Unsupervised deep learning for text line segmentation



Berat Kurar Barakat

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.



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.

Acknowledgment

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

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



Results

Extraction

Data and Training



Different pairs



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

VML-AHTE dataset				ICDAR2017 dataset			ICFHR2010 dataset			
	LinelU	PixelIU		LinelU	PixellU			DR	RA	FM
UTLS	98.55	88.95	UTLS	88.41	87.40		UTLS	73.22	72.38	72.36
Mask-RCNN	93.08	86.97	Winner	97.86	97.05	1	Winner	97.54	97.72	97.63
					and the ar ben con ordine correcto en terfulaniete un tre aprente qual fichati lammo 7 hiter to amor Thi unte numa in acquiftare quele corme o Tulto genter une martina teneres the none labuona teneres offer no fa luom fefter e fettetta none labuona na togni un finagie pre tre mereti ome o terpartito fi ragiona olo accio de un ce te ne cierti Cap sum puegatori to aura fine al fuo ragionameto alto toetore che actento guardina ella ma unta lio presa corento cui nuova fete ancor frugana or tacca de ta ce fo fo lognua or tacca de tenere frugana ella ma unta lio presa corento cui nuova fete ancor frugana or tacca de tenere for forte ope timanetre che factor forte ope time che forte che factor forte ope time che forte che factor forte ope time che forte che forte che forte che forte che forte che che che che che che che che che ch				lie nativische Nordyr hidische Szcan das s die chinespische auf Myanmar (Birma) u im Indischen Ozen Indian ist ein multi Einwichwern (2009) d der Lovollurrungevet Name Indian ist Vor Seetahrer bezeichne Kolonial zait reduzier Dis auf die neutige Branglahrsch, um S Stoatsgründung se	enze Indians, in Süden i Staatsgebied, Indian gr and Banglodeide, Warts and Banglodeide, Warts and Sri Laner und die iethnischer Staat und as zweitheweiterungsvei chsie demokratische Si a Star India algebein ten ganz Ormginn all be sich die Eszeichwarg a Gehiete von Indian, aufließlich bei der indi and hentige Bedeut

- The score *s* measures the relative amount of foreground pixels in a pair of patches.
- As $s \rightarrow 1$ 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.

	fc5	5 (1)								
	fc4 (*	fc4 (1024)								
fc3 (1024)										
tc2 (51	2)	tc2 (512)								
fc2 (512)			fc2 (512)							
fc1 (512)			fc1 (512)							
pool4 (512, 2x2, 2)		-	pool4 (512, 2x2, 2)							
conv5 (512, 3x3, 1)			conv5 (512, 3x3, 1)							
conv4 (512, 3x3, 1)			conv4 (512, 3x3, 1)							
pool3 (256, 2x2, 2)			pool3 (256, 2x2, 2)							
conv3 (256, 3x3, 1)		conv3 (256, 3x3								
pool2 (128, 2x2, 2)			pool2 (128, 2x2, 2)							
conv2 (128, 5x5, 1)			conv2 (128, 5x5, 1)							
pool1 (64, 2x2, 2)			pool1 (64, 2x2, 2)							
conv1 (64, 5x5, 1)		conv1 (64, 5x								
patch1 150x150		-	patch1 150x150							

Siamese network

