo se korišćenje konvolucionih neuronskih mreža. One su napravile revoluciju u oblasti prepoznavanja slika, pružajući visoku preciznost koja je ranije bila nedostižna. Konkretno, neki od relevantnih modela konvolucionih neuronskih mreža su ResNet50, VGG16 i VGG19. U ovom radu je objašnjena generalna struktura i princip rada konvolucionih neuronskih mreža kroz njihove slojeve. Takođe, dat je i prikaz arhitekture i performansi pomenutih modela koji se koriste u zadacima klasifikacije slika. Takođe, objašnjeni su koncepti kao što su prenosno učenje i augmentacija podataka.*

Gljučne reči: *duboko učenje, konvolucione neuronske mreže, ResNet50, VGG16, VGG19.*

1. Uvod

Čovek se u svom životu oslanja na nekoliko osnovnih ljudskih čula, među kojima se nalazi i čulo vida. Ovo čulo se koristi radi upoznavanja i analize okruženja kroz slanje slika ka mozgu na obradu, što ga čini značajnim u odvijanju svakodnevnih životnih i radnih aktivnosti čoveka. Ovaj proces je bolje poznat kao ljudski vid (vizija). Analogno tome, računarska vizija (*computer vision*) predstavlja korišćenje senzora za prikupljanje informacija u obliku slike i slanje ka računaru na dalju obradu. Time se omogućava računarima da ekstrahuju značajne informacije iz digitalnih zapisa (slike, video snimci i sl.) i da na osnovu njih preduzmu određene aktivnosti ili da pruže preporuke. Kako bi se to postiglo, računari koriste kamere, velike skupove podataka, razne zahtevne algoritme i sl. To je proces koji se naziva obukom računarskog vida za rešavanje raznih problema.

Obuka se realizuje kroz algoritme mašinskog učenja nad skupovima podataka sa ciljem imitacije načina na koji ljudi uče i postepenog poboljšavanja preciznosti. Izbor algoritama zavisi od tipa podataka i problema koji je potrebno rešiti. Nakon obuke, ideja je da se dođe do predikcija nad novim skupom podataka sličnog tipa. Polje mašinskog učenja koje je veoma aktuelno se odnosi na duboko učenje. Razlog za to jesu zavidni

rezultati [1][2] koji su dobijeni pri rešavanju problema kao što su prepoznavanje raznih objekata poput lica, oboljenja, ljudskih emocija, itd. Svi ovi problemi se svode na posmatranje slika, detekciju karakterističnih osobina i klasifikaciju. Glavni razlog za razmatranje ovih problema jeste taj što ljudska procena ne može biti pouzdana u potpunosti. Zato se prilazi sa ciljem da računarska vizija postane superiornija u odnosu na ljudski vid tj. da mašina može klasifikovati slike sa većom preciznošću nego što bi to mogao da uradi čovek. U okviru dubokog učenja koriste se konvolucione neuronske mreže koje su pogodne za rešavanje ovog tipa problema. U ovom radu, fokus će biti na predstavljanju tri postojeća modela konvolucionih neuronskih mreža.

Ostatak rada je organizovan na sledeći način. U drugom poglavlju, objašnjavaju se duboke neuronske mreže, gde je fokus na konvolucionim neuronskim mrežama, njenoj strukturi i principu rada, ali i podacima koji se koriste za njihovu obuku. Cilj trećeg poglavlja je prikaz izabranih modela konvolucionih neuronskih mreža i njihovih karakteristika. Četvrto poglavlje govori o konceptima prenosnog učenja i augmentacije podataka. Poslednje poglavlje daje zaključna razmatranja rada.

2. Konvolucione neuronske mreže

Duboko učenje (*deep learning*), kao podoblast mašinskog učenja, se konstantno menja. To se najviše odnosi na modele, funkcije i algoritme koji se koriste u ovoj podoblasti. Duboko učenje se posmatra kroz jedan od tri glavnih tipova [3]: Višeslojni perceptron (*Multi-Layer Perceptron*, MLP), Konvolucione neuronske mreže (*Convolutional Neural Networks*, CNN), kao i Rekurentne neuronske mreže (*Recurrent Neural Networks*, RNN).

Konvolucione neuronske mreže, kao glavni fokus ovog rada, su inspirisane biološkim nervnim sistemom (npr. ljudski mozak). One se sastoje od različitih slojeva, gde svaki sloj ima određeni broj čvorova (neurona). Čvorovi su međusobno povezani tako da izlaz nekog čvora iz sloja predstavlja ulaz čvora u narednom sloju. Jedan od najranijih radova koji se odnosi na korišćenje konvolucionih neuronskih mreža objavljen je 1998. godine [4]. U tom radu, predstavljen je jednostavan CNN model (*LeNet-5*) koji omogućava ekstrahovanje jednostavnijih karakterističnih osobina podataka i njihovu agregaciju čime se dobijaju složenije karakteristične osobine podataka. Primena ovog CNN modela je ograničena na rešavanje problema prepoznavanja ljudskog rukopisa tj. ručno napisanih cifara sa slike. To se postiže kroz obuku modela tako što mu se proslede slike cifara iz kojih on može da ekstrahuje karakteristične osobine. Od modela se očekuje tačna predikcija cifre koja je prikazana na slici koja se ne nalazi u skupu za obuku. Kako bi se sve ovo realizovalo, potrebno je posedovati hardver visokih specifikacija jer je obuka modela veoma računarski zahtevna. Kao glavni računarski resurs koji se ističe za obuku modela jesu grafičke kartice. One pružaju mogućnost paralelnog izvršavanja više operacija zbog velikog broja jezgara za obradu (*Graphical Processing Unit*, GPU) od kojih se sastoje.

2.1. Podaci za obuku

Glavni preduslov koji je potrebno zadovoljiti kako bi se koristili CNN modeli jeste postojanje odgovarajućih baza podataka. Ovo ne treba zanemariti jer one igraju

ključnu ulogu u obuci svakog CNN modela. Neke od korišćenijih i poznatijih baza podataka će biti objašnjene u nastavku.

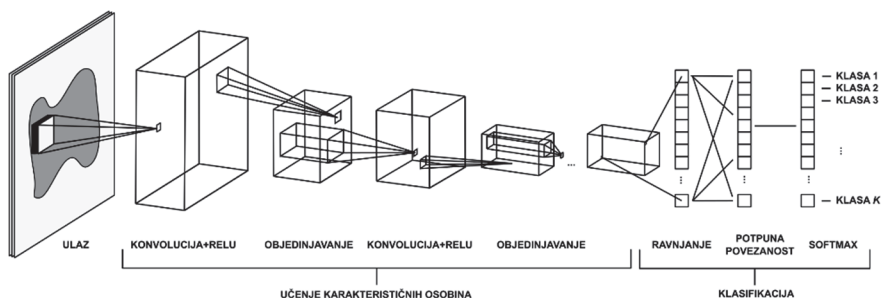
Proces obuke prethodno spomenutog *LeNet-5* modela je sproveden pomoću MNIST (*Modified National Institute of Standards and Technology*) baze podataka [5]. Ova baza je objavljena još 1998. godine i sastoji se od slika na kojima se nalaze ručno napisani jednocifreni brojevi. Te slike su uparene sa odgovarajućom oznakom (labelom) kojih ima ukupno 10 (od 0 do 9). Ova baza podataka se deli na skup za obuku i skup za testiranje. Prvi skup se sastoji od 60 hiljada slika, dok drugi skup ima 10 hiljada slika. Brojevi na slikama su centrirani. Slike su binarne (crno-bele) i njihova visina i širina iznosi 28 piksela.

Sledeća baza podataka koja će biti predstavljena jeste CIFAR (*Canadian Institute For Advanced Research*) objavljena 2009. godine [6]. Ova baza podataka dolazi sa 10 (CIFAR-10) ili 100 (CIFAR-100) klasa. CIFAR-10 se sastoji od 60 hiljada slika gde se u svakoj klasi nalazi 6 hiljada slika. U skupu za obuku se koristi 50000 slika, dok se u skupu za testiranje nalazi 10 hiljada slika. Slike su u RGB (*Red Green Blue*) formatu i njihova visina i širina iznosi 32 piksela. Klase predstavljaju slike određenih predmeta i životinja. Sa druge strane, CIFAR-100 je ista kao i CIFAR-10 osim što se sastoji od 100 klasa gde svaka ima 600 slika. Skupovi za obuku i testiranje sadrže 500 i 100 slika po svakoj klasi, respektivno. Svih 100 klasa je grupisano u 20 super klasa. Super klase sačinjavaju slike klasa iste vrste (istog tipa).

Poslednja baza podataka je *ImageNet* koja je objavljena 2009. godine od strane Univerziteta u Princetonu [7]. Ova baza podataka se konstantno ažurira. Sastoji se od preko 20 hiljada klasa i od preko 14 miliona slika. Međutim, najčešće se koristi skraćena verzija ove baze podataka koja se sastoji od hiljadu klasa i preko 1,2 miliona slika za obuku, 50 hiljada slika za validaciju i 100 hiljada slika za testiranje. Skraćena verzija se naziva *ImageNet-1K*. Slike su u RGB formatu i variraju po pitanju visine i širine. Za neke klase, dimenzije najveće slike su 4.288 i 2.848 piksela, dok su dimenzije najmanje slike 75 i 56 piksela. Slike se obično svode na iste pre njihovog učitavanja u model.

2.2. Struktura i princip rada

CNN se sastoji od dela za izdvajanje karakterističnih osobina (učenje svojstava) ulaznih podataka i od dela za određivanje odgovarajuće klase kojoj ulazni podaci pripadaju (rešavanje problema klasifikacije). Zajedničko im je to što se grade od posebnih elemenata koji se nazivaju slojevima. Kada je u pitanju problem klasifikacije sa više klasa, uopštena struktura CNN modela je data Slikom 1.



Slika 1. Struktura modela konvolucione neuronske mreže

Kroz primer klasifikacije slike može se objasniti struktura CNN modela. Na ulaz u CNN model se prosledi slika. Nakon toga, nastupa učenje karakterističnih osobina ulaznih podataka koja obuhvata kombinaciju konvolucionih slojeva i slojeva za objedinjavanje (*pooling*). Jedna od najčešćih aktivacionih funkcija koja se koristi u konvolucionim slojevima je *Rectified Linear Unit* (ReLU) [8]. ReLU funkcija je opisana sledećom matematičkom formulacijom:

$$\text{relu}(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (1)$$

Ova funkcija vraća vrednost 0 ukoliko joj je prosledena negativna vrednost, dok u svim ostalim slučajevima vraća onu vrednost koja joj je prosledena. Sloj za objedinjavanje omogućava spajanje nekoliko vrednosti u jednu (najčešće najveću). Deo CNN modela koji se odnosi na učenje karakterističnih osobina se može sastojati od više konvolucionih slojeva i slojeva za objedinjavanje. Od različitih načina kombinovanja ovih slojeva zavisi koliko dobro će model moći da nauči podatke. Druga celina CNN modela se odnosi na klasifikaciju i sastoji se od sloja za ravnjanje (*flatten*), potpuno povezanih slojeva i odgovarajuće aktivacione funkcije. Kada je u pitanju rešavanje problema kod kojeg se javlja više od 2 klase koristi se *softmax* aktivaciona funkcija [9]. Suština ove aktivacione funkcije jeste da joj se prosleđuje vektor čije se vrednosti transformišu na interval od 0 do 1. Kako je suma transformisanih vrednosti vektora jednaka 1, te vrednosti se mogu tumačiti kao verovatnoće pripadnosti klasama. Najveća dobijena verovatnoća predstavlja predikciju CNN modela da slika pripada klasi kojoj odgovara ta verovatnoća. Matematička formulacija vektorske *softmax* aktivacione funkcije je data jednačinom:

$$\text{softmax}(\mathbf{x}) = \frac{1}{\sum_{j=1}^K e^{x_j}} [e^{x_1} \quad \dots \quad e^{x_K}]^T, \quad (2)$$

gde je \mathbf{x} vektor definisan kao $\mathbf{x} = [x_1 \quad \dots \quad x_K]$, pri čemu je K broj klasa.

Slojevi se dele na više grupa. U ovom radu, fokus je na slojevima koji su korišćeni u izradi aktuelnih CNN modela, a pripadaju grupi [10]: glavnih slojeva, konvolucionih slojeva, slojeva za objedinjavanje, slojeva za regularizaciju i slojeva za preoblikovanje.

Iz grupe glavnih slojeva, korišćen je potpuno povezan (*dense*) sloj. On se koristi za povezivanje neurona između dva različita sloja. Izlaz jednog neurona iz prethodnog sloja predstavlja ulaz svih neurona iz narednog sloja tj. postoji veza između svakog para neurona. Slojevi ovog tipa se najčešće mogu pronaći na kraju kao poslednji slojevi CNN modela. Zbog toga, oni se mogu posmatrati i kao izlazni slojevi. Kod njih, kada se javlja više od 2 klase, koristi se *softmax* aktivaciona funkcija. U suprotnom, kod binarne klasifikacije koristi se *sigmoidna* funkcija (logistička kriva) [9]. Karakteristika izlaznog sloja jeste da ima onoliko neurona koliko problem koji se rešava ima klasa. Izuzevši izlazni sloj modela, skup više *dense* slojeva se naziva skrivenim slojem. Uvod skrivenog sloja omogućava jednostavniji i pregledniji grafički prikaz strukture nekog CNN modela.

Grupa konvolucionih slojeva se odnosi na slojeve koji se koriste za ekstrahovanje karakterističnih osobina ulaznih podataka. Ovi slojevi su najbitniji u CNN modelu i koriste se u kombinaciji sa slojevima za objedinjavanje. Konvolucija predstavlja

korišćenje receptivnih polja (filtera) određenih dimenzija, sa željenim korakom, kako bi se izvršila detaljnija analiza ulaznih podataka tj. ekstrakovanje karakterističnih osobina. Takođe, ovim postupkom se redukuje količina podataka koja nastavlja ka narednom sloju CNN modela. Nakon konvolucionih slojeva primenjuju se slojevi za objedinjavanje. Ovi slojevi se koriste za dalje redukovanje količine podataka. Postoje razni načini za objedinjavanje, pri čemu je pristup maksimalnog objedinjavanja najkorišćeniji. On uglavnom koristi 2x2 filter koji prolazi dobijene podatke iz konvolucionih slojeva i vraća maksimalnu vrednost koju obuhvata filter. Tako da, kako filter 2x2 obuhvata ukupno 4 podatka po koraku, na njihovo mesto biće vraćen samo 1 podatak i to onaj koji ima najveću vrednost od njih 4. Drugi korišćeni načini obuhvataju pristup traženja prosečne vrednosti ili sume vrednosti podataka koje pokriva filter.

Manje bitne grupe su slojevi za regularizaciju i za preoblikovanje. Među slojevima za regularizaciju najpoznatiji je sloj ispuštanja (*dropout*). On poništava doprinos nekih slučajno izabranih neurona sledećem sloju. Obično se definiše verovatnoća p da će neuron biti isključen u epohi obuke, odnosno verovatnoća da neuron neće biti isključen iznosi $1 - p$. Slučajno isključivanje neurona se koristi kako bi se smanjila mogućnost pojave preterane obuke (*overfitting*) CNN modela. Preterana obuka dovodi do gubitka sposobnosti CNN modela da generalizuje podatke tj. da naučene koncepte primeni nad nepoznatim podacima. Najkorišćeniji sloj iz grupe za preoblikovanje je sloj za ravnjanje. On se koristi najčešće posle sloja za objedinjavanje. Izvršava konvertovanje svih rezultujućih dvodimenzionalnih nizova u jedan linearni vektor. Ukoliko se na ulaz u sloj za ravnjanje dovede matrica dimenzija 20x20 na izlazu će se dobiti vektor koji se sastoji od 400 elemenata. Ovakav vektor se dovodi kao ulaz u potpuno povezani sloj radi klasifikacije podatka.

3. CNN modeli

Usled brzog razvijanja ove oblasti kreiraju se novi CNN modeli koji se koriste. Takođe, pojavom novih koncepata modeli mogu brzo zastareti i biti izbačeni iz upotrebe. Zbog toga, poželjno je da modeli budu podložni konstantnim promenama i unapređenjima. Ovi modeli su korisni jer pružaju mogućnost obuke, validacije i testiranja na osnovu ulaznog skupa podataka za bilo koji problem. U ovom radu izabrana su tri modela, a to su ResNet50, VGG16 i VGG19, koji su etablirani kao pouzdani i biće objašnjeni u nastavku.

3.1. ResNet

Rezidualna mreža (*Residual Network*, ResNet) predstavlja CNN model koji je razvijen u okviru kompanije *Microsoft* [11]. ResNet modeli mogu varirati u odnosu na broj rezidualnih blokova od kojih se sastoje. Rezidualni blok je skup više slojeva u modelu. Kao najkorišćenija varijacija ResNet-a, javlja se ResNet50 model koji se sastoji od ukupno 50 slojeva (48 konvolucionih slojeva, 1 sloj maksimalnog objedinjavanja i 1 sloj prosečnog objedinjavanja). Međutim, bitno je napomenuti da je ResNet152 prvi model ove vrste koji je dobio veliku pažnju. Razlog za to jeste pobeda na ILSVRC takmičenju 2015. godine gde je ostvario impresivnu grešku pri klasifikaciji slika koja je iznosila 3,6% [12].

Razumevanje jedinice građe ResNet modela (rezidualnog bloka) je prilično jednostavno. Neuronske mreže sa propagacijom signala unapred funkcionišu na način da izlaz iz svakog sloja predstavlja ulaz u naredni sloj. U mreži sa rezidualnim blokovima, izlaz iz svakog sloja se šalje kao ulaz u naredni sloj kao i u slojeve udaljene za najčešće 2 ili 3 sloja (*skip* konekcije). Ovakav pristup je omogućio rešavanje postojećih problema nestajućeg gradijenta i problema degradacije [13][14]. Problem nestajućeg gradijenta se javlja kod modela neuronskih mreža sa velikim brojem slojeva. U tom slučaju, postoji i veliki broj težina koje je potrebno podešavati tako da se gubici modela minimizuju. To se postiže kroz proces koji se zove propagacija unazad. Svaku težinu modela je potrebno ažurirati, ali zbog velikog broja težina, njihove vrednosti se menjaju u jako maloj meri ili se uopšte i ne menjaju. Problem degradacije se odnosi na opadanje preciznosti modela kao posledica uvođenja dodatnih slojeva. Intuitivno, dodavanje novih slojeva u model bi trebalo da doprinese boljem razumevanju ulaznih podataka, pa samim tim i većoj preciznosti. Međutim, dodatni slojevi dovode do većih grešaka tokom obuke, pa posledično i tokom validacije i testiranja modela [11]. Nakon uvođenja *skip* konekcija kod ResNet-a, oba problema su pokrivena i omogućeno je korišćenje većeg broja slojeva bez posledica po preciznost modela. Arhitektura ResNet50 modela je prikazana Slikom 2.

3.2. VGG

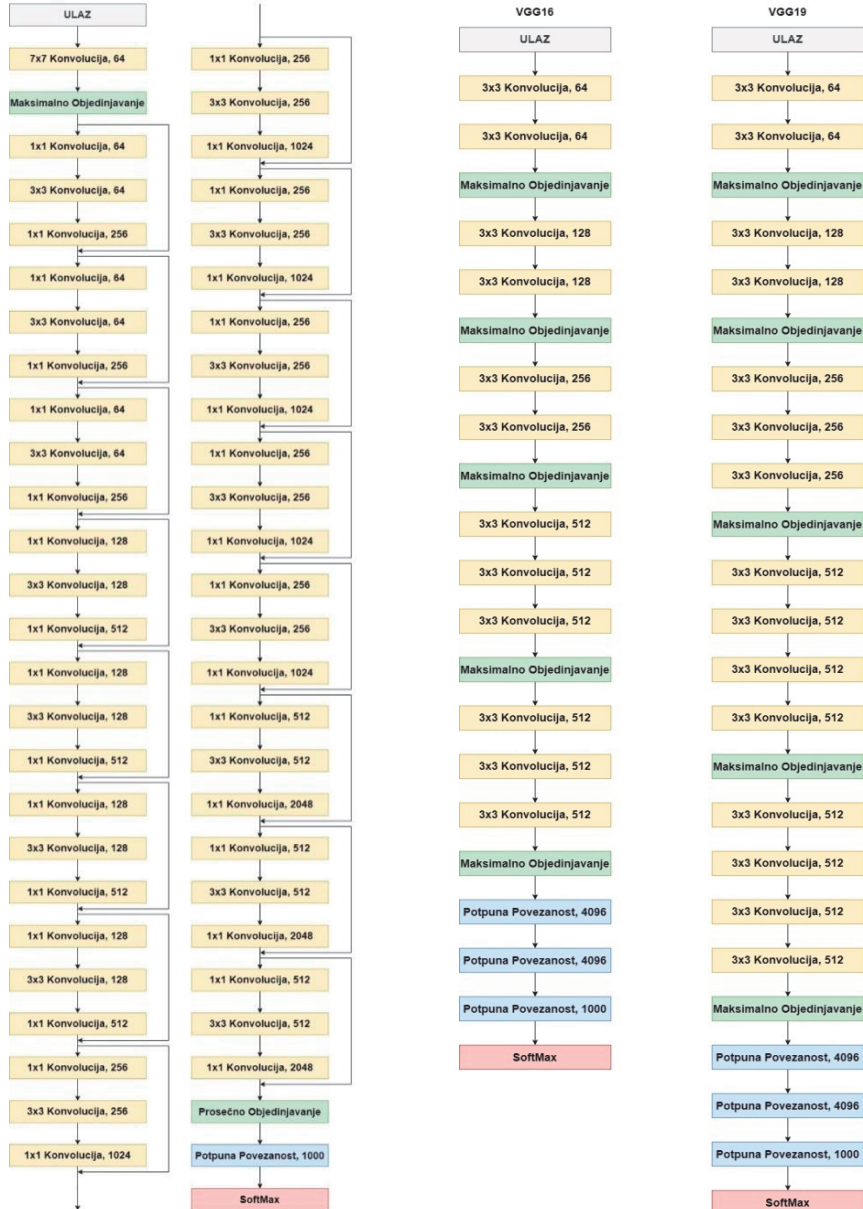
Nova grupa modela je razvijena od strane Vizuelne Geometrijske Grupe (*Visual Geometry Group*, VGG) sa Univerziteta u Oksfordu [15], po kojoj je i dobila ime. VGG modeli su predstavljeni na ILSVRC takmičenju 2014. godine, gde su završili na drugom mestu što je zavidan rezultat ako se uzme u obzir da su ostvarili grešku pri klasifikaciji od 7,3%, dok je na prvom mestu ta greška iznosila 6,7% [16]. Iako su zasnovani na *AlexNet* modelu [17], imaju nekoliko razlika koje ih izdvajaju od ostalih. Umesto da koriste velika receptivna polja kao što je slučaj kod *AlexNet*-a (11x11 sa korakom 4), oni koriste mala receptivna polja (3x3 sa korakom 1). Slaganjem više manjih polja dobijaju se bolji rezultati što je i najbitniji doprinos ovih modela [15]. Takođe, sastoje se od manje parametara u odnosu na *AlexNet* što rezultuje uštedom resursa pri obuci.

Glavna dva predstavnika su VGG16 i VGG19 modeli. Postoje dve varijacije VGG16 modela (C i D). Razlike između njih su minimalne, neka receptivna polja su veličine 1x1 i zbog toga je konačan broj parametara 134 i 138 miliona, respektivno. Brojevi 16 i 19 u imenima modela se odnose na broj ponderisanih slojeva. To znači da VGG19 sadrži tri ponderisana sloja više u odnosu na VGG16. U ponderisane slojeve spadaju konvolucionni slojevi i potpuno povezani slojevi. VGG16 se sastoji od 13 konvolucionih slojeva i 3 potpuno povezana sloja, a VGG19 od 16 konvolucionih slojeva i 3 potpuno povezana sloja. Arhitektura ova dva spomenuta CNN modela je prikazana na Slici 2. Moguće je napraviti i modele ove grupe koji se sastoje od još većeg broja slojeva sa težinama (VGG50, VGG60, itd.). Utvrđeno je da ovo nije dobro rešenje jer se u tim slučajevima, zbog jako velikog broja otežanih slojeva, javlja problem nestajućeg gradijenta koji je ResNet model pokrio.

4. Koncepti pripreme CNN modela i skupa slika

Bitna stavka u rešavanju problema klasifikacije jeste priprema jednog ili više CNN modela koji će se koristiti, ali naravno i priprema izabranog skupa slika koji

predstavlja ulaz u te modele. Kod pripreme CNN modela, fokus će biti na objašnjavanju ključnog koncepta koji se naziva prenosno učenje. Ovaj koncept omogućava bržu obuku CNN modela, što dalje pruža lakše eksperimentisanje sa modelom. Takođe, za pripremu izabranog skupa slika, potrebno je pregledati slike i razmotriti da li je taj skup adekvatan za korišćenje i da li je moguće izvršiti neko poboljšanje. Poboljšanja se postižu korišćenjem transformacija koje spadaju u okvire augmentacije slika.



Slika 2. Arhitektura ResNet50 modela (levo), VGG16 i VGG19 modela (desno)

4.1. Prenosno učenje

Koncept prenosnog učenja (*transfer learning*) je postao široko zastupljen u oblasti CNN-a. Glavna ideja koja stoji iza ovog koncepta jeste, kao i što ime kaže, da izvrši transfer (prenos) znanja nekog već obučenog CNN modela tj. da učita model koji je već obučen pomoću ogromnih baza podataka (najčešće se radi o skupu slika). Uglavnom su poznati CNN modeli obučeni korišćenjem više od milion slika iz *ImageNet* baze podataka. Unapred obučeni model može da klasifikuje slike u 1000 različitih klasa. Kao rezultat toga, model u startu ima prednost jer je naučio razne karakteristike širokog spektra slika (formirao znanje). Sledeći korak se odnosi na korišćenje tog znanja kao osnovu za dalju obuku i rešavanje nekog drugog problema klasifikacije. Bitno je napomenuti da koncept prenosnog učenja znatno skraćuje vreme obuke CNN modela jer nije potrebno obučavati model od početka, već samo za željeni skup slika. Takođe, ovaj koncept se obično koristi za zadatke u kojima se javljaju jako mali (ograničeni) skupovi slika sa kojima nije moguće izvršiti obuku modela od početka.

Realizacija CNN modela i koncepta prenosnog učenja se može postići kroz razne programske jezike, od kojih je najčešće korišćen programski jezik *Python*. U tu svrhu se koriste specijalizovane biblioteke od kojih su najznačajnije *Keras* i *PyTorch*. U okviru *Keras* biblioteke na raspolaganju je 38 CNN modela, dok je kod *PyTorch* biblioteke moguće koristiti njih 33. Za sve ove CNN modele iz obe biblioteke moguće je postaviti inicijalne težine kroz tri opcije. Prva opcija jeste da se inicijalno polje za težine ostavi prazno što bi značilo da model kreće sa obukom od početka. Međutim, preostale dve opcije su znatno interesantnije. Druga opcija podrazumeva učitavanje težina koje su formirane prilikom obuke modela nad *ImageNet* bazom podataka. To znači da su svi CNN modeli koji su dostupni ovim bibliotekama prethodno obučeni i spremni za dalju upotrebu. Na kraju, treća opcija se odnosi na učitavanje proizvoljnih težina u CNN model tj. težina koje su formirane po završetku obuke nekog CNN modela nad nekom bazom podataka. Dakle, drugi i treći slučaj ukazuju na jednostavnu primenu koncepta prenosnog učenja između CNN modela u okviru ove dve biblioteke.

Postoje dve strategije sa kojima je moguće obučavati CNN modele i primeniti koncept prenosnog učenja, a to su: ekstrahovanje karakteristika i fino podešavanje [18]. Prva strategija se obično primenjuje kada se radi o malom skupu slika, dok se druga strategija koristi za srednje i velike skupove slika. Ekstrahovanje karakteristika se vrši kroz zamrzavanje slojeva obučenog CNN modela kako bi se osiguralo očuvanje karakteristika koje nose tokom buduće obuke, a zatim se dodaju novi slojevi preko zamrznutih. Na ovaj način, dodati slojevi će tokom obuke naučiti da pretvore postojeće znanje (karakteristike iz zamrznutih slojeva) u predikciju nad novim skupom slika. Fino podešavanje se služi konceptom zamrzavanja prvih nekoliko slojeva obučenog CNN modela, dok su ostali slojevi podložni obuci i izmenama. Nakon nekog vremena, slojevi se odmrzavaju i obuka se sprovodi do kraja. Obuka se tokom celog ovog procesa finog podešavanja sprovodi sa niskom stopom učenja.

4.2. Augmentacija slika

Jedna od glavnih poteškoća koja se može pojaviti pri obuci CNN modela jeste pronalaženje adekvatne baze podataka. Pod adekvatnom bazom podataka misli se najpre na to da je ona dovoljno velika i da su podaci ispravni. Veliki broj baza podataka je

dostupan na sajtu *Kaggle*, mada za neke ozbiljnije radove i istraživanja potrebno je kontaktirati kreatora tih baza podataka. U zavisnosti od problema koji se rešava, baze podataka za obuku se mogu sastojati od slika, teksta, zvuka i sl. Vezano za problem klasifikacije slika, mora se koristiti baza podataka sa skupom slika koje su podeljene prema klasama. Nad ovakvim skupom slika, ili bilo kojim drugim, može se primeniti augmentacija podataka radi njegovog proširenja i poboljšanja. Augmentacija predstavlja izvršavanje različitih transformacija nad podacima. Može se koristiti u svrhu proširivanja skupa podataka tj. od jedne slike se može dobiti više sličnih (augmentovanih) slika. Dalje, to se preslikava i na ceo skup slika, odnosno primenom augmentacije podataka dobija se veći skup. Korisno je primeniti ovakav pristup jer dobijanje većeg skupa slika značajno može doprineti u obuci CNN modela i njegovoj preciznosti. Potrebno je voditi računa na to koliko i koje transformacije se izvršavaju nad slikama kako slike ne bi postale neupotrebljive, pa čak i štetne, pri obuci CNN modela. Neke transformacije nad slikama koje se mogu primeniti se odnose na: uveličavanje, promenu osvetljenosti, promenu kontrasta, isecanje, rotiranje, horizontalno i vertikalno okretanje, itd. Slučajnost kod ovih transformacija se javlja tokom obuke CNN modela. Ukoliko se koristi više transformacija u modelu onda neka transformacija može biti primenjena nad slikom, a neke ne. Ukoliko je skup dovoljno velik i sadrži već uređene slike, nema potrebe koristiti transformacije u velikoj meri.

5. Zaključak

Potrebno je sagledati realne mogućnosti primene veštačke inteligencije u cilju dolaženja do novih saznanja, unapređivanja svakodnevnog života, prevazilaženja ljudskih ograničenja, itd. U skladu sa tim, ovaj rad se bavio primenom dubokog učenja koje pripada oblasti mašinskog učenja. Duboko učenje zasnovano na konvolucionim neuronskim mrežama je izvršilo revoluciju u pogledu analize digitalnih zapisa. Glavni cilj rada jeste da približi i objasni rad konvolucionih neuronskih mreža, ali i modela kroz koje se one primenjuju. Modeli koji su bili u fokusu i koji su predstavljeni su ResNet50, VGG16 i VGG19. Ovi modeli su dobra osnova za razumevanje funkcionisanja konvolucionih neuronskih mreža i za proučavanje drugih modela. Kao dodatak tome, potrebno je posvetiti pažnju i konceptima koji se koriste radi pripreme modela i baza podataka kao što su prenosno učenje i augmentacija podataka. Buduća istraživanja u ovoj oblasti se mogu odnositi na pronalazak novih aktivacionih funkcija, optimizaciju parametara modela, inovacije u arhitekturi modela, stvaranje novih baza podataka, unapređivanje koncepta pripreme modela i baza podataka, itd.

Literatura

- [1] M. Bakator, and D. Radosav, "Deep Learning and Medical Diagnosis: A Review of Literature", *Multimodal Technologies and Interaction*, 2018.
- [2] G. Hu, et. al., "When Face Recognition Meets With Deep Learning: An Evaluation of Convolutional Neural Networks for Face Recognition", *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, 2015.
- [3] A. Shrestha, and A. Mahmood, "Review of Deep Learning Algorithms and Architectures", *IEEE Access*, vol. 7, pp. 53040-53065, 2019.

- [4] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
- [5] MNIST database. [Online]. Available at: <https://yann.lecun.com/exdb/mnist/>
- [6] CIFAR database. [Online]. Available at: <https://www.cs.toronto.edu/~kriz/cifar.html>
- [7] ImageNet database. [Online]. Available at: <https://www.image-net.org/>
- [8] F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)", 2018.
- [9] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning", 2018.
- [10] Keras Layers. [Online]. Available at: <https://keras.io/api/layers/>
- [11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.
- [12] ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015. [Online]. Available at: <https://image-net.org/challenges/LSVRC/2015/results>
- [13] A. Arohan, A. Koustav, and S. Abhishek, "A Review of Convolutional Neural Networks", *International Conference on Emerging Trends in Information Technology and Engineering*, 2020.
- [14] S. Genevieve, and M. Wasfy, "An Overview of Recent Convolutional Neural Network Algorithms for Image Recognition", *IEEE Xplore*, 2018.
- [15] K. Simonyan, and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", *International Conference on Learning Representations*, 2015.
- [16] ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014. [Online]. Available at: <https://image-net.org/challenges/LSVRC/2014/results>
- [17] R. Kaur, and R. Kumar, "M. Gupta, Review on Transfer Learning for Convolutional Neural Network", *International Conference on Advances in Computing, Communication Control and Networking*, pp. 922-926, 2021.
- [18] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", *Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.

Abstract: *Computer vision training is implemented using deep learning. Deep learning can be applied to solve a variety of classification problems. For this type of problems, the use of convolutional neural networks is proved to be the best solution. They have revolutionized the field of image recognition, providing high precision previously unattainable. Specifically, some of the relevant convolutional neural network models are ResNet50, VGG16, and VGG19. This paper explains the general structure and working principle of convolutional neural networks through their layers. Also, the architecture and performance of the mentioned models used in image classification tasks are presented. Concepts such as transfer learning and data augmentation are also explained.*

Keywords: *deep learning, convolutional neural networks, ResNet50, VGG16, VGG19.*

DEEP LEARNING MODELS FOR IMAGE CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS

Andreja Samčović, Nikola Matijašević, Marko Đogatović