Transfer Learning to improve Arabic handwriting text Recognition

Zouhaira Noubigh
University of Sousse,
ISITCom, 4011, Sousse, Tunisia
zouhaira.noubigh@gmail.com

Anis Mezghani
University of Sfax
Higher Institute of Industrial Management, Tunisia
Anis.mezghani@gmail.com

Monji Kherallah
University of Sfax
Faculty of Sciences of Sfax, Tunisia
Monji.kherallah@gmail.com

***Abstract*—In recent years, the leveraging of deep learning approaches allows a great progress in text recognition task. But they usually need a considerable amount of training examples to learn a new model. Therefore, lack of data can be an issue when developing a new recognition model, especially for handwriting Arabic text recognition where the lack of databases is stilling an interested problem. In this context, the main contributions of this paper is based on transfer learning the parameters learned with a bigger mixed-fonts printed Arabic text database to handwriting one. Experiments shows the good improvement provide with this technique.**

Keywords—Deep learning; handwriting Arabic text; printed Arabic text; transfer learning

I. Introduction

The text recognition problem is one of the most complex tasks. The difficulty of recognition is related to the writing style; more writing is legible and regular, more the resolution is easy[1]. The performance obtained by recognition system varies greatly with the clarity of the images provided. Three writing styles in ascending order of difficulty can be distinguished: the printed word, handwriting but capital letters or characters sticks, and handwriting cursive. Handwriting remains an important means of communication, even if the printed takes a significant part in various applications thanks to its legibility, safety and speed of communication[2].

In recent years, the Deep Learning (DL) technologies proved itself as the most important models in computer vision and pattern recognition due to the great performance improvement they have provided [3]. Firstly, the Convolutional Neural Network (CNN) has provided an efficient solution for handwritten characters and digits recognition and has been employed in many researches [4]. Secondly, Recurrent Neural Network (RNN), especially deep Bidirectional LSTM and Multidimensional LSTM, is successfully used for sequence modeling task especially for text line recognition[5].

Although, important results have been recorded in the handwriting Latin text recognition, till recent date, the most Arabic handwriting recognition systems are limited to the isolated characters, digits or word with a limited vocabulary [6][7]. Very few researches are presented for recognition of unconstrained Arabic text. The recognition of Arabic text presents many challenges because of many characteristics of this script. Mainly, the cursive nature of this script and its characters connectivity make its recognition very sophisticated process. In addition, an interested problem related to the handwriting Arabic text recognition is the lack of big database that is freely available for researchers. Large standard databases are an essential requirement for handwriting recognition research and development. Many Arabic databases have been produced but fewer have become publicly available. In fact, researchers have developed their private databases representing generally small dictionaries with limited lexicon or just isolated forms of letters or/and digits. As a result, until now, there is no robust standard comprehensive database devoted to off-line Arabic handwriting text recognition.

In order to deal with this, we use in this work recent approach related to the deep learning namely transfer learning technique [8].

Transfer learning is a beneficial method for transferring trained system parameters to a new but related task. It is based in the theory that the learned parameters in deep layers do not change much as the features keep similar for the new data. More broadly, transfer learning based in utilizing model pre-trained on a source domain as a starting point and convert it for target domain in order to reduce the data and training time [8].

In this paper, the used recognition model is based on combining CNN with DBLSTM (CNN-DBLSTM) and connectionist temporal classification (CTC) beam search decoder [9].

The rest of this paper is organized as follows: in section 2, we present the architecture of the trained model for text recognition, section 3 deposit and explain the experimental results with an overview of the used databases and in section 4 we conclude our work.

II. Recognition Model

In this work, we propose using the CNN-BLSTM-CTC model that provides its performance for text recognition task [10]. An overview of the proposed architecture is illustrated in Fig.1. It is composed by two main steps that are features extraction and sequences modeling.

The selection of features extraction method remains the most important step in the recognition process. The recent deep learning networks, especially the Convolutional Neural

978-1-7281-8855-3/20/$31.00 ©2020 IEEE

Networks (CNNs) provide efficient solutions for features extraction where deep layers act as a set of features extractors [9]. They extract learning features which are generic and independent of any specific classification task.

The sequence modeling step of our handwriting recognition model is based on combining BLSTM recurrent network and the Connectionist Temporal Classification (CTC). Long Short Term Memory (LSTM) cells have been introduced by Hochreiter et al. [13] to overcome the vanishing gradient problem of RNN. LSTM present a memory cell controlled by the input, forget and output gates [10] which allowed the learning of long term memory dependencies.

The input is a grayscale image of size 64x1024. The first layer in CNN is the convolution layer. In this layer, a sliding matrix called filter is used to find features everywhere in the image. CNN multiplies each pixel in the image with each value in the filter for each filters and the output of this layer will be a set of filtered images. The architecture of our system consists of 10 convolution layers. The filters are of size 3x3. The Rectified Linear Unit (ReLU) activation function is implemented to produce an output after each convolution. A max pooling Layer is used to summarize image regions and outputs a downsized version of the previous layer. CNN output is introduced to a BLSTM layer.

It based on one BLSTM layer with 512 neurons in each direction. A dropout layer is applied to the LSTM cells before inputted in the two hidden layers process the input sequence in two directions. The BLSTM output is a matrix of size T×(C+1), T denotes the time step length and C is the number of characters with a pseudo- character called blank. This matrix is fed into the CTC beam search decoding algorithm.

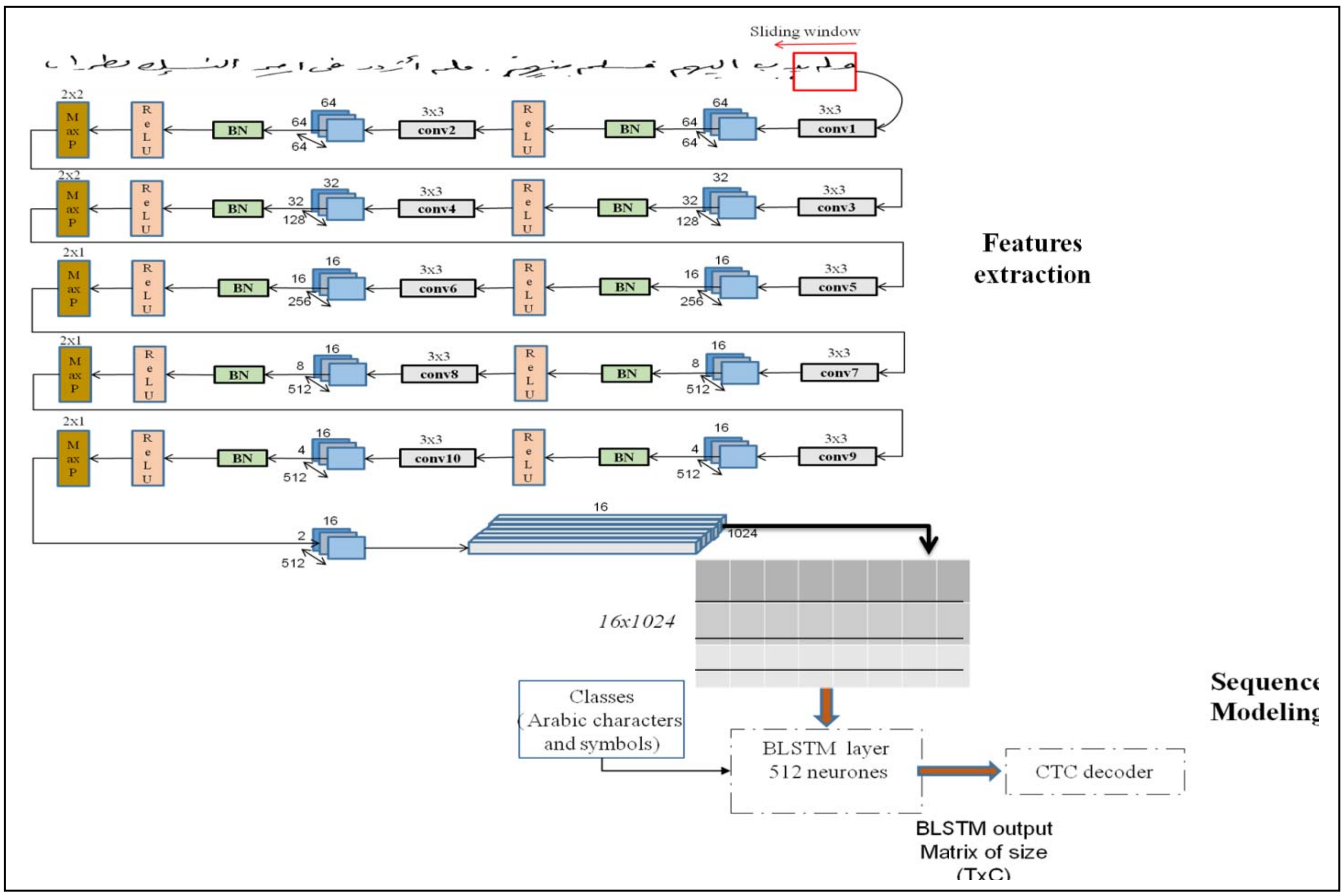

Fig.1. The recognition model architecture

III. Experiments setup

A. Databases

The experiments in this paper are conducted using three Arabic text line databases. P-KHATT contains printed Arabic text line images chosen for training and two handwriting databases namely KHATT and AHTID using for validation. They will be introduced in this section and Table.1 gives useful statistics about them.

1) KHATT database

The offline Handwritten Arabic Text database KHATT [11] contains 4000 grayscale paragraphs images and its ground-truth grouped in two categories.

The first present 2000 of these images contain similar text each covering all Arabic characters and shapes whereas the remaining. The second present 2000 images contain free texts written by the writers on any topic of their choice in an unrestricted style. Those paragraphs are divided in 9494 text line images for training, 2007 for testing and 1901 for validation.

KHATT database is known as a challenging handwriting Arabic database. It presents text line images that are hardly recognized by human. Additionally, the train set of this database don't present enough examples, about the half of the text line images are unique. Fig.2 shows examples of images extracted from KHATT database.

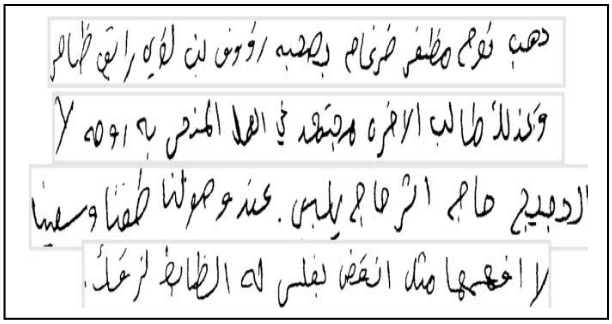

Fig.2. Examples of images extracted from KHATT database

2) P-KHATT database

The P-KHATT corpus is proposed by Ahmad et al. [12] for research in the area of printed text recognition. It contains images with the same ground truth with those existent in the KHATT database in eight different fonts. In this work we use only four fonts that appear similar to the handwriting script namely Thuluth, Andalus, Tahoma and Naskh. Fig.3 presents samples of text line images from the P-KHATT database in the four fonts.

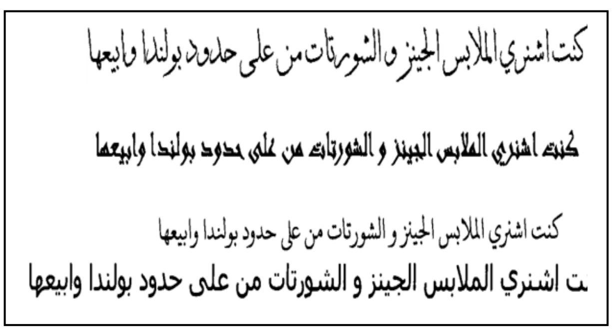

Fig.3. Samples of text line image from the P-KHATT database in four fonts

3) AHTID/MW database

The Arabic Handwritten Text Images Database AHTID/MW has been built at the MIRACL Lab, ISIMS, and University of Sfax - Tunisia in join collaboration with the Institute for Communications Technology, Braunschweig Germany [13].

The AHTID/MW contains 3710 text lines and 22,896 words written by 53 native writers of Arabic. These images are divided into five equilibrated sets. The four first sets are available for scientific community. The database is freely available for worldwide researchers. Fig.4 shows examples of text line images extracted from the database.

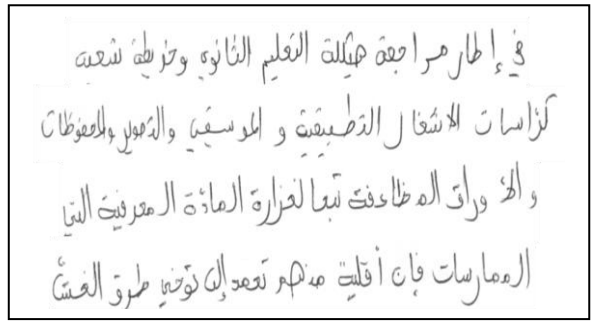

Fig.4. Samples of text line image from the AHTID/MW database

TABLE.1. STATISTICS ABOUT THE USING DATABASES

Dataset	*Train*	*Test*	*Validation*
KHATT	*9,494*	*2,007*	*1,901*
P-KHATT (for each font)	*6472*	*1414*	*1424*
AHTID	*2699*	*901*	*901*

B. Transfer learning for AHTR

Transfer learning aims at improving the performance of a model by transferring the knowledge contained in different but related domains. Due to the wide application prospects, transfer learning has become a popular and promising area in machine learning [14][15]. In this section, the objective is to verify that a recognition model for a large printed Arabic text lines dataset can be transferred directly to limited handwriting Arabic text lines database, resulting in a good performance.

This would proof the robustness of the generated model and allow extending them to different data, without needing a large training set. We supposed that architecture that performed well with large multi-fonts printed Arabic text lines should also be able to attain reasonably good results for small handwriting database.

In the previous subsections, we expose and analyze the results in the training step and the transfer learning step.

1) Training with P-KHATT

The architecture of our system and fundamental parameters have been detailed in section 2. No language models or dictionary were used for the proposed model. The first experiment is performed for training the recognition model. The training is realized through the whole network parts: feature extraction with the CCN, sequence modeling with the BLSTM, and decoding with the CTC. Then, the learned weights will be used for the second experiment which involves directly testing the transfer learning on handwriting databases. Four different fonts from the P-KHATT database are mixed to obtain 25888 text line examples for training and 5656 for validation. For each font, there are 6472 different text line images in training set and 1414 in validation set. To evaluate the performance of our system, we used the Character Error Rate and the Word Error Rate (WER).

The proposed recognition model present good results on the printed Arabic database in terms of CER and WER. After only 10 epochs the character error rate is about 2% in both training and testing sets. The WER tends to 9.75% if it is evaluated over the training set and it is over 10% when evaluating the validation set. It can be concluded that there is enough number of samples in training and the recognition model is validated for this database. The convergence of WER and CER for both training and validation sets are depicted in Fig.5. Table.2 presents the best recorded results and the number of used images for each set.

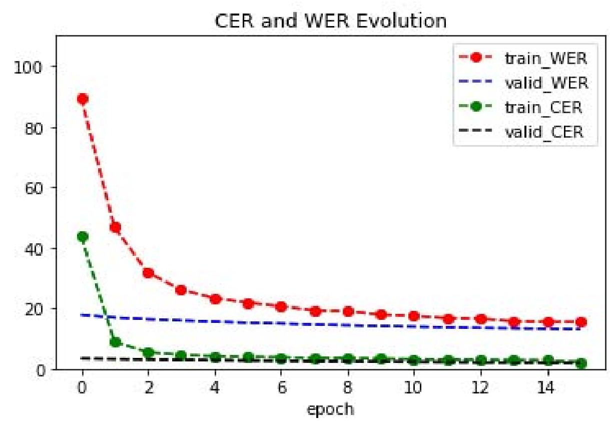

Fig.5. CER and WER evolution in testing and training sets of P-KHATT database

TABLE.2. BEST RECORDED RESULTS WITH P-KHATT

	Number of images	CER%	WER%
Train set	25.888	1.03	9.75
Validation set	5656	1.29	10.18

2) Transfer learning

We kept the same architecture and initialized its parameters to the values trained with the P-KHATT database, and then re-trained the whole network, using text lines from the KHATT database.

In this section, we investigate the evaluation of the performance of TL strategy for handwriting text recognition problem. Firstly, the experiments are performed using the whole KHATT database with 9494 text line images. Secondly, it was done over the set of unique text lines contains only 4823 examples noted as KHATT_unique. Finally, we validate the pre-trained model with AHTID handwriting database that is smaller than KHATT.

We compare the results obtained using TL strategy with those recorded when training from scratch. When training from scratch, the model is initialized with random weights and trained for a number of epochs.

When training the CNN-BLSTM-CTC architecture from scratch for the all KHATT database, the CER tends to 5% if it is evaluated in the training set and it is about 29% evaluating the validation set. The results are better when using the learned weights.

We reduced the training set to unique text line images and repeated the experiments. In this case, the results for the training and validation sets were greatly improved using the TL. We illustrate the convergence along epochs for the two cases in Fig.6. After 25 epochs, the CER tends to 10% and the WER is about 60% using the training set when training from scratch. The WER and CER are improved by about 30% and 8% respectively.

Those results are interesting for Arabic handwriting recognition. We can see that, after a few numbers of epochs on a small set of training data, the WER and CER already surpasses the first experiment we had conducted without TL. We can conclude that TL helps to ameliorate results in term of training time and improve greatly the WER.

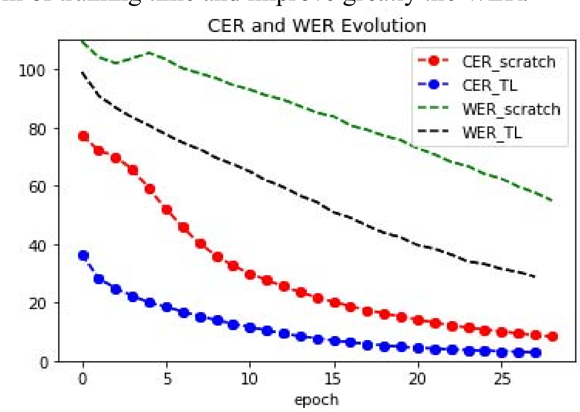

Fig.6. WER and CER evolution with and without TL on the unique set of KHATT database

These results could be considered promising to apply this model for other handwriting database containing small training set.

We transfer the learned model to another handwriting database. AHTID as described previously is small database contains 2699 text line images written by 53 different writers. As depicted in Fig.7, the performance of the recognition system is significantly enhanced by using the TL strategy. Fig.8 present the CER and WER evolutions using TL strategy on the validation set of AHTID database. The CER tends to 2% and the WER is about 15%. It can be concluded that the pre-trained model is sufficient to obtain good recognition accuracy for this database. We are not needed to spot time in training with this small database that doesn't present enough examples of images in training sets.

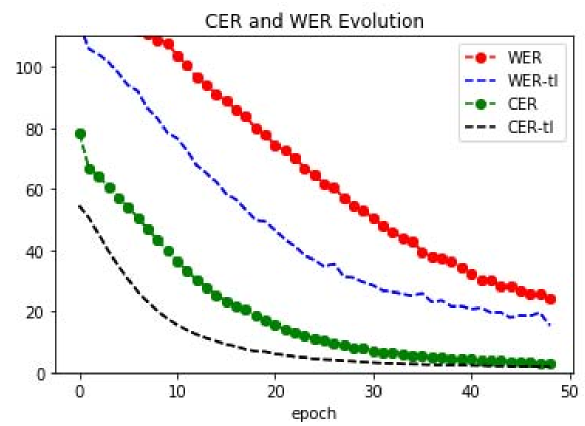

Fig.7. WER and CER evolution with and without using TL strategy on training set of AHTID database

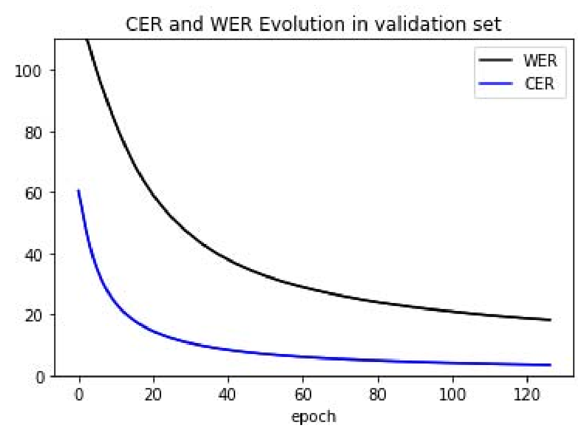

Fig.8. CER and WER evolution using TL on the validation set of AHTID database

The obtained results for the two handwriting databases show the robustness of the pre-trained model and the success of the transfer learning technique from mixed-font printed text to handwriting text recognition.

The experiments prove that handwriting recognition systems rapidly specialize on learning parameter. Results demonstrate the importance of performing TL as the right way to train handwriting text solutions based on deep neural networks that are able to generalize well over small databases. The results of this analysis are included in Table.3.

TABLE.3. OBTAINED RESULTS USING TRANSFER LEARNING FOR HANDWRITING DATABASES

Databases		CER %	WER%
KHATT	Unique set	2.74	17.84
	All	1.64	10.22
AHTID		2.03	15.85

IV. CONCLUSION

The main contribution of this paper is to improve the performance of text recognition model proposed for Arabic script using transfer learning technique. The recognition system is based on CNN-BLSTM-CTC architecture. The experiments exhibited good performance of the TL scheme. Parameters learned on big mixed-fonts printed text lines are transferred to handwriting text recognition problem. Robust results are obtained by transferring learning from the printed P-KHATT database to the handwriting KHATT and AHTID databases.

REFERENCES

[1] T. M. Ghanim, M. I. Khalil, and H. M. Abbas, "*Comparative Study on Deep Convolution Neural Networks DCNN-Based Offline Arabic Handwriting Recognition*," *IEEE Access*, 2020.

[2] Rami Ahmed et al., " *Offline Arabic Handwriting Recognition Using Deep Machine Learning: A Review of Recent Advances".*, *International Conference on Intelligent Systems and Computer Vision (ISCV)*, 2020.

[3] M. Arif Wani , Farooq Ahmad Bhat, Saduf Afzal, Asif Iqbal Khan, "*Advances in Deep Learning", Book, Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland, 2020*

[4] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review,", *Comput. Intell. Neurosci.*, 2018.

[5] S. Sengupta *et al.*, "A review of deep learning with special emphasis on architectures, applications and recent trends," *Knowledge-Based Syst.*, 2020.

[6] Ahmed Lawgali, "A Survey on Arabic Character Recognition,", I*nternational Journal of Signal Processing, Image Processing and Pattern Recognition,*.2015.

[7] I. Yousi and A. Shaout, "Off-Line Handwriting Arabic Text Recognition: A Survey," , I*nternational Journal of Advanced Research in Computer Science and Software Engineering*, 2014.

[8] F. Zhuang ,Z.Qi ,K. Duan,D. xi,Y Zhu, H. Zhu, H.Xiong, and Q. HE, "A Comprehensive Survey on Transfer Learning," , *JPROC,* 2020.

[9] N.Zouhaira, M.Anis and K.Monji, " Contribution on Arabic Handwriting Recognition Using Deep Neural Network", *HIS,* 2019

[10] V.M.Safarzadeh and P.Jafarzadeh " Offline Persian Handwriting Recognition with CNN and RNN-CTC", *25th International Computer Conference, Computer Society of Iran (CSICC),* 2020

[11] S. Khan and S. P. Yong, "A comparison of deep learning and hand crafted features in medical image modality classification," *3rd Int. Conf. Comput. Inf. Sci. ICCOINS,* 2016.

[12] F. Visin, K. Kastner, K. Cho, M. Matteucci, A. Courville, and Y. Bengio, "ReNet: A Recurrent Neural Network Based Alternative to Convolutional Networks,", *CVPR*, 2016.

[13] M. Alshayeb *et al.*, “KHATT: An open Arabic offline handwritten text database,” *Pattern Recognition* , 2013

[14] I. Ahmad, S. A. Mahmoud, and G. A. Fink, “Open-vocabulary recognition of machine-printed Arabic text using hidden Markov models,” *Pattern Recognit.*, 2016.

[15] A. Mezghani, S. Kanoun, M. Khemakhem, and H. El Abed, “A Database for Arabic Handwritten Text Image Recognition and Writer Identification,”, *Int. Conf. Front. Handwrit. Recognition,* ICFHR, 2012.

[16] J. C. Aradillas Jaramillo, J. J. Murillo-Fuentes, and P. M. Olmos, “Boosting handwriting text recognition in small databases with transfer learning,”, *Int. Conf. Front. Handwrit. Recognition, ICFHR*, 2018.

[17] A. Granet, E. Morin, H. Mouchère, S. Quiniou, and C. Viard-Gaudin, “Transfer Learning for a Letter-Ngrams to Word Decoder in the Context of Historical Handwriting Recognition with Scarce Resources,”, *27th Int. Conf. Comput. Linguist*, 2018.