Contribution on Arabic Handwriting recognition using deep Neural Network

Zouhaira Noubigh[1], Anis Mezghani[2], Monji Kherallah[3],

[1]Higher Institute of Computer Science and Communication Technologies, University of Sousse, Tunisia
[2]Higher Institute of Industrial Management, University of Sfax, Tunisia
[3]Faculty of Sciences of Sfax, University of Sfax,Tunisia
{zouhaira.noubigh, anis.mezghani, monji.kherallah}@gmail.com

Abstract. Arabic handwriting recognition is considered among the most important and challenging recognition research subjects due to the cursive nature of writing and the similarities between different characters shapes. In this paper, we investigate the problem of handwritten Arabic recognition. We propose a new architecture combining CNN and BLSTM based on character model approach with CTC decoder. The handwriting Arabic database KHATT is used for experiments. The results demonstrate a net advantage of performance for the CNN-BLSTM combining approach compared to the approaches used in the literature.

Keywords: Deep learning;CNN; LSTM; Arabic database; handwriting recognition

1 Introduction

In recent years, offline Handwriting recognition is considered as a very important research area for several pattern recognition applications. Various handwriting recognition systems were proposed and the difficulty of these systems is dependent on the writing styles of the recognizing units[1]. In fact, recognizing characters or digits is significantly easier than recognizing cursive words or text lines. Therefore, the previous handwriting recognition systems are usually able to recognize single characters with very small vocabularies[2]. Nowadays, the accelerating progress and availability of low-cost computer hardware and the growing difficulty of the tackled problems encouraged the use of computationally expensive techniques. Therefore, recent recognizers were developed to deal with continuous sequence in order to recognize isolated words and text lines extracted from handwritten documents. Recent researches are focused on open vocabulary recognition with less constrained type of document[3].

In the last few decades, Arabic handwriting recognition (AHR) has attracted considerable attention and has become one of the challenging areas of research in the field of document image processing. The cursive nature of the Arabic script, the similarity between many Arabic character shapes and the unlimited variation in human handwriting make the AHR a complicated task and present some specific challenges[1]. Therefore, few efforts were provided in the recognition of Arabic text compared to the recognition of text in other scripts like Latin and Chinese. Most of the recent approaches for Arabic handwritten text/word recognition have used HMM-based techniques or shallow Artificial Neural Networks (ANN)[6]. Recently the Deep Learning (DL), subfield of machine learning [7][8], proved a great performance improvement for a robust classification, recognition, and segmentation. The most famous deep learning techniques are Convolutional Neural Network (CNN) and different variations of Recurrent Neural Network (RNN) like Long Short-Term Memory (LSTM), Bidirectional LSTM , and Multidimensional LSTM [8]. Deep convolutional neural networks has provided an efficient solution for handwritten characters and digits recognition [9][10]. LSTM showed promising performance and has proved as an efficient model with a combination of output CTC layers for sequence labeling over Hidden Markov Models (HMM) and other models[11][12][13].In this paper, a new contribution for Arabic handwriting recognition based on the combination of two famous deep learning techniques cited below, CNN and BLSTM,is presented.

The paper is organized as follows: section II reports related works based on deep learning approach for text recognition. Section III details the proposed CNN-BLSTM based method for Arabic character recognition. Experimental results obtained on KHATT database are presented in section IVwith a comparison study. Finally, Section V presents the conclusion of the paper.

2 Related works

Recent research works investigate combining deep learning technologies to improve recognition results. Shi et al. [14] were the first ones proposed the combination of deep CNN and RNN with CTC decoder for image based Sequence Recognition. Afterwards, many approaches for handwritten text recognition were inspired from this deep architecture. In this section, we present the important works based on this approach proposed for Arabic handwriting recognition. The same architecture based on combining CNN and BLSTM was used by Suryani et al. [15] with a hybrid HMM decoder instead of CTC. CNNs were applied on both isolated characters and text lines processed by a sliding window technique. The proposed approach was tested on offline Chinese handwriting datasets.

Rawls et al. [16]published a CNN-LSTM model where CNN is used for feature extraction, and bidirectional LSTMs for sequence modeling. In this work, authors presented a comparison stage between features provided types. It is proved that CNN model is better than both existing handcrafted features and a simpler neural model consisting entirely of Fully Connected layers. Results are presented on English and Arabic handwritten data, and on English machine printed data.

For Arabic handwriting recognition, AL-Saffar et al. proposed a review that presented Deep Learning Algorithms for Arabic Handwriting Recognition [2]. Authors improve that the first successful systems based DL proposed for Arabic character were based on Convolutional Neural Network (CNN) [17][18] and Deep Belief Networks [19][20].

BenZeghiba proposed a comparative study based on four different optical modeling units for offline Arabic text recognition[21]. These units are the isolated characters, extended isolated characters with the different shapes of Lam-Alef, the character shapes within their contexts and, the recently proposed subcharacter units that allow sharing similar patterns in the different character shapes. All systems, for the four models, use the same architecture of the MDLSTM-RNN based optical model. Experiments are conducted using Maurdor and Khatt databases. Optical models using isolated characters generally perform better, and they have the advantage of being the set with the smallest number of units.

Ahmad et al. [22] proposed an MDLSTM based Arabic character recognition system. Connectionist Temporal Classification (CTC) is used as a final layer to align the predicted labels according to the most probable path. Authors apply preprocessing on text-lines to prune extra white regions, and deskew the text lines for accurate height normalization. KHATT datasets was used for experiments.

Jemni et al. [23]proposed an Arabic handwriting recognition system based on multiple BLSTM-CTC combination. The paper presented a comparative study of different combination levels of BLSTM-CTC recognition systems trained on different feature sets. Three combination levels were compared low-level fusion, Mid-level combination methods and the high-level fusion. The experiments were conducted on the Arabic KHATT dataset.

3 Proposed method

In this section, we describe the architecture of the proposed system. It is a hybrid approach based on combining CNN and BLSTM for Arabic handwriting text lines recognition. It consists of three main steps as presented in Fig.1. The first step is the preprocessing of the input image. The preprocessing stage is necessary in order to reduce the generated noise and eliminate any variability resource that occurred during the images scanning phase, especially with the challenging issues related to text-lines KHATT dataset. In this work, we applied the same preprocessing used in [30],including the discard of any additional white region, the Binarization and the skew detection and correction.

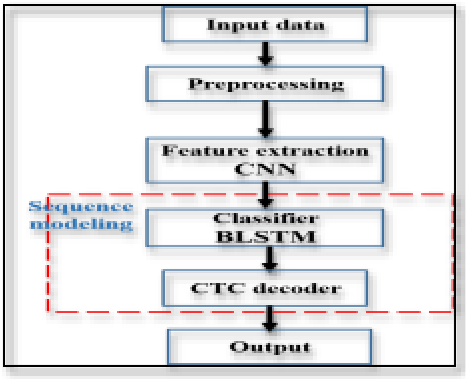

Fig. 1. Proposed approach steps

The two principal steps in the proposed recognition system are the feature extraction with CNN and the sequences modeling based on BLSTM and CTC.

3.1 Feature extraction

In handwriting recognition, the purpose of the feature extraction step is to capture the essential characteristics of the character or the word which make it different from another. Features extraction techniques differ from one application to another dependent on the complexity of studied script and image quality. Therefore, the selection of features extraction method remains the most important step in the recognition process. The Features extraction techniques used for handwritten texts can be classified into two global categories; handcrafted features methods and non-handcrafted or learned features methods. The recent deep learning networks, especially the Convolutional Neural Networks (CNNs) provide efficient solutions for feature extraction where deep layers act as a set of feature extractors. They extract a non-handcrafted features named learning features which are generic and independent of any specific classification task. The convolution operation generate many maps that present different features extracted from the original image. The idea behind this approach is to discover multiple levels of representation so that higher level features can represent the semantics of the data, which in turn can provide greater robustness to intra-class variability.

In this paper, we use the learning features for our handwriting recognition system. 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 consist of 6 convolution layers. The filters are of size 3x3. Second layer is called “Nonlinearity layer”. The Rectified Linear Unit (ReLU) activation function [a verifier] is implemented to produce an output after each convolution. The ReLU objective is to introduce non-linearity in the network since convolution is a linear operation. A max pooling Layer is used to summarize image regions and outputs a downsized version of the previous layer.

The CNN is applied over a sequence of images of size 64x64 obtained from the text line image using a horizontal sliding window scanning the image from right to left. It result a multi-channel output of dimension 1x16x256, where 256 is the number of filter maps in the last convolution layer, and the two other dimensions depend on the amount of pooling in the CNN. The architecture details are illustrated in table 1 and Fig.2.

Table 1. CNN layers configuration and dropout

Type	Configuration
Input	64x64 gray scale image
Conv1	#maps:32 k :3x3
Max pooling	Window :2x2, s :2
Conv2	#maps:64 k:3x3
Max pooling	Window :2x2, s :2
Conv3	#maps:128 k:3x3
Max pooling	Window :2x1, s :2
Conv4	#maps:128 k:3x3
Max pooling	Window :2x1, s :2
Conv5	#maps:256 k:3x3
Max pooling	Window :2x1, s :2
Conv6	#maps:256 k:3x3
Max pooling	Window: 2x1
Output	1x16x256
Dropout	Dropout ratio = 0.7

3.2 Sequence modeling

A few of works for Arabic handwriting recognition are based on BLSTM although this model proves its performance for other scripts. The successful results of deep BLSTM networks in several applications motivating us to use it for Arabic text recognition. The deep BLTSM networks for text recognition is usually combined with the Connectionist Temporal Classification (CTC) function. This loss function is a variant of the Forward Backward algorithm that enables to train LSTM networks by inferring the ground truth at the frame level from the word transcription level. CTC allows the network to predict the sequences of output labels directly without the need to segment the input.

The first and simplest approximation of decoding the RNN output is the best path decoding presented in [6]. It is based on the selection of the most probable character per time-step, the most probable sequence will correspond to the most probable labelling. This approach is not sufficient to satisfy the needs of many sequence tasks although it can already provide useful transcriptions. Other decoding algorithm called beam search is described in the paper of Hwang and Sung [3]. Multiple candidates for the final labeling are iteratively calculated and are called beams. At each time-step, each beam-labeling is extended by all possible characters. Additionally, the original beam is also copied to the next time-step. The Beam Width (W) is defined to give the number of beams to keep (the best beams). The beam width determines the complexity and the accuracy of the algorithm. If W is big enough, the probability will be one and the algorithm will be too complex. But if W is too small, the probability of using beam search to find the correct answer will be too small. So there is a tradeoff between the size of W and the accuracy. In our system, we use the beam search decoding algorithm with BLSTM and W is fixed experimentally to 30.

The proposed system present three BLSTM layers with 512 neurons in each layer and direction. The first layer get their input from the preceding CNN features extraction stage. The RNN 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 added to the RNN called blank. This matrix is fed into the CTC beam search decoding algorithm. The probability of a path is defined as the product of all character probabilities on this path. A single character from a labeling is encoded by one or multiply adjacent occurrences of this character on the path, possibly followed by a sequence of blanks.

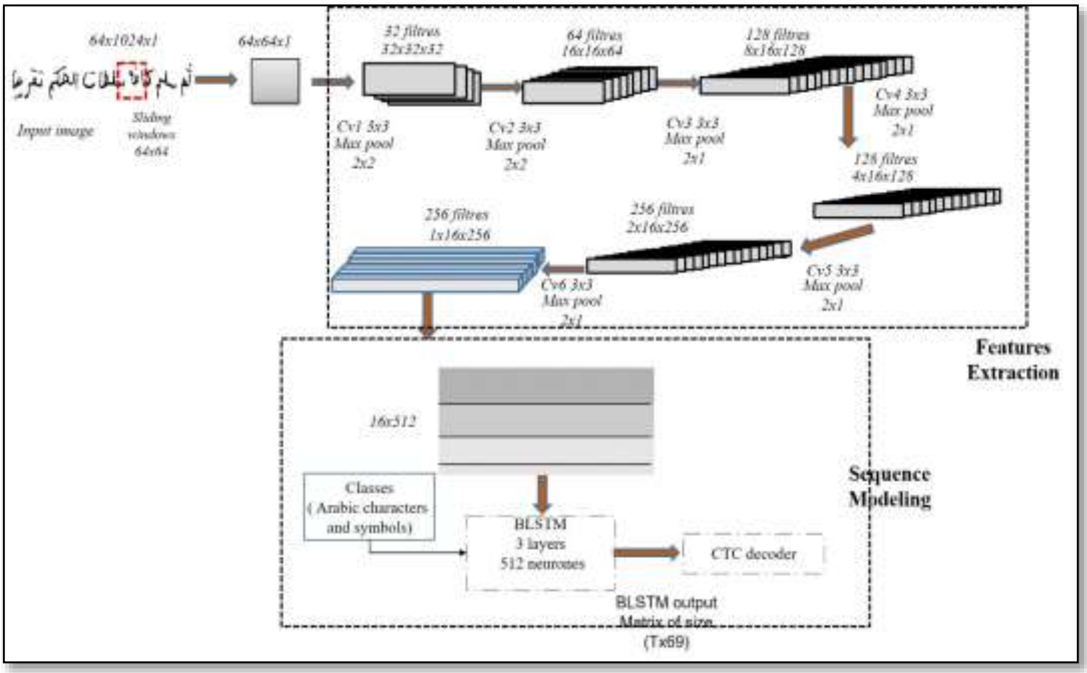

Fig. 2.CNN-BLSTM architecture

4 Experiments and discussion

4.1 KHATT database

In this approach, we used the offline Handwritten Arabic Text database KHATT, which was created by King Fahd University of Petroleum & Minerals, Technical University of Dortmund and Braunschweig University of Technology[24]. The KHATT database contains 4000 grayscale paragraph images and its ground-truth as described in [29]. It consists of scanned Arabic handwriting at different resolutions (200, 300 and 600 dpi) from 1,000 distinct male and female writers representing diverse countries, age groups, handedness and education levels. 2000 of these images contain similar text each covering all Arabic characters and shapes whereas the remaining 2000 images contain free texts written by the writers on any topic of their choice in an unrestricted style.

4.2 System settings

Sliding windows

A horizontal sliding window is used to scan the image from right to left, the direction of written of Arabic texts. The height of the window is equal to the height of the text-line image, which has been normalized to 64 pixels. İn the last convolutional layer, 16 feature vectors are extracted from the feature maps for each window of size 64×64. Those features are the inputs of the BLSTM layer.

Dropout

Dropout is a regularization approach used in neural networks. It prevents over fitting and helps reducing interdependent learning between the neurons. The term “dropout” refers to dropping out (shut down) units in a neural network. A unit is dropping out mean that it is temporarily removed from the network and the choice of this unit is randomaly. All the incoming and outgoing connections from this unit are ignored. Applying dropout to a neural network amounts to sampling a “thinned” network from it. The thinned network consists of all the units that not dropped[25]. Dropout improves the performance of neural networks has been reported to have achieved success and on supervised learning tasks on several benchmark databases.

For our recognition system, a first dropout layer is applied after the CNN layers with dropout ratio 0.5 and a second layer applied to the RNN cells with dropout ratio 0.8.

4.3 Results and discussion

In these experiments, the inputs were preprocessing text line images extracted from the KHATT database.These images were passed through the five CNN layers followed by three BLSTM layers.The experiments are carried out on full-text-lines images of KHATT dataset. The training set has 8505 text-lines, test set has 1867 text-lines, and validation set contains 1584 text-line images. In this work, we report the Word Error Rate (WER) and Character Error Rate(CER) as presented in Table2. Furthermore, Fig.4 shows the plot of errors corresponding to each epoch during training and validation.

Table 2: Performance regarding CER and WER

Heading level	CER	WER
Training set	8 %	20.1%
Test set	15..8 %	30.6%
Validation set	9.6 %	25.5%

Fig. 4. Errors corresponding to each epoch during training and testing on KHATT database

Table3 present a comparison study between our recognition system and the best results reported so far on KHATT dataset. In this study, we introduce for each system the used part of KHATT dataset, the features extraction techniques, the used vocabulary, the language models and the results.It appears clear from the results of those approaches that using language model improves the system performance.Our contribution for Arabic handwriting text line recognition, is the first approach based CNN-BLSTM considering the KHATT dataset as a train and test case.

The proposed approach, in this paper, is based on CNN-BLSTM architechture with CTC beam search decoder and use only KHATT training set for trainainig. The results demonstrate a net advantage of performance for the CNN-BLSTM combining approach. Furthermore, the use of CNN-BLSTM instead of MDLSTM, that is the most used for Arabic handwriting recognition, is less expensive in memory and computing time and prove more performance.

Table 3: Comparison study

Paper reference	Model	Database	Features	Vocabulary	Language model	Result	
						Character error rate	Word error rate
BenZeghiba et al. [26]	MDLSTM + CTC	Unique-text-lines of KHATT dataset 4, 428 for train, 959 for test 876 for validation	Raw pixels	About 20K - 30K word (A hybrid vocabulary that incorporates both the most frequent words and the resulted PAWs)	LM based word 3_grams	--	37,8 %
					LM based Part-Of-Arabic-Word 3_grams	--	30,9 %
					LM based hybrid word/PAW 3_grams	--	31,3 %
BenZeghiba [21]	MDLSTM + CTC	Full KHATT dataset	Raw pixels	About 20K - 30K word (A hybrid vocabulary that incorporates both the most frequent words and the resulted PAWs)	LM based hybrid word/PAW 3_grams	--	24,1%
Jemni et al. [23]	BLSTM + CTC	Khatt database train: 9475 validation: 1901 Test : 2007	Segment Based Feature extraction + Distribution-Concavity (DC) based features	23K distinct words (extracted from the KHATT corpus)	3_grams LM	7,85%	13.52%
				18k words running in the training corpus.	3_grams LM	16,27%	29,13%
Our approach	CNN + BLSTM + CTC	Full KHATT dataset	CNN for features extraction	23K distinct words (extracted from the KHATT Corpus)	No	8%	20.1%

5 Conclusion

A new contribution for Arabic Handwriting recognition is submitted in this paper. The proposed architecture is a deep CNN-BLSTM combination based on character model approach. KHATT dataset, which is one of the challenging datasets that contains Arabic handwritten text lines, is used in experiments for training and testing. The obtained results are very promising and encouraging. In fact, other preprocessing are possible to improve the quality of input images. Furthermore, it will be interesting to improve the proposed approach performance with LMs and to test it for a large vocabulary Arabic corpus.

References

[1] Y. M. Alginahi, "A Survey on Arabic Character Segmentation", International Journal on Document Analysis and Recognition, (2002), pp. 1-22.

[2] A. Al-saffar, S. Awang, W. Al-saiagh, S. Tiun and 3A. S. Al-khaleefa, "Deep Learning Algorithms for Arabic Handwriting Recognition ", International Journal of Engineering & Technology, 7 (3.20) (2018) 344-353.

[3] C. Wigington, S. Stewart, B. Davis, B. Barrett, B. Price, and S. Cohen, "Data Augmentation for Recognition of Handwritten Words and Lines using a CNN-LSTM Network", ICDAR,, 2017.

[4] M. Cai *et al.*, "An Open Vocabulary OCR System with Hybrid Word-Subword Language Models," , ICDAR, 2017.

[6] M. T. Parvez and S. A. Mahmoud, "Offline arabic handwritten text recognition," *ACM Comput. Surv.*, vol. 45, no. 2, pp. 1–35, 2013.

[7] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[8] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] N. Ly, "Deep Convolutional Recurrent Network for Segmentation-free Offline Handwritten Japanese Text Recognition", ICDAR, 2017.

[10] M. A. Mudhsh and R. Almodfer, "Arabic handwritten alphanumeric character recognition using very deep neural network," *Inf.*, vol. 8, no. 3, 2017.

[11] R. Messina and J. Louradour, "Segmentation-free Handwritten Chinese Text Recognition with LSTM-RNN", ICDAR, pp. 171–175, 2015.

[12] E. Sabir, M. Del Rey, S. Rawls, M. Del Rey, and M. Del Rey, "Implicit Language Model in LSTM for OCR", ICDAR, 2017.

[13] Y. Wu, F. Yin, Z. Chen, and C. Liu, "Handwritten Chinese Text Recognition Using Separable Multi-Dimensional Recurrent Neural Network", ICDAR, 2017.

[14] B. Shi, X. Bai, and C. Yao, "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298–2304, 2017.

[15] D. Suryani, P. Doetsch, and H. Ney, "On the benefits of convolutional neural network combinations in offline handwriting recognition," *Proc. Int. Conf. Front. Handwrit. Recognition, ICFHR*, pp. 193–198, 2017.

[16] S. Rawls, H. Cao, S. Kumar, and P. Natarajan, "Combining Convolutional Neural Networks and LSTMs for Segmentation-Free OCR," *2017 14th IAPR Int. Conf. Doc. Anal. Recognit.*, pp. 155–160, 2017.

[17] M. Elleuch, R. Mokni, and M. Kherallah, "Offline Arabic Handwritten Recognition System with Dropout applied in Deep Networks based-SVMs", IEEE, 2016.

[18] M. Amrouch and M. Rabi, "Deep Neural Networks Features for Arabic Handwriting Recognition," *Int. Conf. Adv. Inf. Technol. Serv. Syst.*, pp. 138–149, 2017.

[19] U. Porwal, Yingbo Zhou, and V. Govindaraju, "Handwritten Arabic text recognition using Deep Belief Networks," *21st Int. Conf. Pattern Recognit.*, no. November, pp. 302–305, 2012.

[20] J. H. Alkhateeb, "DBN – Based learning for Arabic Handwritten Digit Recognition Using DCT Features", 6th International Conference on CSIT, pp. 222–226, 2014.

[21] M. F. Benzeghiba, "A Comparative Study On Optical Modeling Units For Off-line Arabic Text Recognition", ICDAR, 2017.

[22] R. Ahmad, S. Naz, M. Z. Afzal, S. F. Rashid, M. Liwicki, and A. Dengel, "The Impact of Visual Similarities of Arabic-like Scripts Regarding Learning in an OCR System", ICDAR, 2017.

[23] S. K. Jemni, Y. Kessentini, S. Kanoun, and J. M. Ogier, "Offline Arabic handwriting recognition using blstms combination," *Proc. - 13th IAPR Int. Work. Doc. Anal. Syst. DAS 2018*, pp. 31–36, 2018.

[24] M. Alshayeb *et al.*, "KHATT: An open Arabic offline handwritten text database," *Pattern Recognit.*, vol. 47, no. 3, pp. 1096–1112, 2013.

[25] A. Cicuttin *et al.*, "A programmable System-on-Chip based digital pulse processing for high resolution X-ray spectroscopy," *2016 Int. Conf. Adv. Electr. Electron. Syst. Eng. ICAEES 2016*, vol. 15, pp. 520–525, 2017.

[26] M. F. Benzeghiba, J. Louradour, and C. Kermorvant, "Hybrid word/Part-of-Arabic-Word Language Models for Arabic text document recognition," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2015, vol. 2015-Novem, pp. 671–675.

[27] R. Ahmad, S. Naz, M. Z. Afzal, S. F. Rashid, M. Liwicki, and A. Dengel, "KHATT : a Deep Learning Benchmark on Arabic Script", ICDAR, 2017.