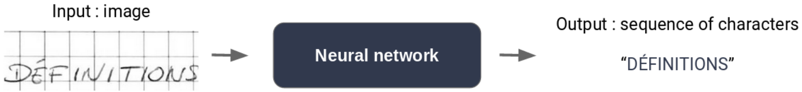
Have convolutions already made recurrence obsolete for unconstrained handwriting recognition ?

Denis Coquenet, Yann Soullard, Clément Chatelain, Thierry Paquet

LITIS Laboratory - EA 4108 Normandie University - University of Rouen, France

ICDAR-WML, 21st September 2019

Deep learning handwriting recognition system



Sequence alignment

Connectionist Temporal Classification (CTC)

- Focus on optical model only without language model nor lexicon constraints

State of the art

Recurrent models

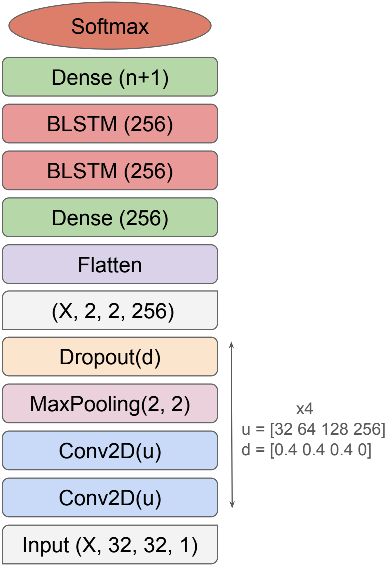
- Multi-Dimensional Long-Short Term Memory (MDLSTM) [Pham2014]
- Convolutional Neural Network + Bidirectional Long-Short Term Memory (CNN+BLSTM) [Puigcerver2017]

Non-recurrent models

- Fully Convolutional Networks (FCN) [Ptucha2018]
- FCN with gating mechanism [Yousef2018; Ingle2019]

Do we really need recurrence for handwritten text recognition ?

Our baseline model - CNN+BLSTM

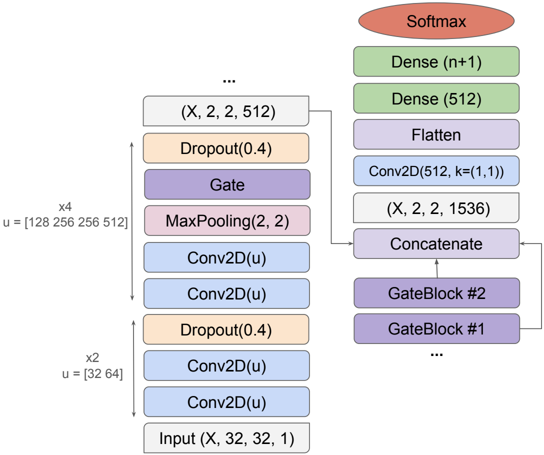


Features

- From [Soullard2019] (state-of-the-art results)
- Recurrent model
- 8 convolutions
- 2.5 million of parameters

n : number of characters in the alphabet

Our Gated-CNN - overview

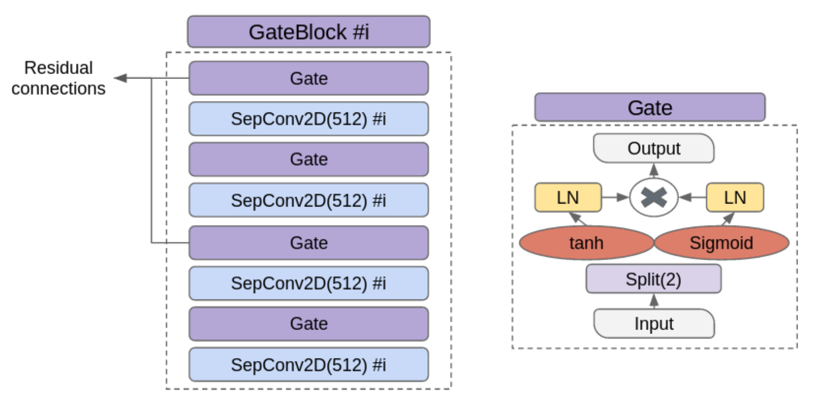


Features

- Based on baseline model
- Non-recurrent model
- 21 convolutions
- 6.9 million of parameters
- Shared weight layers
- Depthwise Separable Convolutions
- Residual connections

Our G-CNN - gates

Gating mechanism



RIMES dataset

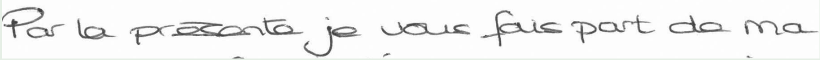
Dataset characteristics

- +1,300 writers
- French writings
- 12,723 pages segmented into lines

RIMES dataset split

Training	Validation	Test	Alphabet
9,947	1,333	778	100

Example



First experiment : Raw comparison

Architecture	CER(%) validation	CER (%) test	Training time	Parameters (M)
CNN+BLSTM	**6.98**	**6.88**	1d22h59	4.1
CNN+Dense only	17.73	19.03	1h10	1.5
G-CNN	9.92	10.03	10h00	6.9

- BLSTM layers responsible for a large amount of parameters (2.6 M)

First experiment : Raw comparison

Architecture	CER(%) validation	CER (%) test	Training time	Parameters (M)
CNN+BLSTM	**6.98**	**6.88**	1d22h59	4.1
CNN+Dense only	17.73	19.03	1h10	1.5
G-CNN	9.92	10.03	10h00	6.9

- BLSTM layers responsible for a large amount of parameters (2.6 M)
- BLSTM layers increase performance dramatically (-12.15 in test)

First experiment : Raw comparison

Architecture	CER(%) validation	CER (%) test	Training time	Parameters (M)
CNN+BLSTM	**6.98**	**6.88**	1d22h59	4.1
CNN+Dense only	17.73	19.03	1h10	1.5
G-CNN	9.92	10.03	10h00	6.9

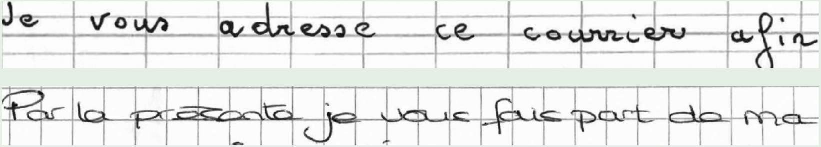
- BLSTM layers responsible for a large amount of parameters (2.6 M)
- BLSTM layers increase performance dramatically (-12.15 in test)
- G-CNN : more parameters but training time shorter (parallel computing)

Second experiment - Robustness against complexified data

Modified version of RIMES dataset

Lined paper background addition

Examples



- Similar behavior - CER increased by 2.39 for the CNN+BLSTM and 2.52 for the G-CNN

Third experiment - Impact of data augmentation

7-time augmented training set

- Raw
- contrast alteration
- sign flipping
- long/short scaling
- width/height dilation

- CER decreased by 1.3 for the G-CNN and 0.94 for the CNN+BLSTM

Conclusion

CNN+BLSTM

- Better performance
- Longer training time

G-CNN

- Deeper networks, bigger receptive fields
- Architecture and tuning complex
- Gating mechanism almost enables to reach the same performance

Future works: exploring other alternatives

- Toward an even lighter network with FCN
- Attention models [Michael2019]

References

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

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

[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).

[Soullard2019] Y. Soullard et al. *CTCModel: a Keras Model for Connectionist Temporal Classification*. 2019.

[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.