Toward Classification of Arabic Manuscripts Words Based on the Deep Convolutional Neural Networks

Marouan ELMANSOURI
IA Laboratory, Computer science dept.
Faculty of Sciences, Meknes
My Ismail University, Morocco
elmansourimarouan@gmail.com

Noureddine EL MAKHFI
PMIC Laboratory, Computer science dept.
FST, Al-Hoceima
Abdelmalek Essaadi University, Morocco
nelmakhfi@uae.ac.ma

Badraddine AGHOUTANE
IA Laboratory, Computer science dept.
Faculty of Sciences, Meknes
My Ismail University, Morocco
b.aghoutane@umi.ac.ma

***Abstract*—Deep learning is an area that has seen many developments in recent years. One of these algorithms that have provided good results is Deep Convolutional Neural Networks (DCNN). It is proven to be effective in various fields such as natural language processing, pattern recognition, computer vision, object detection in images, etc. Despite the development of these technologies, Arabic manuscripts in digital libraries still use traditional indexing methods based on metadata, annotation or transcription. In this article, we propose two methods of word classification based on deep learning, the first one uses a simple Neural Network (DNN) and the last one uses a Convolutional Neural Network (DCNN). The idea is to segment words of Arabic manuscripts images and predict the class of each word. The experimental results show the efficient of this classification system based on the DCNN. By comparing the results obtained, we can observe that the DCNN method provides excellent results than those obtained with the DNN method.**

***Index Terms*—Arabic Manuscripts, Classification, Deep Convolutional Neural Network, Digital Libraries, Segmentation, Pattern Recognition**

I. Introduction

Statistically speaking, Arabic is one of the five most widely spoken languages in the world today, where it is used by hundreds of millions of people, It is an official language in more than 16 countries and is widely spoken in a number of other countries, as shown in the recent study in [1]. In addition to this, Arabic is an ancient language of many parts of the world, hence the appearance of a large number of Arabic manuscripts. The condition of some manuscripts is aggravated by storage conditions. The overriding objective of all institutions is to be able to preserve this cultural heritage. Digitization is a solution adopted by several libraries around the world with the aim of limiting direct access to manuscripts. Arabic manuscripts as well as other manuscripts (such as Latin, Chinese, etc.) still have problems of access to content in the form of images. Currently, optical character recognition (OCR) software in images is effective for the recognition of printed text. However, such software does not allow recognition of handwritten text. Several researches have been developed in this field, but still the problem of access to the content of manuscript images persists.

978-1-7281-8041-0/20/$31.00 ©2020 IEEE

The handwritten technologies have been developed for many years in different tasks. Despite this development that has been shown by the researchers, these methods still record an accuracy not enough for practical applications. These records refer to several challenges shown through a multitude of confusing characters and excessive cursiveness in Arabic handwritings. In addition, the historical and modern Handwritten documents always suffer from the variability in writing style, which can not only be manifested from different authors but also from documents of the same writer [2]. Figure 1 shows some examples of Arabic words, written by the same writer, that are very similar and often confusing. Arabic contains a lot of such confusing characters.

Among recent techniques, Deep Convolutional Neural Networks (DCNN) could achieve good results and outperform many techniques in different computer vision tasks like classification, image recognition and detection. thanks to their layers, advances in computing technologies and the availability of large-scale training data, DCNN can model and extract much more complicated features than shallow networks, which lead to adopting DCNN technologies in many problem domains. The experiments have shown that the methods based on deep convolutional neural network have shown a state-of-the-art performance in both word recognition and word spotting. However, in Handwritten Arabic, several methods have been developed for the word recognition, providing not very efficient results. Until today none of these existing methods are based on the deep convolutional neural network techniques. In this context, we have implement two methods based on DCNN, in order to predict the class of each Arabic manuscripts word. We have used the famous HADARA80P dataset to evaluate ours proposed methods.

This document is organized as following: section II, introduces briefly some projects that have been developed for different handwritten kinds. Section III, describes our proposed methods and the HADARA80P dataset. Section IV, presents the experimental results of our method. Finally, the last section V concerns conclusion and perspective of this work.

Fig. 1. Examples of Arabic manuscripts "Source HADARA80P dataset"

II. Related Work

Regarding handwritten documents, there are several challenges that need to be addressed including the defective editions of many documents, low quality of historical manuscript images, the semi-cursive nature, the diversity of writing styles used and complexity of handwritten Arabic writing, etc. Over the years, Handwritten researchers, even that of Arabic handwritten, have developed and proposed several techniques for automatic and semi-automatic transcription [3], searching in manuscripts using metadata and annotation [4], word spotting [5], [6], character and word recognition [7], [8]. In general, we can group these techniques into two categories: traditional and modern techniques. Most traditional methods are based on feature extractors such as Scale Invariant Feature Transform algorithm (SIFT) [9] [10], Speeded Up Robust Features algorithm (SURF) [11], geometric features and HOG-based descriptors. In contrast the modern methods are based on deep convolutional neural networks [7], [8] [12]. In this section, we mention some projects whose treat the different problems identified in some handwritten kinds such as Handwritten Arabic, Hangul, Chinese.

Using the SIFT or SURF algorithm as a word feature extractor in the study [3], both algorithms are based on extracting the interest points. The authors propose a method for semi-automatic transcription of Arabic manuscripts. First, their method is based on the word segmentation in images of manuscripts. Secondly, the search for words is carried out in a database of word images. The results of the image word search are in the form of equivalent text. The objective is to find all occurrences of an image word in a manuscript to facilitate transcription.

inspired by [13], in [7] the authors presented a CNN architecture designed for word spotting called PHOCNET. They could achieve great results in different dataset kinds such as Goerge Washington dataset (GW), IAM Handwritten Database, and IFN/ENIT database. Moreover [8], [12] also have worked on character recognition by using convolutional neural network, the first one for Handwritten Hangul and last one for Handwritten Chinese.

III. Proposed Method

Given the problem of the image content of Arabic manuscripts which is difficult to explore. We present a new method, based on new Deep Learning techniques, allowing more in-depth searches in the images. Our method aims to classify images by words, in order to help search engines to access the content of these images and to facilitate the recognition of their texts afterwards. We use In the first contribution a simple Neural Network (DNN) and in the second one a Convolutional Neural Network (DCNN) . Our proposed method for the word classification in Arabic manuscript images is illustrated by the following sequence of steps.

A. Preprocessing and Word Segmentation

The first stage after the acquisition of images of the manuscripts is the preprocessing phase. Most professional manuscript scanners offer a series of preprocessing steps when scanning images. These scanners offer several algorithms such as: compression of images in different formats, calibration and resolution change, etc. In addition to these preprocessing, we can apply different preprocessing to our images such as: curvature correction, straightening, detail enhancement, contrast enhancement, etc. Our goal is to have better quality images, in order to apply them the word segmentation algorithm and other classification algorithms. The second stage after the preprocessing is the word segmentation in the images of Arabic manuscripts. In this step, we first propose the segmentation of the lines [14] [15] and then the segmentation of the words [16]. The idea is to locate the words in the image to prepare their classification. Indeed, in the line segmentation step, we will convert the color image to a grayscale image in a first step. Our goal is to convert the 2D image into a 1D signal to facilitate line detection in a second step. We use the projection algorithm [17] afterwards. This algorithm is sensitive to noise but is known for its efficiency and extremely short processing time. In this case of line segmentation, we prefer to use grayscale images because the projection of binary images, as in the study [18], is sensitive to line overlap. However, in word segmentation, we apply a dilation based on morphological filtering of the binary image along each segmented line. This filtering based on a horizontal dilation of the image causes the merging of characters and pseudowords. Finally, we reapply the projection algorithm in the vertical direction at the level of each line to locate the words.

B. Neural Networks

A multilayer perceptron "MLP" (or neural network) is a class of feedforward artificial neural networks. A MLP is usually composed of at least three layers of perceptrons (neurons, nodes or units) which are: an input layer, a hidden layer and an output layer (See Figure 2). The output layer can in principle contain several perceptrons. We may apply a different activation function as for the hidden layers depending on the type of problems we have at hand: regression or classification.

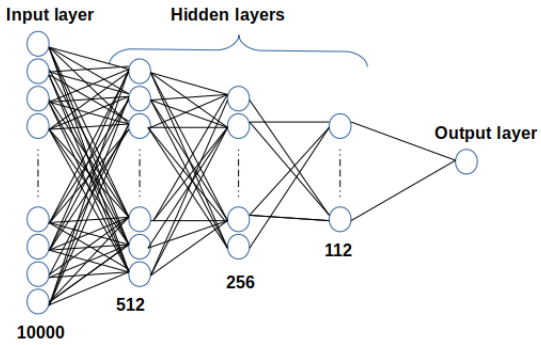

Fig. 2. Our proposed architecture.

C. Convolutional Neural Networks

Convolutional neural networks are a specialized type of multilayer perceptron often used in computer vision tasks like image recognition and object detection. As their name indicates, this type of networks uses the mathematical discrete convolution operation. Given a two-dimensional image I and a two-dimensional kernel K, the result of discrete convolutional operation is defined by:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n)$$

Convolutional neural networks are fundamentally composed of three kinds of layers: a convolutional layer, pooling layer and fully connected layer. See the figure 3 below.

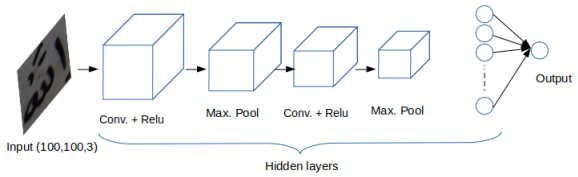

Fig. 3. Hidden cascading layers.

Convolutional layer: It's the heart of a convolutional neural network that allow it to learn linear features. At that stage we apply the convolution operation between the matrix of a portion of its inputs and the matrix of learnable parameters known as kernel. Then, the kernel is shifted by a number s of pixels, s is called the stride. We also add a *zero padding*, which is a margin of size p containing zero values around the image in order to control the size of the output. Depending on the situation, hyper-parameters p and s needs to be chosen in order to perform the convolution operation. If the input of this layer is of size $W \times H \times C$ and the kernel is of size $f \times f$, then the size of the output is governed by this formulas:

$$H_{output} = \frac{H + 2p - f}{s} + 1$$

$$W_{output} = \frac{W + 2p - f}{s} + 1$$

Pooling layer: Also called *subsampling*, replaces the output of the nearby outputs. By doing so, it can help to reduce the spatial size of the representation, and more importantly it helps to make the output less sensitive to small translations of the input. This is very useful when we focus on whether some feature is present in the input than its exact location. Among the popular types of pooling layers we find max pooling, which reports the maximum output within a neighborhood, average pooling and a weighted average based on the distance from a central pixel. Convolutional layers are often followed by a non-linearity to allow network to learn more complex and non-linear features.

Fully connected layer: A convolutional neural network generally ends up with fully connected layers. This type of layers helps to connect every neuron in one layer to every neuron in another layer. The resulting output of the previous layers is flattened into a vector and then we add perceptron layers as in regular fully connected neural network.

D. Dataset

The dataset we used is called HADARA80P. It's published in [19]. This dataset is extracted from an ancient Arabic manuscript "*badalu almaaun fi fadlu altaaun*", which was written by the author EL Hafid Ibn Hajr El Askalani and was published on February 1430 AD, corresponding to the Islamic calendar Jumada al-Awwal 833 AH. The book contains about 250 text pages grouped into five chapters.

The HADARA80P dataset consists of 80 images of a fully handwritten book. Each image corresponds to one page of the book as well as the lining cover. For this work, we could obtain 400 word samples from [20]. This database includes 200 samples of each word, which are "*Allah*" and "*Altaaun*".

IV. Results and Discussion

In this step, we will show the results provided by the DNN and DCNN classifiers without using a descriptor. Experimental Models are implemented in Python 3.6 language. We used a microcomputer with the following characteristics: an Intel i5-2520M CPU microprocessor, a 2.50GHz $\times$ 4 frequency, a 3MB cache, and a 4GB RAM memory. We used a database of 400 images. We dedicated 3/4 images as training samples. The rest of the images are used as test samples. The images are resized to a scale of 100x100 pixels. For the first proposed method, we defined the images as a first step with gray scale pixel values then we flatten them to vectors, which gives us vectors of size 10000. For the second method, we provide the images as is it.

A. Results

In the first step, we tried to use a simple neural network containing the following layers:

- An input layer presented by a vector of size 10000;
- Three hidden layers of size 512, 256 and 128 respectively. We apply at each neuron a Relu activation function.
- An output layer contains one unit (perceptron), since our objective is to predict the class of each word, either "*Allah*" or "*Altaaun*". For the prediction, we apply a sigmoid activation function at the output.

Regarding the learning phase, we used a total of 300 images, 150 for each category, including 210 samples for training and 90 for validation. We obtained good results, the figure 4 shows the evolution of the precision of our model

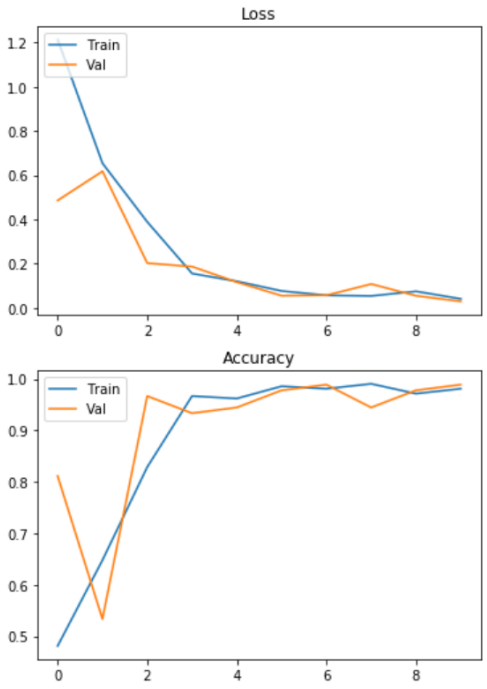

Fig. 4. Accuracy and loss of our neural network

Our second work consist of using a DCNN to predict the class of each word. Our model contains 7 layers : input layer of size $(100, 100, 3)$, 5 hidden layers; two convolutional layers, each one followed by a Relu activation function and a pooling layer "Max-Pool"; and a Fully connected layer, and output layer containing only one perceptron (See Figure 3).

The figure 5 shows the evolution of our model.

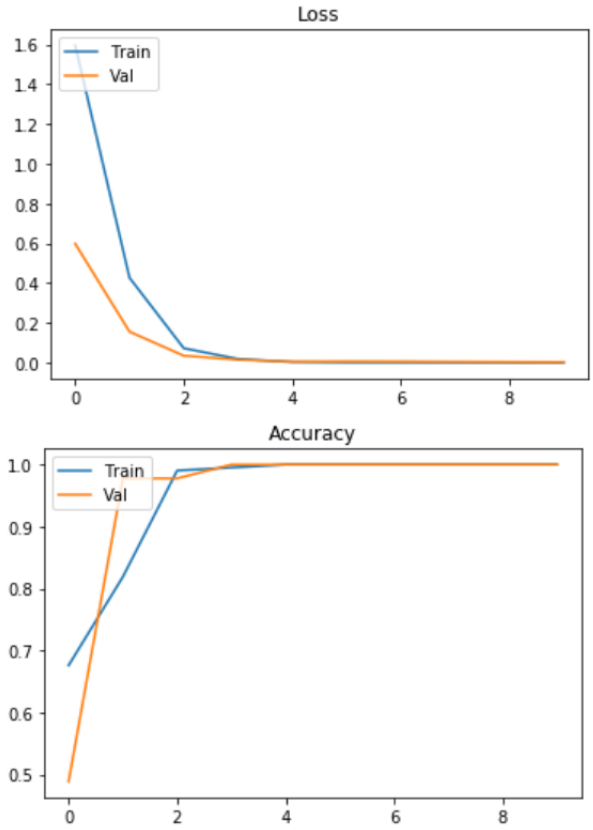

Fig. 5. Accuracy and loss of our deep convolutional neural network using 10 epochs

B. Discussion

In this work, we set out to build two systems of classification based on DNN and DCNN for Arabic words handwritten using the HADARA80P database. The results we have achieved above are based on the classification of the data by the DNN and DCNN algorithms. These data are from the HADARA80P dataset. Regarding the training phase, we have chosen the words: "*Allah*" and "*Altaaun*". In order to evaluate the model formed, we used a cross-validation method to estimate the classification results. The results observed by our classification model are very excellent. We could achieve an accuracy of around 98.42%. The model obtained with the DCNN is more efficient than that of the DNN. As regards the validation of our method, the evaluation criterion between the other existing methods must be the same and the evaluation protocol must be known in advance. Most methods do not share their evaluation protocols. This makes comparison with

these methods difficult. On the other hand, we were able to perform the comparison with one of the existing methods to validate the performance of our method. The work carried out by [21] and concerning the same objective, the authors used an SVM model for classifying words according to 24 classes. They have validated their model with a precision rate that does not displace 92%.

V. Conclusion and future work

In this paper, we proposed two methods for Arabic manuscript words classification. The first one based on deep neural network and the last one on deep convolutional neural network. The aims is to predict the class of each word either "*Allah*" or "*Altaaun*" which are obtained from the HADARA80P dataset. The results obtained show empirically that our proposed methods are able to achieve state-of-the-art performance of 98.42% on the HADARA80P datasets.

In future work, we will suggest a novel approach based on deep convolutional neural networks to predict the class of a huge number of words and to face several challenges in the field of Regarding manuscripts documents, specifically historical Arabic manuscripts.

References

[1] A. Ashiquzzaman and A. K. Tushar,"*Handwritten Arabic Numeral Recognition using Deep Learning Neural Networks*" ,International Conference on Imaging, Vision and Pattern Recognition (icIVPR),Dhaka 3–6, (2017). DOI: 10.1109/ICIVPR.2017.7890866

[2] A. P. Giotis, G. Sfikas, B. Gatos, and C. Nikou, "*A survey of document image word spotting techniques Angelos*" Pattern Recognition, v. 68, 310–332, (2017).

[3] N. El Makhfi and R. Benslimane, "*Feature extraction in segmented words for semi-automatic transcription of handwritten Arabic documents*", J. of Theoretical and Applied Information Technology, v. 70, n. 1, 68–75, (2014).

[4] O. El Bannay, R. Benslimane, N. El Makhfi, B. Aghoutane, and N. Rais, " *Searching in Arab Manuscripts Using Metadata and Annotation* ", European J. of Scientific Research, ISSN 1450-216X, v. 28, n. 1, 155-164 , (2009).

[5] M. Kassis and J. El-Sana, " *Automatic Synthesis of Historical Arabic Text for Word-Spotting* " Proc. - 12th IAPR Int. Work. Doc. Anal. Syst. DAS 2016, 239–244, (2016).

[6] T. Faisal and S. AlMaadeed, "*Enabling indexing and retrieval of historical Arabic manuscripts through template matching based word spotting*", In 1st International Workshop on Arabic Script Analysis and Recognition (ASAR),IEEE, Nancy, France, 3-5 April (2017). DOI: 10.1109/ASAR.2017.8067760

[7] S. Sudholt, G. A. Fink, and A. W. Spotting, "*PHOCNet: A Deep Convolutional Neural Network for Word Spotting in Handwritten Documents*" In 15th International Conference on Frontiers in Handwriting Recognition (ICFHR),IEEE, Shenzhen, China, 23-26 Oct. (2016). DOI: 10.1109/ICFHR.2016.0060

[8] In-Jung Kim and Xiaohui Xie, "*Handwritten Hangul recognition using deep convolutional neural networks*", International J. on Document Analysis and Recognition (IJDAR), v. 18, 1–13, (2015). DOI: 10.1007/s10032-014-0229-4

[9] David. G. Lowe, "*Distinctive Image Features from Scale-Invariant Keypoints*", v.60, n. 2, 91–110, (2004). [Online]. Available: https://link.springer.com/article/10.1023/B:VISI.0000029664.99615.94.

[10] David G. Lowe, "*Object Recognition from Local Scale-Invariant Features*", Proc. of the International Conference on Computer Vision, Kerkyra, Greece, (1999). DOI: 10.1109/ICCV.1999.790410.

[11] Bay H., T.Tuytelaars, L.V. Gool, "*SURF : Speeded Up Robust Features*" , in 9th European Conference on Computer Vision, Graz Austria, 404-417, (2006). DOI: 10.1016/j.cviu.2007.09.014.

[12] Weixin Yang, Lianwen Jin, Zecheng Xie, Ziyong Feng "*Improved Deep Convolutional Neural Networl(For Online Handwritten Chinese Character Recognition using Domain-Specific Knowledge*", In 13th International Conference on Document Analysis and Recognition, Tunis, Tunisia, 23-26 Aug. (2015). DOI: 10.1109/ICDAR.2015.7333822

[13] Arjun Sharma ; Pramod Sankar K. "*Adapting off-the-shelf CNNs for Word Spotting and Recognition*" In 13th International Conference on Document Analysis and Recognition (ICDAR),IEEE, 23-26 Aug. (2015).

[14] Noureddine EL MAKHFI, "*Handwritten Text Segmentation Approach in Historical Arabic Documents*", Springer Nature Singapore Pre Ltd. Embedded Systems and Artificial Intelligence, Advances in Intelligent Systems and Computing, v. 1076, 645-654. (2020).

[15] Noureddine EL MAKHFI, Badraddine AGHOUTANE, Omar EL BANNAY and El bachir Tazi, "*Text Line and Pseudo-Word Segmentation of Arabic Manuscripts Based on Scale-Space Method*", International Symposium on Computer Sciences and Applications, Errachidia, Morocco, (2017).

[16] N. El Makhfi and O. El Bannay. "*Scale-space approach for character segmentation in scanned images of Arabic documents*". J. of Theoretical and Applied Information Technology,v. 94, n. 2, 444-450 , (2016).

[17] R. Manmatha, N. Srimal, "*Scale-space technique for word segmentation in handwritten manuscripts*" In 2nd International Conference on Scale Space Theory in Computer Vision, v. 1682, n. 2, 22-33, Corfu, GRECE, (1999).

[18] J. Ha, R. M. Haralick and I. T. Phillips, "*Document page decomposition by the bounding-box projection technique*" The Third International Conference on Document Analysis and Recognition. 1119-1122, Montreal, Quebec, Canada, (1995).

[19] W. Pantke, M. Dennhardt, D. Fecker, V. Märgner, and T. Fingscheidt, " *An Historical Handwritten Arabic Dataset for Segmentation-Free Spotting - HADARA80P*". 14th International Conference on Frontiers in Handwriting Recognition, Heraklion, 15-20, (2014). DOI : 10.1109/ICFHR.2014.11

[20] HADARA80P Dataset,Technische Universität Braunschweig, [Online]. Available: https://www.ifn.ing.tu-bs.de/research/data/HADARA80P (Date accessed: April 30, 2020).

[21] Alalshekmubarak, A., Hussain, A., and Wang, Q.-F. "*Off-Line Handwritten Arabic Word Recognition Using SVMs with Normalized Poly Kernel*". Lecture Notes in Computer Science, 85–91, (2012).