FAKULTA APLIKOVANÝCH VĚD ZÁPADOČESKÉ UNIVERZITY V PLZNI

STUDENTSKÁ VĚDECKÁ KONFERENCE 2019

Reading text in images with EAST and Kaldi

Martin Bulín[1]

1 Introduction

Thinking of the motivation for using smart artificial systems to help or even substitute people when dealing with a specific task, one of the first aspects that comes on my mind is the speed computers are capable of working at. These systems are often solemnly called ”artificial intelligence”, which is wrong in my point of view, as they cannot perform the complex cognitive behavior as humans can do.

However, when a specific (not complex) behavior is required to deal with a certain task, machines can be effectively used by taking advantage of their working pace and moreover of the fact they never get bored or tired. One of such applications is image processing in general. No living human is capable of browsing and processing thousands of images per second, but a well-designed artificial system possibly could be.

This work is focused on reading text in images, especially in scanned documents, and motivated by a call for digitization of huge amount of archival scans and searching in them. The desired functionality of the developed system is what the commonly known shortcut `CTRL+F` does, however, applied on a scanned document instead of a digital text.

2 Methods and Results

The process of reading a text in an image consists of two main tasks (see Fig. 1):

1. *text localization*;
2. *optical character recognition (OCR)*.

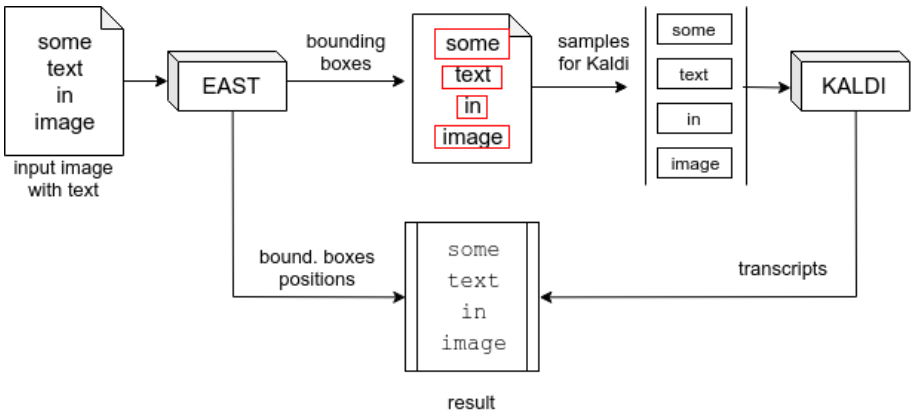


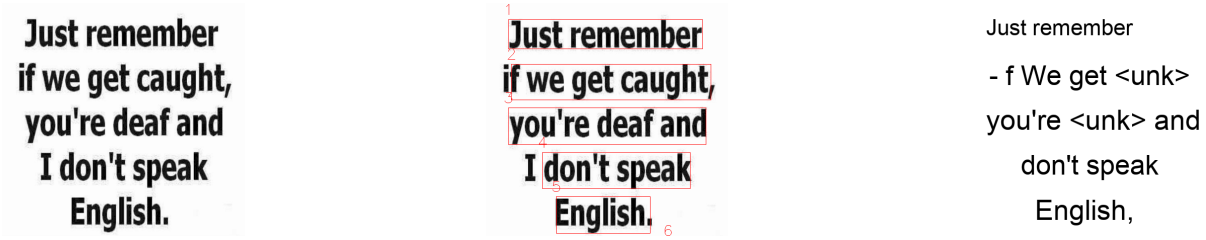
Figure 1: The overall process of reading a text in an image using the EAST algorithm (Zhou et al. (2017)) for text localization and the Kaldi toolkit (Povey et al. (2011)) for OCR.

[1] PhD student of Applied Sciences and Computer Engineering, field of study Cybernetics and Control Engineering, focused on Neural Networks. E-mail: bulinm@kky.zcu.cz

The main goal of this work is to show how to combine the EAST algorithm (Zhou et al. (2017)), capable of text localization, with a Kaldi (Povey et al. (2011)) model responsible for OCR. An open-sourced implementation of the EAST algorithm available from (Argman. (2017)), specifically the pre-trained model called `resnet`, was used.

The Kaldi toolkit is commonly known as the state-of-the-art ASR (Automatic Speech Recognition) toolkit, however, as there is an analogy between ASR and OCR data, the same methods can be used for OCR in a very similar way. The context-dependency is the key feature allowing to apply the well-known ASR approaches including the use of a language model. The only part that is being handled in a different way is the parametrization process at the beginning.

Fig. 2 shows an example input image (2a), output of the text localization process (2b) and a Kaldi transcript positioned using the information from EAST (2c). No measure of how well the procedure works is defined yet. The presentation corresponding to this abstract is mainly aimed to show how to combine the two main parts - text localization with EAST and OCR with Kaldi.



(a) **(b)** **(c)**

Figure 2: An example image with text. (a) input image; (b) output of the EAST text localization algorithm (red: bounding boxes); (c) output of the Kaldi pre-trained model, using positions from EAST.

Acknowledgement

Special thanks to Jan ”Yenda” Trmal, Associate Research Scientist at Johns Hopkins University and my supervisor during my stay at CLSP, JHU. The internship was supported by the Department of Cybernetics, University of West Bohemia (project nb. 52240).

References

Povey, D., Ghoshal, A., Boulianne, G., Burget L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K. (2011) The Kaldi Speech Recognition Toolkit. *In IEEE 2011 Workshop on Automatic Speech Recognition and Understanding.* IEEE Signal Processing Society, 2011.

Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., Liang, J. (2017) EAST: An Efficient and Accurate Scene Text Detector. *In Proc. IEEE Conference Computer Vision Pattern Recognition*, pp. 2642-2651

Argman (2017) A tensorflow implementation of EAST text detector. `https://github.com/argman/EAST`