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

Denis Coquenet, Clément Chatelain, Thierry Paquet
LITIS Laboratory - EA 4108 Normandie University - University of Rouen, France

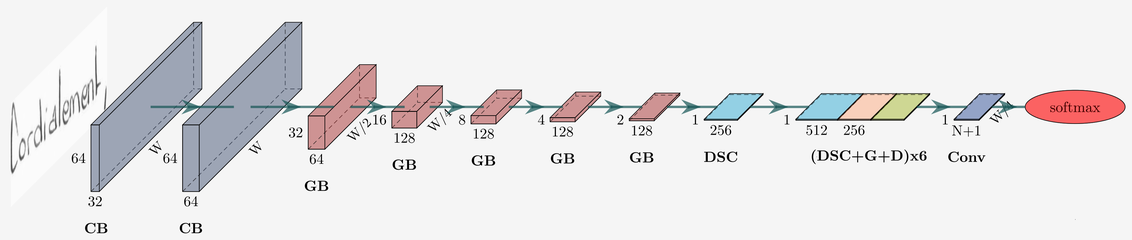
Introduction

Unconstrained handwritten text recognition is a major step in most document analysis tasks. Recurrent models are usually used for handwritten text recognition (CNN + LSTM, MDLSTM, seq2seq). The main drawbacks of these components are the large number of parameters involved and their sequential execution during training and prediction.

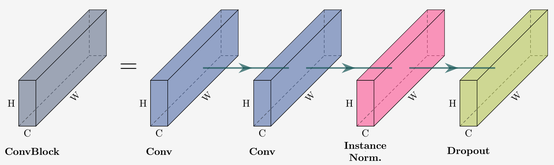
Objective: propose a recurrence-free model reaching competitive results for that task, a Gated Fully Convolutional Network (GFCN).

Source code: `https://github.com/FactoDeepLearning/LinePytorchOCR`.

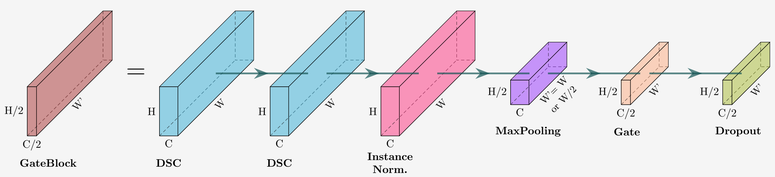
Architecture - GFCN



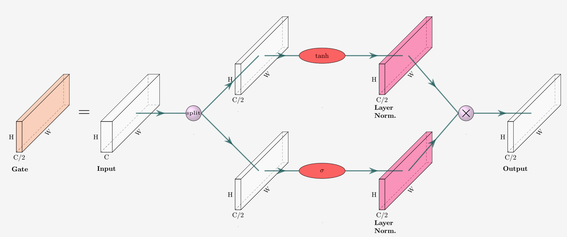
(a) Overview of the GFCN.



(b) ConvBlock (CB) definition.



(c) GateBlock (GB) definition.



(d) Gating mechanism (G). White elements are just representations of the tensors at a given step (no specific operation is performed).

CB : ConvBlock, GB: GateBlock, DSC: Depthwise Separable Convolution, G: Gate, D: Dropout.

Key components

- State of the art: LSTM based models
  - Imply a lot of parameters
  - Sequential execution, longer training and prediction times
- How to avoid recurrence ?
  - Stack a lot of convolutional layers to increase the receptive field and thus the context taken into account
  - Use a gating mechanism to compensate the selection operated in LSTM cells
- How to reduce even more the number of parameters ?
  - Do not use densely connected layers: being fully convolutional
  - Use Depthwise Separable Convolutions which are lighter than standard ones
  - Goals: being faster, more stable / better generalization / fewer data required.

Datasets

Recognition of text lines of 2 datasets : IAM and RIMES.

	Train	Validation	Test	Charset size	Language
RIMES	9,947	1,333	778	100	French
IAM	6,482	976	2915	79	English

Split, charset size and language of the different datasets.

Preprocessing: images are resized to obtain a height of 64 pixels, preserving the original width.

Results

- No language model, data augmentation nor lexicon constraints

Architecture	CER (%) IAM	WER (%) IAM	CER (%) RIMES	WER (%) RIMES	Parameters
2D-LSTM [1]	8.88	29.15	4.94	16.03	0.8 M
2D-LSTM-X2 [1]	8.86	29.31	4.80	16.42	3.3 M
CNN + 1D-LSTM [1]	8.2		4.39	14.05	9.6 M
CNN + 1D-LSTM [2]	**7.73**	25.22	**3.3**		9.6 M
Ours [3]	7.99	28.61	4.35	18.01	1.4 M

Comparative results on the RIMES and IAM test set.

Conclusion

We propose a neural network for handwritten text recognition combining multiple advantages:

- Recurrence-free: parallel computations, smaller training and prediction times
- Only use convolutional components: less parameters
- Use of a gating mechanism: select the relevant information throughout the model
- Very deep: bigger receptive field (196 in height, 240 in width)
- Competitive results on IAM and RIMES datasets

References

[1] B. Moysset et al. "Are 2D-LSTM really dead for offline text recognition?" In: *IJDAR* 22 (June 2019), pp. 1–16.
[2] J. Puigcerver. "Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition?" In: *ICDAR* (2017), pp. 67–72.
[3] D.Coquenet, C. Chatelain, and T.Paquet. "Recurrence-free unconstrained handwritten text recognition using gated fully convolutional network". In: *17th ICFHR*. 2020, pp. 19–24.