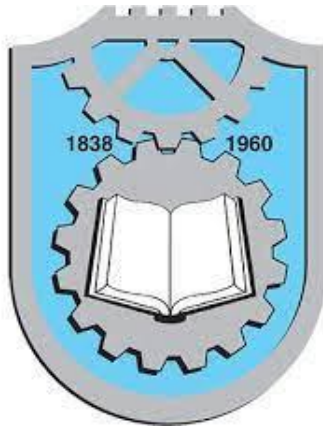


**Fakultet inženjerskih nauka
Univerzitet u Kragujevcu**



I domaći: Konvolucione neuronske mreže

- Neuronske mreže –

Student:

Ilija Todorović 313/2023

Predmetni nastavnik:

Prof. Dr. Vesna Ranković

Kragujevac, 2024

Rezime

U ovom radu biće teoretski predstavljen model konvulacionih neuronskih mreža. U sklopu toga biće objašnjen generalni opis arhitekture u koji spadaju: Ulazni sloj, sloj konvolucije, sloj sažimanja, sloj aktivacione funkcije i potpuno povezan sloj. Takođe biće predstavljen i pregled različitih pristupa i arhitektura.

Ključne reči: *Neuronske mreže, Konvulacione neuronske mreže, CNN, ReLu, Neuron.*

Sadržaj

Uvod.....	3
Generalni opis arhitekture.....	4
Pregled različitih pristupa arhitektura CNN.....	8
Zaključak.....	13
Literatura.....	14

1. Uvod

Razvoj veštačkih neuronskih mreža započeo je četrdesetih godina prošlog veka, kada su McCulloch i Pitts predstavili prvi model veštačkog neurona. Od tada, razvijeno je mnogo različitih modela i arhitektura, specijalizovanih za razne tipove zadataka. [1]

Konvolucione neuronske mreže (CNN - u daljem tekstu) su vrsta dubokih neuronskih mreža koje su našle primenu u najrazličitijim oblastima nauke i industrije. Preteča modernih CNN jeste Neokognitron, veštačka neuronska mreža dizajnirana za prepoznavanje obrazaca. Ideja njenog funkcionisanja potekla je iz istraživanja vizuelnog korteksa mozga mačaka koje su sprovedi naučnici Hubel i Wiesel 1962. godine. U istraživanju je dokazano da se određeni neuroni mozga pobuđuju isključivo kad su svetlosne trake postavljene u određenom pravcu, dok se promenom pravca pobuđuju drugi neuroni. To znači da se aktivacijom određenih neurona stvara jedna informacija o objektu posmatranja i daljim prenošenjem se stvara celokupna slika. Algoritam obučavanja sa propagacijom greške unazad za obučavanje CNN prvi put je upotrebljen 1989. godine sa ciljem da mreža raspozna rukom pisane cifre. Kako je vreme prolazilo razvijeni su razni modeli poput ResNet, VGGNet, GoogLeNet. CNN se koriste u poljoprivredi, medicini, stomatologiji itd. kao za unapređenje procesa zato što su veoma efikasno sredstvo kompjuterske vizije. [1]

Konvolucione neuronske mreže se sastoje iz sledećih slojeva:

- Ulazni sloj
- Sloj konvolucije
- Sloj sažimanja
- Sloj aktivacione funkcije(ReLU)
- Potpuno povezan sloj

Konvolucione neuronske mreže se u oblasti računarskog vida (vizija) najčešće primenjuju nad tenzorima čije su dimenzije [2]:

- Visina
- Širina
- Broj kanala boja (nijanse sive boje ili RGB tonovi)

Tenzor sa navedenim dimenzijama predstavlja sliku. Slike iz skupa podataka se mogu posmatrati kao instance ili primeri klase iz skupa podataka i kao takve predstavljaju primer za dalju diskusiju u okviru ovog domaćeg zadatka. Pored slika, podaci mogu biti govor, trodimenzionalne scene, snimci... [2]

2. Generalni opis arhitekture

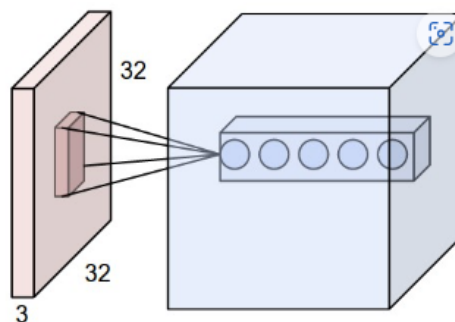
CNN se sastoji iz nekoliko slojeva:

Ulazni sloj

Ulazni sloj nije specifičan za CNN arhitekturu, odnosi se na bilo koju neuronsku mrežu kao mesto kroz koje podaci ulaze u mrežu. Kod CNN-a, ulazni podaci su najčešće slike, a koncept ulazne zapremine direktno je povezan s ovim slojem. [3]

Sloj konvolucije

Sloj konvolucije se smatra temeljem CNN-a. Ovaj sloj sadrži skup filtera koji se mogu trenirati. Svaki filter je prostorno mali, na primer tipičan filter u prvom sloju ConvNet-a može imati dimenzije $5 \times 5 \times 3$ (5 piksela u širinu i visinu i 3 jer slike imaju dubinu 3 - za RGB kanale). Tokom prolaska unapred, svaki filter vrši konvoluciju preko širine i visine ulaznog volumena (slike) i računa skalarni proizvod između vrednosti filtera i ulaza u toj poziciji. Kako se filter pomera po širini i visini ulaznog volumena, proizvede 2D mapu aktivacije koja prikazuje odgovore filtera na svakoj prostornoj poziciji. Intuitivno, mreža će naučiti filtere koji se aktiviraju kada prepoznaju određene vizuelne karakteristike, kao što su ivice određenih orijentacija ili boje u prvom sloju, ili složenije šablone u višim slojevima. U svakom konvulacionom sloju imamo skup filtera i svaki od njih proizvodi zasebnu 2D mapu aktivacije. Zatim ćemo ove mape aktivacije poredati duž dimenzije dubine i proizvesti izlazni volumen. [4]



slika 1 - Primer ulaznog volumena i prvog sloja [4]

Na slici 1 sa leve strane može se videti ulazni volumen tj. slika dimenzija $32 \times 32 \times 3$ i primer volumena neurona u prvom konvulacionom sloju. Svaki neuron povezan je samo lokalnim regionom u ulaznom volumenu prostorno, ali kroz celu dubinu. Postoji više neurona duž dubine i svi posmatraju isti region, ali imaju različite težine.

Dimenzije izlazne mape zavise od sledećih hiperparametara[5]:

Dubina - definiše broj filtera koji se koristi. Prednost veće dubine jeste to što svaki filter može tražiti svoj zasebni deo slike i tako zapaziti više uzoraka.

Skok (eng. stride) - definiše skok u broju piksela koji vrši filter nakon konvolucije. Na taj način skok veličine 1 pomera filter za jedno polje u desno (ili na dole, zavisno od pozicije).

Punjenje (eng. padding) - utiče na širinu i visinu izlaznog volumena. Najčešće se koristi zero-padding koji određene piksele s kraja popuni nulom. Izostanak punjenja smanjuje izlazni volumen zavisno o širini i visini samog filtera.

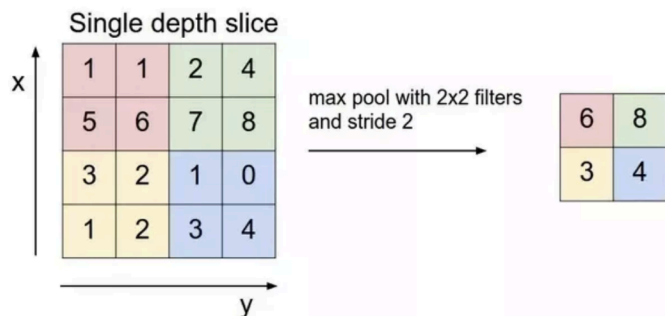
Dimenzije nove slike se računaju po formuli[5]:

$$D_{nova} = \frac{D - F + 2P}{S} + 1 \quad (1)$$

Gde je D stara dimenzija, F širina filtera, P broj piksela punjenja, S skok između dva filtera. Dubina izlaznog volumena jednaka je broju filtera koji se koriste.

Sloj sažimanja

Sloj sažimanja(eng. pooling layer) sadrži filter koji sažima dimenzije slike. Ovaj filter je drugačiji od onog u konvolucionom sloju s obzirom da ne sadrži težine. Njegova uloga je odabir vrednosti u okviru koji taj filter tih dimenzija obuhvata. Uglavnom se koristi Max Pooling operacija koja bira maksimalnu vrednos između odabranih piksela. Cilj ove operacije jeste očuvanje bitnijih karakteristika slike i njihovo dalje propagiranje u mreži. [5]



slika 2 - Primer sažimanja[5]

Kao i u sloju konvolucije i ovde se može koristiti skok(eng. stride).

U sloju sažimanja izlazne dimenzije se računaju po formuli[5]:

$$D_{nova} = \frac{D - F}{S} + 1 \quad (2)$$

Gde je D stara dimenzija, F širina filtera a S skok između dva odabira vrednosti.

Sloj aktivacione funkcije(ReLU)

Pravi izbor aktivacione funkcije može u značajnoj meri poboljšati performanse CNN koja se primenjuje u određenim zadacima. LeNet-5 arhitektura ne podrazumeva korišćenje aktivacione funkcije iza svakog sloja dok novije i savremenije arhitekture podrazumevaju takav pristup. Jedna od najčešće primenjivanih aktivacionih funkcija je ReLU (eng. Rectified Linear Unit), pa se zato u nekim literaturama može naći pojam ReLU sloj, što predstavlja sloj na kome se primenjuje ova funkcija. [3] Ova funkcija ima oblik [6]:

$$f(x) = \max(0, x) \quad (3)$$

Što znači da ako je ulazna vrednost pozitivna funkcija vraća ulaznu vrednost, a ako nije pozitivna ili je nula, funkcija vraća nulu.

Prednosti funkcije[6]:

- Jednostavnost računanja.
- Ublažava problem nestajanja gradijenta(s obzirom da je gradijent ove funkcije ili 0 ili 1 to sprečava da on postane premali).

Mane[6]:

- Mogućnost pojavljivanja mrtvih neurona.

Pored ReLU postoje i Leaky ReLU, Parametric ReLU, Randomized ReLU, Exponential Linear Unit (ELU) i druge [3]

Potpuno povezan sloj

Ovaj sloj se obično koristi u završnim slojevima mreže, ali to ne mora uvek da bude tako. Razlog korišćenja ovog sloja je taj što se dimenzije slike tokom propagacije nastoje smanjiti. Potpuna povezanost znači da su svi neuroni ovog sloja povezani sa svim izlazima prethodnog sloja. Nakon što konvolucioni i slojevi sažimanja izvuku značajne karakteristike iz slike, potpuno povezani sloj kombinuje sve te karakteristike i koristi ih za donošenje odluke o kojoj klasi pripada ulazni podatak. [5]

Izlazi iz CNN za klasifikaciju slika većinom predstavljaju klase kojima slika može pripadati i za aktivaciju se obično koristi SoftMax funkcija [5]:

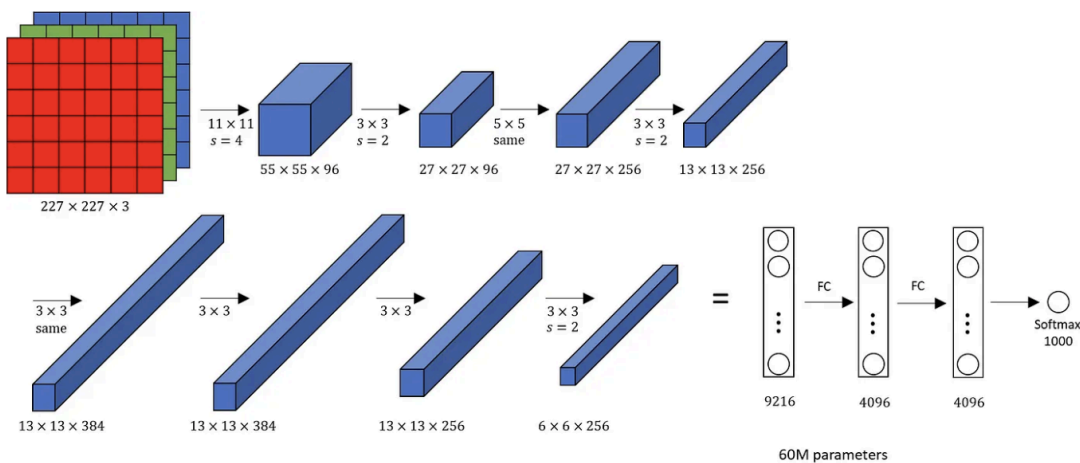
$$s(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}} \quad (4)$$

Ova funkcija normalizuje rezultate dobijene na izlazima transformišući ih da budu pogodni za klasifikaciju u rasponu od 0 do 1.

3. Pregled različitih pristupa i arhitektura CNN

AlexNet

Ovo je bila prva arhitektura koja je koristila GPU za ubrzanje treniranja. AlexNet se sastoji od 5 konvolucionih slojeva, 3 max-pooling sloja, 2 normalizovana sloja, 2 potpuno povezana sloja i 1 SoftMax sloja. Svaki konvolucionni sloj ima filter za konvoluciju i ne-linearnu ReLU funkciju aktivacije. Max-pooling slojevi smanjuju dimenzije, dok fiksna ulazna veličina (zbog potpuno povezanih slojeva) iznosi $227 \times 227 \times 3$, iako se često navodi kao $224 \times 224 \times 3$ zbog dodavanja paddinga. AlexNet sadrži preko 60 miliona parametara.[7]



slika 3 - AlexNet arhitektura [7]

Ključne karakteristike[7]:

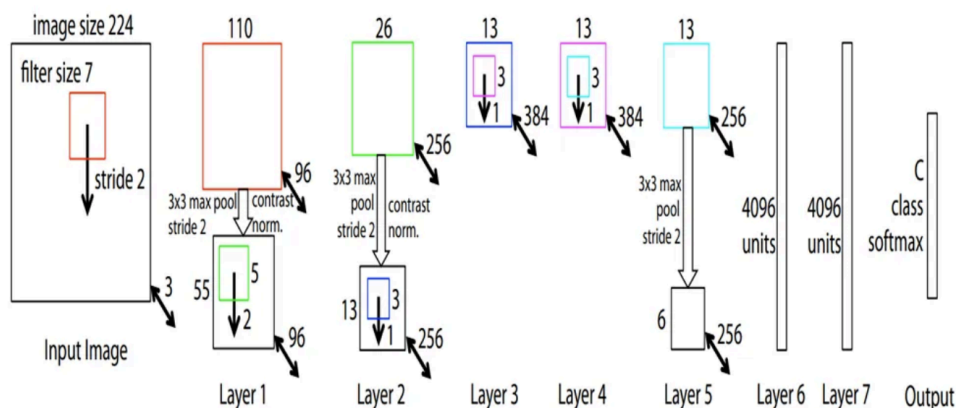
- *ReLU* je korišćen kao aktivaciona funkcija umesto *tanh*.
- Veličina serije podataka (batch) iznosi 128.
- Korišćen je algoritam *SGD Momentum* za optimizaciju učenja.

ZF Net

Rob Fergus i Matthew D. Zeiler predstavili su ZFNet, nazvan po njihovim prezimenima, Zeiler i Fergus. ZFNet je doneo blago poboljšanje u odnosu na AlexNet. On je omogućio vizualizaciju rada svakog sloja AlexNet mreže i prilagođavanje parametara za postizanje veće tačnosti. On ima istu osnovni strukturu kao AlexNet ali uvodi manje promene posebno u prvim slojevima. ZF Net koristi 7×7 filtere i 3×3 u kasnijim slojevima dok AlexNet koristi 11×11 , 5×5 i 3×3 . Razlikuju se i po skokovima u ulaznom sloju AlexNet “skače” za 4 dok ZF za 2. [8]

Ključne karakteristike ZFNet arhitekture[8]:

- **Konvolucioni slojevi:** Koriste se za izdvajanje važnih karakteristika primenom konvolucionih filtera.
- **MaxPooling slojevi:** Smanjuju prostorne dimenzije mape karakteristika pomoću agregacione funkcije (maksimum).
- **ReLU:** Nakon svakog konvolucionog sloja dodaje nelinearnost i osigurava pozitivne vrednosti na mapi karakteristika.
- **Potpuno povezani slojevi:** U završnim delovima mreže koriste se za izdvajanje obrazaca sa aktivacionom funkcijom ReLU.
- **SoftMax aktivacija:** U poslednjem sloju daje verovatnoće za 1000 klasa.
- **Dekonvolucioni slojevi:** Omogućavaju vizualizaciju naučenih karakteristika projekcijom aktivacija nazad u prostor piksela.



slika 4 - ZF Net arhitektura [9]

VGG Net

VGG predstavlja tipičan duboki CNN dizajn sa velikim brojem slojeva, pri čemu *VGG* označava *Visual Geometry Group*. "Duboka" arhitektura odnosi se na broj slojeva, gde VGG-16 i VGG-19 imaju 16 i 19 konvolucionih slojeva. VGGNet se koristi za izgradnju inovativnih modela za identifikaciju objekata i postiže izuzetne rezultate na raznim zadacima i skupovima podataka, ne samo na ImageNet-u. I dalje je jedna od najčešće korišćenih arhitektura za prepoznavanje slika. [10]

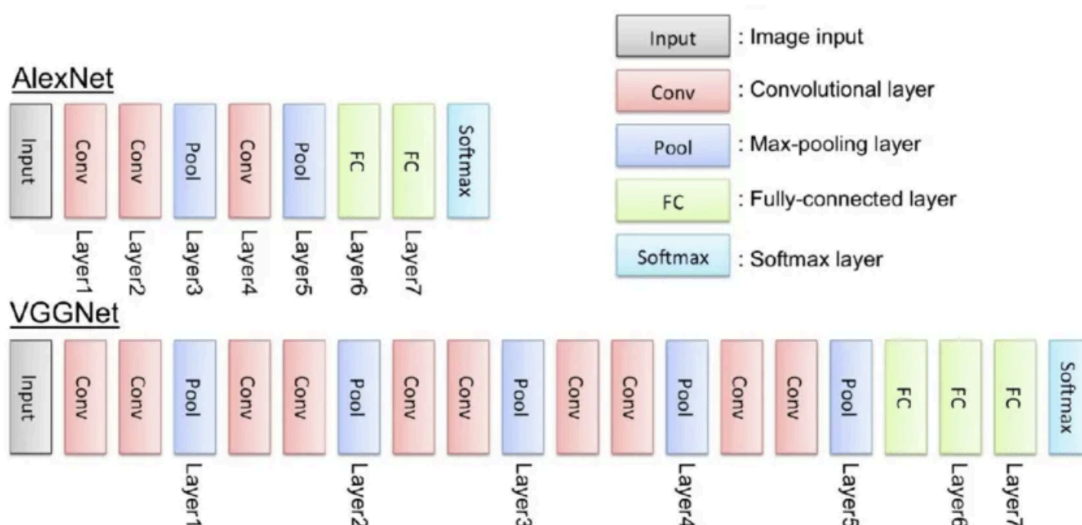
VGG mreže koriste veoma male konvolucione filtere. VGG-16 se sastoji od trinaest konvolucionih slojeva i tri potpuno povezana sloja.

Pogledajmo ukratko arhitekturu VGG mreže[10]:

- **Ulazi:** VGGNet prihvata slike dimenzija 224x224 piksela kao ulaz.
- **Konvolucionni slojevi:** U konvolucionim slojevima VGG koriste se najmanja moguća receptivna polja, veličine 3x3, kako bi se zabeležili pokreti s leva na desno i od gore prema dolje. Takođe, koriste se filteri od 1x1 za linearnu transformaciju ulaza. Sledeća komponenta je ReLU jedinica, što je značajno unapređenje u odnosu na AlexNet, jer skraćuje vreme obuke. Korak konvolucije je fiksiran na 1 piksel da bi se očuvala prostorna rezolucija nakon konvolucije.
- **Skriveni slojevi:** Svi skriveni slojevi VGG mreže koriste ReLU. Lokalna normalizacija odgovora (LRN) obično se ne koristi jer povećava potrošnju memorije i vreme obuke, a ne doprinosi značajnije tačnosti.
- **Potpuno povezani slojevi:** VGGNet se sastoji od tri sloja sa potpunom povezanošću. Prva dva sloja imaju po 4096 kanala, dok treći sloj ima 1000 kanala, po jedan za svaku klasu.

Ograničenja VGG-16 [10]:

- **Sporost obuke:** Obuka originalnog VGG modela trajala je 2–3 nedelje na Nvidia Titan GPU-u.
- **Velika veličina težina:** Težine VGG-16 obučene na ImageNet-u zauzimaju 528 MB, što zahteva značajan prostor na disku i propusnost, čime se smanjuje efikasnost.
- **Eksplodirajući gradijenti:** Model ima 138 miliona parametara, što može dovesti do problema sa eksplodirajućim gradijentima.



slika 5 - Poređenje AlexNet i VGGNet arhitekture [10]

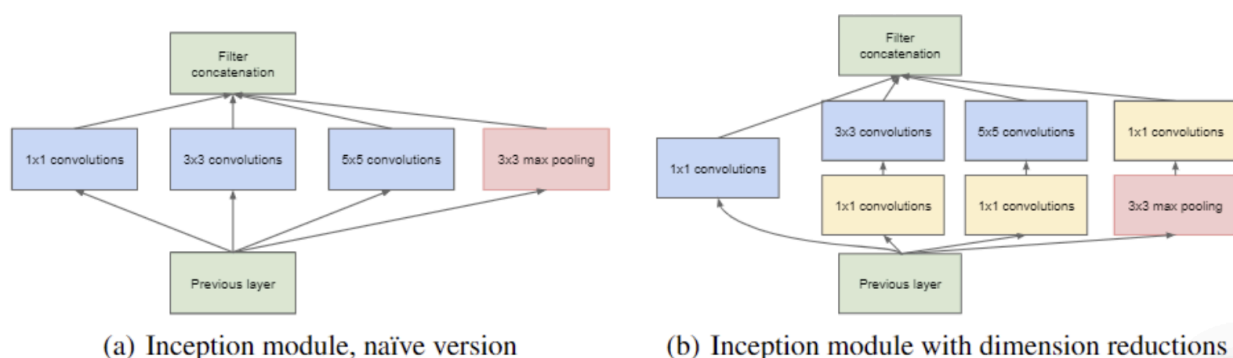
GoogLeNet

GoogLeNet predložili su istraživači iz Google-a u saradnji sa različitim univerzitetima 2014. godine, u istraživačkom radu pod naslovom "Going Deeper with Convolutions". Ova arhitektura je pobedila na ILSVRC 2014 takmičenju za klasifikaciju slika, donoseći značajno smanjenje stope greške u poređenju sa prethodnim pobednicima, AlexNet (pobednik 2012) i ZF-Net (pobednik 2013), a sa znatno nižom stopom greške od VGG (drugoplasirani 2014). Ova arhitektura koristi tehnike kao što su 1×1 konvolucije u sredini arhitekture i "global average pooling". [11]

1×1 konvolucija: Inception arhitektura koristi 1×1 konvoluciju. Ove konvolucije se koriste za smanjenje broja parametara (težina i pristrasnosti) arhitekture. Smanjenjem broja parametara povećava se dubina arhitekture.[11]

Global Average Pooling: U prethodnim arhitekturama kao što je AlexNet, potpuno povezani slojevi korišćeni su na kraju mreže. Ovi potpuno povezani slojevi sadrže većinu parametara mnogih arhitektura, što dovodi do povećanja troškova računanja. U GoogLeNet arhitekturi koristi se metoda nazvana **global average pooling** na kraju mreže. Ovaj sloj uzima karakterističnu mapu dimenzija 7×7 i uproseči je na 1×1 . To takođe smanjuje broj obučivih parametara na 0 i poboljšava top-1 tačnost za 0,6%.[11]

Inception Module: Inception modul se razlikuje od prethodnih arhitektura kao što su AlexNet i ZF-Net. U ovoj arhitekturi postoji fiksna veličina konvolucije za svaki sloj. U Inception modulu, konvolucije veličine 1×1 , 3×3 i 5×5 , kao i 3×3 max pooling, izvode se paralelno na ulazu, a rezultati se spajaju kako bi se generisao konačni izlaz. Ideja je da konvolucijski filteri različitih veličina bolje obrade objekte različitih skala.[11]



slika 6 - Vrste inception modula[11]

Arhitektura GoogLeNet modula se sastoji iz 22 sloja i prikazana je na slici 7:

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

slika 7 - Arhitektura GoogLeNet [11]

4. Zaključak

U ovom seminarskom radu istražene su konvolucione neuronske mreže (CNN) kao ključna komponenta modernih rešenja za obradu slike i analizu podataka. Detaljno su opisani različiti slojevi CNN-a, uključujući ulazni sloj, konvolucioni sloj, sloj sazimanja, sloj aktivacione funkcije i potpuno povezani sloj. Svaki od ovih slojeva igra vitalnu ulogu u procesu ekstrakcije karakteristika i klasifikacije podataka, omogućavajući mrežama da efikasno uče složene obrasce.

Pored toga, razmatrane su i različite arhitekture konvolucionih mreža, poput AlexNet, ZF-Net, VGGNet i GoogleNet. Svaka od ovih arhitektura donela je inovacije koje su unapredile performanse i efikasnost u zadacima klasifikacije slika. Na primer, AlexNet je postavio temelje za korišćenje GPU-a u obuci modela, dok je GoogleNet uveo Inception module, omogućavajući istovremeno korišćenje konvolucija različitih veličina.

Razvoj konvolucionih neuronskih mreža i njihovih arhitektura značajno je uticao na napredak u oblasti mašinskog učenja i veštačke inteligencije, otvarajući nove mogućnosti u istraživanju u raznim naukama. U budućnosti, očekuje se dalji napredak u ovim tehnologijama, sa novim pristupima i optimizacijama koje će dodatno unaprediti sposobnosti konvolucionih mreža.

5. Literatura

- [1] - Konvolucione neuronske mreže u sistemima automatskog upravljanja – pregled stanja u oblasti istraživanja - Natalija Pešić, Radiša Jovanović - Mašinski fakultet u Beogradu - <https://scindeks-clanci.ceon.rs/data/pdf/0040-2176/2023/0040-21762304433P.pdf> - pristupljeno 26.10.2024.
- [2] - Domaći zadatak: Konvolucione neuronske mreže - Krarišić Đorđe - Fakultet inženjerskih nauka Kragujevac
- [3] - Duboke konvolucione neuronske mreže - koncepti i aktuelna istraživanja - Marko M. Dabović, Igor I. Tartalija - https://www.etrn.rs/common/pages/proceedings/ETRN2017/VI/IcETRN2017_paper_VII_1.pdf - pristupljeno 27.10.2024.
- [4] - CS231n Convolutional Neural Networks for Visual Recognition - GitHub - <https://cs231n.github.io/convolutional-networks/#conv> - pristupljeno 28.10.2024.
- [5] - PREPOZNAVANJE RUKOPISA STROJNIM UČENJEM I PRIMJENA U IDENTIFIKACIJU POPUNJENIH POLJA NA UPLATNICI - Josip Mrden - <https://bib.irb.hr/datoteka/988518.MrdenJosip.pdf> - pristupljeno 28.10.2024.
- [6] - ReLu funkcija - <https://bs.martech.zone/acronym/relu/> - pristupljeno 29.10.2024.
- [7] - AlexNet Architecture Explained - Medium - <https://medium.com/@siddheshb008/alexnet-architecture-explained-b6240c528bd5> - pristupljeno 29.10.2024.
- [8] - A Brief Overview Of ZFNet Architecture - Medium - <https://medium.com/@ibtedaazeem/a-brief-overview-of-zfnet-architecture-c56aa015d20f> - pristupljeno 29.10.2024.
- [9] - Visualizing and Understanding Convolutional Networks - Matthew D. Zeiler, Rob Fergus - <https://arxiv.org/pdf/1311.2901> - pristupljeno 29.10.2024.
- [10] - VGG-Net Architecture Explained - Medium - <https://medium.com/@siddheshb008/vgg-net-architecture-explained-71179310050f> - pristupljeno 29.10.2024.
- [11] - Understanding GoogLeNet Model – CNN Architecture - GeeksforGeeks - <https://www.geeksforgeeks.org/understanding-googlenet-model-cnn-architecture/> - pristupljeno 29.10.2024.