Text Baseline Recognition using a Recurrent Convolutional Neural Network

Matthias Wödlinger, Robert Sablatnig
Computer Vision Lab, TU Wien

INTRODUCTION

The detection of baselines of text is a necessary preprocessing step for many modern methods of automatic handwriting recognition.
In this work, we present a two-stage system for the automatic detection of text baselines of handwritten text. In a first step, we perform pixel-wise segmentation on the document image to classify pixels as baselines, start points, end points and background. This segmentation is then used to extract the start points of lines. Starting from these points we extract the baseline using a recurrent convolutional neural network that directly outputs the baseline coordinates.

This method allows the direct extraction of baseline coordinates as the output of a neural network without the use of any post-processing steps.
We evaluate the model on the cBAD dataset from the ICDAR 2019 competition on baseline detection.

METHODS

Page segmentation:
Perform pixel wise segmentation to detect start points, angles and end points using a U-Net inspired CNN. Additionally use segmentation for baseline and text to aid the coordination extraction network by combining the segmentation map with the original image (see Figure 1).

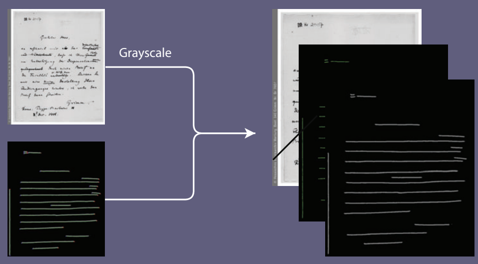

Figure 1: Combine Combine segmentation with grayscale input image

Coordinate extraction:
Starting from extracted start points follow the baselines and extract the coordinates. For every predicted baseline coordinate and angle we extract an image patch from the page image. This image patch is then used as input for two CNNs. The first CNN determines the angle of the baseline which is used to compute the next coordinate. The second CNN determines if the end of the baseline is reached and if so, predicts the end of the last segment. The step is repeated until the end of the baseline. The whole process (Figure 2) is differentiable and can be optimized with backpropagation from start to end.

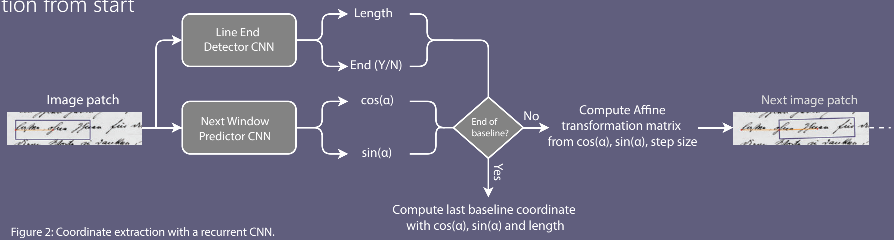

Figure 2: Coordinate extraction with a recurrent CNN.

RESULTS

We evaluate out approach on the cBAD 2019 competition for historical baseline detection [1]. We also evaluate our method with ground truth start points to investigate where errors appear (start point detection/coordinate extraction). When compared with the winner of cBAD 2019 our approach performs worse. However, when using GT start points and angles (Method A and Method B) we find a significant performance boost indicating that the majority of errors are due to wrong or undetected start points and angles.

Name	Precision	Recall	F1 score
Ours	0.872	0.890	0.881
ARU-NET [2] (Winner cBAD 2019)	0.937	0.926	0.931
Method A (With GT start points and angles and with segmentation)	0.953	0.974	0.963
Method B (With GT start points and angles and without segmentation)	0.882	0.972	0.924

Figure 3: Typical examples for the results of our methods.

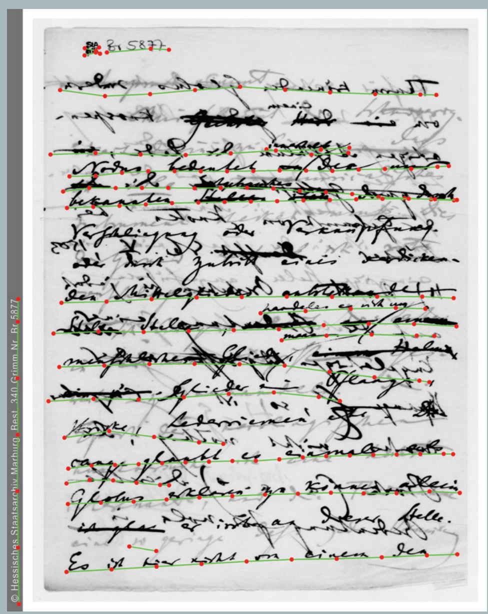

Figure 4: A failure case. Some lines are not detected due to an error in the start point detection. Other lines overlap due to an error in the end point detection

References:
[1] M. Diem, F. Kleber, R. Sablatnig, and B. Gatos, "cbad: Icdar2019 competition on baseline detection," in ICDAR 2019 vol. 1. IEEE, 2019, pp. 1494–1499.
[2] T. Gruning, G. Leifert, T. Strauß, J. Michael, and R. Labahn, "A two- stage method for text line detection in historical documents," in IJDAR, vol. 22, no. 3, pp. 285–302, 2019.

mwoedlinger@cvl.tuwien.ac.at
This poster is presented at the International Conference for Pattern Recognition (ICPR) 2020

CVL Computer Vision Lab · TU WIEN · Informatics