	Comparison with state-of-the-art 0000	Ablation study 00	

Multiple Document Datasets Pre-training Improves Text Line Detection With Deep Neural Networks

#### Mélodie Boillet<sup>12</sup>, Christopher Kermorvant<sup>12</sup> and Thierry Paquet<sup>2</sup>

<sup>1</sup>Teklia SAS, Paris, France <sup>2</sup>LITIS, Rouen-Normandy University, France

SIFED - 6th July 2020



M. Boillet

Multiple document datasets pre-training improves text line detection with DNN 1 / 1

Context 00000	Model and data	Comparison with state-of-the-art	Ablation study	Conclusion
	00000	0000	00	0

### Table of contents



- 2 Model and data
- 3 Comparison with state-of-the-art
- 4 Ablation study



M. Boillet

Context	Comparison with state-of-the-art	Ablation study	
<b>●</b> 0000			

### Text line segmentation

- ► **Goal:** detect the text lines of an image;
- Application: apply a text recognition system on the detected text lines.



Context 00000	Comparison with state-of-the-art 0000	Ablation study 00	

### Text line segmentation

- ► **Goal:** detect the text lines of an image;
- Application: apply a text recognition system on the detected text lines.



Context 00000	Comparison with state-of-the-art 0000	Ablation study 00	

### State-of-the-art

### ► dhSegment [Oliveira2018]:

- CNN + Resnet-50 pre-trained on ImageNet;
- Multi-task: text line detection, ornament detection...;
- Good results on various datasets: DIVA, cBAD...

### $\blacktriangleright$ Yang et al. [Yang2017]:

- Multimodal FCN + use of the text content;
- Good results on modern documents: DSSE, SectLabel...

#### ► Moysset et al. [Moysset2015]:

- Recurrent network;
- Text line segmentation of paragraphs;
- 2 labels: line and interline.

Context	Model and data	Comparison with state-of-the-art	Ablation study	Conclusion
00000	00000	0000	00	0

### Problems of dhSegment

- ▶ Needs a lot of annotated data;
- ► Good results but can still be improved;
- ► Too long to analyse a whole corpus: ~66 days for 2M images (Balsac dataset).

Is pre-training on natural scene images the most suitable for working on document images ?

Context	Model and data	Comparison with state-of-the-art	Ablation study	Conclusion
0000●	00000	0000	00	0
Main	goal			

### Analyse the impact of a pre-training step on the line segmentation task.

We want to develop a model:

- ▶ Containing no pre-trained part learnt on natural scene images;
- Having less parameters than dhSegment and a reduced prediction time;
- ▶ Yielding higher accuracy than dhSegment.

Context	Model and data	Comparison with state-of-the-art	Ablation study	Conclusion
	00000			

### Yang's architecture [Yang2017]



M. Boillet

Multiple document datasets pre-training improves text line detection with DNN 6 / 17

→ ∃ →

E >

315

Context	Model and data	Comparison with state-of-the-art	Ablation study	Conclusion
	0000			

### Architecture of our model



M. Boillet

A B + 
 A B +
 A
 B
 A
 B
 A
 B
 A
 A
 B
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A

ELE DQC

ヨト・イヨト

Context 00000	Model and data	Comparison with state-of-the-art	Ablation study	Conclusion
	00●00	0000	00	0

### Implementation details

- Input image size:  $384 \times 384 \ px$ ;
- ▶ Batch normalization + dropout layers after all convolutions;
- ▶ Use of concatenations to help the detection of small objects;
- Post-processing: thresholding + removal of small connected components.

Model and data 000●0	Comparison with state-of-the-art 0000	$\begin{array}{c} \text{Ablation study} \\ \text{oo} \end{array}$	

### Datasets

#### **Balsac**: 913 annotated images

I alla per Buckland La mornin a ch alla Join fat Kory & Hothall, de Rober Sight' Willow John attached in the X9) and fin bytamic guardings to the Andrews the distance to the Andrews to guardings with the standard to the memory day phase of the first of the standard signal serve they to the full of the first set signal serve they the first full of the first set signal serve the first full of the first set signal serve the first set of the second second second second second the first second Le calles the will we found and the four plan Hend proved a Vinlamating, he sensinger Black proved a Vinlamating, he sensinger Black provent and soil and son star path in to vigt new In which will be lighted he parai a to William them it to more manual Carta them on the signi non them; the para-a you igno, taken take Jugh Plan, Marine Jallante, William of Williams II arend

# Horae [Boillet2019]: 557 annotated images



< □ > < 🗇 >

医下颌 医下颌

11 DQC

Model and data 0000●	Comparison with state-of-the-art 0000	Ablation study 00	

### Datasets II

# **READ-BAD** [Grüning2017]: 2036 annotated images



# **DIVA-HisDB** [Simistira2016]: 120 annotated images

	S
	To prove for an - 6
Con Contraction	por ward a rea by
A naturgit externerative are	Funday 2 fay 194 2
be well never some Or the lange recently Simple	and many the part of the
white alinger myrange Aline formiles Shineano	par 2 grain itogre 120.
har anone page and D was to remain abacade ware	the program and the
the the pulper which foren the lafter the of the to the the	
frait his i type and a grate pless amb to palme	
come But for any Baleno to chalme	
Ce Lano aver 1 tenolamere	
de ferr idea 30 voir Artete	
Foguina la per tueto finno streo	
R agriffe om lietor fon fibro alieno	
the wells a Born bin tato berile	
) wit quetto exelecto gotile	
quali aja cha viguatian inc	
1 to an 2 and and Sur I would be Dar Bart Affecter	
the and shaper how or exercise presente Belle pute for	
and the server of the provide the server of the server fort	
a ver man handle with - Levelle warderen - regulate	
"The second party of the first of the first one water from officers barries press	
fin ten Ofer partes all want worth a pares raines Gringe behan brit	
the state of the s	
the state of the s	1.1

ELE DOG

- E - E

M. Boillet

Context	Model and data	Comparison with state-of-the-art	Ablation study	Conclusion
		●000		

### Comparison with dhSegment

DATASET	Model	IoU	Pr	Rec	F1	Time1
Balaac	dhSegment	73.78	92.07	78.76	84.81	66.3
Daisac	Our	83.79	94.80	87.86	91.11	9.2
Horae	dhSegment Our	<b>65.22</b> 63.95	71.70 <b>78.38</b>	<b>89.29</b> 80.45	82.32 <b>84.93</b>	18.8 <b>2.3</b>
READ-Simple	dhSegment Our	<b>64.55</b> 64.03	<b>85.04</b> 81.76	71.85 <b>75.60</b>	<b>77.25</b> 76.66	$8.4^2$ <b>1.0</b> <sup>2</sup>
READ-Complex	dhSegment Our	52.91 <b>54.40</b>	79.28 <b>83.62</b>	59.16 <b>61.97</b>	69.27 <b>73.16</b>	$10.6^2$ <b>1.3</b> <sup>2</sup>
DIVA-HisDB	dhSegment Our	74.24 <b>75.71</b>	<b>92.41</b> 92.14	79.10 <b>80.88</b>	85.19 <b>86.09</b>	NA NA

	dhSegment	Our
Number of	22 2M(0.26M)	4 1M
PARAMETERS	52.81v1(9.301v1)	4.11/1

 $^1\mathrm{Prediction}$  time (GPU GeForce RTX 2070 8G) in days to analyse the whole corpus.

<sup>2</sup>Estimation based on the manuscripts sizes without *BHIC* and *Unibas*.  $\langle \square \rangle$   $\langle \square \rangle$   $\langle \square \rangle$ 

M. Boillet Multiple document datasets pre-training improves text line detection with DNN 11 / 1

ELE SQA

Context	Model and data	Comparison with state-of-the-art	Ablation study	Conclusion
		0000		

### Multiple document dataset

# Does pre-training on document images improve the performances ?



M. Boillet

Multiple document datasets pre-training improves text line detection with DNN 12/17

Context	Model and data	Comparison with state-of-the-art	Ablation study	Conclusion
		0000		

### Comparison with dhSegment

Data	Model	IoU	Pr	Rec	F1
	dhSegment	73.78	92.07	78.76	84.81
Palana	dhSegment PT	74.02	91.89	79.09	84.95
Daisac	Our	83.79	94.80	87.86	91.11
	Our PT	84.87	94.25	89.49	91.75
	dhSegment	65.22	71.70	89.29	82.32
Hanaa	dhSegment PT	60.69	80.94	73.65	81.99
погае	Our	63.95	78.38	80.45	84.93
	Our PT	68.81	80.31	84.80	88.62
	dhSegment	64.55	85.04	71.85	77.25
DEAD Simple	dhSegment PT	65.07	88.34	71.56	80.72
READ-Simple	Our	64.03	81.76	75.60	76.66
	Our PT	68.14	83.19	78.05	79.45
	dhSegment	52.91	79.28	59.16	69.27
PEAD Complex	dhSegment PT	53.34	85.51	57.80	68.47
READ-Complex	U-FCN	54.40	83.62	61.97	73.16
	U-FCN PT	60.28	81.03	68.17	78.30
	dhSegment	74.24	92.41	79.10	85.19
DIVA HigDB	dhSegment PT	73.00	91.56	78.28	84.32
DIVA-IIISDD	U-FCN	75.71	92.14	80.88	86.09
	U-FCN PT	74.72	89.43	82.20	85.44

M. Boillet

Multiple document datasets pre-training improves text line detection with DNN

イロト イヨト イヨト イヨト

3 / 17

3 2

		Comparison with state-of-the-art $0000$	Ablation study 00	
Conch	usion			

### Does pre-training on document images improve the performances ? YES

Intersection-over-Union :

- ✓ +5 percentage points on Horae and READ-Complex;
- $\checkmark$  +4 percentage points on READ-Simple;
- $\approx$  Similar performances on Balsac;
- ✗ −1 percentage point on DIVA-HisDB.

Our results are overall better than dhSegment's (except for the precision metric).





M. Boillet Multiple document datasets pre-training improves text line detection with DNN

15 / 17

		Comparison with state-of-the-art 0000	Ablation study ⊙●	
Ablati	on study II			

- Adding Batch Normalization and Dropout layers after all convolutions improves the performances;
- Using dilation rates [1, 2, 4, 8, 16] allows to have a bigger receptive field;
- Adding more training images and using bigger input images also improves the results.

BUT still better results (77.42 %) than SOTA dhSegment (73.78 %) with only 90 images.

		Comparison with state-of-the-art 0000	Ablation study 00	$\bigcirc$ Conclusion
Conch	usion			

We designed a model:

- ► Lighter than dhSegment;
- Giving most of the time better results;
- ▶ Having a reduced prediction time: up to 8 times faster.

Future work:

- ▶ Test our architecture on datasets with more than 2 classes;
- ▶ Build an historical model trained on various historical documents.

	Comparison with state-of-the-art 0000	Ablation study 00	

### Bibliography

[Oliveira2018]	Sofia Ares Oliveira, Benoit Seguin, and Frederic Kaplan. "dhSegment: A generic deep-learning approach for document segmentation". In: Frontiers in Handwriting Recognition (ICFHR), 2018 16th International Conference on. IEEE. 2018, pp. 7–12
[Yang2017]	Xiao Yang et al. "Learning to Extract Semantic Structure from Documents Using Multimodal Fully Convolutional Neural Network". In: vol. abs/1706.02337. 2017. arXiv: 1706.02337. URL: http://arxiv.org/abs/1706.02337
[Moysset2015]	B. Moysset et al. "Paragraph text segmentation into lines with Recurrent Neural Networks". In: 2015 13th ICDAR. Aug. 2015, pp. 456-460. DOI: 10.1109/ICDAR.2015.7333803
[Boillet2019]	Mélodie Boillet et al. "HORAE: An Annotated Dataset of Books of Hours". In: Proceedings of the 5th International Workshop on Historical Document Imaging and Processing. HIP '19. Sydney, NSW, Australia: Association for Computing Machinery, 2019, 7-12. ISBN: 9781450376686. DOI: 10.1145/3352631.3352633. URL: https://doi.org/10.1145/3352631.3352633
[Grüning2017]	Tobias Grüning et al. "READ-BAD: A New Dataset and Evaluation Scheme for Baseline Detection in Archival Documents". In: <i>CoRR</i> abs/1705.03311 (2017). arXiv: 1705.03311. URL: http://arxiv.org/abs/1705.03311
[Simistira2016]	F. Simistira et al. "DIVA-HisDB: A Precisely Annotated Large Dataset of Challenging Medieval Manuscripts". In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). Oct. 2016, pp. 471–476. DOI: 10.1109/ICFHR.2016.0093

M. Boillet Multiple document datasets pre-training improves text line detection with DNN 17 / 17

Post-processing after dhSegment:

- ▶ Probabilities thresholding to keep the highest;
- ▶ Removal of the small connected components.

To be comparable, we used the same post-processing after our model.

Does this thresholding have a real impact on our results ? If so, how to optimize this threshold ?

## Thresholding impact



M. Boillet

Multiple document datasets pre-training improves text line detection with DNN

## Thresholding impact

0.50.60.70.80.9

M. Boillet

## Thresholding impact II

- Choice of the threshold essential for dhSegment;
- ► Thresholding step obsolete for U-FCN:
  - $\rightarrow$  no need to choose a threshold;
  - $\rightarrow$  assign the class having the highest probability.