



## KLASIFIKACIJA SLIKA PRIMENOM FASTAI BIBLIOTEKE

### CLASSIFICATION OF IMAGES USING FASTAI LIBRARY

Jovana Antić, *Fakultet tehničkih nauka, Novi Sad*

#### Oblast – ELEKTROTEHNIKA I RAČUNARSTVO

**Kratak sadržaj** – U ovom radu je razmatran problem klasifikacije slika primenom konvolucionih mreža realizovanih u sklopu fastai biblioteke. Trebalo je prikupiti podatke, obučiti model i prikazati rezultate.

**Ključne reči:** fastai, CNN, ResNet, learning rate, confusion matrix.

**Abstract** – The problem of image classification using CNN algorithm realize as part of fastai library is considered in this paper. It was requested to collect the data, train the model and display the results.

**Keywords:** fastai, CNN, ResNet, learning rate, confusion matrix.

#### 1. UVOD

Klasifikacija slika je složen proces na koji mogu uticati mnogi faktori. Ovaj rad ispituje proces klasifikacije slike korišćenjem fastai biblioteke, kao i probleme i perspektive klasifikacije slika. Prolazeći kroz poglavlja rada steći će se šira slika o samoj klasifikaciji slika, funkcionisanja konvolucijskih neuronskih mreža, jupiter okruženja i fastai biblioteke.

Da bi sam proces bio jasniji prikazana su i dva primera, gde se može videti sa kakvim problemima se možemo suočiti tokom samog procesa kako ih rešiti ali i unaprediti.

#### 2. KLASIFIKACIJA SLIKA I KONVOLUCIONA NEURONSKA MREŽA

Klasifikacija objekata za ljude je lak zadatak, ali pokazalo se da predstavlja složen problem za mašine [1]. Porast računara velikog kapaciteta, dostupnost visokokvalitetnih i jeftinih video kamera, kao i sve veća potreba za automatskom video analizom, izazvali su interesovanje za algoritme klasifikacije objekata.

Danas je uz pomoć učenja iz podataka, prepoznavanje slika postalo veoma efikasno. Ova je metoda mašinskog učenja dizajnirana po uzoru na funkcionisanje ljudskog mozga.

Tako se računari uče da prepoznaju vizuelne elemente unutar slike.

Konvolucijske neuronske mreže (engl. Convolutional neural network - CNN) predstavljaju aktuelan i istraživački atraktivan domen u širokoj oblasti veštačke inteligencije (eng. Artificial Intelligence – AI).

Konvolucijske neuronske mreže su biološki inspirisana verzija višeslojnog perceptrona (eng. Multilayer Perceptron – MLP). Smatra se da su od 2014. godine duboke konvolucijske neuronske mreže postale „glavni tok“ i predmet istraživačkih napora i zapaža se da od tada postižu značajne prednosti u raznim testovima [2].

#### 3. KONVOLUCIONA NEURONSKA MREŽA

Konvolucionna neuronska mreža je dobila ime po konvoluciji, široko korišćenom operatoru, koji se primenjuje u obradi slike i signala.

Ideja konvolucione neuronske mreže jeste da se postavi veći broj slojeva za otkrivanje bitnih osobina ulaznih podataka. U skladu sa tom idejom se konvolucionni filteri primenjuju na sliku kako bi se izvukle korisne karakteristike i kreirale njihove mape (eng. feature maps).

Primenom filtera na ulaznu sliku, pored otkrivanja značajnih karakteristika, vrši se i redukcija rezolucije.

Nedostatak konvolucione neuronske mreže jeste potreba za značajnim hardverskim resursima, kao i to što su dosadašnji modeli obučavani velikim brojem slika, ali noviji rezultati [3] sugerišu da se visoke performanse mogu postići i sa relativnom malim brojem uzoraka.

#### 4. FASTAI BIBLIOTEKA

Fastai su osnovali Rachel Thomas i Jeremy Howard. Rachel Thomas je profesor na univerzitetu u San Francisku, dok je Jeremy Howard naučnik koji se bavi istraživanjem na pomenutom univerzitetu.

Na osnovu PyTorch-a (biblioteka mašinskog učenja otvorenog koda zasnovana na biblioteci Torch, a koristi se za aplikacije kao što su računarski vid i obrada govora), fastai sadrži neke od najpopularnijih algoritama za klasifikaciju slika i zadatke na prirodnom jeziku. U praksi to znači da se modeli mogu kreirati i pokretati u samo nekoliko linija koda.

##### 4.1 Realizacija konvolucione mreže u fastai -u

Osnovna ideja pri realizaciji konvolucionih mreža u sklopu fastai-a je da crvene, zelene i plave vrednosti piksela koji su u prevedenom značenju brojevi od 0 do 255, unose u jednostavno izračunavanje (tj. prvi sloj) a rezultat tog proračuna, odnosno izlaz iz prvog sloja, prelazi u drugi sloj, a rezultat toga prelazi u treći sloj i tako dalje. Može biti do hiljadu slojeva neuronske mreže.

#### NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bila dr Mirna Kapetina, docent.

Prvi sloj u osnovi nalazi grupe piksela koje imaju malu dijagonalnu liniju, drugi pronalazi dijagonalnu liniju u drugom pravcu, treći pronalazi gradijente koji idu od žute do plave boje i tako dalje.

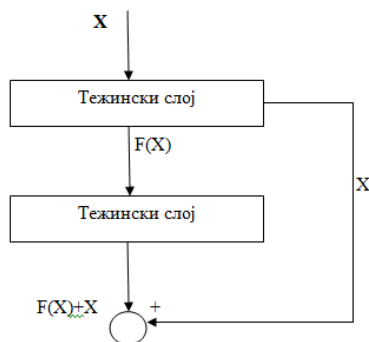
U prvom sloju imamo stvari koje mogu pronaći samo jednu liniju, a u drugom sloju možemo pronaći stvari koje imaju dve linije spojene.

Treći je mogao da preuzme sve stvari sa sloja 2 i kombinuje ih zajedno, dok četvrti sloj kombinuje stvari iz trećeg sloja

U sloju pet imamo nešto na čemu se mogu naći očne jabučice ptica i guštera ili lica određenih tipova pasa i tako dalje.

#### 4.1.1. ResNet

ResNet je kratak naziv za *residual network*. Duboke konvolucione neuronske mreže postigle su rezultat klasifikacije slike na nivou čoveka. Duboke mreže izvlače karakteristike i klasifikatore niskog, srednjeg i visokog nivoa na višeslojni način, a broj složenih slojeva može obogatiti „nivo“ karakteristika.



Slika 1. Gradivni blok

##### 4.1.1.1 Datasets

ImageNet je skup milion označenih slika visoke rezolucije, koje se mogu svrstati u otprilike 22 hiljada kategorija. Slike su sakupljene sa interneta i ljudi su ih etiketirali pomoću alata za prikupljanje publike.

Počevši od 2010. godine, u okviru Pascal Visual Object Challenge-a (PASCAL VOC), održava se godišnje takmičenje pod nazivom ImageNet Large-Scale Visual Recognition Challenge (ILSVRC2013). ILSVRC koristi podskup ImageNet-a sa otprilike 1000 slika u svakoj od 1000 kategorija.

Postoji otprilike 1,2 miliona slika za trening, 50k validacije i 150k testiranja slika.

PASCAL VOC pruža standardizovane skupove podataka o slikama za prepoznavanje klase objekata. Takođe pruža standardni set alata za pristup skupovima podataka i napomenama, omogućava procenu i poređenje različitih metoda i izvršava izazove za ocenjivanje performansi na prepoznavanju klase objekata.

##### 4.1.1.2 Arhitektura

Plain baselines uglavnom su inspirisane filozofijom VGG mreža (konvolucionarna neuronska mreža koja se obučava na više od milion slika iz baze ImageNet). Konvolucionarni slojevi uglavnom imaju  $3 \times 3$  filtera i slede dva jednostavna pravila:

- Za istu mapu izlaznih karakteristika, slojevi imaju isti broj filtera;
- Ako se veličina karte karakteristika prepolovi, broj filtera se udvostručuje kako bi se sačuvala vremenska složenost svakog sloja. Vredno je primetiti da ResNet model ima manje filtera i nižu složenost od VGG mreža.

#### Residual Network:

Na osnovu gore navedene jednostavne mreže ubačena je veza za prečicu koja mrežu pretvara u preostalu verziju. Prečice za identitet  $F(x\{V\} + x)$  mogu se direktno koristiti kada su ulaz i izlaz istih dimenzija. Kada se dimenzije povećavaju), razmatra dve mogućnosti:

1. Prečica vrši mapiranje identiteta, sa dodatnim nulnim unosima podložnim za povećanje dimenzija. Ova opcija ne uvodi dodatne parametre.
2. Prečica projekcije u  $F(x\{V\} + x)$  koristi se za usklađivanje sa dimenzijama.

#### 4.1.1.3 Implementacija

Slika se menja u veličinu tako da je kraća strana nasumično uzorkovana u  $[256,480]$  radi povećanja skale. "224x224 crop" nasumično se uzima iz slike ili njenog horizontalnog preokreta, pri čemu se oduzima srednja vrednost po pikselu. Stopa učenja počinje od 0,1 i deli se sa 10 kada se platoi grešaka i modeli obučavaju do  $60 \times 10000$  iteracija. Koriste opadanje težine od 0,0001 i momentum 0,9. napona mreže.

#### 4.1.1.4 Rezultat

Mreža sa 18 slojeva samo je potprostor u mreži sa 34 nivoa i još uvek deluje bolje. ResNet ima bolji napredak u slučaju da je mreža dublja.

## 4.2 JUPYTER okruženje

Za realizaciju CNN mreže i primenu algoritama iz fastai biblioteke korišteno je jupyter okruženje. Jupyter notebook (ranije IPython Notebooks) je veb aplikacija otvorenog koda. To je interaktivno računarsko okruženje za kreiranje Jupyter notebook dokumenata. Izraz notebook može kolokvijalno da se odnosi na mnogo različitih entiteta, u zavisnosti od konteksta, uglavnom na Jupyter veb aplikaciju, Jupyter Python veb server ili Jupyter document format. Jupyter notebook dokument je json dokument, koji sadrži listu ulaznih / izlaznih ćelija koje mogu sadržati kod, tekst matematiku, grafike i druge medije, obično sadrže „ipinb“ ekstenziju.

## 5. KLASIFIKACIJA PRIMENOM "FASTAI" BIBLIOTEKE

U ovom poglavlju će biti opisani svi koraci koji su se koristili prilikom izgradnje modela.

### 5.1 Formiranje baze podataka

Prva i osnovna stvar u klasifikaciji slika je ništa drugo do prikupljanje samih podataka, tj. slika.

Postoji više mesta sa kojih možemo da prikupimo podatke, neki od njih su akademski skupovi podataka, *Kaggle competition datasets* i mnogi drugi.

Mesto sa kog su prikupljeni podaci koji su korišćeni u ovom radu je *Google*.

Prvi korak u procesu prikupljanja podataka je odlazak na <https://images.google.com/> i pronalazak se odgovarajućih slika. Sledeća stvar koja treba da se uradi jeste da se dobije spisak svih URL adresa. Da bi to bilo učinjeno, unutar prozora sa google slikama se pritisne komanda pristupi Javascript konzoli i zadavanjem putanje:

```
urls = Array.from(document.querySelectorAll('.rg_di .rg_me
ta')).map(el=>JSON.parse(el.textContent).ou);
window.open('data:text/csv;charset=utf-
8,' + escape(urls.join("\n")));
```

Na ovaj način se dobija URL slika koje smo želeli. U ovom slučaju to su slike tri vrste majmuna (orangutan, šimpanza i gorila ) i ljudi (muškarac, žena, dete ).

## 5.2 Kreiranje skupova podataka za obradu i validaciju

Kao što je već napomenuto pošto uzimamo podatke sa interneta, mogu da se podkadu neki neispravni podaci ili podaci koji nisu relevantni za naše potrebe. Te je naredna stvar pored brisanja neispravnih podataka sledeća stvar koja treba da se uradi je da se uklone slike koje u stvari i nisu slike. To se dešava stalno. U svakoj seriji uvek se nalazi nekoliko slika koje su iz bilo kog razloga oštećene. Vrlo često, kada se preuzima skup podataka iz nekog akademskog skupa podataka, postoje folderi nazvane trening, validacija i test [4] koji sadrže različite skupove podataka.

U ovom slučaju ne postoji zaseban skup provera validacije jer su slike uzete sa “Google“ pretrage. Ali kako bismo znali da li naš model ide u dobrom pravcu potreban nam je validacioni set.

## 5.3 “data.show\_batch”

Da bi imali uvid u podatke sa kojima radimo nije na odmet prikazati iste. Naredba “data.show batch” se može koristiti za prikazivanje nekih od sadržaja u grupi podataka

## 5.4 Treniranje modela

Posle pripreme podataka može se pristupiti obučavanju modela. Odnosno njegovom treniranu, što se u fastai-ju realizuje uz pomoć funkcije *learner*.

## 5.5 Otključavanje (Unfreezing), fine-tuning i learning rates

Kako poboljšati model? Može se poboljšati upotrebom precizne prilagodbe. Do sada su opremljene četiri epohe i to je prošlo prilično brzo. Razlog zbog kojeg je brzo prošla je taj što je korišćen mali trik. Ove konvolutivne mreže imaju mnogo slojeva.

Ono što je urađeno jeste da je dodato nekoliko dodatnih slojeva do kraja i samo su obučeni. Uglavnom je ostavljena većina modela upravo onakav kakav je bio, tako da je to zaista brzo. Ako se model gradi na nečemu što je slično, originalnom, unapred obučenom modelu, to deluje prilično dobro.

Ali ono što stvarno želimo je da se vratimo i obučimo čitav model. To je razlog zašto se prilično često koristi ovaj proces u dve faze. Kada podrazumevano, na

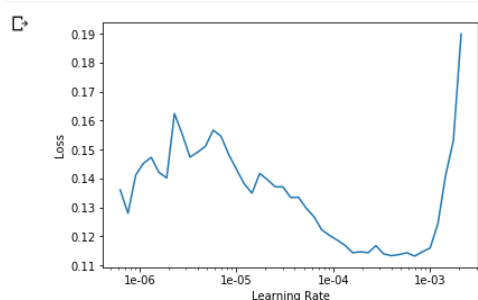
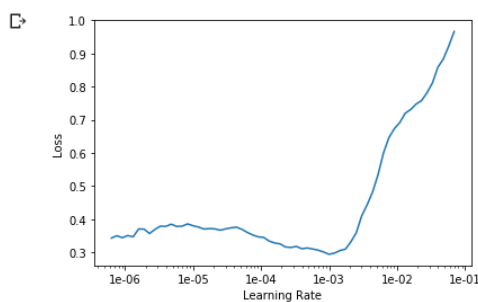
ConvLearner-u zovemo fit ili `fit_one_cycle`, jednostavno će se prilagoditi ovih nekoliko dodatnih slojeva koji su dodani do kraja i pokrenuće se vrlo brzo.

To u osnovi nikada neće biti prekomerno, ali da bi se to zaista postiglo, mora se pozvati *unfreezing* odnosno otključavanje . Otključavanje je stvar koja će da obučiti ceo model.

## 5.6 Pronalaženje *learning rate-a*

Konkretno, stopa učenja je podesivi hiperparametar koji se koristi u treningu neuronskih mreža koji ima malu pozitivnu vrednost, često u rasponu između 0,0 i 1,0. Stopa učenja kontroliše koliko brzo se model prilagođava problemu.

Iscrtaćemo dva grafika. Prvi grafik je grafik iz klase ljudi a drugi majmuna.



Korišćenjem ključne reči u Python-u koja se zove *slice* i koja može uzeti početnu i stop vrednost i u osnovi ono što kaže je treniranje prvih slojeva sa stopom učenja od  $3e-5$ , a poslednji slojevi brzinom od  $3e-4$  u prvom slučaju i  $3e-6$  i  $3e-3$  u drugom, i rasporedite sve ostale slojeve preko toga (tj. Između te dve vrednosti podjednako). Ispostavi se da se najblji rezultati za klasu ljudi dobiju u prvoj tabeli a majmuna u drugoj.

epoch	train_loss	valid_loss	error_rate	time
0	0.126556	0.311108	0.097345	02:51
1	0.082458	0.323523	0.097345	02:51

epoch	train_loss	valid_loss	error_rate	time
0	0.142270	0.342863	0.080357	03:49
1	0.148123	0.273799	0.053571	03:51

## 5.7 Interpretacija

Da bi se videlo šta izlazi iz našeg modela, može da se koristi ovu klasu za interpretaciju klase.

Koristiće se ova fabrička metoda od učenika, tako da se prelazi u naučeni objekt.

Predmet učenja zna dve stvari:

1. Koji su naši podaci
2. Koji je naš model. Sada to nije samo arhitektura, već je obučeni model

To su sve informacije koje su nam potrebne za tumačenje tog modela.

### 5.7.1 Confusion matrix i Most confused

U oblasti mašinskog učenja, a posebno problema statističke klasifikacije, matrica konfuzije, poznata i kao matrica grešaka. Matrica konfuzije [9] je tabela koja se često koristi za opisivanje modela klasifikacije (ili „klasifikatora“) na skupu podataka ispitivanja za koje su poznate prave vrednosti.

Omogućava vizuelizaciju performansi algoritma. Omogućava lako prepoznavanje konfuzije između klasa, npr. jedna klasa se obično pogrešno označava kao druga. Većina mera performansi izračunava se iz matrice konfuzije.

Broj tačnih i pogrešnih predviđanja sažima se sa vrednostima brojanja i raščlanjuje ih o svakoj klasi. To nam daje uvid ne samo u greške koje radi klasifikator, već je važnije i vrste grešaka koje se čine.

Pored matrice konfuzije koristi se i komanda Most confused. Krajnji cilj obe stvari je isti, ali ako imamo mnogo klasa pogodnije je koristiti Most confused. ost confused će jednostavno izdvojiti iz matrice konfuzije posebne kombinacije predviđenog i stvarnog kod kojih se najčešće javlja greška .

## 5.8 Stavljanje modela u produkciju

Ono što bi moglo da bude interesantno je da na osnovu nasumično odabrane slike vidimo da li je i koliko naš model u stanju da pogodi o kojoj slici je reč.

U nastavku je dat po jedan primer za svaku od opisanih klasa prikazane na slikama 2 i 3. U prvom slučaju otvorena je slika gorile što je naš model pogodio, u drugom primeru je u pitanju devojčica što je sa jednakom tačnošću i utvrđeno.



Slika2. Pogadanje iz mreže koje je svrstalo sliku u kategoriju 'gorila'



Slika 3. Pogadanje iz mreže koje je svrstalo sliku u kategoriju 'dete'

## 6. ZAKLJUČAK

Klasifikacija slika je kao tema veoma interesantna i često primenjivana u današnjem vremenu.

Kao što je u radu napomenuto domen primene same teme je veoma širok. Takođe smo se upoznali sa nekim od primera primene same klasifikacije, a koliko je tema rasprostranjena govori to da se koristi i za biomedicinsko snimanje, biometriju, video-nadzor, navigaciju vozila, industrijski vizuelni pregled, navigaciju robota i daljinsko istraživanje.

U ovom radu je prikazano kako možemo klasifikovati slike primenom fastai biblioteke.

Mogli smo da primetimo da korišćenje pomenute biblioteke dosta ubrzava i olakšava samu klasifikaciju, a da se pri tom ne gubi na tačnosti i pouzdanosti modela.

Prolazilo se kroz sve faze procesa klasifikacije, prikazani su i neki probleme do kojih se može doći tokom rada, takođe je pokazano kako te probleme rešiti kao i na koji način unaprediti sam model.

## 7. LITERATURA

[1] Image Classification. Dostupno na:

<https://www.sciencedirect.com/topics/computer-science/image-classification>

[2] Rethinking the Inception Architecture for Computer Vision. Dostupno na:

<https://arxiv.org/abs/1512.00567/>

[3] Dictionary of computer vision and image processing.

[4] About Train, Validation and Test Sets in Machine Learning. Dostupno na:

<https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7>

### Kratka biografija:



**Jovana Antić** rođena je u Beogradu 1995. god. Master rad na Fakultetu tehničkih nauka iz oblasti Elektrotehnike i računarstva – Adaptivno i napredno upravljanje odbranila je 2019.god.

Kontakt: anticj9@gmail.com