

České vysoké učení technické v Praze

Fakulta jaderná a fyzikálně inženýrská

Katedra softwarového inženýrství

Akademický rok 2021/2022

ZADÁNÍ VÝZKUMNÉHO ÚKOLU

Student: Bc. Anna Gruberová

Studijní program: Aplikace informatiky v přírodních vědách

Název práce česky: Optické rozpoznávání znaků na naskenovaných historických plakátech pomocí nejmodernějších metod

Název práce anglicky: Optical Character Recognition on Scanned Historical Posters Using the State-of-the-Art Methods

Pokyny pro vypracování:

1. Seznamte se s problematikou optického rozpoznávání znaků. Na základě rešerše vyberte několik metod, se kterými budete dále pracovat a vyhodnocovat úspěšnost jejich detekce.
2. Stáhněte několik volně dostupných datasetů, které jsou využívány v literatuře k porovnání jednotlivých metod na OCR. Dále vytvořte svůj vlastní dataset z obdržených dat.
3. Nastudujte techniky porovnání OCR výstupů s ground-truth.
4. U vybraných metod prostudujte jejich chování na jednotlivých datasetech při různém nastavení parametrů.
5. Navrhněte také možnosti filtrování výstupů jednotlivých metod za účelem snížení falešných detekcí.

Doporučená literatura:

- [1] R. C. Gonzalez, R. E. Woods, Digital Image Processing (4th ed.). Pearson, 2018. ISBN 9353062985.
- [2] GOODFELLOW, Ian, Yoshua BENGIO a Aaron COURVILLE. Deep learning. Cambridge, Massachusetts: The MIT Press, [2016]. ISBN 0262035618.
- [3] SMITH, R. An Overview of the Tesseract OCR Engine. In: Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2 [online]. IEEE, 2007, 2007, s. 629-633. ISBN 0-7695-2822-8. ISSN 1520-5363. Dostupné z: doi:10.1109/ICDAR.2007.4376991
- [4] CHEN, Xiaoxue, et al. Text Recognition in the Wild. ACM Computing Surveys [online]. 2021, 54(2), 1-35 [cit. 2021-10-2]. ISSN 0360-0300. Dostupné z: doi:10.1145/3440756

Jméno a pracoviště vedoucího práce:

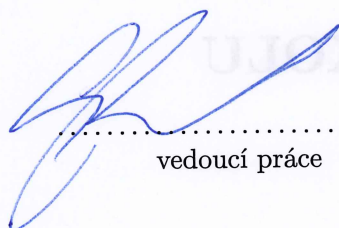
Ing. Adam Novozámský, Ph.D.

Computer Vision Lab, Institute of Visual Computing & Human-Centered Technology,
TU Wien - Faculty of Informatics

Datum zadání výzkumného úkolu: 15. 10. 2021

Termín odevzdání výzkumného úkolu: 31. 8. 2022

V Praze dne 15. 10. 2021



.....
vedoucí práce

.....
vedoucí katedry