Handwritten Text Recognition using TensorFlow 2.0

Computer Vision

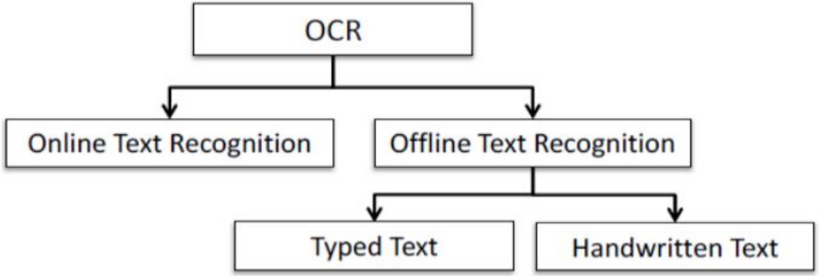
Arthur Flor
afsn@ecomp.poli.br

Content

Introduction

Offline Handwritten Text Recognition Systems (HTR) has the purpose of transcribing cursive text contained in images to the digital medium (ASCII, Unicode)

(BUNKE; ROTH; SCHUKAT-TALAMAZZINI, 1995)



Classification of Optical Character Recognition.
(SONKUSARE; SAHU, 2016)

Objectives

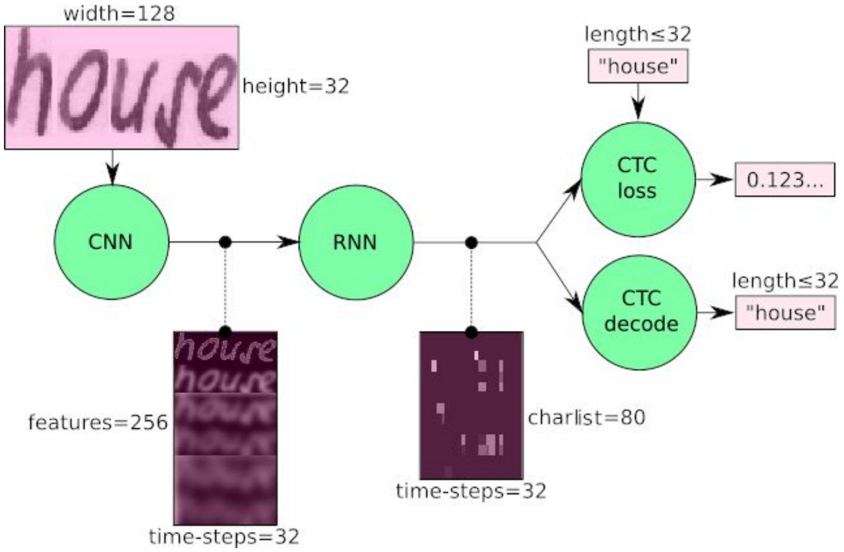
- ➔ To develop HTR system using TensorFlow 2.0;
- ➔ To implement optical models of the literature;
- ➔ To implement a new optical model inspired by the literature;
- ➔ To provide easy code for development for HTR systems (code, documentation, tutorial).

Handwritten Text Recognition System

CNN: these layers are trained to extract **relevant features** from the image.

RNN: these layers propagates relevant information through the CNN output. The output sequence is mapped to a matrix of size **time-steps**x**charlist**.

CTC: with the output of the RNN, calculates the **loss** and also **decodes** into final text.



Example of HTR system and the passage of data through the optical model (SCHEIDL, 2018)

Handwritten Text Recognition System

charlist is a character dictionary used to encode the ground truth into a sequence of numbers. The charlist is usually created from the characters of the dataset that will be trained.

For example, IAM dataset has 80 characters (CTC blank label added):

!"#&'()*+,-./0123456789:;?ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz

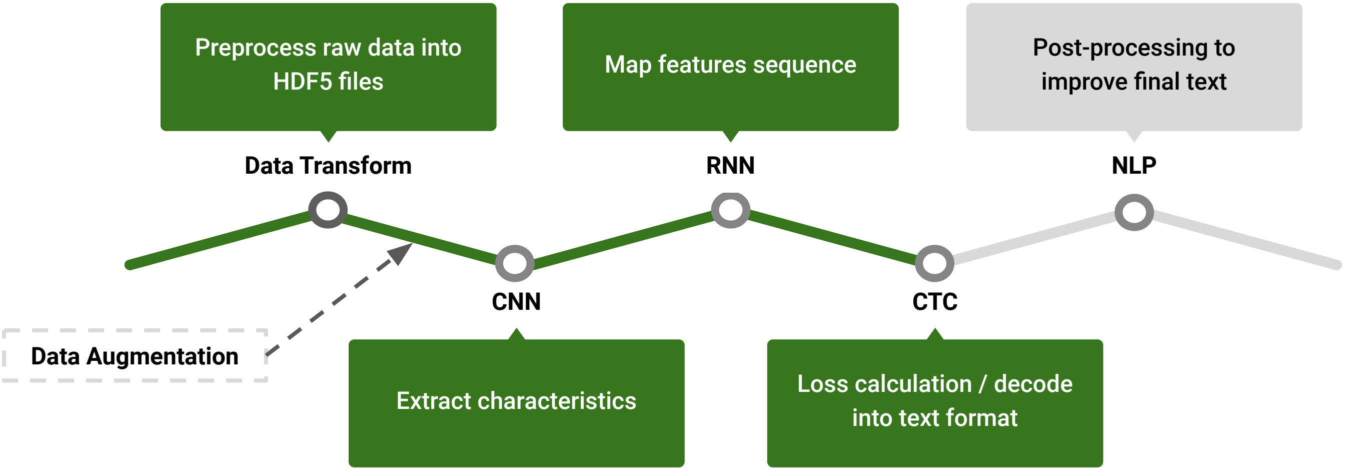
house = [60, 67, 73, 71, 57]

Tools

- ➔ Python 3.x
- ➔ OpenCV 4.x
- ➔ TensorFlow 2.x Beta

- ➔ Google Colaboratory/Drive

HTR Project Workflow



Datasets

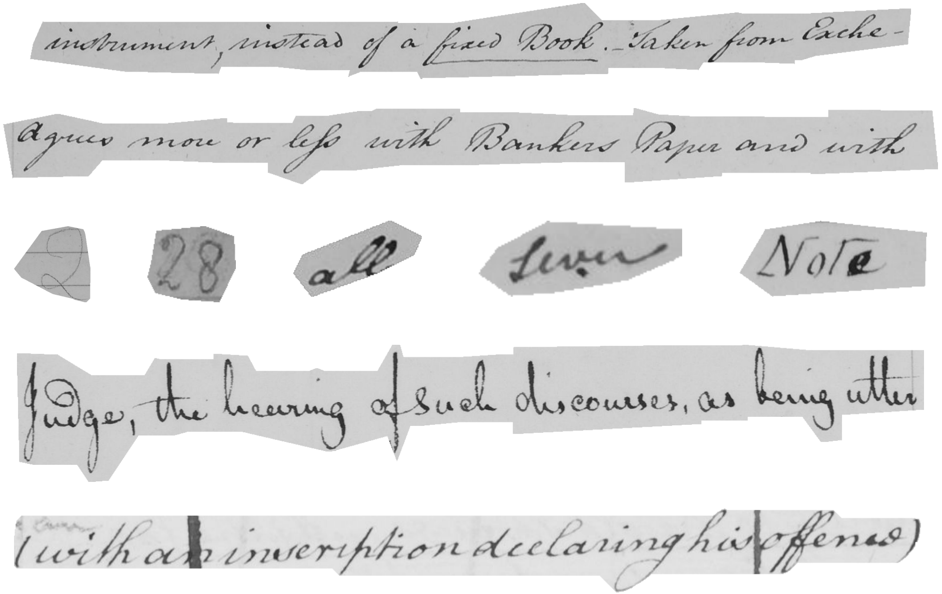
➔ **Bentham** (DIEM et al., 2014)
http://transcriptorium.eu/datasets/bentham-collection

➔ **IAM** (MARTI; BUNKE, 2002)
http://www.fki.inf.unibe.ch/databases/iam-handwriting-database

➔ **Rimes** (GROSICKI et al., 2008)
http://www.fki.inf.unibe.ch/databases/iam-historical-document-database/saint-gall-database

➔ **Saint Gall** (FISCHER et al., 2010)
http://www.a2ialab.com/doku.php?id=rimes_database:start

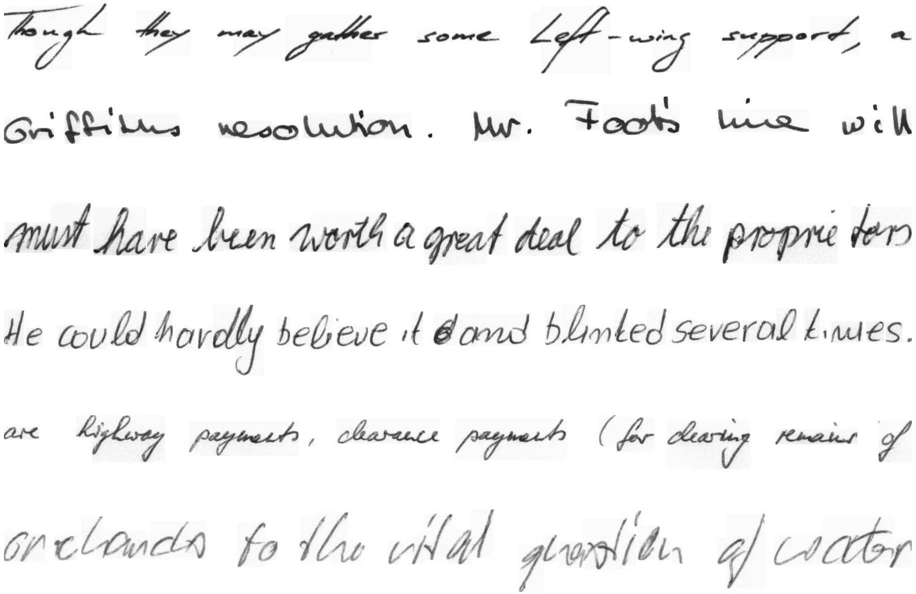
Bentham



Partitions	Characters	Words	Lines
Train	424,199	86,015	9,195
Validation	64,679	12,958	1,415
Test	38,429	6,960	860
Total	527,307	105,933	11,470

Number of images per dataset partition

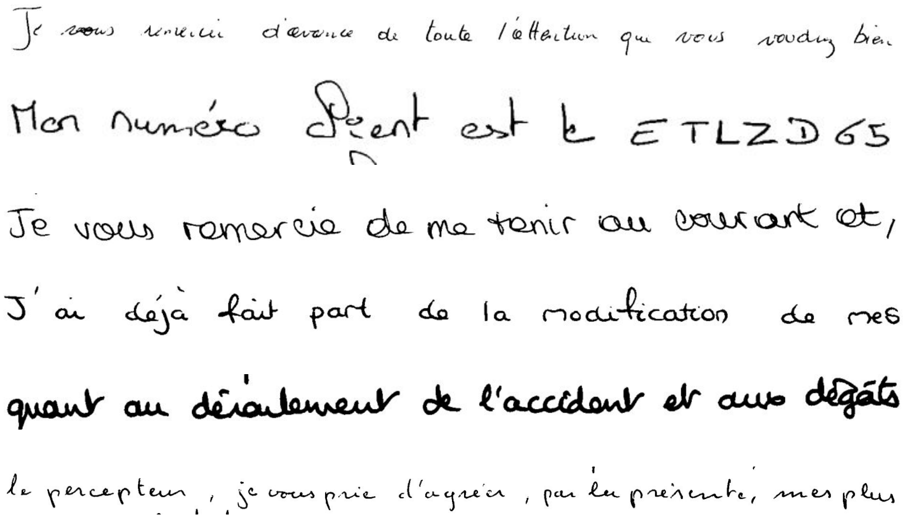
IAM



Partitions	Characters	Words	Lines
Train	274,964	53,728	6,161
Validation	40,770	7,899	900
Test	81,676	17,616	1,861
Total	397,410	79,246	8,922

Number of images per dataset partition

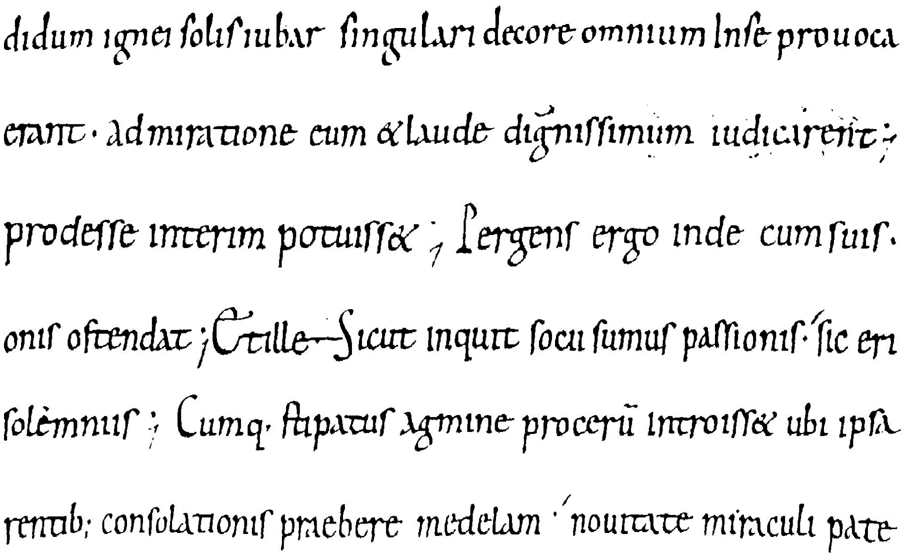
Rimes



Partitions	Characters	Words	Lines
Train	458,737	73,798	10,193
Validation	51,968	8,404	1,133
Test	35,194	5,639	778
Total	545,899	87,841	12,104

Number of images per dataset partition

Saint Gall



Partitions	Characters	Words	Lines
Train	26,909	3,933	468
Validation	13,107	1,982	235
Test	40,243	6,160	707
Total	80,259	12,075	1,410

Number of images per dataset partition

Preprocessing (Images)

➔ Resize with padding: 1024x128

➔ Illumination Compensation (CHEN et al., 2012)

➔ Deslant Cursive Images (VINCIARELLI; LUETTIN, 2001)

Preprocessing (Images)

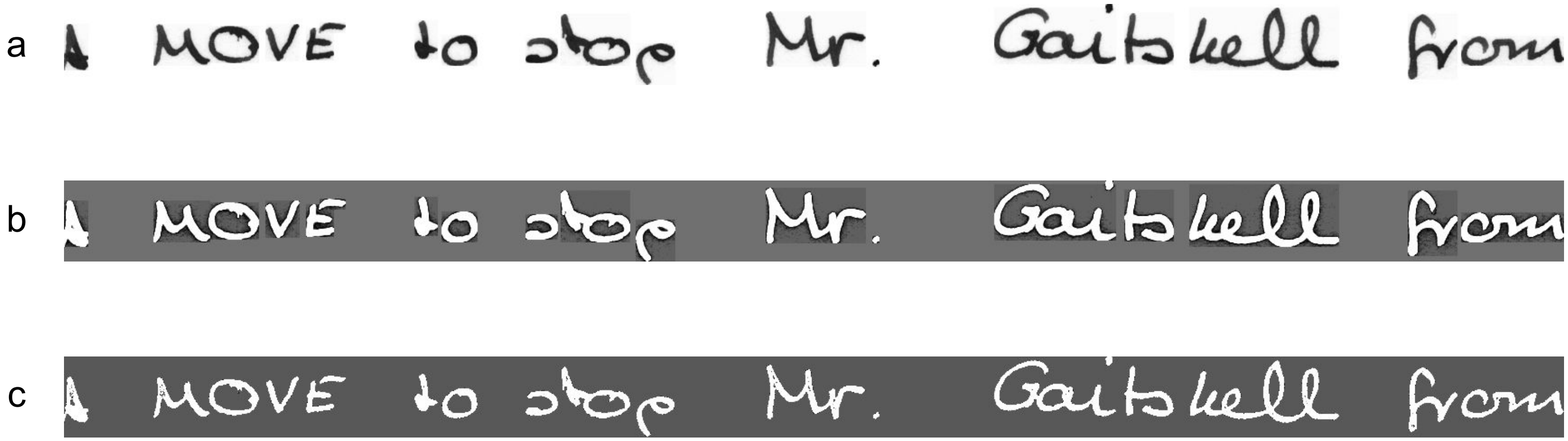
a

b

c

Bentham example: original (a), normalization (b), illumination + deslanting + normalization (c)

Preprocessing (Images)



IAM example: original (a), normalization (b), illumination + deslanting + normalization (c)

***Note**: memory allocation (int / float)*

Preprocessing (Text)

1. **Accentuation normalization**
2. **charlist**: 95 printable characters from ASCII (letters of the latin alphabet, punctuation marks and math signs) + CTC blank label added:

 !"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~

GT: et les documents nécessaires à cette adhèsion .

PP: et les documents necessaires a cette adhesion .

Sparse: [69, 84, 0, 76, 69, 83, 0, 68, 79, 67, 85, 77, 69, 78, 84, 83, 0, 78, 69, 67, 69, 83, 83, 65, 73, 82, 69, 83, 0, 65, 0, 67, 69, 84, 84, 69, 0, 65, 68, 72, 69, 83, 73, 79, 78, 0, 14]

Data Augmentation

- ➔ Rotation and Scale
- ➔ Height and width shift
- ➔ Morphological transformations

***Note**: keras generator / openCV*

Optical Models (Related work)

2017 - Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition

2017 - Gated Convolutional Recurrent Neural Networks for Multilingual Handwriting Recognition

2018 - Are 2D-LSTM really dead for offline text recognition

2018 - Word Beam Search A Connectionist Temporal Classification Decoding Algorithm

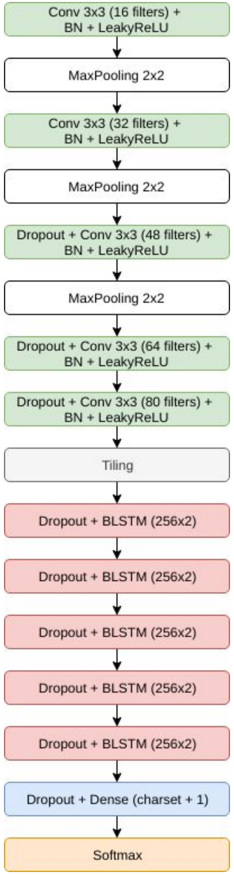
2018 - Desarrollo y experimentación de un sistema de aprendizaje profundo para redes neuronales convolucionales y recurrentes

2018 - Medicine Box, Doctor's Prescription Recognition Using Deep Machine Learning

2018 - An Efficient End-to-End Neural Model for Handwritten Text Recognition

2019 - A Scalable Handwritten Text Recognition System

Puigcerver Model



- ➔ 5 convolutional layers with LeakyReLU activations
- ➔ 5 BLSTM (256x2 units) with dropout system (PHAM et al., 2014)
- ➔ Dense + Softmax (output size equal to charlist + 1)

- 9,590,176 total parameters
- RMSProp optimizer with learning rate equal to 3e-4

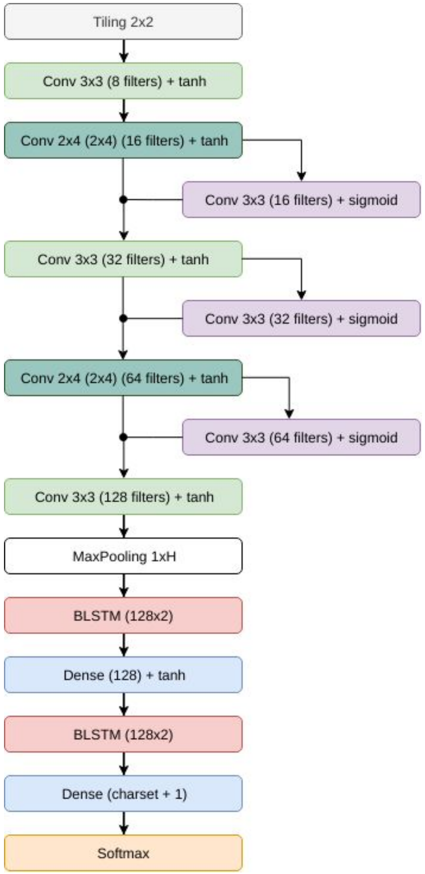
Bluche Model

➔ 8 convolutional layers (3 Gated) with tanh activations
  - Gated Convolutional from Bluche and Messina (2017)

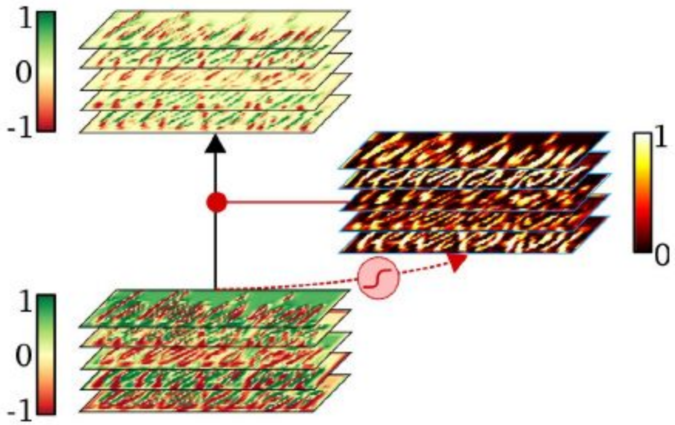
➔ 2 BLSTM (128x2 units)

➔ Dense + Softmax (output size equal to charlist + 1)

- 728,680 total parameters
- RMSProp optimizer with learning rate equal to 4e-4



Gated Convolutional (Bluche and Messina)



Gated Convolutional Recurrent Neural Networks for Multilingual Handwriting Recognition (BLUCHE; MESSINA, 2017)

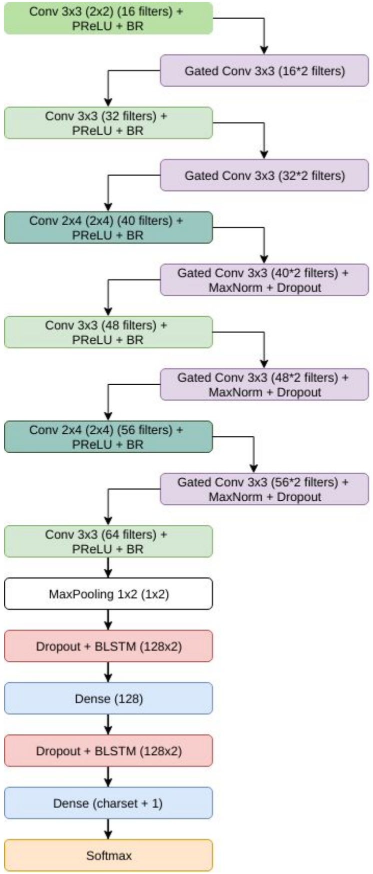
Flor Model

➔ 11 convolutional layers (5 Gated) with PReLU activations
  ○ Gated Convolutional from Dauphin et al. (2017)

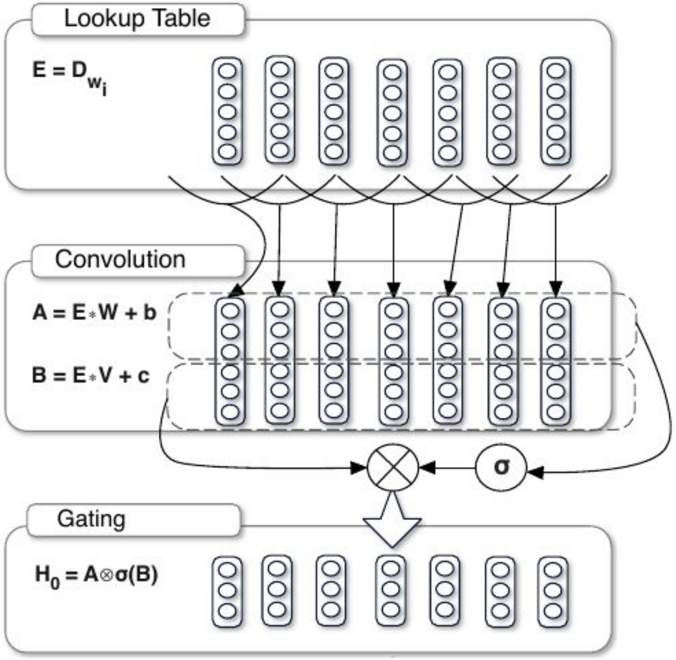
➔ 2 BLSTM (128x2 units) with dropout system (PHAM et al., 2014)

➔ Dense + Softmax (output size equal to charlist + 1)

- 822,140 total parameters
- RMSProp optimizer with learning rate equal to 5e-4



Gated Convolutional (Dauphin et al.)



Language Modeling with Gated Convolutional Networks
(DAUPHIN; FAN; AULI; GRANGIER, 2017)

Connectionist Temporal Classification (CTC)

- ➔ Token Passing (GRAVES et al., 2009)
- ➔ **Vanilla Beam Search (HWANG; SUNG, 2016)**
- ➔ Word Beam Search (SCHEIDL; FIEL; SABLATNIG, 2018)

~ **No** language model applied in post-processing ~

Experiment Setup

- ➔ Default models and learning rates
- ➔ Mini-batch: 16
- ➔ Reduce Learning Rate on Plateau (10 epochs)
- ➔ EarlyStopping if validation loss doesn’t improve after 20 epochs
- ➔ Metrics: Character Error Rate (CER) and Word Error Rate (WER)

- ➔ GPU: Nvidia Tesla T4 16gb (Google Colab default settings)

Bentham Results

860 images (test)

Model	Default		No punctuation	
	CER	WER	CER	WER
Puigcerver	0.1175	**0.3931**	0.0916	0.2443
Bluche	0.1658	0.4665	0.1424	0.3581
Flor	**0.1131**	0.3966	**0.0849**	**0.2404**

IAM Results

1,861 images (test)

Model	Default		No punctuation	
	CER	WER	CER	WER
Puigcerver	0.0939	0.2934	0.0932	0.3213
Bluche	0.1430	0.4117	0.1422	0.4452
Flor	**0.0858**	**0.2790**	**0.0852**	**0.3058**

Rimes Results

778 images (test)

Model	Default		No accentuation		No punctuation	
	CER	WER	CER	WER	CER	WER
Puigcerver	0.1050	0.3919	0.0856	0.3129	0.1053	0.3818
Bluche	0.1331	0.4640	0.1140	0.4054	0.1330	0.4527
Flor	**0.0874**	**0.3781**	**0.0678**	**0.2975**	**0.0870**	**0.3657**

Saint Gall Results

707 images (test)

Model	Default	
	CER	WER
Puigcerver	0.0965	0.4199
Bluche	0.1267	0.5248
Flor	**0.0906**	**0.4033**

Conclusions

- ➔ Bluche model and dropout (MOYSSET; MESSINA, 2018)
- ➔ Flor model:
  - PReLU and LeakyReLU
  - Batch Renormalization and Batch Normalization
- ➔ Reduce LR on Plateau x Warms restarts x EarlyStopping
- ➔ Convolutional alternatives:
  - SeparableConv2D and DepthwiseConv2D (HOWARD et al., 2017)
  - Octave Convolution (CHEN et al., 2019)
- ➔ Repository with code, documentation and tutorial (Jupyter Notebook):
  - https://github.com/arthurflor23/handwritten-text-recognition

References

➔ BLUCHE, T. Deep Neural Networks for Large Vocabulary Handwritten Text Recognition. Tese (Doutorado) — Université Paris-Sud, 05 2015.

➔ BLUCHE, T. Joint line segmentation and transcription for end-to-end handwritten paragraph recognition. Advances in Neural Information Processing Systems, p. 838–846, 2016.

➔ BLUCHE, T.; MESSINA, R. Gated convolutional recurrent neural networks for multilingual handwriting recognition. 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), p. 646–651, 11 2017.

➔ BUNKE, H.; ROTH, M.; SCHUKAT-TALAMAZZINI, E. G. Off-line cursive handwriting recognition using hidden markov models. Pattern Recognition, v. 28, p. 1399–1413, 09 1995.

➔ CHEN, Kuo-Nan; CHEN, Chin-Hao; CHANG, Chin-Chen. Efficient illumination compensation techniques for text images. Digital Signal Processing, v. 22, p. 726-733, 2012;

➔ CHEN, Y.; FAN, H.; XU, B.; YAN, Z.; KALANTIDIS, Y.; ROHRBACH, M.; YAN, S.; FENG J. Drop an Octave: Reducing Spatial Redundancy in Convolutional Neural Networks with Octave Convolution, 2019.

References

- ➔ CHOWDHURY, A.; VIG, L. An efficient end-to-end neural model for handwritten text recognition. In: 29th British Machine Vision Conference (BMVC). Newcastle upon Tyne: BMVA, 2018.
- ➔ DIEM, M. et al. Competition on handwritten digit string recognition in challenging datasets (hdsrc 2014). 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), v. 2014, 12 2014.
- ➔ FISCHER, A. et al. Ground truth creation for handwriting recognition in historical documents. ACM International Conference Proceeding Series, p. 3–10, 01 2010.
- ➔ GEORGE, C. et al. Survey on handwritten character recognition using artificial neural network. IJSTE - International Journal of Science Technology & Engineering, v. 2, p. 287–290, 04 2016.
- ➔ GRAVES, A.; FERNáNDEZ, S.; GOMEZ, F. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. International Conference on Machine Learning, p. 369–376, 2006.
- ➔ GRAVES, A.; FERNáNDEZ, S.; SCHMIDHUBER, J. Multidimensional recurrent neural networks. International Conference on Artificial Neural Networks, 2007.

References

➔ GRAVES, A. et al. A novel connectionist system for unconstrained handwriting recognition. IEEE transactions on pattern analysis and machine intelligence, v. 31, p. 855–68, 06 2009.

➔ GROSICKI, E. et al. Rimes evaluation campaign for handwritten mail processing. In: ICFHR 2008 : 11th International Conference on Frontiers in Handwriting Recognition. Montreal, Canada: Concordia University, 2008. p. 1–6. Disponível em: .

➔ HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. Neural computation, v. 9, p. 1735–80, 12 1997.

➔ HOWARD A.G.; ZHU, M.; CHEN B.; KALENICHENKO D.; WANG W.; WEYLAN T.; ANDREETTO M.; ADAM H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017.

➔ HWANG, K.; SUNG, W. Character-level incremental speech recognition with recurrent neural networks. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 3 2016. Disponível em: .

➔ IOFFE, S.; SZEGEDY, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2015.

References

- ➔ IOFFE, S. Batch Renormalization: Towards Reducing Minibatch Dependence in Batch-Normalized Models. 2017.
- ➔ KAMALANABAN, E.; GOPINATH, M.; PREMKUMAR, S. Medicine box: Doctor's prescription recognition using deep machine learning. International Journal of Engineering and Technology(UAE), v. 7, p. 114–117, 09 2018.
- ➔ LECUN, Y. et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, v. 86, n. 11, p. 2278–2324, 11 1998. ISSN 0018-9219.
- ➔ LOSHCHILOV, I.; HUTTER, F. SGDR: Stochastic Gradient Descent with Warm Restarts. 2016.
- ➔ MAAS, A. L.; HANNUN, A. Y.; NG, A. Y. Rectifier nonlinearities improve neural network acoustic models. In: ICML Workshop on Deep Learning for Audio, Speech and Language Processing. Atlanta, Georgia: JMLR.org, 2013.
- ➔ MARTI, U.-V.; BUNKE, H. The iam-database: An english sentence database for offline handwriting recognition. International Journal on Document Analysis and Recognition, v. 5, p. 39–46, 11 2002.

References

- ➔ MOYSSET, B.; MESSINA, R. O. Are 2d-lstm really dead for offline text recognition? CoRR, abs/1811.10899, 2018.
- ➔ PALEHAI, D.; FANANY, M. I. Handwriting recognition on form document using convolutional neural network and support vector machines (cnn-svm). 5th International Conference on Information and Communication Technology (ICoIC7), 05 2017.
- ➔ PHAM, V. et al. Dropout improves recurrent neural networks for handwriting recognition. 2014 14th International Conference on Frontiers in Handwriting Recognition, IEEE, Sep 2014. Disponível em: .
- ➔ PUIGCERVER, J. Are multidimensional recurrent layers really necessary for handwritten text recognition? 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), p. 67–72, 11 2017.
- ➔ SCHEIDL, H.; FIEL, S.; SABLATNIG, R. Word beam search: A connectionist temporal classification decoding algorithm. In: 16th International Conference on Frontiers in Handwriting Recognition (ICFHR). Niagara Falls, United States: IEEE Computer Society, 2018. p. 253–258.
- ➔ SCHUSTER, M.; PALIWAL, K. Bidirectional recurrent neural networks. Trans. Sig. Proc., IEEE Press, Piscataway, NJ, USA, v. 45, n. 11, p. 2673–2681, 11 1997.

References

➔ SINGH, S. et al. Handwritten words recognition for legal amounts of bank cheques in english script. ICAPR 2015 - 2015 8th International Conference on Advances in Pattern Recognition, p. 1–5, 02 2015.

➔ SONKUSARE, M.; SAHU, N. A survey on handwritten character recognition (hcr) techniques for english alphabets. Advances in Vision Computing: An International Journal, v. 3, p. 1–12, 03 2016.

➔ SOULLARD, Y.; RUFFINO, C.; PAQUET, T. CTCModel: a Keras Model for Connectionist Temporal Classification. 2019.

➔ SáNCHEZ, J.-A. et al. Icfhr2016 competition on handwritten text recognition on the read dataset. 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), p. 630–635, 10 2016.

➔ SáNCHEZ, J. A. et al. A set of benchmarks for handwritten text recognition on historical documents. Pattern Recognition, v. 94, p. 122–134, 2019. ISSN 0031-3203.

➔ VINCIARELLI, Alessandro; LUETTIN, Juergen. A New Normalization Technique for Cursive Handwritten Words. Pattern Recognition Letters 22, 2001;