

# End-to-end Handwritten Paragraph Text Recognition Using a Vertical Attention Network

Denis Coquenot<sup>1,3</sup>, Clément Chatelain<sup>1,2</sup>, Thierry Paquet<sup>1,3</sup>

LITIS

<sup>1</sup>Normandie Université, Normandie, France

<sup>2</sup>INSA de Rouen, Normandie, France

<sup>3</sup>Université de Rouen, Normandie, France

SIFED 2021



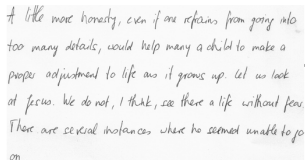
# Handwritten text recognition

## Task

- Goal: to recognize the text from a scanned document, in the correct order
- Constraints: writing diversity (character shape, slant, size, spacing), input of variable sizes (number of pixels, characters, words, lines), arbitrary document layout with strong variabilities

## Paragraph recognition - overview

Input: paragraph image



A little more honesty, even if one refrains from going into too many details, would help many a child to make a proper adjustment to life as it grows up. Let us look at Jesus. He do not, I think, see there a life without fear. There are several instances where he seemed unable to go on.

Deep neural network

Output: sequence of characters

"A little more [...] to go on."

# State of the art

## Explicit line segmentation, followed by text line recognition

- Segmentation based on object detection approach: [1], [2]
  - Segmentation based on start-of-line prediction: [3], [4]
- Needs line-level segmentation label and cumulates errors

## End-to-end implicit segmentation & recognition

- Attention-based models: line-level [5], character-level [6]
- FCN models: [7], [8]

# Datasets

## Characteristics

Dataset	Level	Training	Validation	Test	charset size	Language	# lines
RIMES [9]	Line	10 532	801	778	100	French	2-18
	Paragraph	1 400	100	100			
IAM [10]	Line	6 482	976	2 915	79	English	2-13
	Paragraph	747	116	336			
READ 2016 [11]	Line	8 349	1 040	1 138	89	Early Modern	1-26
	Paragraph	1 584	179	197		German	

- Handwritings
- Resolution of 300 dpi

## Datasets

Il a bien reçu votre lettre concernant mes affaires  
antérieures, d'un montant de 2000.  
Toutefois, j'éproue actuellement des difficultés  
financières. Aussi je vous demande de bien vouloir  
me retourner les paiements mensuellement sur une  
période de 10 mois.  
Je régleme les mensualités de 5 de chaque mois  
sur un compte bancaire, sur vos instructions  
adéquates, à nos conventions. Je régleme  
en une seule fois les 200 de frais de commission.  
En espérant que vos affaires prospèrent à la  
manière, je vous prie d'agréer Madame, l'assurance de  
mes sentiments distingués.

Je vous informe que je vous ai mis votre enregistrement  
dans un dossier et dans votre ma nouvelle adresse:  
M. J. J. J.  
1 rue de la  
BISSEZELLE  
A. de la B. N. H. R. E. O.

Je vous remercie de bien vouloir à jour vos données et vos  
avis et vos nouvelles coordonnées.  
Je vous en remercie de bien vouloir.

Je vous prie de bien vouloir me faire parvenir vos  
documents, s'il y a lieu, à l'adresse ci-dessus.  
Bonne nuit.

RIMES

This figure has been reported only on the one of the  
of Geneva documents on July 19, 1952. And official  
from it may be too much for the city's 47-page copy.  
They will accept and print them will have to be used.

He made these changes Mr. Weaver's  
alleged association with organizations State-  
led by the Government, immediately Mr.  
Weaver upon a letter to Senator Tolson  
saying the Federal Bureau of Investigation had  
reported on Mr. Weaver. He believed  
he would perform "outstanding service"  
in his post. Senator Tolson's committee  
was to pass Mr. Weaver's nomination before it  
can be completed by the full Senate.

It said these concerns Mr. Weaver's  
alleged association with organizations State-  
led by the Government, immediately  
Mr. Weaver upon a letter to Senator  
Tolson saying the Federal Bureau of In-  
vestigation had reported on Mr. Weaver.  
He believed he would perform "outstanding  
service" in his post. Senator Tolson's  
committee was to pass Mr. Weaver's  
nomination before it can be com-  
pleted by the full Senate.

IAM

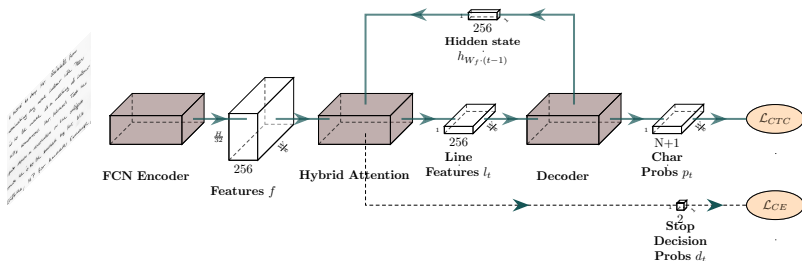
Je vous prie d'agréer Madame, l'assurance de  
mes sentiments distingués.  
M. J. J. J.  
1 rue de la  
BISSEZELLE  
A. de la B. N. H. R. E. O.

Je vous prie d'agréer Madame, l'assurance de  
mes sentiments distingués.  
M. J. J. J.  
1 rue de la  
BISSEZELLE  
A. de la B. N. H. R. E. O.

39

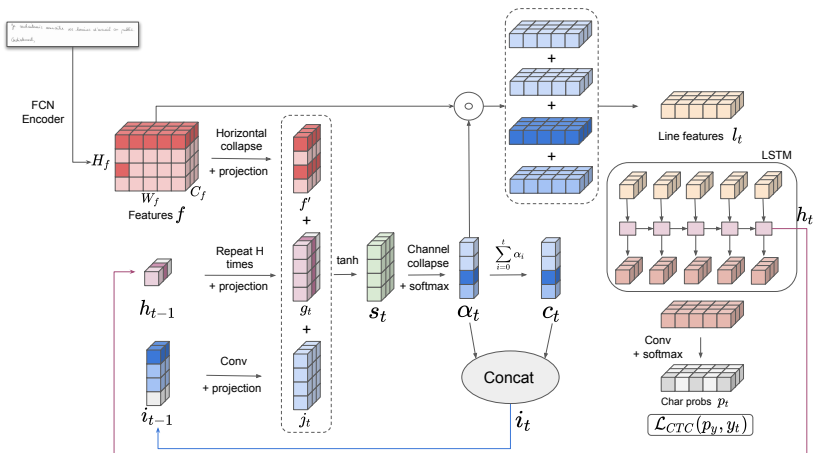
READ 2016

# VAN: Vertical Attention Network [12]

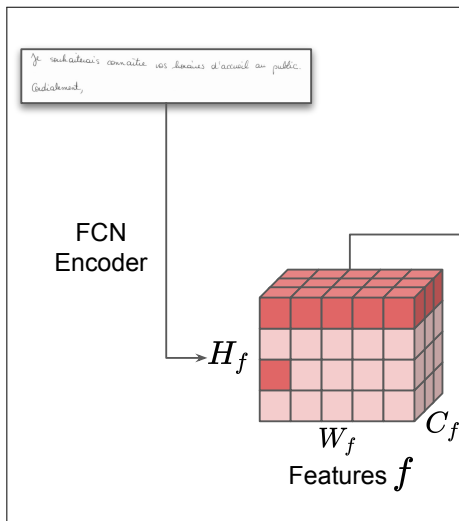


- Seq2seq architecture with attention (recurrent process)

# VAN - Focus on the hybrid attention mechanism

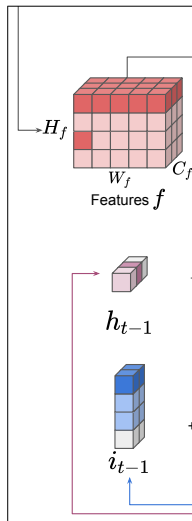


# VAN - Focus on the hybrid attention mechanism





# VAN - Focus on the hybrid attention mechanism



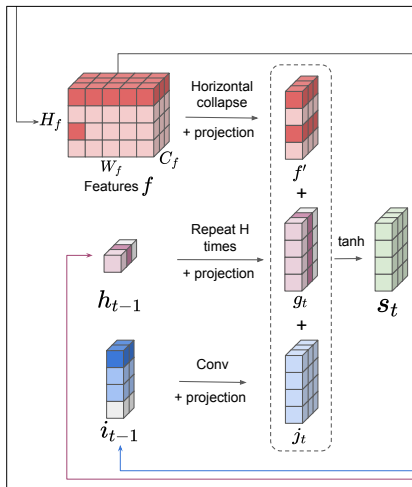
Hybrid attention:

→ location-based + content-based

Dimension mismatching:

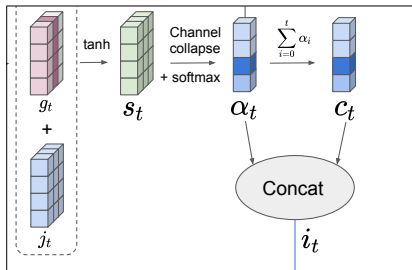
- Features  $f \in \mathbb{R}^{H_f \times W_f \times C_f}$
- Decoder LSTM hidden state  $h_{t-1} \in \mathbb{R}^{1 \times 1 \times C_h}$
- Information from previous attention weights  $i_{t-1} \in \mathbb{R}^{H_f \times 1 \times 2}$

# VAN - Focus on the hybrid attention mechanism



$$s_{t,i} = \tanh(f'_i + g_{t,i} + j_{t,i})$$

# VAN - Focus on the hybrid attention mechanism



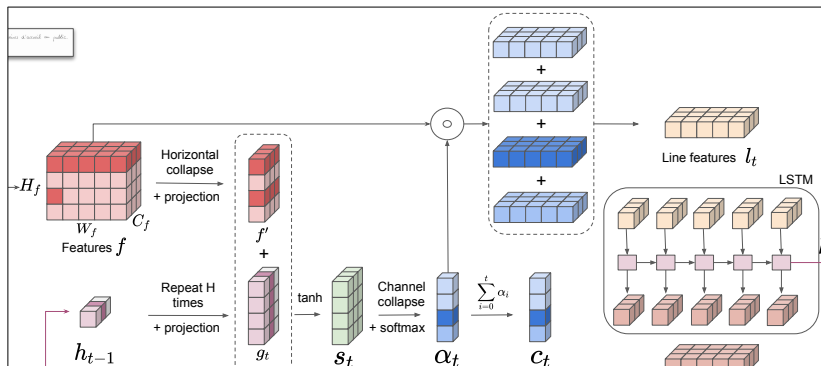
Attention score:  $e_{t,i} = W_a \cdot s_{t,i}$

Attention weights:

$$\alpha_{t,i} = \frac{\exp(e_{t,i})}{H_f \sum_{k=1} \exp(e_{t,k})}$$

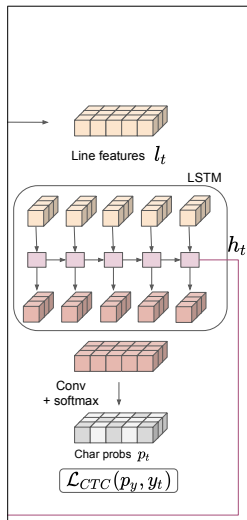
Coverage vector:  $c_t = \sum_{k=0}^t \alpha_k$

# VAN - Focus on the hybrid attention mechanism



$$l_t = \sum_{i=1}^{H_f} \alpha_{t,i} \cdot f_i$$

# VAN - Focus on the hybrid attention mechanism

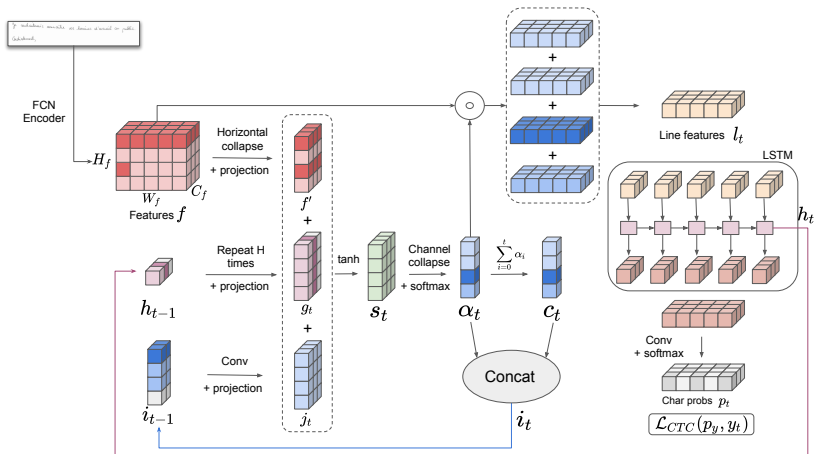


Line-by-line alignment with the CTC loss:

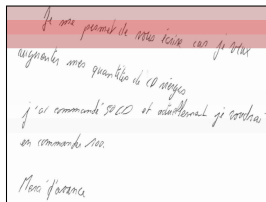
$$\mathcal{L} = \sum_{k=1}^L \mathcal{L}_{\text{CTC}}(p_k, y_k)$$

where  $L$  is the number of lines.

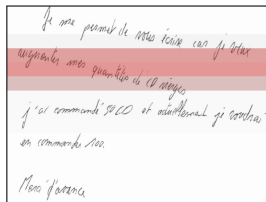
# VAN - Focus on the hybrid attention mechanism



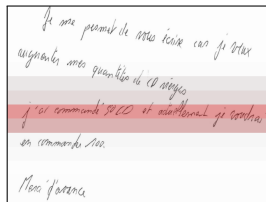
# VAN - Visualization



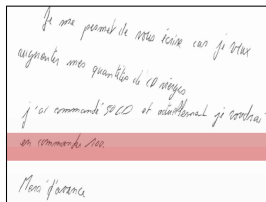
Je me permets de vous  
écrire car je vMux



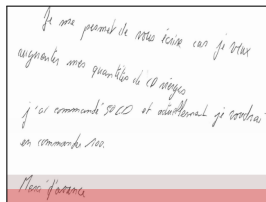
augmente mes quantités  
de CD vierges



j'ai commandé 50 CD et  
aduellement je voudrai



en commander 100.



Merci d'avance

# Results

Architecture	Attention	IAM		RIMES		READ 2016		# Param.
		CER (%)	WER (%)	CER (%)	WER (%)	CER (%)	WER (%)	
<b>Two-step approaches</b>								
[1] RPN+CNN+BLSTM	✗	15.6						
[2] RPN+CNN+BLSTM	✗	8.5						
[13] RPN+CNN+BLSTM	✗	6.4	23.2	2.1	9.3			
<b>End-to-end approaches</b>								
[7] FCN	✗	4.7						16.4 M
[8] FCN	✗	5.45	19.83	4.17	15.61	6.20	25.69	19.2 M
[6] CNN+MDLSTM	character	16.2						
[14] CNN+Transformer	character	6.7						27 M
[5] CNN+MDLSTM	line	7.9	24.6	2.9	12.6			
[12] Ours (VAN) - FCN+LSTM	line	<b>4.45</b>	<b>14.55</b>	<b>1.91</b>	<b>6.72</b>	<b>3.59</b>	<b>13.94</b>	2.7 M



# Conclusion

## VAN

- Seq2seq architecture using hybrid attention
- Results beyond the state of the art on three datasets

## Limitations

- Pretraining requirements (isolated lines / cross dataset)
- Limited to single-column document

## Paper

- Under review (Arxiv: "End-to-end Handwritten Paragraph Text Recognition Using a Vertical Attention Network")
- Available source code and pretrained weights: <https://github.com/FactoDeepLearning/VerticalAttentionOCR>

# Perspective

## Handle complex layout

- Multi-column of texts, non-textual items...
- Use transformer architecture with attention at character level [15, 14]

## New challenges

- Lack of public datasets
- Images of whole documents more voluminous
- Doubtful reading order (maps, schema)  
→ inadequate metrics / losses
- From line to character attention  
→ the number of iterations increases dramatically

# Reference I

- [1] Manuel Carbonell et al. “End-to-End Handwritten Text Detection and Transcription in Full Pages”. In: [Workshop on Machine Learning, WML@ICDAR](#). 2019, pp. 29–34.
- [2] Jonathan Chung and Thomas Delteil. “A Computationally Efficient Pipeline Approach to Full Page Offline Handwritten Text Recognition”. In: [Workshop on Machine Learning, WML@ICDAR](#). 2019, pp. 35–40.
- [3] Bastien Moysset, Christopher Kermorvant, and Christian Wolf. “Full-Page Text Recognition: Learning Where to Start and When to Stop”. In: [International Conference on Document Analysis and Recognition, ICDAR](#). 2017, pp. 871–876.
- [4] Chris Tensmeyer and Curtis Wigington. “Training Full-Page Handwritten Text Recognition Models without Annotated Line Breaks”. In: [International Conference on Document Analysis and Recognition, ICDAR](#). 2019, pp. 1–8.
- [5] Théodore Bluche. “Joint Line Segmentation and Transcription for End-to-End Handwritten Paragraph Recognition”. In: [Advances in Neural Information Processing Systems 29](#). 2016, pp. 838–846.

## Reference II

- [6] Théodore Bluche, Jérôme Louradour, and Ronaldo O. Messina. "Scan, Attend and Read: End-to-End Handwritten Paragraph Recognition with MDLSTM Attention". In: [International Conference on Document Analysis and Recognition](#).
- [7] Mohamed Yousef and Tom E. Bishop. "OrigamiNet: Weakly-Supervised, Segmentation-Free, One-Step, Full Page Text Recognition by learning to unfold". In: [Conference on Computer Vision and Pattern Recognition, CVPR. 2020](#), pp. 14698–14707.
- [8] Denis Coquenot, Clément Chatelain, and Thierry Paquet. "SPAN: A Simple Predict & Align Network for Handwritten Paragraph Recognition". In: [16th International Conference on Document Analysis and Recognition](#). Vol. 12823. Lecture Notes in Computer Science. Springer, 2021, pp. 70–84.
- [9] Emmanuele Grosicki and Haikal El Abed. "ICDAR 2011 - French Handwriting Recognition Competition". In: [International Conference on Document Analysis and Recognition, ICDAR. 2011](#), pp. 1459–1463.

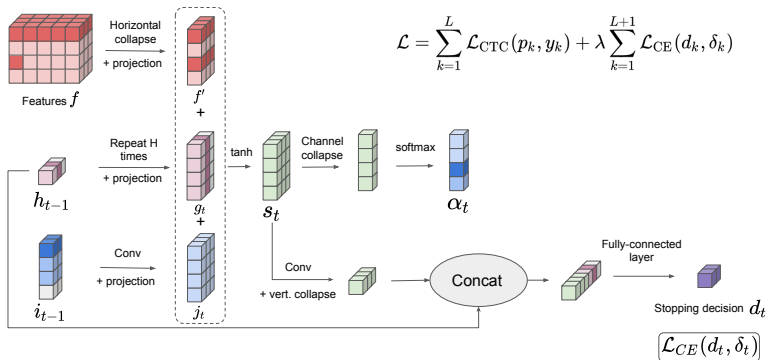
## Reference III

- [10] Urs-Viktor Marti and Horst Bunke. “The IAM-database: an English sentence database for offline handwriting recognition”. In: [International Journal on Document Analysis and Recognition, IJDAR 5.1 \(2002\)](#), pp. 39–46.
- [11] Joan-Andreu Sánchez et al. “ICFHR2016 Competition on Handwritten Text Recognition on the READ Dataset”. In: [15th International Conference on Frontiers in Handwriting Recognition, ICFHR. 2016](#), pp. 630–635.
- [12] Denis Coquenat, Clément Chatelain, and Thierry Paquet. “End-to-end Handwritten Paragraph Text Recognition Using a Vertical Attention Network”. In: [CoRR \(2020\)](#). URL: <https://arxiv.org/abs/2012.03868>.
- [13] Curtis Wigington et al. “Start, Follow, Read: End-to-End Full-Page Handwriting Recognition”. In: [European Conference on Computer Vision. Vol. 11210. Lecture Notes in Computer Science. 2018](#), pp. 372–388.
- [14] Sumeet S. Singh and Sergey Karayev. “Full Page Handwriting Recognition via Image to Sequence Extraction”. In: [CoRR \(2021\)](#). URL: <https://arxiv.org/abs/2103.06450>.

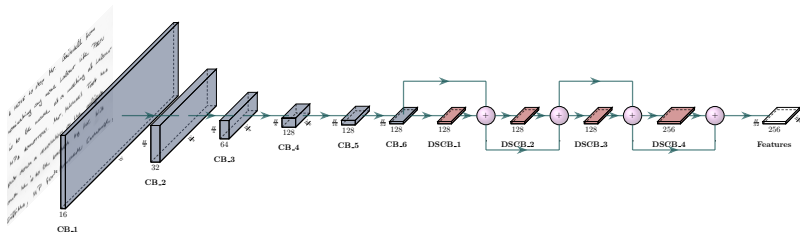
# Reference IV

- [15] Ashish Vaswani et al. "Attention is All you Need". In: Annual Conference on Neural Information Processing Systems. 2017, pp. 5998–6008.

# VAN - End-of-paragraph detection



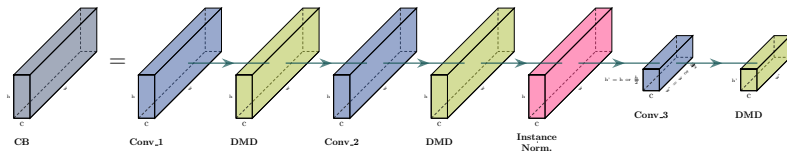
# VAN - FCN Encoder



CB : Convolution Block, DSC : Depthwise Separable Convolution Block

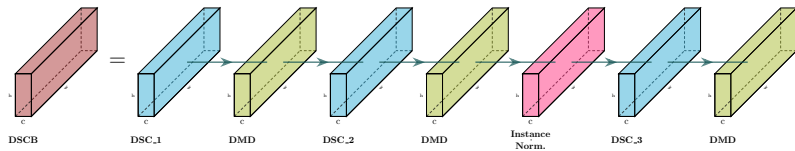


# CB : Convolution Block



DMD : Diffused Mix Dropout

# DSCB : Depthwise Separable Convolution Block



DMD : Diffused Mix Dropout

# Training details

- Preprocessing: resolution reduced to 150 dpi + normalization
- Data augmentation: contrast, lightness, resolution, perspective, projection, elastic distortion, dilation and erosion
- Optimizer : Adam
- Initial learning rate:  $10^{-4}$
- Mini-batch size: 16 for lines, 8 for the VAN

# CTC

h h e  $\epsilon$   $\epsilon$  l l l  $\epsilon$  l l o

h e  $\epsilon$  l  $\epsilon$  l o

h e l l o

h e l l o

Connectionist Temporal Classification

Image from Hannun, "Sequence Modeling with CTC", Distill, 2017.