Recurrence-free unconstrained handwritten text recognition using gated fully convolutional network

Denis Coquenet [1,3], Clément Chatelain [2], Thierry Paquet [3]

LITIS

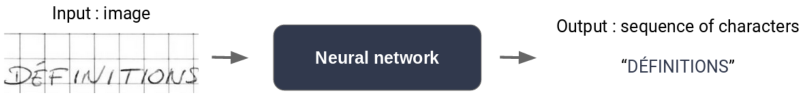
[1]Normandie Université, Normandie, France

[2]INSA de Rouen, Normandie, France

[3]Université de Rouen, Normandie, France

ICFHR, September 2020

Deep learning handwriting recognition system



Constraints

- Images (input) of variable size
- Sequence of characters (output) of variable length

Sequence alignment

Connectionist Temporal Classification (CTC) [Graves2006]

State of the art

Recurrent models (recurrent layers)

- Multi-Dimensional Long-Short Term Memory (MDLSTM) [Pham2014; Voigtlaender2016]
- Convolutional Neural Network + Bidirectional Long-Short Term Memory (CNN+BLSTM) [Puigcerver2017]

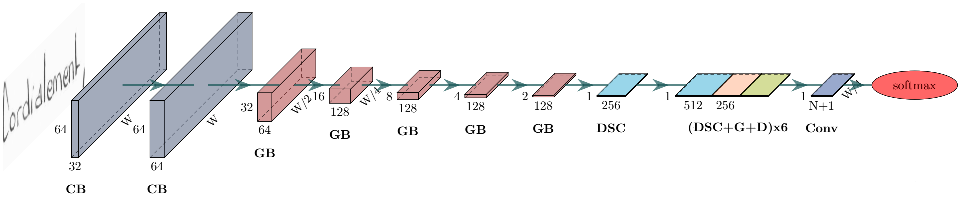
Non-recurrent models

- Convolutional Neural Networks (CNN) [Ptucha2018]
- Gated Fully Convolutional Network (GFCN) [Yousef2018; Ingle2019]

Attention models (recurrent process)

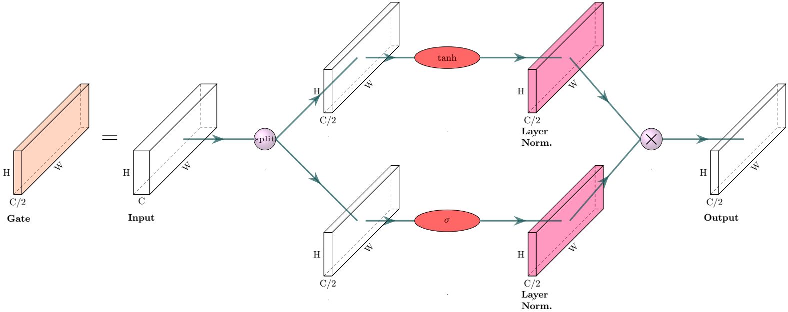
- Encoder-decoder architecture with soft attention [Chowdhury2018; Michael2019]

Our GFCN model



- G (Gate)
- DSC (Depthwise Separable Convolution)
- CB (Convolution Block) = Conv + Conv + Instance Norm. + Dropout
- GB (Gate Block) = DSC + DSC + Instance Norm. (+ MaxPooling) + Gate + Dropout
- (H, W, C) = (Height, Width, Feature maps)
- N = charset size (+ 1 for the CTC blank)

Gate


H
C/2
W
Gate
=
H
C
W
Input
split
H
C/2
W
H
C/2
W
tanh
σ
H
C/2
W
Layer
Norm.
H
C/2
W
Layer
Norm.
H
C/2
W
Output

Details

Architecture

- Deep: 22 convolutional layers
- Parameters: 1.4 M
- Receptive field: (196, 240)
- Input: Fixed-height image (64px) preserving the original width

Hyperparameters

- Framework: Pytorch
- Optimizer: Adam(0.0001)
- loss: CTC

Datasets

- grayscaled line text images (300dpi)

Dataset characteristics

Dataset	Training	Validation	Test	Alphabet	Language
RIMES	9,947	1,333	778	100	French
IAM	6,482	976	2,915	79	English

Example

[RIMES]

[IAM]

Normalization techniques - visualisation

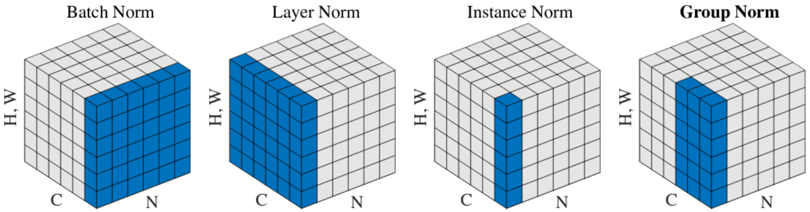


Image from [Wu2018]

- N: Mini-batch size
- H: Height
- W: Width
- C: Feature maps

Normalization techniques - results

Normalization	CER (%) 50 epochs	CER (%) 100 epochs	CER (%) 150 epochs	CER (%) 200 epochs	Time (/epoch)
Instance	6.87	**5.03**	**4.47**	**4.28**	**8.5 min**
Layer	**6.75**	5.04	**4.47**	**4.28**	15 min
Group (32)	7.10	5.30	4.86	4.32	8.75 min
Batch	9.6	5.7	5.4	4.8	**8.5 min**

Table: Effect of type of normalization for our GFCN with the RIMES dataset (for a mini-batch size of 2). CER is computed on the valid set.

Impact of ending blocks

Number of ending blocks (DSC+G+D)	CER (%) 100 epochs	CER (%) 200 epochs	Parameters	Receptive Field (h, w)
6 (baseline)	6.82	**5.80**	1,375,792	(196, 240)
5	**6.69**	5.97	1,241,904	(196, 212)
4	8.14	7.48	1,108,016	(196, 184)
3	6.93	6.23	974,128	(196, 156)
2	7.35	6.63	840,240	(196, 128)
1	8.30	7.83	706,352	(196, 100)

Table: Impact of the receptive field on the IAM dataset. CER is computed over the valid set.

IAM

Architecture	CER (%) validation	WER (%) validation	CER (%) test	WER (%) test	Parameters
2D-LSTM [Moysset2019]	5.41	20.15	8.88	29.15	0.8 M
2D-LSTM-X2 [Moysset2019]	5.40	20.40	8.86	29.31	3.3 M
CNN + 1D-LSTM [Puigcerver2017]	5.1		8.2		9.6 M
CNN + 1D-LSTM [Moysset2019]	**4.62**	17.31	**7.73**	25.22	9.6 M
Ours	5.23	21.12	7.99	28.61	1.4 M

Table: Comparative results on the IAM dataset without LM, lexicon nor data augmentation.

RIMES

Architecture	CER (%) validation	WER (%) validation	CER (%) test	WER (%) test	Parameters
2D-LSTM [Moysset2019]	3.32	13.24	4.94	16.03	0.8 M
2D-LSTM-X2 [Moysset2019]	3.14	12.48	4.80	16.42	3.3 M
CNN + 1D-LSTM [Moysset2019]	**2.9**	11.68	4.39	14.05	9.6 M
CNN + 1D-LSTM [Puigcerver2017]	3.0		**3.3**		9.6 M
Ours	3.82	15.60	4.35	18.01	1.4 M

Table: Comparative results on the RIMES dataset without LM, lexicon, nor data augmentation.

Conclusion

Our model

- A recurrent-less fully convolutional network
- Deep, with a large receptive field
- Competitive results on both RIMES and IAM datasets

Future works

Improving the performances

- Implement a data augmentation strategy

Toward paragraph-level text recognition

- Seq2Seq model with attention [Bluche2016; Bluche2017]

References I

[IAM] U. Marti et al. "The IAM-database: An English sentence database for offline handwriting recognition". In: *IJDAR* 5 (Nov. 2002), pp. 39–46. DOI: 10.1007/s100320200071.

[Graves2006] A. Graves et al. "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks". In: *ICML*. Vol. 2006. Jan. 2006, pp. 369–376.

[RIMES] Emmanuele Grosicki and Haikal El Abed. "ICDAR 2011-French Handwriting Recognition Competition". In: Sept. 2011, pp. 1459–1463. DOI: 10.1109/ICDAR.2011.290.

[Pham2014] V. Pham et al. "Dropout Improves Recurrent Neural Networks for Handwriting Recognition". In: *ICFHR* (2014).

[Voigtlaender2016] Voigtlaender et al. "Handwriting Recognition with Large Multidimensional Long Short-Term Memory Recurrent Neural Networks". In: *ICFHR*. 2016, pp. 228–233.

[Bluche2016] Théodore Bluche. *Joint Line Segmentation and Transcription for End-to-End Handwritten Paragraph Recognition*. 2016. eprint: 1604.08352.

References II

[Bluche2017] T. Bluche, J. Louradour, and R. Messina. "Scan, Attend and Read: End-to-End Handwritten Paragraph Recognition with MDLSTM Attention". In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. Vol. 01. 2017, pp. 1050–1055.

[Puigcerver2017] J. Puigcerver. "Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition?" In: *ICDAR*. 2017, pp. 67–72.

[Chowdhury2018] C. Arindam et al. *An Efficient End-to-End Neural Model for Handwritten Text Recognition*. 2018.

[Yousef2018] M. Yousef et al. *Accurate, Data-Efficient, Unconstrained Text Recognition with Convolutional Neural Networks*. 2018.

[Ptucha2018] Felipe Petroski Such et al. "Intelligent Character Recognition using Fully Convolutional Neural Networks". In: *Pattern Recognition* 88 (Dec. 2018).

[Wu2018] Yuxin Wu and Kaiming He. *Group Normalization*. 2018. arXiv: `1803.08494 [cs.CV]`.

References III

[Ingle2019] R. Ingle et al. *A Scalable Handwritten Text Recognition System*. 2019. arXiv: 1904.09150.

[Michael2019] Johannes Michael et al. *Evaluating Sequence-to-Sequence Models for Handwritten Text Recognition*. 2019. eprint: 1903.07377.

[Moysset2019] B. Moysset and R. Messina. "Are 2D-LSTM really dead for offline text recognition?" In: *IJDAR* 22 (June 2019), pp. 1–16. DOI: 10.1007/s10032-019-00325-0.

Standard Convolution

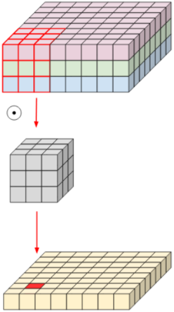
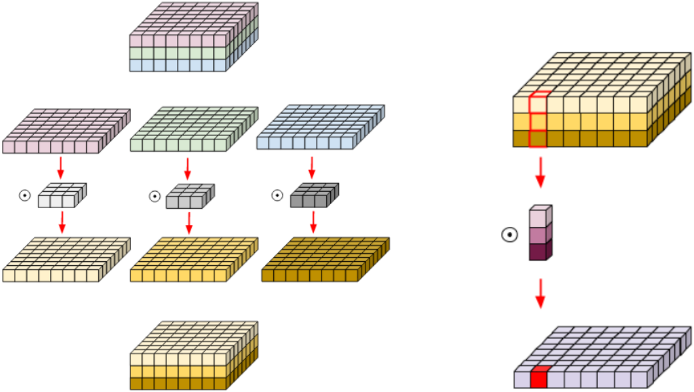


Image from
https://eli.thegreenplace.net/2018/depthwise-separable-convolutions-for-machine-learning/

Depthwise Separable Convolution



(a) Depthwise convolution (b) Pointwise convolution

Image from
https://eli.thegreenplace.net/2018/depthwise-separable-convolutions-for-machine-learning/

CTC

Image from `https://distill.pub/2017/ctc/`