Unsupervised deep learning for text line segmentation

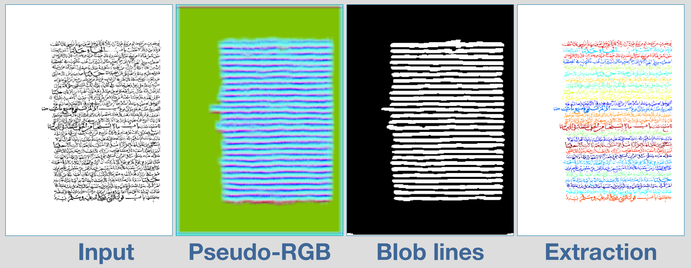
Berat Kurar Barakat Ahmad Droby Reem Alasam Boraq Madi Irina Rabaev
Raed Shammes Jihad El-Sana

Ben-Gurion University of the Negev

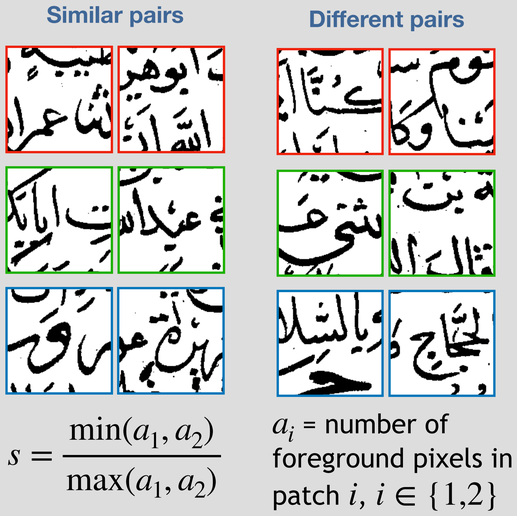
SCE SHAMOON COLLEGE OF ENGINEERING

Introduction

- Humans can segment the text lines written in a language they do not speak.
- Gestalt principles state that the proximity and similarity of the visual elements forms the basis for the unsupervised text line segmentation ability.
- We present an unsupervised deep learning method **(UTLS)** that emulates this principle.



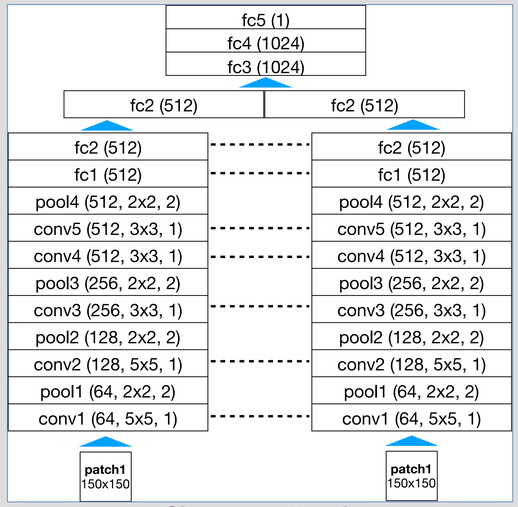
Data and Training



$$s = \frac{\min(a_1, a_2)}{\max(a_1, a_2)}$$

a_i = number of foreground pixels in patch i, $i \in \{1,2\}$

- The score s measures the relative amount of foreground pixels in a pair of patches.
- As $s \to 1$ a pair of patches is labelled as similar.
- As $s \to 0$ a pair of patches is labelled as different.
- Siamese network is trained to discriminate whether two patches are similar or different.



Siamese network

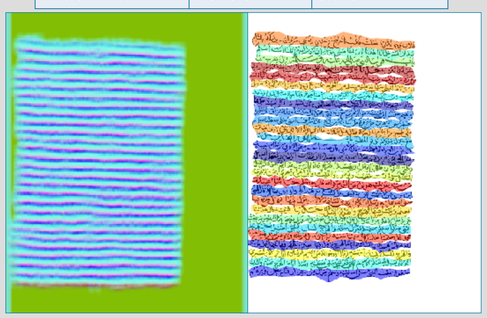
Prediction

- Penultimate layer of a single branch is used to extract features of document image patches.
- For every patch, the first 3 principles of its features are used to generate the pseudo-RGB image.
- Pseudo-RGB images are thresholded into blob lines.
- Blob lines assist an energy minimization function for extracting the text lines.

Results

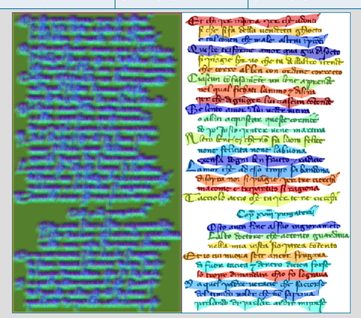
VML-AHTE dataset

	LineIU	PixelIU
UTLS	**98.55**	**88.95**
Mask-RCNN	93.08	86.97



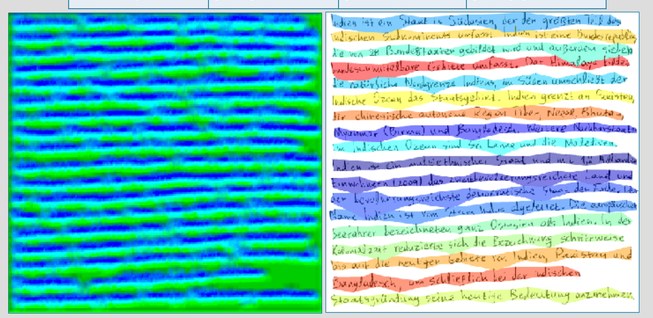
ICDAR2017 dataset

	LineIU	PixelIU
UTLS	88.41	87.40
Winner	**97.86**	**97.05**



ICFHR2010 dataset

	DR	RA	FM
UTLS	73.22	72.38	72.36
Winner	**97.54**	**97.72**	**97.63**



Acknowledgment
The authors would like to thank Gunes Cevik for annotating the data.