

Prepoznavanje karaktera sa tablica

Filip Jovašević, Luka Milošević, Stefan Lazović

Matematički fakultet
Univerzitet u Beogradu

25. Maj 2019.



Pregled

1 Uvod

- Opis problema?

2 Preprocesiranje

- Početni izgled podataka
- Obrada slika

3 Neuronske mreže

- Konvolutivne neuronske mreže - CNN
- Rekurentne neuronske mreže - RNN

4 Treniranje modela i rezultati

- Izgled modela
- Rezultati treninga
- Testiranje dobijenog modela

Opis problema?

- Primena neuronskih mreža i fazi logike u prepoznavanju karaktera sa tablica
- Skup podataka koji koristimo sastoji se od tablica koje su izgenerisane automatski, to jest ne predstavljaju prave tablice
- Na slikama se nalaze samo tablice

Pregled

- 1 Uvod
 - Opis problema?
- 2 Preprocesiranje
 - Početni izgled podataka
 - Obrada slika
- 3 Neuronske mreže
 - Konvolutivne neuronske mreže - CNN
 - Rekurentne neuronske mreže - RNN
- 4 Treniranje modela i rezultati
 - Izgled modela
 - Rezultati treninga
 - Testiranje dobijenog modela

Početni izgled podataka



- Anotacije o svakoj od tablica u .json formatu:

```
“tags”: [“test”],  
“description”: “A007HA50”,  
“objects”: [],  
“size”: “height”: 34, “width”: 152
```

Izgleđ podataka nakon fazifikacije



- Pomoću fazi logike izoštrićemo brojeve i slova na tablici
- INT algoritam, ili algoritam intenziteta, je algoritam koji spada u grupu fazi algoritama i koristi se za popravljjanje slike isticanjem kontrasta
- Učenje modela se izvršava nad slikama koje su crno bele tj. nalaze se u grayscale obliku

Izgled podataka nakon fazifikacije

- Vrednosti ulazne slike X , dimenzija $M \times N$, fazifikuju se primenom funkcije μ :

$$\mu_{ij} = \frac{x_{ij} - x_{min}}{x_{max} - x_{min}},$$

gde je $0 \leq i < M, 0 \leq j < N$, x_{max} maksimalna, a x_{min} minimalna vrednost piksela na slici X i x_{ij} vrednost piksela na mestu ij slike X

Izgled podataka nakon fazifikacije

- Nakon što smo napravili fazi skup prelazimo na fazu modifikacije, tu se primenjuje druga karakteristična funkcija koja će poboljšati kontrast

$$\bar{\mu}_{ij}(\mu_{ij}) = \begin{cases} 2 \cdot \mu_{ij}^2 & 0 \leq \mu_{ij} \leq \mu_c \\ 1 - (2 \cdot (1 - \mu_{ij})^2) & \mu_c < \mu_{ij} \leq 1 \end{cases} \quad (1)$$

- Sada je neophodno vratiti se iz fazi domena u domen slike, izlazna slika Y se dobija primenom inverzne funkcije μ^{-1} :

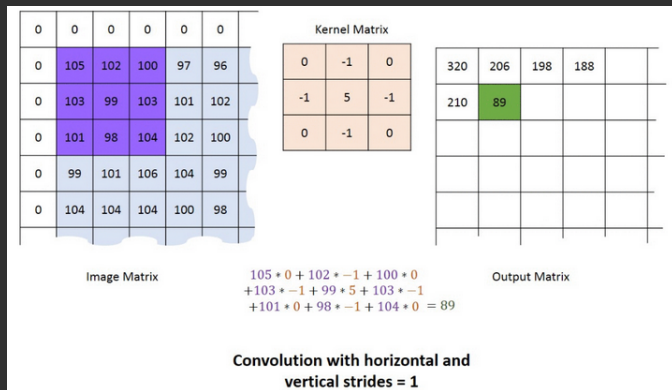
$$y_{ij} = \mu^{-1}(\bar{\mu}_{ij}) = x_{min} + \bar{\mu} \cdot (x_{max} - x_{min}).$$

Pregled

- 1 Uvod
 - Opis problema?
- 2 Preprocesiranje
 - Početni izgled podataka
 - Obrada slika
- 3 **Neuronske mreže**
 - Konvolutivne neuronske mreže - CNN
 - Rekurentne neuronske mreže - RNN
- 4 Treniranje modela i rezultati
 - Izgled modela
 - Rezultati treninga
 - Testiranje dobijenog modela

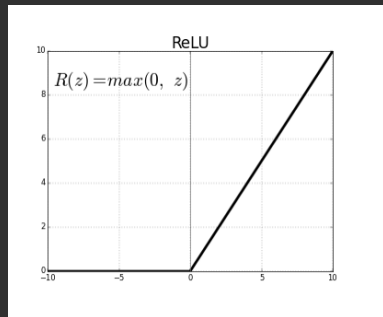
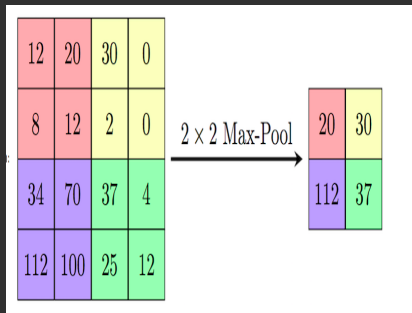
Konvolutivne neuronske mreže - CNN

- Pogodna za obradu slika
- Rad konvolutivne mreže prikazan na slici, korišćen kernel (3, 3)
- Neki od primenjivanih parametara učenja su: SGD, RMSprop i Nadam



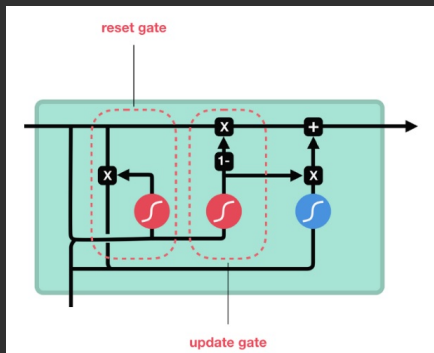
Konvolutivne neuronske mreže - CNN

- Korišćen je MaxPooling sa veličinom (2, 2)
- Korišćena funkcija aktivacije je Relu (ispravljena linearna funkcija aktivacije), jer ona radi po principu dodeljivanja verovatnosnih vrednosti za svaki od n izlaza



GRU - Gated Recurrent Unit

- Ima ulogu da obrađuje dalje rezultate CNN mreža
- Pogodne su za prepoznavanje karaktera
- Određuju verovatnoću da će se neki karakter naći na nekom mestu



Pregled

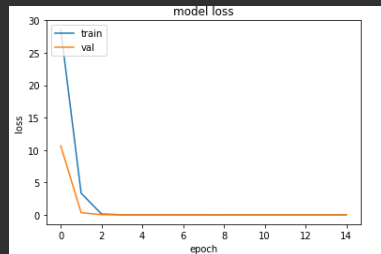
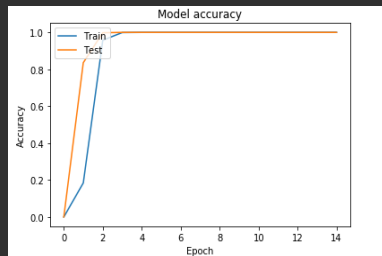
- 1 Uvod
 - Opis problema?
- 2 Preprocesiranje
 - Početni izgled podataka
 - Obrada slika
- 3 Neuronske mreže
 - Konvolutivne neuronske mreže - CNN
 - Rekurentne neuronske mreže - RNN
- 4 **Treniranje modela i rezultati**
 - Izgled modela
 - Rezultati treninga
 - Testiranje dobijenog modela

Izgled modela

Layer (type)	Output Shape	Param #	Connected to
the_input (InputLayer)	(None, 128, 64, 1)	0	
conv1 (Conv2D)	(None, 128, 64, 16)	160	the_input[0][0]
max1 (MaxPooling2D)	(None, 64, 32, 16)	0	conv1[0][0]
conv2 (Conv2D)	(None, 64, 32, 16)	2320	max1[0][0]
max2 (MaxPooling2D)	(None, 32, 16, 16)	0	conv2[0][0]
reshape (Reshape)	(None, 32, 256)	0	max2[0][0]
dense1 (Dense)	(None, 32, 32)	8224	reshape[0][0]
gru1 (GRU)	(None, 32, 512)	837120	dense1[0][0]
gru1_b (GRU)	(None, 32, 512)	837120	dense1[0][0]
add_1 (Add)	(None, 32, 512)	0	gru1[0][0] gru1_b[0][0]
gru2 (GRU)	(None, 32, 512)	1574400	add_1[0][0]
gru2_b (GRU)	(None, 32, 512)	1574400	add_1[0][0]
concatenate_1 (Concatenate)	(None, 32, 1024)	0	gru2[0][0] gru2_b[0][0]
dense2 (Dense)	(None, 32, 23)	23575	concatenate_1[0][0]
softmax (Activation)	(None, 32, 23)	0	dense2[0][0]

Preciznost i gubici

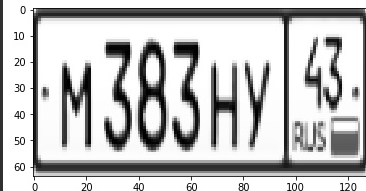
- Rezultati dobijeni treniranjem mreže kroz 15 epoha
- U svakoj od epoha prosledjivano po oko 100 slika



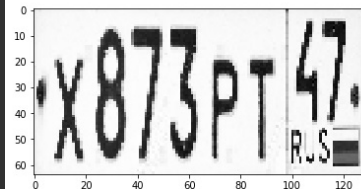
- Kasnije smo se opredelili za 2 epohe kojima prosleđujemo sve slike
- Radi uštede vremena a dobijanja još manjih gubitaka

Test primer

Predicted characters: M383HY43



Predicted characters: X87PT42



Slika: Test nad tablicom iz test skupa

Slika: Test nad tablicom iz realnog života

Hvala na pažnji.
Imate li pitanja?

Literatura

- Illustrated Guide to LSTM's and GRU's: A step by step explanation. on-line at: <https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21/>
- Image ocr. on-line at: https://keras.io/examples/image_ocr/
- What is kernel size, 2017. on-line at: <https://stats.stackexchange.com/questions/296679/what-does-kernel-size-mean>
- Max-pooling / Pooling, 2018. on-line at: https://computersciencewiki.org/index.php/Max-pooling/_/_Pooling.
- ANPR OCR dataset, 2019. on-line at: <https://supervise.ly/explore/projects/anpr-ocr-21232/datasets>
- Model visualization, 2019. on-line at: [https://keras.io/ visualization/](https://keras.io/visualization/)
- Nebojša Perić. Neke primene teorije fazi skupova i fazi logike u procesi- ranju slika, 2014. on-line at: http://www.racunarstvo.matf.bg.ac.rs/MasterRadovi/2014_01_29_Nebojsa_Peric/rad.pdf