

**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

files/zadanie.pdf

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 Lúčny, 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	10
4 Modely	11
4.1 LeNet	11
4.2 ResNet	11
4.3 Custom	11
5 Gro	12
6 Výsledky	13
7 Záver	14

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 písmena 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.

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 nasledovnom súbore: Charakteristika

treba doplniť typy fonrov

Ukážka dát

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