# Recognition of Handwritten Arabic Words with Dropout Applied in MDLSTM

Rania Maalej[1(✉)], Najiba Tagougui[2,3], and Monji Kherallah[4]

[1] Research Group on Intelligent Machines, National School of Engineers of Sfax, University of Sfax, Sfax, Tunisia
rania.mlj@gmail.com
[2] Higher Institute of Management of Gabes, Gabes, Tunisia
najiba.tagougui@isggb.rnu.tn
[3] Faculty of Computer Science and Information Technology, Al Baha University, Al Bahah, Kingdom of Saudi Arabia
[4] Faculty of Sciences, University of Sfax, Sfax, Tunisia
monji.kherallah@enis.rnu.tn

**Abstract.** Offline handwriting recognition is the ability to decode an intelligible handwritten input from paper documents into digitized format readable by machines. This field remains an on-going research problem especially for Arabic Script due to its cursive appearance, the variety of writers and the diversity of styles. In this paper we focus on the Intelligent Words Recognition system based on MDLSTM, on which a dropout technique is applied during training stage. This technique prevents our system against overfitting and improves the recognition rate. To evaluate our system we use IFN/ENIT database.

**Keywords:** RNN · LSTM · MDLSTM · Dropout · Offline Arabic handwritten recognition

## 1 Introduction

In offline handwriting recognition, researchers have usually extracted a sequence of features from the input image either manually or automatically then they use a classifier such as a Hidden Markov Model (HMM) or a Neural Network (NN) to classify got features. But the trend is toward combining those two sequential and visual aspects, in order to get a generic system robust to distortion, scale and rotation. The most successful system in this Holistic approach is based on Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM) and Connectionist Temporal Classification (CTC). For example, a powerful system [1] based on Multidimensional Long Short Term Memory (MDLSTM) and CTC is successfully applied to recognize offline Arabic Handwriting script. This system is tested by a large database "IFN/ENIT" [2]. But with the huge number of parameters, overfitting can occur. In order to protect the network against this problem, dropout is applied. This technique consists in temporarily removing some units from the network. Those removed units are randomly selected only during the training


A. Campilho and F. Karray (Eds.): ICIAR 2016, LNCS 9730, pp. 746–752, 2016.
DOI: 10.1007/978-3-319-41501-7_83

stage. With such regularization method, we can improve network performance and the error rate is significantly reduced.

This paper is organised as follows. Section 2 presents relevant previous works. Section 3 describes our contribution and in Sect. 4 we report on experiment results. Finally, conclusions and futures work are presented in Sect. 5.

## 2 Related Work

The traditional procedure of recognition is based on six steps Image Acquisition, Preprocessing, Segmentation, Features Extraction, Classification and Post Processing. The Features Extraction stage requires considerable time and expertise because they must be redesigned for each alphabet. To overcome this complex step, a system trained on pixel data is proposed. This kind of system, which is subsumed into Holistic approaches, has the same difficulty degree to recognize several languages. And, the main interest of using those raw images in training is the ability to learn both the visual and the sequential aspect of cursive handwriting at the same time.

Lately, most of the works done are based on HMM [3] or on hybridization of HMM with neural networks [4]. Although HMM's success, it has several drawbacks among which we can mention the poor discrimination and the lack of power to manage the long-term dependencies in sequences because they follow a first-order equation.

The suitable solution adapted by some researchers was the use of Recurrent Neural Networks RNN [5]. In fact, RNNs prove its efficiency for modeling times series. They can be trained discriminatively and they do not require a prior knowledge of data.

On the other hand, RNN suffer from the vanishing gradient and the burden of exploding. Those problems can be solved with a particular node called the Long Short-Term Memory, this node holds better results either in speech recognition [6] or in online handwriting recognition [7]. For the latter field, Bidirectional LSTM was proposed because it provides the possibility to integrate context in both sides of each given letter in the input sequence. This architecture is not the right choice for offline handwriting recognition because the input data is no longer one-dimensional, so we opt for applying the MDLSTM.

The concept of Multidimensional Long Short Term Memory (MDLSTM) [8, 9] is to combine a Multi-Dimensional Recurrent Neural Network (MDRNN) with the LSTM nodes. MDRNN is a recurrent network on which a single recurrent connection is replaced by many connections to represent all spatio-temporal dimensions of input data.

Due to the large number of hidden layers and the huge number of parameters, overfitting can occur on MDLSTM network. This drawback can be corrected by using dropout [9]. This technique was successfully applied with several types of deep neural networks [10–12] and it shows a significant improvement for a recognition rate (see Table 1). It was also successfully used in RNN, specifically in BLSTM and it has shown its effectiveness by reducing label error by more than 8 % on ADAB Dataset for online Arabic handwriting recognition [13].

**Table 1.** Error recognition rate reduced with dropout

| Authors | Network | Dataset | Error rate reduction w/dropout |
|---|---|---|---|
| Srivastava et al. [10] | CNN | CIFAR-10 | 37 % |
| Miao and Metze [11] | DNN | TIMIT | 11,6 % |
| Zhang et al. [12] | DNN | LVCSR | 12,3 % |
| Maalej et al. [13] | BLSTM | ADAB | 8,12 % |

## 3 System Overview

In this section, we present the architecture of our offline Arabic handwriting recognition system based on MDLSTM and CTC and on which the dropout technique is applied during training stage.

MDLSTM is a robust method that allows a flexible modeling of this multidimensional context by providing recurrent connections for every spatio-temporal dimensions existing in the input data. These connections make MDLSTM strong against local distortion in image input (e.g. rotation, shears …). The issue of this approach is how to get one-dimensional label sequences from the two-dimensional images. The proposed solution is to push data through a hierarchy of MDLSTM layers as well as sub-samples windows added after each level to incrementally collapse the two-dimensional images into one-dimensional sequences to be finally labelled by the output layers.

Output layers are based especially on the CTC method [14]. This technique involves a Softmax layer to compute the probability distribution for each step throughout the input sequence.

This distribution covers the 120 Target labels incremented by one extra blank symbol to represent a non-output. So, in total, the size of this Softmax layer achieves 121.

At every timestep the network chooses to emit a label or not. All these decisions define a distribution over alignments between the input and target sequences. Afterwards, and due to forward-backward algorithm, CTC sum over all possible alignments and finally it normalizes the probability of the target sequence given the input sequence [1]. Thus CTC is the best choice for unsegmented cursive handwriting recognition.

The architecture of our proposed network contains 28 layers, three of which are hidden and fully connected. As we can see in Fig. 1, these hidden layers are composed by respectively 2, 10 and 50 LSTM units [15]. This node has the advantage of temporarily saving information.

Concerning the Dropout technique, it is applied only in feed-forward connections after some MDLSTM layers, more precisely, before sub-sample layers which not fully-connected. The selection of those layers is carefully done to not affect the recurrent connections in order to keep the RNN's ability to model long input sequences (see Fig. 1). Dropout layers return the same input except dropped nodes that return null. In our system, 50 % of nodes are randomly dropped.

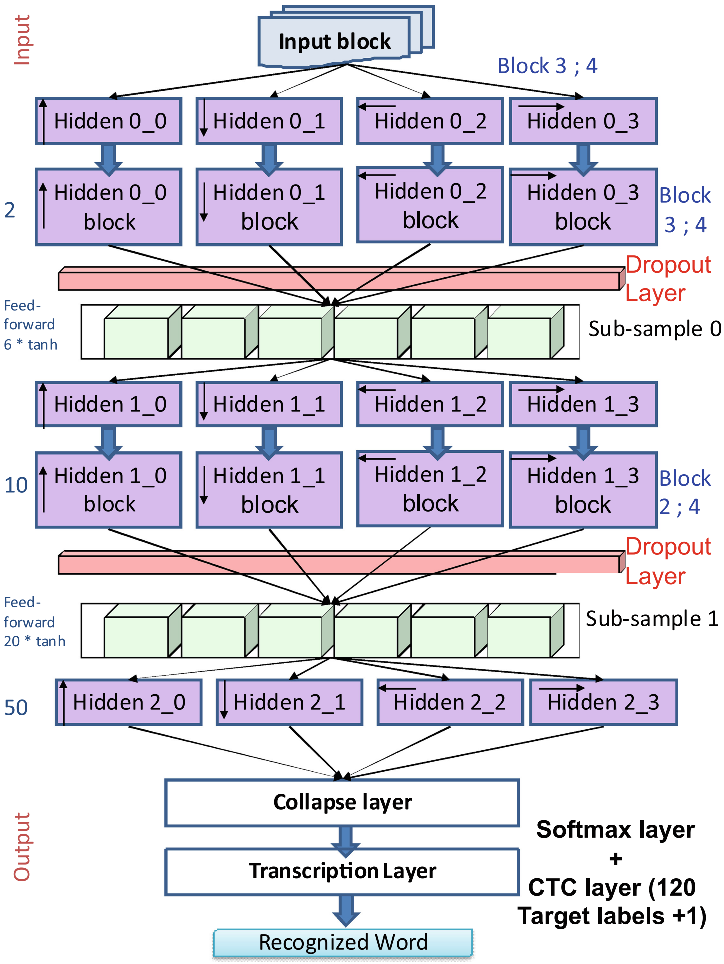


**Fig. 1.** Architecture of proposed recognition system.

## 4 Experiments Results

To validate our system, we use IFN/ENIT [2] Database, with 32492 images of Arabic words written by more than 1000 writers. Those words are 937 Tunisian town/village names. IFN/ENIT Database is divided in 5 sets (see Table 2) and it was successfully used by more than 50 research groups [16–18] as well in Offline Arabic handwriting recognition competition in ICDAR 2009 [19].

**Table 2.** The IFN/ENIT database

| Sets | Words | Characters |
|---|---|---|
| a | 6537 | 51984 |
| b | 6710 | 53862 |
| c | 6477 | 52155 |
| d | 6735 | 54166 |
| e | 6033 | 45169 |
| TOTAL | 32492 | 257336 |

We train our system with 19724 words collected in *seta*, *set b* and *setc*, however we use *set d* and *set e,* that contains 12768 words, for testing. We fix network's parameters as mentioned in Table 3.

**Table 3.** Different parameters for training

| Parameters | Values |
|---|---|
| Max tests no best | 20 |
| Hidden size | 2, 10, 50 |
| Sub-sample size | 6, 20 |
| Hidden blocks | 3, 4; 2, 4 |
| Input blocks | 3, 4 |
| Learn rate | 1e-4 |
| Momentum | 0.9 |
| Dropout layer Size | 6, 20 |
| Dropout percentage | 50 % |

Our network training finished after 224 epochs in total and it takes more than 96 h (see Fig. 2). In 218$^{th}$ epoch we find the best network with least label error.

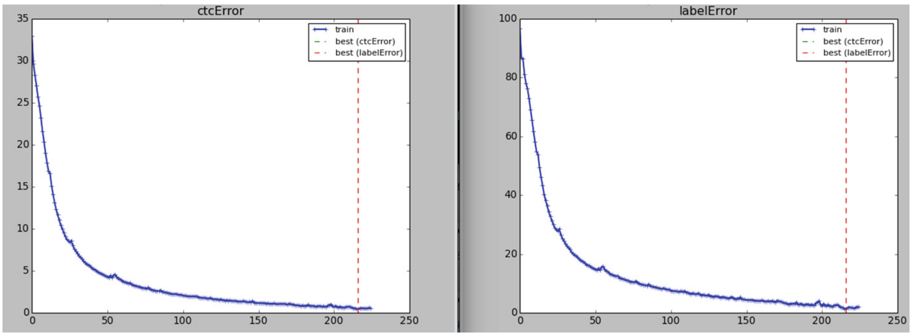


**Fig. 2.** Label Error and CTC Error during 224 epochs for training

Label error is a Levenshtein distance [20], this string metric measures the difference between two words. It is equal to the minimum number of characters that must be deleted, inserted or substituted to change one sequence into the other.

After training, we test our best obtained network with *set d* and *set e*. We get an impressive label error rate which does not exceed 12,09 % compared to 16,97 % obtained with the same architecture without dropout, and 16,30 % obtained with CDBN approach (See Table 4).

**Table 4.** Label error recognition rate comparison using the IFN/ENIT database with no features extraction

| Authors | Approach | (LER) |
|---|---|---|
| Present work | MDLSTM w/CTC w/dropout | 12,09 % |
| Graves [22] | MDLSTM w/CTC | 16,97 % |
| Elleuch et al. [21] | CDBN | 16,30 % |

## 5 Conclusion

In this manuscript, we propose a new offline Arabic handwriting recognition system based on MDLSTM with CTC and on which dropout technique is applied in some hidden layers. This architecture makes our system able to recognize an unsegmented input data in the form of raw pixel. We notice that this MDLSTM's regularization, using dropout, significantly improves the offline Arabic handwriting recognition. Our contribution is to add two dropout layers, on which 50 % of nodes are dropped. This choice has successfully decreased label error rate by more than 4,88 %.

**Acknowledgment.** We would like to express our great appreciation to Mr. Alex Graves for making RNNLIB library as an open source available on internet [22].

## References

1. Graves, A.: Offline Arabic handwriting recognition with multidimensional recurrent neural networks. In: Märgner, V., El Abed, H. (eds.) Guide to OCR for Arabic Scripts, pp. 297–313. Springer, London (2012)
2. Pechwitz, M., Maddouri, S.S., Märgner, V., Ellouze, N., Amiri, H.: IFN/ENIT-database of handwritten Arabic words. In: Proceedings of CIFED, vol. 2, pp. 127–136, October 2002
3. Pechwitz, M., Maergner, V.: HMM based approach for handwritten Arabic word recognition using the IFN/ENIT-database, p. 890. IEEE, August 2003
4. Dreuw, P., Doetsch, P., Plahl, C., Ney, H.: Hierarchical hybrid MLP/HMM or rather MLP features for a discriminatively trained gaussian HMM: a comparison for offline handwriting recognition. In: 2011 18th IEEE International Conference on Image Processing (ICIP), pp. 3541–3544. IEEE, September 2011
5. Graves, A., Liwicki, M., Bunke, H., Schmidhuber, J., Fernández, S.: Unconstrained on-line handwriting recognition with recurrent neural networks. In: Advances in Neural Information Processing Systems, pp. 577–584 (2008)

6. Graves, A., Mohamed, A.R., Hinton, G.: Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6645–6649. IEEE, May 2013
7. Kozielski, M., Doetsch, P., Ney, H.: Improvements in RWTH's system for off-line handwriting recognition. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR), pp. 935–939. IEEE, August 2013
8. Graves, A.: Supervised Sequence Labelling, pp. 5–13. Springer, Berlin (2012)
9. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors (2012). arXiv preprint arXiv:1207.0580
10. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
11. Miao, Y., Metze, F.: Improving low-resource CD-DNN-HMM using dropout and multilingual DNN training (2013)
12. Zhang, S., Bao, Y., Zhou, P., Jiang, H., Dai, L.: Improving deep neural networks for LVCSR using dropout and shrinking structure. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6849–6853. IEEE, May 2014
13. Maalej, R.,Tagougui, N., Kherallah, M.: Online Arabic handwriting recognition with dropout applied in deep recurrent neural networks. In: 2016 12th IAPR International Workshop on Document Analysis Systems (DAS), pp. 418–421. IEEE, April 2016
14. Graves, A., Fernández, S., Gomez, F., Schmidhuber, J.: Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 369–376. ACM, June 2006
15. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
16. Kessentini, Y., Paquet, T., Hamadou, A.B.: Off-line handwritten word recognition using multi-stream hidden Markov models. Pattern Recogn. Lett. **31**(1), 60–70 (2010)
17. AlKhateeb, J.H., Ren, J., Jiang, J., Al-Muhtaseb, H.: Offline handwritten Arabic cursive text recognition using hidden markov models and re-ranking. Pattern Recogn. Lett. **32**(8), 1081–1088 (2011)
18. Pechwitz, M., El Abed, H., Märgner, V.: Handwritten Arabic word recognition using the IFN/ENIT-database. In: Märgner, V., El Abed, H. (eds.) Guide to OCR for Arabic Scripts, pp. 169–213. Springer, London (2012)
19. El Abed, H., Märgner, V.: ICDAR 2009-Arabic handwriting recognition competition. Int. J. Doc. Anal. Recogn. (IJDAR) **14**(1), 3–13 (2011)
20. Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions, and reversals. In: Soviet physics Doklady, vol. 10, no. 8, pp. 707–710, February 1966
21. Elleuch, M., Tagougui, N., Kherallah, M.: Deep learning for feature extraction of Arabic handwritten script. In: Azzopardi, G., Petkov, N. (eds.) CAIP 2015. LNCS, vol. 9257, pp. 371–382. Springer, Heidelberg (2015). doi:10.1007/978-3-319-23117-4_32
22. Graves, A.: Rnnlib: a recurrent neural network library for sequence learning problems (2008). http://sourceforge.net/projects/rnnl/