

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



Experimenty s OCR na báze hlbokého učenia

DIPLOMOVÁ PRÁCA

2020

Peter Kulcsár Szabó

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

Experimenty s OCR na báze hlbokého učenia

DIPLOMOVÁ PRÁCA

Študijný program: Aplikovaná informatika
Študijný odbor: 11378 Aplikovaná informatika
Školiace pracovisko: Katedra aplikovanej informatiky
Vedúci práce: RNDr. Andrej Lúčny, PhD.

Bratislava 2020

Peter Kulcsár Szabó



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Peter Kulcsár Szabó
Študijný program: aplikovaná informatika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Experimenty s OCR na báze hlbokého učenia
Experiments with OCR using Deep Learning

Anotácia: Práca má výskumný charakter a je z oblasti hlbokého učenia a vychádza z jednoduchšej architektúry hlbkej neurónovej siete použitej pre OCR individuálneho písma. Skúša analógiu triku, ktorý pomáha zvýšiť úspešnosť klasického OCR na OCR pomocou hlbokého učenia.

Cieľ: Cieľom práce je vyskúšať tréning OCR na báze hlbkej neurónovej siete a vyhodnotiť či rozšírenie klasifikácie znakov o príznaky ako sú počet veľkých dier v stredovej vertikálnej línii, prítomnosť vonkajších a vnútorných oblúkov a rohov, zúženie v strede zľava a sprava, záseky v obvodovej línii zľava v dolnej polovici a sprava v dolnej polovici, spojitost' vnútra s ľavou či pravou stranou a podobne zlepši dosiahnutú úspešnosť. Pritom príznaky možno požadovať na výstupnej vrstve zo siete alebo priamo v jej latentom priestore. Cieľom práce nie je dosiahnuť čo najvyššiu úspešnosť, ale zistiť či danou úpravou architektúry úspešnosť klesá alebo stúpa.

Literatúra: Chollet, F.: Deep learning v jazyku Python, Grada, 2019
Learning OpenCV 3, Computer Vision in C++ with the OpenCV Library By Gary Bradski, Adrian Kaehler, O'Reilly Media, 2016
Smith, R.: An Overview of the Tesseract OCR Engine. ICDAR '07 Proceedings of the Ninth International Conference on Document Analysis and Recognition, Volume 02, pp. 629-633
Lúčny, A: Čítanie textu na pneumatike, KUZ 2018, M.U. Brno
learnopencv.com

Poznámka: Platforma: OpenCV, Keras

Kľúčové slová: OCR, hlboké učenie, počítačové videnie, OpenCV, Keras

Vedúci: RNDr. Andrej Lúčny, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: prof. Ing. Igor Farkaš, Dr.
Dátum zadania: 08.10.2019

Abstrakt v štátnom jazyku

KULCSÁR SZABÓ, Peter: Experimenty s OCR na báze hlbokého učenia [Diplo-
mová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky
a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: RNDr. An-
drej Lúčný, PhD., Bratislava, 2021, XX s

Lorem ipsum

Kľúčové slová: xx

Abstract

KULCSÁR SZABÓ, Peter: Experiments with OCR using Deep Learning [Master Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: RNDr. Andrej Lúčny, PhD., Bratislava, 2021, XX p.

Lorem ipsum

Key words: xx

Obsah

Úvod	7
1 Popis problému	8
2 Dáta	9
3 Prostredie	11
4 Modely	12
4.1 LeNet	12
4.2 ResNet	12
4.3 Custom	12
5 Gro	13
6 Výsledky	14
7 Záver	15

Úvod

Lorem ipsum

1 Popis problému

Optické rozpoznávanie znakov (OCR - optical character recogniotion) je rozšírená technológia na rozpoznávanie textu na obrázkoch, ako sú naskenované dokumenty a fotografie. Technológia OCR sa používa na konverziu prakticky všetkých druhov obrázkov obsahujúcich písaný text (písaný, písaný rukou alebo vytlačený) do digitálnej formy.

Technológia OCR sa stala populárnou na začiatku 90. rokov, keď sa pokúšali digitalizovať historické noviny. Odvtedy táto technológia prešla niekoľkými zlepšeniami. V súčasnosti poskytujú riešenia takmer dokonalú presnosť.

Pravdepodobne najznámejším prípadom použitia OCR je konverzia tlačенých papierových dokumentov na strojovo čitateľné textové dokumenty. Keď naskenovaný papierový dokument prejde spracovaním OCR, text dokumentu sa dá upravovať textovými procesormi ako Microsoft Word alebo Google Docs. Predtým, ako bola dostupná technológia OCR, bola jedinou možnosťou digitalizácie tlačенých papierových dokumentov ručné prepisovanie textu. Nielenže to bolo časovo náročné, ale aj nepresnosti a preklepy boli časté.

OCR sa často používa ako „skrytá“ technológia, ktorá poháňa mnoho známych systémov a služieb v našom každodennom živote. Medzi menej známe, ale rovnako dôležité prípady použitia technológie OCR patrí automatizácia zadávania údajov, indexovanie dokumentov pre vyhľadávače (google), automatické rozpoznávanie ŠPZ, ako aj pomoc nevidiacim a slabozrakým osobám.

Technológia OCR sa ukázala ako nesmierne užitočná pri digitalizácii historických novín a textov, ktoré sa teraz zmenili na plne prehľadateľné formáty a uľahčili a urýchlili prístup k týmto historickým textom.

2 Dáta

Diplomová práca sa zaoberá s optickým rozpoznávaním znakov, pre ktorú som vytvoril dve skupiny dát. Obe skupiny obsahujú veľké písmená od a po z a čísla od 0 po 9. V prvej skupine sú čiernobiele dáta. Na bielom ale zašumenom pozadí sú čierne písmená alebo čísla s rôznymi typmi písma a niektoré sú hrubým písmom alebo zakriveným. V druhej skupine sú farebné dáta. Z farebných vzorov som generoval pozadie, a následne písal na farebné písmená a čísla.

Vzorky z prvého datasetu:



Vzorky z druhého datasetu:



K písmenám a číslam som potom vytvoril súbor kde sú popísané ich charakteristiky. Nasledovné charakteristiky som popísal: či má znak roh vpravo hore / dole alebo vľavo hore / dole, zaoblený roh, počet dier, zúženie v stredovej línii a ****. Dáta sú dostupné v nasledovnej tabuľke 1

Let	C TL	C TR	C BL	C BR	R TL	R TR	R BL	R BR	# holes	Red L	Red R	Int L	Int R
A									1				
B	1		1			1		1	2		1		
C					1		1						1
D	1		1			1		1	1				
E	1		1										1
F	1												1
G					1	1	1						1
H													
I													
J							1	1					
K											1		
L			1										
M													
N													
O					1	1	1	1	1				
P	1		1			1			1				
Q					1	1	1	1	1				
R	1					1			1		1		
S					1	1	1	1				1	1
T												1	1
U							1	1					
V													
W													
X										1	1		
Y										1	1		
Z												1	1
0					1	1	1	1	1				
1													
2					1	1						1	1
3					1	1	1	1				1	
4									1				
5	1						1	1				1	1
6					1	1	1	1	1				1
7												1	
8					1	1	1	1	2	1	1		
9					1	1	1	1	1			1	

Tabuľka 1: Charakteristika písmen

Vysvetlenie k tabuľke:

- C - corner (roh)
- R - rounded corner (zaoblený roh)
- Red - reduction (***)
- Int - intersection (***)
- TL - top left (hore vľavo)
- TR - top right (hore vpravo)
- BL - bottom left (dole vľavo)
- BR - bottom right (dole vpravo)
- L - left (vľavo)
- R - right (vpravo)

3 Prostredie

Diplomová práca je programovaná v jazyku Python. Pri programovaní sme využili známe knižnice numpy, keras a matplotlib. Pomocou knižnice keras sme programovali hlboké neurónové siete. Knižnica matplotlib bola využitá na vykresľovanie grafov.

Pre diplomovú prácu som vytvoril prostredie v ktorom sa lahko pridáva nový model s novou architektúrou hlbokého učenia. Prostredie pozostáva z troch hlavných komponentov:

- models.py
- data.py
- helpers.py

V models súbore sa nachádzajú modely ktoré sme testovali počas diplomovej práci. Nachádzajú sa tu vlastné architektúry ale aj známe ako napríklad LeNet alebo ResNet.

V data súbore nájdeme funkcie, ktorými načítavame dáta zo súborov a pridávanie rôznych šumov (Gausovský, PRIDAŤ ĎALŠIE)

V helpers súbore nájdeme pomocné funkcie, ale aj funkcie ktoré vykresľujú a ukladajú grafy do results priečinka.

4 Modely

4.1 LeNet

Lorem ipsum

4.2 ResNet

Lorem ipsum

4.3 Custom

Lorem ipsum

5 Gro

Lorem ipsum

6 Výsledky

Lorem ipsum

7 Závěr

Lorem ipsum