

Bird species recognition using deep learning

Projekat iz Računarske inteligencije
Matematički fakultet
Univerzitet u Beogradu

Valentina Tošić
mi19212@alas.matf.bg.ac.rs

Maj 2024

Sadržaj

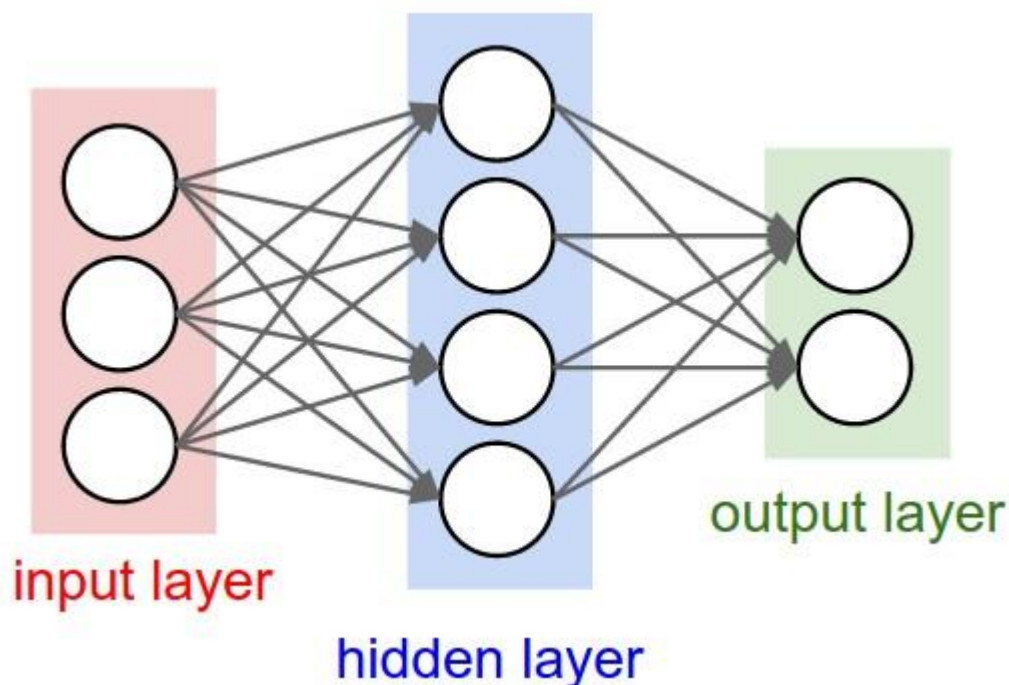
1 Uvod	3
2 Opis	4
3 Rešenje	4
3.1 Pretprocesiranje	4
3.2 Moj model	5
3.3 ResNet18	7
3.4 VGG19	9
4 Zaključak	11
5 Literatura	11

1 Uvod

Duboko učenje (eng. Deep Learning), je grana mašinskog učenja koja se sastoji od neuronske mreže sa tri ili više slojeva:

- Ulazni sloj: podaci ulaze kroz ovaj sloj.
- Sakriveni slojevi: obrađuju i prenose podatke na druge slojeve.
- Izlazni sloj: krajnji rezultat se formira u izlaznom sloju.

Neuronska mreža je računarski model koji funkcioniše na sličan način kao neuroni u ljudskom mozgu. Svaki neuron uzima ulaz, izvodi neke operacije, a zatim prosledjuje izlaz sledećem neuronu.



Treniranje dubokih mreža je zahtevan proces, pri kome se koristi dosta resursa, a često je potrebna i velika količina podataka za obuku neuronske mreže. Dok jedan sloj neuronske mreže može da donese korisna, približna predviđanja i odluke, dodatni slojevi mreže pomažu da se preciziraju i optimizuju rezultati radi veće tačnosti.

2 Opis

Tema projekta je prepoznavanje ptica korišćenjem neuronskih mreža. Skup podataka je sačinjen od 100 različitih vrsta ptica. Za rešavanje problema sam koristila svoju mrežu, kao i neke već gotove arhitekture i pretrenirane mreže.

3 Rešenje

3.1 Pretprocesiranje

Pretprocesiranje je veoma bitan korak zbog transformacije slika u format pogodan za model. Podaci su podeljeni u dve kategorije, podaci za obuku i za testiranje. Podaci za obuku se koriste za učenje CNN mreže, dok se na kraju za performanse tačnosti mreže koriste podaci za testiranje, sa kojima se mreža prvi put susreće. Na slike je primenjena rotacija za 10 stepeni, na slucajan način odabrane, neke slike su horizontalno okrenute i skraćene, skaliranje slika na 80x80. Slike su učitane kao RGB(3 kanal).



3.2 Moj model

Jednostavna konvolutivna neuronska mreža (CNN) koja se sastoji iz nekoliko slojeva, i to:

- Konvolutivni slojevi: Ovi slojevi primjenjuju konvolutivne operacije na ulaznu sliku kako bi se izdvojile njene karakteristike. Prvi sloj (conv1) prima ulaz sa 3 kanala (RGB slika) i generiše 32 mape karakteristika. Drugi sloj (conv2) prima 32 mape iz prethodnog sloja i generira 64 mape.

- Pooling sloj: služi za smanjenje prostornih dimenzija mapa uz zadržavanje važnih informacija.

- Potpuno povezani slojevi: potpuno (gusto) povezani slojevi koji izimaju “izravnat” izlaz iz poslednjeg konvolutivnog sloja i vrše klasifikaciju.

- Flatten sloj: pravi vektor od izlaza poslenjeg konvolutivnog sloja. Za aktivacionu funkciju je korišćen ReLu, koja se primenjuje nakon svakog konvolutivnog i potpuno povezanog sloja kako bismo uveli određenu nelinearnost u mrežu.

ConvNet(

(conv1): Conv2d(3, 32, kernel_size=(5, 5), stride=(1, 1))

(pool): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)

(conv2): Conv2d(32, 64, kernel_size=(5, 5), stride=(1, 1))

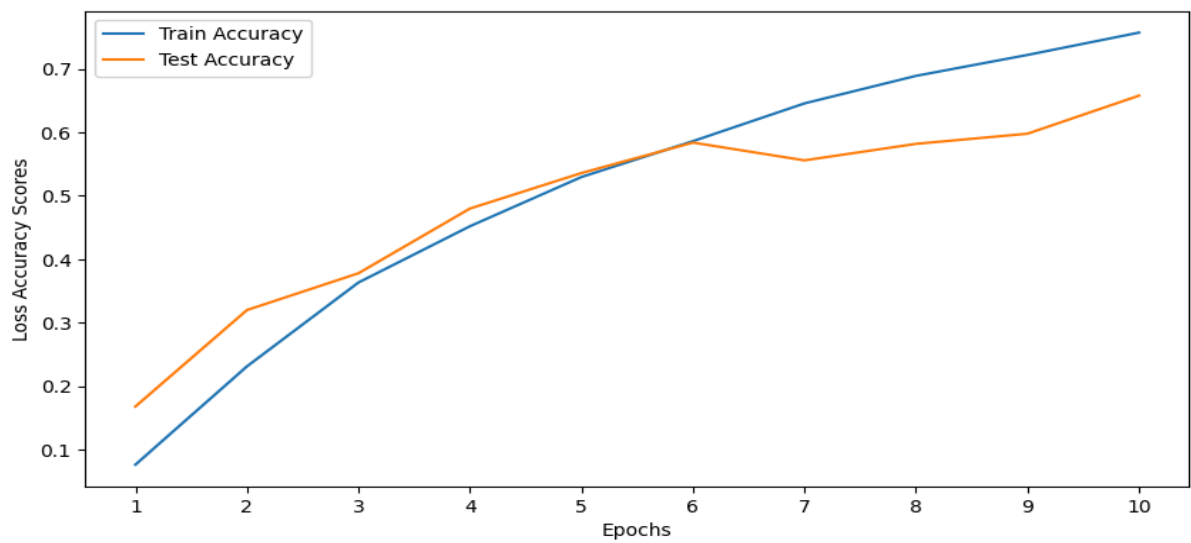
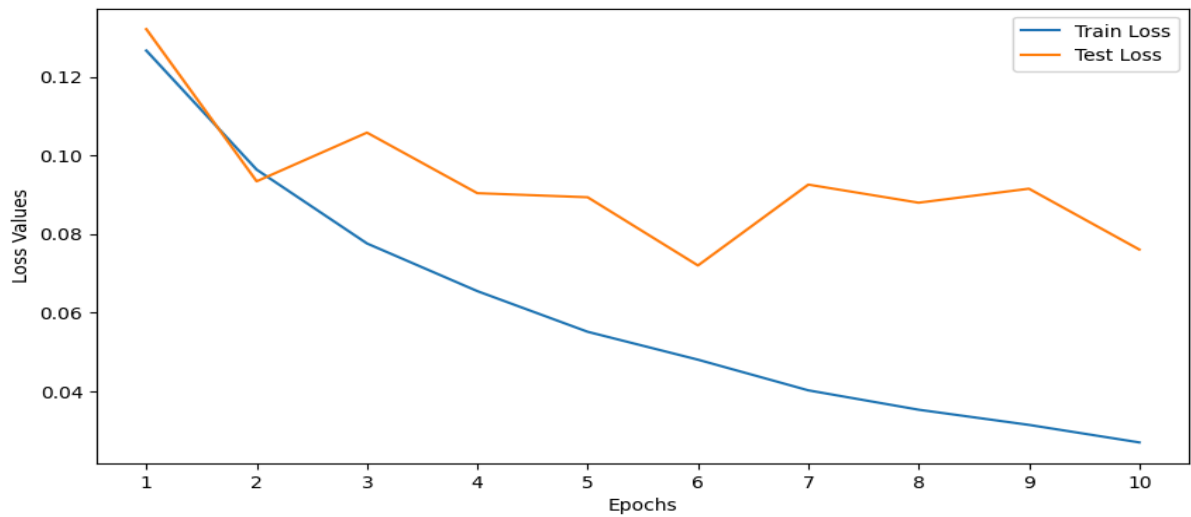
(flatten): Flatten(start_dim=1, end_dim=-1)

(fc1): Linear(in_features=18496, out_features=1600, bias=True)

(fc2): Linear(in_features=1600, out_features=512, bias=True)

(fc3): Linear(in_features=512, out_features=100, bias=True)

Mreža ima ukupno 7 slojeva i 30.5 miliona parametara.



Average loss: 0.13211749839782716
 Accuracy: 0.168
 Average loss: 0.09340506887435913
 Accuracy: 0.32
 Average loss: 0.10579317617416382
 Accuracy: 0.378
 Average loss: 0.09038523030281066
 Accuracy: 0.48
 Average loss: 0.08934893298149109
 Accuracy: 0.536
 Average loss: 0.07199175083637238
 Accuracy: 0.584
 Average loss: 0.092569819688797
 Accuracy: 0.556
 Average loss: 0.08796060156822205
 Accuracy: 0.582
 Average loss: 0.09152399110794067
 Accuracy: 0.598
 Average loss: 0.0760469069480896
 Accuracy: 0.658

3.3 ResNet18

S obzirom da ResNet radi sa slikama koje su velicine 224x224, nije bilo potrebe skalirati ih. Svaki sloj ResNet-a se sastoji od nekoliko blokova. ResNet ide dublje u mrežu povećavanjem broja operacija u okviru bloka, dok ukupan broj slojeva ostaje isti. Operacija se ovde odnosi na konvoluciju, normalizaciju serije i ReLu aktivaciju na ulazne podatke. U PyTorch implementaciji razlikujemo blokove koji uključuju dve operacije(Basic Block) i tri operacije(Bottleneck Block).

Layer Name	Output Size	ResNet-18
conv1	$112 \times 112 \times 64$	$7 \times 7, 64, \text{stride } 2$
conv2_x	$56 \times 56 \times 64$	$3 \times 3 \text{ max pool, stride } 2$
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
conv3_x	$28 \times 28 \times 128$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
conv4_x	$14 \times 14 \times 256$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
conv5_x	$7 \times 7 \times 512$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
average pool	$1 \times 1 \times 512$	$7 \times 7 \text{ average pool}$
fully connected	1000	$512 \times 1000 \text{ fully connections}$
softmax	1000	

ResNet je već obučen model, za treniranje sam izmenila njegov poslednji potpuno povezan sloj i trenirala samo njega.

```
Epoch 0/4
-----
train Loss: 0.6713 Acc: 0.8968
test Loss: 0.4694 Acc: 0.9380

Epoch 1/4
-----
train Loss: 0.5640 Acc: 0.9089
test Loss: 0.4101 Acc: 0.9400

Epoch 2/4
-----
train Loss: 0.4949 Acc: 0.9186
test Loss: 0.3682 Acc: 0.9440

Epoch 3/4
-----
train Loss: 0.4450 Acc: 0.9237
test Loss: 0.3315 Acc: 0.9440

Epoch 4/4
-----
train Loss: 0.4121 Acc: 0.9333
test Loss: 0.3290 Acc: 0.9520
```

ResNet je dao veoma dobre rezultate, sa tačnošću 95%. Neki od promašaja koje je imao je zbog velike sličnosti nekih vrsta prica, kao na primer:

ABBOTTS BABBLER
AZARAS SPINETAIL

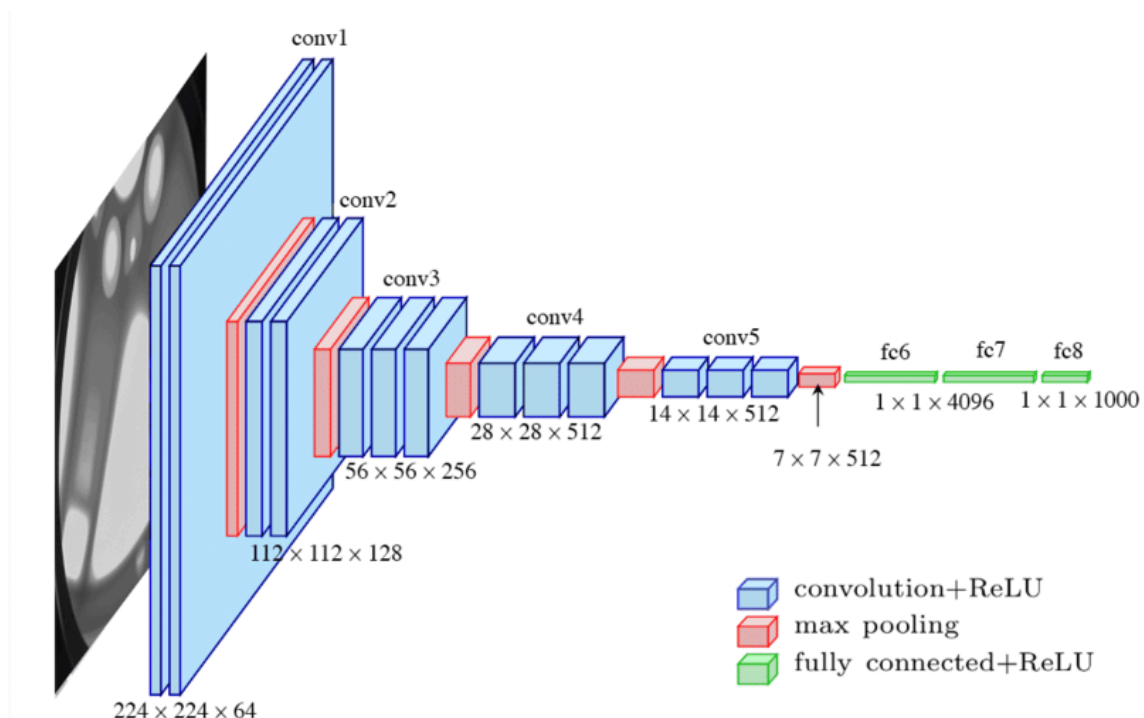


ABBOTTS BABBLER
ABBOTTS BABBLER

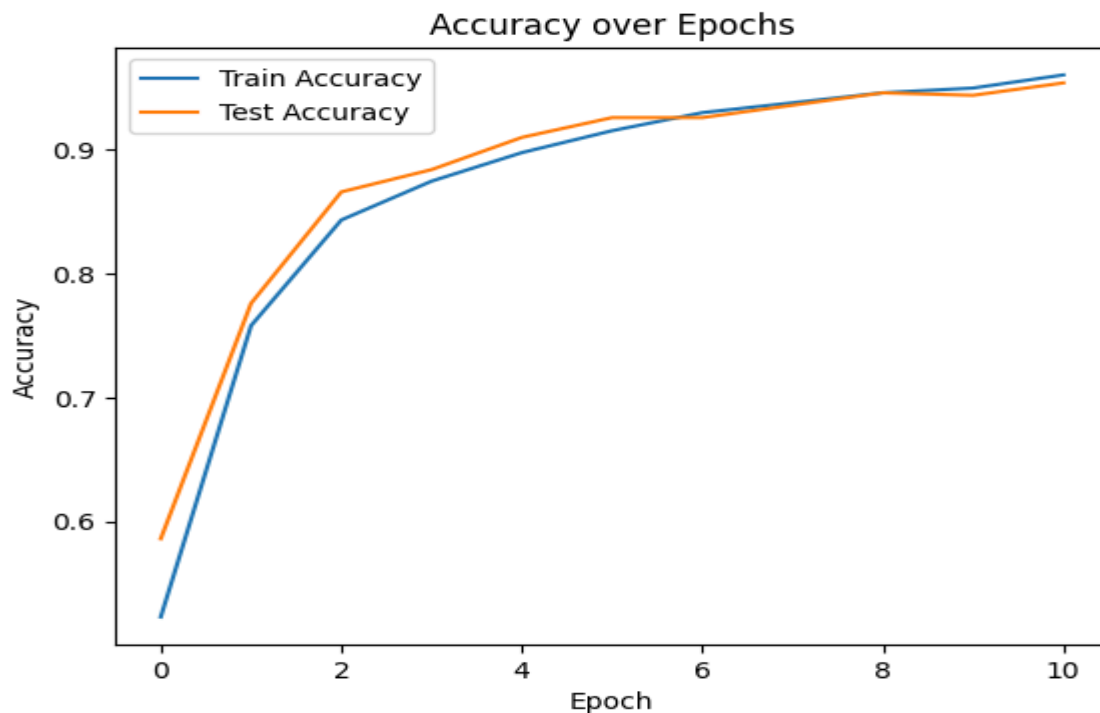


3.4 VGG19

Model konvolutivnih neuronskih mreža (CNN), sa brojnim slojevima od kojih su 19 konvolutivni slojevi i sa 70.5 miliona parametara. Na ulazu prima slike veličine 224×224 . Svi skriveni slojevi u VGG mreži koriste ReLu, ima tri potpuno povezana sloja, od kojih poslednji ima 1000 kanala. Sastoji se iz blokova, gde je svaki blok sastavljen od 2DConvolution and MaxPooling slojeva.



Svi modeli su bili trenirani na platformi Google Colab, zbog velikog broja parametara i slojeva.



```
Epoch: 1 Step: 300 Loss: 4.1234297490119936
Test Accuracy: 0.572
Epoch: 2 Step: 600 Loss: 2.822670192718506
Test Accuracy: 0.794
Epoch: 2 Step: 900 Loss: 2.0283568759759265
Test Accuracy: 0.872
Epoch: 3 Step: 1200 Loss: 1.572364919781685
Test Accuracy: 0.896
Epoch: 4 Step: 1500 Loss: 1.327743706703186
Test Accuracy: 0.916
Epoch: 4 Step: 1800 Loss: 1.1127227675914764
Test Accuracy: 0.914
Epoch: 5 Step: 2100 Loss: 0.9873341811696689
Test Accuracy: 0.938
Epoch: 5 Step: 2400 Loss: 0.8678135692079862
Test Accuracy: 0.944
Epoch: 6 Step: 2700 Loss: 0.7655240209897359
Test Accuracy: 0.942
Epoch: 7 Step: 3000 Loss: 0.7106464616457622
Test Accuracy: 0.94
Epoch: 7 Step: 3300 Loss: 0.6383654770751794
Test Accuracy: 0.952
```

4 Zaključak

Na ovom skupu podataka od 100 vrsta ptica, primećen je značajan jaz u performansama između klasičnog modela i pretreniranih modela. Klasičan model se borio da pravilno klasifikuje slike, postigavši samo stopu tačnosti od 65%, performanse koje su nešto superiornije od nasumičnog pogađanja u binarnoj klasifikaciji. Nasuprot ovom modelu, pretrenirani modeli su pokazali odlične performanse, postigavši tačnost veću od 95%. Ova veća preciznost se verovatno može prepisati unapred obučenim konvolutivnim slojevima, koji su naučeni da prepoznaju određene karakteristike na skupu ImageNet.

5 Literatura

- [1] Materijali sa kursa Računarska inteligencija
- [2] Deep Learning Bible - 2. Classification - Eng.