SVEUČILIŠTE U ZAGREBU

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

Deep Dream za vizualizaciju neuronskih mreža

Daniel Bratulić

Voditelj: Tomislav Hrkać

Zagreb, svibanj, 2017

**Sadržaj**

[1. Uvod 3](#_Toc479542515)

[2. Duboko učenje 4](#_Toc479542516)

[2.1 Konvolucijske neuronske mreže 5](#_Toc479542517)

[2.1.1 Lokalna receptivna polja 5](#_Toc479542518)

[2.1.2 Dijeljene težine 6](#_Toc479542519)

[2.1.3 Sloj sažimanja (engl. Pooling layer) 7](#_Toc479542520)

[3. Tensorflow paket 8](#_Toc479542521)

[5. Zaključak 10](#_Toc479542522)

[6. Literatura 11](#_Toc479542523)

[7. Sažetak 12](#_Toc479542524)

# Uvod

Zadatak ovog seminarskog rada je pokušati shvatiti što se događa u skrivenim slojevima dubokih neuronskih mreža. Takav stil strojnog učenja, ili prikladnije nazvanog dubokog učenja je danas sve veći hit u svijetu i jedna od najboljih i najkorištenijih tehnika u području računalnog vida. Uz računalni vid, duboko učenje je vrlo primijenjeno u obradi zvuka i teksta.

Pokušavam riješiti ovaj zadatak jer iako su duboke neuronske mreže vrlo popularne i vrlo uspješne u rješavanju raznih klasifikacijskih zadataka, još uvijek postoji pitanje kako točno izgledaju značajke skrivenih slojeva te što su zapravo te značajke.

Samo nekoliko radova u posljednjih pet godina (u koje je duboko učenje zadobilo veliku popularnost) se dotaklo te teme, a u ovom radu dotaknut ću se blog objave koju su napisali suradnici u Googleu, nazvana Inceptionism (referenca).

Za rješavanje ovog problema koristit ću programski alat Tensorflow(referenca) te jupyter bilježnice(referenca) u kojima ću prikazati svoj kod. Tensorflow je danas jedan od vodećih "open source" programskih paketa za strojno učenje. Dostupan je u jezicima Python i C++ te se trenutno radi na podršci za jezik Java. U ovom radu koristit ću podršku za programski jezik Python.

Organizacija ostatka rada će biti strukturirana na sljedeći način, u prvom poglavlju opisat ću podrobnije opisati principe dubokog učenja i neuronskih mreža te u čemu se oni zapravo razlikuju od dosadašnjih tehnika strojnog učenja, također navesti ću prednosti i mane dubokih neuronskih mreža.

Sljedeće poglavlje biti će zaduženo za pobliži pogled na Tensorflow. Opisat ću na koji način funkcionira te priložiti nekoliko jednostavnih isječaka koda koji predstavljaju neki jednostavan posao. Za to ću koristiti neke prednaučene modele, poput onog modela koji je naučen na skupu podataka za natjecanje ImageNet[20], najveće natjecanje u području računalnog vida. Opisat ću također ukratko i Jupyter Notebook paket/sučelje u sklopu ovog poglavlja pošto je to okolina u kojoj ću raditi.

Treće poglavlje, i posljednje veliko poglavlje će se dotaknuti analize samih pristupa vizualizaciji skrivenih slojeva neuronske mreže, odnosno "Deep Dreaming"-a. Uz to, napravit ću osvrt na nekoliko ostalih radova koji su se bavili tematikom vizualizacije i analize skrivenih slojeva dubokih neuronskih mreža.

Na kraju su, kao i uvijek, zaključak i literatura.

# Duboko učenje

Strojno učenje je u današnje doba zastupljeno svugdje oko nas, vrlo često nismo niti toga svjesni. Svaki put kada utipkamo neki izraz u najpoznatiju svjetsku tražilicu Google, u pozadini se vrti algoritam strojnog učenja koji rangira rezultate pretrage, ako zatim odaberemo neku web trgovinu koja nam se prikazala u rezultatima, vrlo vjerojatno slijedi još jedan algoritam strojnog učenja koji nam preporučuje razne proizvode na temelju naših dosadašnjih odabira.

Postoji i velika primjena u području računalnog vida, od prepoznavanja objekata na slikama, segmentacije slika, grupiranja slika, prepoznavanje rubova, uglova i slično. Gotovo svaki problem se može efikasnije riješiti korištenjem metoda strojnog učenja.

Upravo je zbog računalnog vida područje dubokog učenja sve više zaživjelo i doživjelo veliki procvat u posljednjih pet godina. Metode dubokog učenja su vrlo uspješne u raznim zadacima iz područja računalnog vida zbog toga što mogu puno lakše baratati sa sirovim podacima, odnosno slikama.

Klasične metode strojnog učenja iziskuju pronalaženje i odabir najkvalitetnijih značajki unutar slika, poput rubova, histograma i slično. Tim značajkama bismo predstavili sirovi podatak – sliku, u vektorskom obliku i na taj način bismo mogli krenuti na sljedeći uobičajeni korak u strojnom učenju, odabiru modela.

Duboko učenje je podgrana strojnog učenja. Metode dubokog učenja koriste niz slojeva koji imaju nelinearne jedinice za transformaciju podataka koji se dobivaju iz prethodnih slojeva.

Glavno oružje dubokog učenja su umjetne neuronske mreže koje praktički funkcioniraju na spomenuti način u prethodnom odlomku. O njihovoj uporabi u metodama dubokog učenja ću govoriti kroz ovaj rad te će na njima biti glavni fokus, točnije na još specifičnijim neuronskim mrežama koje su nazvane konvolucijske neuronske mreže.

U slučaju slika, razni slojevi odrađuju različit posao te na taj način simuliraju ekstrakcije značajki kod konvencionalnih, već spomenutih metoda strojnog učenja. Na primjer, prvih nekoliko slojeva neuronske mreže prepoznaju pozicije rubova i njihove orijentacije, a neki dublji slojevi mogu slagati te značajke te sve ostale otkrivene iz prethodnih slojeva i iskoristiti ih za (npr.) detekciju objekata, koji je jedan od čestih zadataka dubokog učenja.

Već spomenute nelinearne jedinice, koje su u neuronskim mrežama prikladno nazvane neuroni, su najveća prednost dubokog učenja. Sama činjenica što se značajke i veze između njih pronalaze u skrivenim slojevima, umjesto da programer mora razmišljati kako i na koji način odabrati te značajke, uklanja puno posla prije samog odabira nekog modela.

Ovakav postupak i pristup pokazao se kao trenutno najbolji u svijetu na natjecanju u prepoznavanju objekata na slikama, ImageNet, spomenuto u uvodnom poglavlju.

Najčešći oblik strojnog učenja u današnje vrijeme jest nadgledano učenje koje zahtjeva postojanje podataka i njihovih oznaka, npr. Zamislimo da imamo dvodimenzionalni prostor i nekoliko točaka unutar njih te je naš zadatak odrediti pripadnost razredima budućih točaka koje se pojavljuju. Tada će se svaki podatak sastojati od koordinata (x,y) te pripadnosti razreda z, koji se mijenja od 1 do N, ovisno koliko postoji razreda. Ovo je vrlo jednostavan primjer nadgledanog učenja.

Neuronske mreže imaju težine koje se podešavaju u ovisnosti o ulaznom podatku i njihovoj oznaci. Težine se podešavaju vrlo poznatim postupkom koji se naziva gradijentni spust, te je standardan postupak u učenju neuronskih mreža danas. On funkcionira na način da se primjeri propuste kroz neuronsku mrežu te se na izlazu odredi pripadnost jedne od mogućih N klasa, najčešće u zadnjem neuronu bude upravo N izlaznih neurona te onaj koji ima najveću vjerojatnost određuje klasu. Zatim u ovisnosti izlaza neuronske mreže počinje prostiranje unatrag gdje se podešavaju težine svih neurona. Postoji više varijanti gradijentnog spusta, primjerice skupni i stohastički gradijentni spust.

Kod arhitektura dubokih neuronskih mreža te jedinice su nelinearne te na taj način omogućuju prepoznavanje sitnih razlika između pojedinih slika, ili objekata na tim slikama.

Ono što je imalo jako puno utjecaja na uspješnost dubokog učenja u posljednjih nekoliko godina jesu grafičke kartice. Mogućnost da se treniranje takvih ogromnih neuronskih mreža vrši na grafičkim karticama znatno ubrzava cjelokupni proces. Na taj način se vrijeme učenja smanjilo s nekoliko tjedana na svega dan ili čak manje. Mnogi programski paketi imaju podršku za učenje modela na grafičkim karticama te je taj pristup danas uobičajen.

## Konvolucijske neuronske mreže

Konvolucijske neuronske mreže su danas najučestaliji tip neuronskih mreža koje se koriste u području računalnog vida. Koriste posebnu arhitekturu koja je prilagođena upravo za klasifikaciju slika i imaju barem jedan sloj koji koristi konvoluciju umjesto matričnog produkta, otkud su dobile ime. Konvolucija je operacija definirana nad dvije funkcije realne varijable.

Optimiranje konvolucijskih mreža se vrši na isti način kao i optimiranje neuronskih mreža, dakle stohastičkim (ili grupnim) gradijentnim spustom, no jednadžbe unutar algoritma nisu jednake zbog različitih arhitektura mreža i različitih priroda tih mreža.

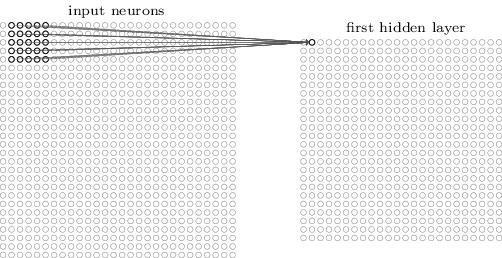
Takve mreže imaju tri osnovne ideje: lokalna receptivna polja (engl. *local receptive fields*), dijeljene težine (engl. *shared weights*) i udruživanje (engl. *pooling*).

### Lokalna receptivna polja ili konvolucije

Svaki skriveni neuron u prvom skrivenom sloju nije spojen sa svim ulaznim neuronima već samo nekim dijelom podataka, odnosno piksela. Svaki skriveni neuron u tom prvom skrivenom sloju pokazuje na dijelove ulazne slike, koji su dimenzija npr. 3x3, 4x4, 5x5 piksela ili nekih drugih dimenzija. Svaki takav dio ulazne slike (koji je obuhvaćen spomenutim dimenzijama) naziva se lokalno receptivno polje za neki skriveni neuron.

Skriveni neuron će pokušati naučiti analizirati to specifično receptivno polje s kojim je povezan. Zbog ovakvog pristupa, ostvaruje se jedna velika prednost konvolucijskih mreža, a to je što skriveni sloj neće imate jednak broj neurona kao ulazni podaci, niti će biti potpuno povezan sa svim ulaznim pikselima, već će tih skrivenih neurona biti manje i na taj način će model biti jednostavniji od običnih potpuno povezanih mreža (barem na tom dijelu mreže).

Taj prozor dimenzija 5x5 ili nekih sličnih dimenzija se pomiče kroz sliku i mapira na ostale skrivene neurone. Na taj način, kao što je u prethodnom odlomku spomenuto, imamo puno manje težina i veza, tj. imamo puno manje parametara.

  
Slika : Lokalna receptivna polja(iz Literatura: 2)

### Dijeljene težine

U konvolucijskim mrežama postoji još jedno poboljšanje u usporedbi s običnim unaprijednim mrežama, za svaki skriveni neuron u prvom konvolucijskom sloju (na konvolucijski sloj se odnose ovo i prethodno svojstvo konvolucijske mreže) vrijedi da ima dijeljene težine i pristranost (engl. *bias)* sa svim ostalim neuronima u tom sloju.

Dakle ako imamo situaciju kao na slici 1 gdje postoje 24x24 skrivena neurona, gdje na svakog od skrivenog neurona pokazuje 5x5 piksela iz ulaznog sloja, imali bismo upravo 5x5 težina koje dijele svi neuroni. To nam kazuje kako će neuroni u skrivenom sloju detektirati istu značajku, ali na drugim lokacijama na slici.

Poanta dijeljenih težina jest da ako skriveni neuron detektira (npr.) neke rubove na specifičnom lokalnom polju, vrlo vjerojatno je da će i neki drugi skriveni neuron htjeti detektirati rubove u svom specifičnom lokalnom polju, a dijeljene težine olakšavaju to. Drugim riječima, ako postoji slika psa te je pomaknemo u neku stranu, ili je zarotiramo, to je još uvijek slika psa. Neće se izgubiti informacije o toj slici.

Takva poveznica između ulaznog sloja i prvog skrivenog sloja se često naziva mapa značajki. Velika prednost korištenja dijeljenih težina jest smanjeni broj parametara, kao što sam prethodno spomenuo, postojati će 5x5 težina te *bias* što zajedno daje 26 parametara.

To je puno manje u usporedbi s potpuno povezanom neuronskom mrežom, gdje bi ukupan broj parametara bio jednak 28 x 28 (ulazne dimenzije slike, lijeva matrica na slici 1) x 28 (neki proizvoljan broj skrivenih neurona) naspram 26 x 10 (proizvoljan broj mapa značajki).  
Uvijek se koristi više mapa značajki kako bi omogućile što bolju potragu i ekstrakciju značajki sa slika.

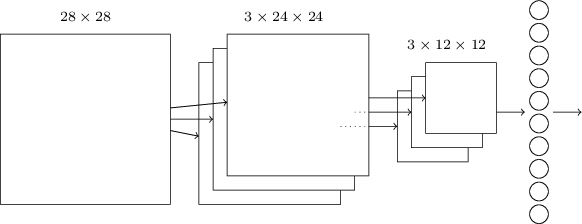
### Sloj sažimanja (*engl. Pooling layer*)

Slojevi sažimanja se nalaze odmah iza konvolucijskih slojeva, oni pojednostavljuju informacije s izlaznih slojeva konvolucijskog sloja. Uzimaju izlaz iz mape značajki te sažmu informacije iz nekog manjeg područja na slici (npr. 2x2) u jedan neuron u svojem sloju. Postoji nekoliko vrsta sažimanja, a najpopularnije tehnike se zasnivaju na osnovnom statističkom agregiranju.

Najpopularnija tehnika jest *max-pooling,* odnosno maksimalno sažimanje, gdje se uzima maksimalna vrijednost iz tog manjeg područja u prethodnom sloju te spremi u neuron u sloju sažimanja.

Sloj sažimanja se sastoji od još manjeg broja neurona, no što ih ima u konvolucijskom sloju (prvi skriveni sloj nakon ulaznog). *Max pooling* možemo zamisliti kao način na koji mreža pita je li pronađena neka značajka negdje u promatranoj regiji na slici.

Postoje ostali pristupi osim *max poolinga*, poput *mean pooling* koji se uzima prosjek vrijednosti unutar regije te *L2 pooling*, koji uzima korijen sume svih kvadrata unutar 2x2 prozora na slici te ostala agregatne funkcije.

  
Slika : Jednostavna konvolucijska neuronska mreža

Na slici 2 je prikazana jednostavna konvolucijska neuronska mreža koja se sastoji od prethodno nabrojanih slojeva. Prvi je ulazni sloj, čije su dimenzije 28x28, zatim slijedi konvolucijski sloj koji ima 3 mape značajki te nakon njega slijedi sloj sažimanja ( *pooling* ). Na kraju je izlazni sloj koji je potpuno povezan sa slojem sažimanja kao i u običnim neuronskim mrežama.

### Sloj ispadanja (engl. Dropout layer)

Sloj ispadanja je praktički postao standard u arhitekturama konvolucijskih neuronskih mreža nakon što je odlično poboljšanje spomenuo Geoffrey Hinton zajedno sa suradnicima u [5]. Taj sloj bi zapravo služio kao oružje protiv prenaučenosti, princip rada tog sloja jest da se nasumično odabere neuroni koji bi se prilikom treniranja jednostavno ugasili, odnosno vrijednost njihovih aktivacijskih funkcija bi se postavila na 0. Udio neurona koji bi se ugasili određen je nekom vrijednošću između 0 i 1. Na taj način mreža nema redundantnih reprezentacija. Dakle, sloj ispadanja služi kao jedna vrsta regularizacije.

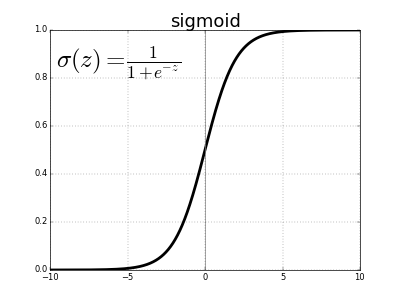
Prilikom testiranja mreže, vrijednost vjerojatnosti ispadanja nekog neurona postavila bi se na 0, dakle niti jedan neuron se ne bi ispustio kako bi se maksimalno iskoristila mreža.



Slika 3 Sloj ispadanja (engl. Dropout layer)

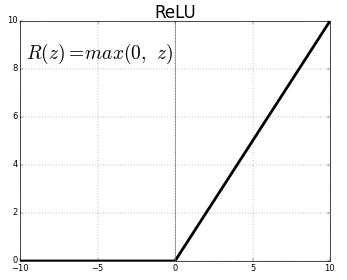
### Razne aktivacijske funkcije

Uz prethodno navedena svojstva konvolucijskih mreža još je jedna velika stvar imala utjecaj na njihovu uspješnost u raznim izazovima računalnog vida, to je raznovrsnost aktivacijskih funkcija. Sve do prije nekoliko godina uobičajena je bila sigmoidalna aktivacijska funkcija.



Slika 4 Sigmoidalna aktivacijska funkcija

Trenutno najbitnija aktivacijska funkcija za duboke neuronske mreže jest ReLU, odnosno zglobnica (*engl. Rectified Linear Unit*). Za razliku od sigmoide ona ne dovodi do zasićenja neurona koji su inicijalizirani na jako niske negativne vrijednosti ili jako visoke pozitivne vrijednosti te konvergira brže. Uz to, računalno je puno efikasnije od sigmoide. Formula je puno jednostavnija i glasi f(x) = max(0, x). Problem kod ove aktivacijske funkcije su mrtvi neuroni koji se mogu javiti ako se vrijednost inicijalizira na vrijednost manje od 0, no to se lako izbjegne manipulacijom inicijalizacije vrijednosti.



Slika 5 Funkcija zglobnice, ReLU aktivacijska funkcija

# TensorFlow

TensorFlow je programski paket za primjenu algoritama strojnog i dubokog učenja koji je izašao u uporabu tokom 2015. godine. Prvo stabilno izdanje, dakle verzija TensorFlow 1.0 izašlo je u 2. mjesecu 2017. godine.

Autor paketa TensorFlow jest kompanija Google, točnije jedan od timova unutar Google-a, Google Brain Team. Google je za svoju unutarnju uporabu prethodno koristio paket DistBelief u istu svrhu, dakle za razvoj modela strojnog i dubokog učenja. Nakon uvida u nedostatke tog paketa, odlučili su nadograditi i pojednostaviti neke dijelove te je iz toga nastao TensorFlow.

TensorFlow je danas vodeći paket u području dubokog učenja, a konkurenti su mu Theano, Caffe, Torch, Keras i mnogi drugi. Ono što ga posebice izdvaja od konkurencije jest njegova srž, računalni graf koji se sastoji od mnoštvo čvorova i bridova. U takvom jednom grafu, čvorovi predstavljaju operacije, dok bridovi između čvorova predstavljaju podatke koji teku između pojedinih operacija. Takav graf se jednostavno izgrađuje povezivanjem više operacija i time možemo razbiti izgradnju modela na više jednostavnijih dijelova.

Sljedeća pozitivna stvar TensorFlowa jest mogućnost jednostavnog i vrlo učinkovitog prikaza složenih modela neuronske mreže pomoću alata koji se zove TensorBoard, taj alat je instaliran zajedno s TensorFlowom. Uz samu vizualizaciju modela, moguće je i pratiti stanje u modelima tokom učenja.

Uz to, TensorFlow je moguće pokretati na više CPU i što je još važnije, na više GPU uređaja. Kao što je spomenuto u poglavlju 2, jačanje grafičkih jedinica je značajno pogodilo razvoju dubokog učenja, TensorFlow automatski prepoznaje instalirano razvojno okruženje za grafičke kartice (nVIDIA predvodi to područje, s CUDA tehnologijom) te automatski omogućava učenje na grafičkim karticama.

Za kraj, posljednja velika stvar koja ističe TensorFlow od ostatka konkurencije jest mogućnost pokretanja na Linuxu, macOS i nekim mobilnim platformama poput Androida i iOS-a. TensorFlow API dostupan je za više programskih jezika, što je također vrlo zgodna stvar. Trenutno dostupni jezici su Python, C++, Haskell, Java i Go. U ovom radu će fokus i primjeri biti napisani u Pythonu zbog brzine i jednostavnosti samog jezika kao i lakoće čitanja.

## Osnove TensorFlowa

Kao što je spomenuto, osnova rada TensorFlowa su računalni grafovi kroz koje se definiraju operacije koje će se izvoditi. Nakon što se definiraju te operacije potrebno je predati vrijednosti na ulaze takvog grafa, ukoliko je to uopće potrebno. Nabrojit ću neke osnovne i neophodne jedinica za rad u TensorFlowu kao i pojmove koji predstavljaju principe po kojima radi TensorFlow.

### Operacije

Operacije su predstavljene svim čvorovima u grafu i zadužene su za transformaciju ili kombiniranje podataka koji se kreću kroz graf. Operacije mogu biti razne stvari, poput običnog zbrajanja, množenja pa zatim matrično množenje ili čak neki dijelovi neuronskih mreža, poput prethodno spomenutih aktivacijskih funkcija. Detaljan prikaz vidljiv je u Tablici 1. Operacije mogu biti nula ili više ulaza te nula ili više izlaza, npr. konstanta je operacija koja nema ulaza i ima jedan izlaz.

|  |  |
| --- | --- |
| **Kategorija** | **Primjer operacije** |
| Matematičke operacije s elementima | Add, Sub, Mul, Div, Exp, Log, Equal, ... |
| Operacije s poljima | Concat, Slice, Split, Rank, Shape, … |
| Operacije s matricama | MatMul, MatrixInverse, … |
| Operacije sa stanjima | Variable, Assign, AssignAdd, … |
| Blokovi za neuronske mreže | SoftMax, Sigmoid, ReLU, MaxPool, ... |

Tablica 1 Prikaz primjera operacija u TensorFlowu

### Tenzori

Tenzori su osnovne jedinice za pohranu podataka unutar TensorFlowa. Tenzori se predaju kao ulaz čvorovima i oni su ujedno izlaz iz pojedinih operacija. To su matrice proizvoljnih dimenzija nekog fiksno određenog tipa. Svaki tenzor ima svoj rang (*engl. rank,* različito od ranga matrice), koji predstavlja dimenziju tenzora. Dakle, matrice bi imale rang 2, vektori bi imali rang 1 dok bi obične skalarne vrijednosti imale rang 0. Tenzori ne drže vrijednosti u svojoj memoriji već pružaju sučelje za dohvat vrijednosti koje referencira taj tenzor.

### Varijable

Tenzori često bivaju uništeni prilikom rada s modelima dubokih neuronskih mreža, što nam je ponekad u redu, no nekad bismo htjeli zapamtiti te vrijednosti kako bismo ih mogli višestruko iskoristit. Za to nam služe varijable (*engl. Variable*), koje su u TensorFlowu označene s tf.Variable. One su operacije koje pamte tenzore te na taj način dozvoljavaju promjenu vrijednosti unutar tenzora.

### Sesije (engl. Sessions)

Nakon što izgradimo naš računalni graf s prethodno spomenutim dijelovima i postupcima, potrebno je izvršiti taj graf. Način na koji je to ostvareno u TensorFlowu je unutar posebnog okruženja, odnosno sesije (*engl. Session*). Graf se izvršava tako što se pozove metoda *run()* nad objektom tipa *tf.Session*. Toj metodi se preda ciljni čvor koji je izlaz grafa te ukoliko treba, predaju se vrijednosti za čvorove koji predstavljaju ulazne vrijednosti(*engl. placeholders*).

### TensorBoard

TensorBoard je alat, te se zapravo logički po opisu razlikuje od prethodnih pojmova, pošto su oni neophodni za rad bilo kojeg programa u TensorFlowu. Prikaz grafova u TensorBoardu je opcionalan, no kao što je spomenuto u samom uvodu u TensorFlow, to svojstvo je vrlo moćno i od velike pomoći. Bez TensorBoarda bi bilo vrlo teško predočiti neke duboke mreže pošto one znaju imati puno više od 10000 čvorova.

Strukture podataka koje TensorBoard može vizualizirati jesu Summary objekti unutar TensorFlowa (tf.summary). To mogu biti slike, tekstovi, histogrami, zvukovi ili obični skalari, poput stope učenja ili vrijednosti funkcije gubitka. Također, svaka operacija je jedan tf.summary objekt čime se olakšava prikaz grafova.

Kako bi se spomenuti objekti tipa tf.Summary mogli prikazati, mora se pozvati objekt tf.summary.FileWriter koji je zadužen za spremanje čvorova koji će se iscrtavati te privremenih vrijednosti koje se dobivaju tokom treniranja.

## Primjer kodova u TensorFlowu

### Jednostavan početni primjer i prikaz grafa



Isječak koda 1 Osnovna funkcionalnost TensorFlowa

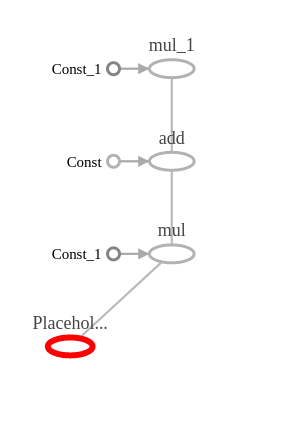
Ovaj primjer pokazuje osnovni rad s TensorFlowom, radi se o običnom zbrajanju i množenju, no kroz funkcionalnost TensorFlowa.

U ovom isječku koda potrebno je objasniti neke notacije koje još nisu spomenute. Tf.constant jest operacija koja sadržava konstantu, to je čvor koji nema ulaza već ima samo jedan izlaz i to skalar koji je zapisan pri definiciji.

Tip podatka koji je vezan uz varijablu x, tf.placeholder je već natuknut u potpoglavlju o sesijama. To je operacija kojoj nije odmah zadana vrijednost već joj se predaje prilikom izvršavanja sesije, dakle ulazna vrijednost, to se može vidjeti prilikom poziva metode run(), ondje se argumentu feed\_dict predaje rječnik koji inicijalizira sve varijable koje su tipa tf.placeholder s njihovim pripadnim vrijednostima, u ovom slučaju to je varijabla x i vrijednost 3.

Na slici 6 je vidljiv graf koji prikazuje ovaj isječak koda. Placeholder je varijabla x kojoj zadajemo vrijednost. Const\_1 je varijabla b, dok je Const varijabla a. Moguće je izmijeniti imena čvorova, no za ovaj osnovni primjer to nije urađeno.

Mul i add su operacije množenja odnosno zbrajanja. Prilikom interaktivnog rada na TensorBoardu moguće je odabrati svaki pojedini čvor te ispitati ulaze i izlaze tih čvorova.



Slika 6 Graf koji prikazuje prethodni isječak koda

### Prikaz jednostavne konvolucijske mreže

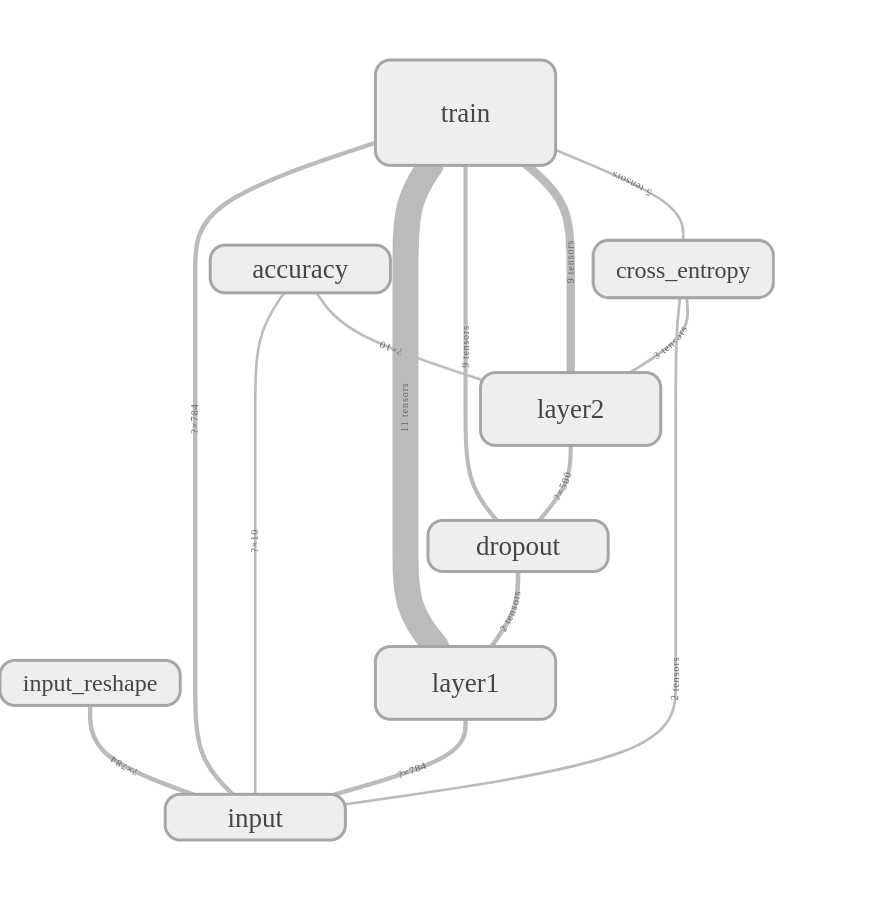


Isječak koda 2 Kreiranje konvolucijske mreže



Isječak koda 3 Kreiranje i pokretanje konvolucijske mreže, 2.dio

U prethodna dva isječka koda napisana je jednostavna konvolucijska neuronska mreža, koja se sastoji od 2 konvolucijska gdje nakon svakog konvolucijskog sloja odmah slijedi sloj sažimanja te dodatni sloj ispadanja koji prima izlaze prvog skrivenog sloja. Na kraju imamo dva potpuno povezana sloja kako bismo mogli odrediti klasu ulazne slike.



Slika 7 Arhitektura jednostavne konvolucijske mreže

## Zaključak o TensorFlowu

U prethodnih nekoliko poglavlja ukratko sam opisao princip rada TensorFlowa te demonstrirao te principe kroz nekoliko isječaka koda. Kroz te isječke je vidljiva jednostavnost i moć TensorFlowa pošto se jednostavna konvolucijska mreža za treniranje rukom pisanih znamenki može zapisati u svega šezdesetak linija koda u Pythonu.

# Vizualizacija neuronskih mreža i Deep Dream

Kao što je spomenuto u poglavljima 2. i 3. duboko učenje, odnosno duboke neuronske mreže su odličan alat za rješavanje teških problema koji su do sada bili teško rješivi. No iako su se pokazale učinkovite pri rješavanju tih problema i veliki broj ljudi i istraživača koristi duboke mreže u svakodnevnom radu još uvijek postoji pitanje kako zapravo one rade te što se zapravo događa s težinama u skrivenim slojevima.

U tu se svrhu sve više radova počelo koncentrirati na samu vizualizaciju neuronskih mreža, odnosno njihovih skrivenih slojeva. U jednom od takvih radova [12] se opisuje pokušaj vizualizacije aktivacija u svakom sloju neuronske mreže kada na ulaz dobije neku određenu sliku ili video, autor tog rada je kreirao grafičko sučelje koje je moguće pokrenuti te na taj način bilo tko može preko svoje vlastite web kamere snimati sebe ili neke objekte te vidjeti kako aktivacije u pojedinim slojevima reagiraju na te objekte. Uz to, opisali su i prikaz značajki u pojedinom sloju mreže pomoću optimizacije u prostoru slike, što je slično Deep Dream algoritmu.

## Inception model

Prije no što krenemo s daljnjom pričom o Deep Dream algoritmu, opisat ću u kratko model Inception kojeg je razvila tvrtka Google u sklopu ImageNet natjecanja 2014. godine te su s tim modelom uvjerljivo pobijedili u više kategorija. Ovaj model nam je bitan jer je on korišten u Deep Dream algoritmu.

Taj model je također konvolucijska mreža, no s nekoliko originalnih preinaka koji su smanjili broj parametra i broj slojeva te samim time i vrijeme treniranja u odnosu na neke slične mreže. Pošto je kreiran za ImageNet natjecanje, učen je na tom skupu podataka u kojem se nalazi tisuću klasa i preko milijun slika. Značajna činjenica je što se u tom skupu podataka nalazi najviše životinja, na primjer, postoje čak 120 klasa za različite pasmine pasa. To će utjecati na slike koje će biti generirane u Deep Dream algoritmu.

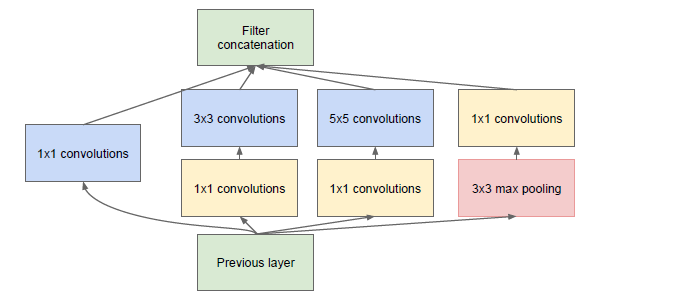
### Inception modul

Ono što je bilo različito u ovom modelu od svih tadašnjih modela jest njihov Inception modul, koji se sastoji od međusobno više konvolucija te slojem sažimanja koji je spojen u jedan izlazni tenzor.

Dakle, u jednom sloju umjesto jedne konvolucije, kao što je uobičajena konvolucijska mreža izgrađena, postoje 3 konvolucije, dimenzija 1x1, 3x3 i 5x5 te uz ta 3 sloja konvolucije dodan je i sloj sažimanja dimenzija 1x1 koji se u svim dosadašnjim mrežama pokazao kao dobar korak.

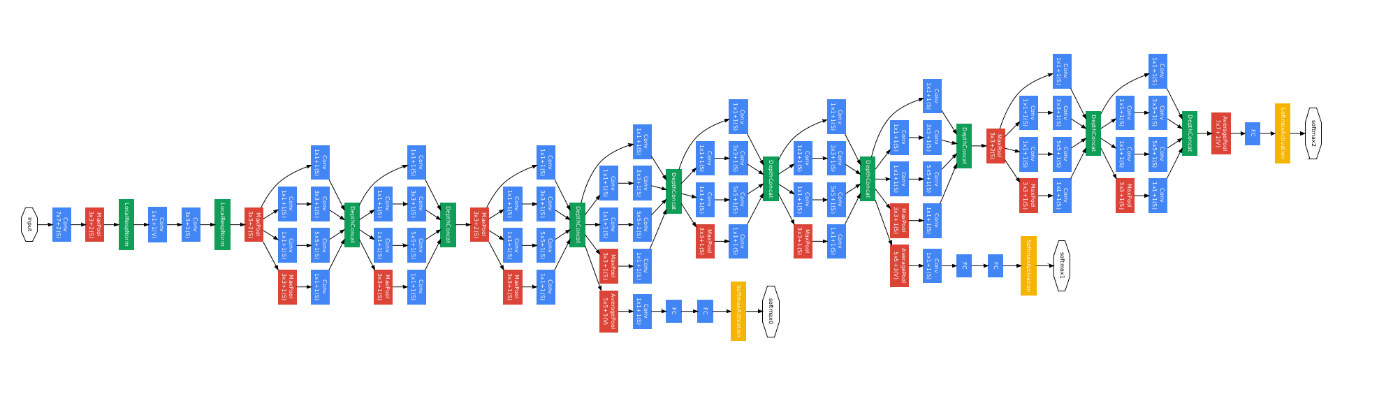
No takva struktura slojeva dovodi do znatnog povećanja broja izlaza iz pojedinog sloja što nikako nije poželjno, kako bi se taj problem riješio, dodane su dodatne konvolucije prije slojeva dimenzija 3x3 i 5x5 i to konvolucije dimenzija 1x1. Za sloj sažimanja je također dodana konvolucija, no nakon njega, ta konvolucija je dimenzija 1x1.

Najveća prednost ovakve arhitekture jest mogućnost znatnog povećanja broja jedinki u svakom sloju bez pretjerane eksplozije složenosti programa.



Slika 8 Inception modul koji zamjenjuje standardne slojeve u konvolucijskim mrežama[13]

Imajući to na umu, u nastavku je slika cjelokupne arhitekture Inception modela koji se sastoji od mnoštva takvih manjih modula.



Slika 9 Cjelokupna Inception mreža, odnosno GoogLeNet[13]

Na slici 9 je prikazana cjelokupna mreža Inception modela, također zvana GoogLeNet, pošto su autori iz tvrtke Google. Kao što se može vidjeti, većinu slojeva u toj mreži sačinjavaju Inception moduli. Plavom bojom su označene konvolucije, crvenom slojevi sažimanja, bijelom izlazi i ulazi, zelenom slojevi koji spajaju tenzore te žutom bojom su označeni slojevi sa *Softmax* aktivacijama. Može se primjetiti da postoje tri takva sloja, koji su direktno povezani s izlazima. Time se postiže evaluacija klasifikacija kroz različite razine u mreži te se na kraju sveukupni gubitak zbroji, no gubitak ranijih izlaza se pomnoži s nekom težinom kako bi se dao manji utjecaj tim klasifikacijama. Prilikom testiranja te glave su ugašene te nema nikakvog doticaja s njima.

Inception model je bio vrlo značajan za svijet računalnog vida i dubokog učenja jer se greška na natjecanju spustila gotovo do razine ljudske pogreške, što je vrlo veliki uspjeh.

## Deep Dream

Nakon kratkog uvida u vizualizaciju neuronskih mreža te opis modela koji se uvelike koristi kod većine korisnika koji se dotaknu Deep Dreama, vrijeme je za opis samog algoritma te ujedno teme ovog seminara.

Kako bi ovaj algoritam funkcionirao potrebno je imati neki prednaučeni model, u ovom slučaju to je model mreže GoogLeNet koja je naučena na skupu podataka ImageNet koji je spominjan nekoliko puta. Uz to, koristit će se i DeepOCR arhitektura tvrtke Microblink[14] koja je naučena na skupu podataka vezanih uz matematički sadržaj i koristi se u proizvodu Photomath[15]. To je mreža koja uz klasifikaciju također radi i detekciju objekata.

Ideja algoritma Deep Dream jest vrlo jednostavna, kao i implementacija istog. Umjesto da se slika provede od početka do kraja neuronske mreže, kao prilikom bilo kojeg postupka treniranja neuronske mreže, ovdje je potrebno dovesti sliku do nekog skrivenog sloja unutar mreže. U sloju za koji smo odlučili da je konačna točka propagiranja slike kroz mrežu uzmemo aktivacije svih neurona u tom sloju, te gradijentima dodijelimo vrijednosti tih aktivacija. Iz tog sloja zatim napravimo prostiranje unatrag (*engl. backpropagation*) sve do ulaza, odnosno slike.

Za razliku od uobičajenog postupka treniranja mreže umjesto podešavanja težina, ovdje se s gradijentom izračunatim kroz prostiranje unatrag djeluje na sliku te se na taj način djeluje na sliku sa značajkama koje su otkrivene u mreži. Na taj način neuronska mreža ističe ono što prepoznaje na slici. Ovo je postupak optimiranja slike, koji je sličan već spomenutom radu [12].

Naravno, svaki sloj neće djelovati jednako uspješno, odnosno neće isticati jasno raspoznatljive objekte na slikama jer to nije uloga svakog sloja. Raniji slojevi, oni koji su bliži ulazu neuronske mreže su apstraktniji te će češće isticati rubove, krugove, teksture ili neki drugi oblik, ovisno o tome koji se neuroni najviše aktiviraju, dok kasniji slojevi, oni bliže izlazu iz mreže će se jače aktivirati na već prepoznatljive objekte i jasne oblike koji može prepoznati bilo koja osoba, u slučaju modela GoogLeNet to će biti najčešće psi pošto je većina klasa u skupu podataka bila neka vrsta pasmine (120 različitih pasmina od ukupno 1000 klasa) ili neka druga životinja pošto one dominiraju podacima.

### Programska implemetnacija Deep Dream algoritma

Sama implementacija Deep Dream algoritma je vrlo kratka, svega stotinjak linija. Ovdje ću prikazati samo najbitniji dio, srž algoritma Deep Dream te opisati neke funkcije koje se pozivaju iz tog pseudokoda.



Isječak koda 4 Implementacija DeepDream algoritma[16]

U prikazu Isječka koda 4 vidljiv je pseudokod algoritma Deep Dream. Kao što je spomenuto, neke funkcije su izostavljene, no iako su i one bitan dio algoritma, poput funkcija koje mijenjaju veličinu slike kako bi se mogla ispravno obraditi te funkcije koja računa gradijent.

### Rezultati nastali primjenom algoritma Deep Dream

Na sljedećoj slici ćemo napraviti nekoliko izvođenja Deep Dream algoritma.



Slika 10 slika muffina[17]



Slika 11 Slika muffina nakon izvođenja Deep Dream algoritma u plitkim slojevima

Na prethodnoj slici vidimo kako zapravo funkcionira Deep Dream algoritam. Ova slika je rezultat izvođenja algoritma gdje je ciljni sloj bio drugi Inception modul, odnosno sloj pod nazivom *mixed3b\_3x3\_bottleneck*. *Bottleneck* ovdje naznačava konvoluciju dimenzija 1x1 koja se javlja prije prave konvolucije dimenzija 3x3 kao što je opisano u poglavlju o Inception modelu. Ovdje se može vidjeti kako je ovaj sloj još uvijek vrlo apstraktan te se samo izvlače neki oblici i konture na slici.

Na sljedeće dvije slike (Slika 12 i Slika 13)se vidi zapravo sva čar algoritma Deep Dream te zašto je bio toliko popularan i zašto je Internet ludovao za ovakvim slikama. Sljedeće dvije slike dolaze iz malo dubljih slojeva. Na tim slikama se može vidjeti kako se značajke koje se javljaju u tim slojevima brinu za neke specifične slojeve. I to u jednom sloju su zadužene za detekciju psa, dok u drugom za detekciju voća.



Slika 12 DeepDream na dubljim slojevima, značajke vezane uz pse su aktivirane



Slika 13 Deep Dream na dubljim slojevima, značajke vezane uz voće su aktivirane

Dosad je spomenuto više puta kako postoji puno različitih pasmina u skupu podataka na kojem je ova mreža učena, na Slici 12 je jasno došlo do toga te uzorci pasa su bili najučestaliji u višestrukim pokušajima izvođenja algoritma Deep Dream.

Na sljedećoj slici je to posebice izraženo, sljedeća slika originalno prikazuje livadu s šest stabala, no kada se ta slika provrti kroz algoritam, rezultati izgledaju potpuno drukčije.

Doduše, ako bismo kao ciljni sloj predali neki raniji sloj u mreži, rezultati bi bilo potpuno drukčiji, te kao u primjeru sa slikom muffina, vrlo vjerojatno ne bi rezultirali prepoznatljivim objektima. No u tome i leži čar ovog postupka.



Slika 14 Livada prepuna pasa

Za ulaz u mrežu nije potrebna neka smislena slika, u nekoliko pokušaja kreirana je prazna crna slika te je predana neuronskoj mreži s nekim ciljnim slojem. Na taj način mogu se dobiti vrlo fascinantne slike, no ukoliko je predan neki sloj koji je po dubini sličan onima kao na slikama 12 i 14, moguće je da se u mreži aktiviraju značajke koje su zadužene za prepoznavanje pasa te dobijemo obrise psa na ulaznoj slici.



Slika 15 Slika generirana kojoj je kao ulaz predana prazna crna slika

## Primjena algoritma Deep Dream

Sve više i više se umjetno generirane slike dodaju skupu za učenju te treniraju neuronsku mrežu s takvim slikama zajedno s pravim primjerima za učenje. U tom segmentu se mogu iskoristiti i slike generirane pomoću Deep Dream algoritma.

Još jedna primjena ovog algoritma jest analiza skrivenih slojeva u neuronskoj mreži kojom možemo doći do saznanja koje slike ne odgovaraju dosad naučenoj neuronskoj mreži te s kakvim primjerima trebamo pojačati učenje mreže, ovo naravno ovisi o našoj domeni primjene, jer ukoliko želimo prepoznavati samo tekst, i ovo se generira na slikama koje su dane za ulaz, tada je takav ishod u potpunosti u redu.

# Zaključak

# Literatura

[1] Neural Networks and Deep Learning: <http://neuralnetworksanddeeplearning.com/>, Michael Nielsen

[2] Deep learning, Review: Yann LeCun, Yoshua Bengio, Geoffrey Hinton, <http://pages.cs.wisc.edu/~dyer/cs540/handouts/deep-learning-nature2015.pdf>

[3] Predavanja iz Dubokog Učenja, FER: Siniša Šegvić, Zoran Kalafatić, Sven Lončarić, Marko Subašić, Marko Čupić

[4] Neural Networks github.io post(slike): <https://ml4a.github.io/ml4a/neural_networks/>

[5] Dropout: A Simple Way to Prevent Neural Networks from Overfitting: Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov, <http://www.jmlr.org/papers/volume15/srivastava14a.old/source/srivastava14a.pdf>

[6]CS231n Convolutional Neural Networks for Visual Recognition, Stanford course, <http://cs231n.github.io/> : Andrej Karpathy, Justin Johnson, Fei Fei Li

[7] Prva Laboratorijska vježba iz dubokog učenja: Siniša Šegvić, <http://zemris.fer.hr/~ssegvic/du/lab1.shtml>

[8] TensorFlow: Large-Scale Machine Learning on Heterogenous Distirbuted Systems: Ian GoodFellow et al

[9] A Tour of TensorFlow: Peter Goldsborough

[10] [https://en.wikipedia.org/wiki/TensorFlow](https://en.wikipedia.org/wiki/TensorFlow%20)

[11] <https://www.tensorflow.org/get_started/>: Upute za rad u TensorFlowu

[12] Understanding Neural Networks Through Deep Visualization: Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, Hod Lipson

[13] Going Deeper With Convolutions: Christian Szegedy et. al. (Google).

[14] [Microblink.com](https://microblink.com/en) – autori mreže DeepOCR

[15] [Photomath.net](https://photomath.net/en/) – proizvod u kojem je korištena arhitektura DeepOCR

[16] Github profil TensorFlowa: <https://github.com/tensorflow/tensorflow>

[17] Slika muffina: <https://static.esea.net/global/images/users/694560.1411342686.jpg>

[18] Slika stabala:

<https://pixabay.com/en/trees-hill-green-blue-nature-park-790220/>

[19] Inceptionism: Going deeper into Neural Networks:  Alexander Mordvintsev, Christopher Olah, and Mike Tyka <https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>

[20] ImageNet natjcanje: <http://www.image-net.org/challenges/LSVRC/>

# Sažetak