EasyChair Preprint

№ 3247

Handwritten Arabic Classification Using Deep Convolutional Neural Networks

Marouan Elmansouri, Noureddine El Makhfi and Badraddine Aghoutane

April 23, 2020

Handwritten Arabic Classification Using Deep Convolutional Neural Networks

Abstract

Handwritten Arabic, like other handwritten (such as Latin, Chinese, etc.), have received increasing attention from several researchers. To preserve and promote wider access to the invaluable cultural and literary heritage held in both public and private collections of manuscripts, the researchers have proposed and developed several approaches based on annotation, metadata, and transcription. The need to access to the manuscript text is increasing on a large scale. For this reason, traditional methods of indexing such as annotation or transcription will be outdated as they require a considerable and unreliable manual effort. It is, therefore, necessary to develop new tools for the identification and recognition of handwritten text contained in images. However, despite the development that has been shown by Convolutional Neural Network (CNN) in different computer vision tasks, the latter has not known many uses in the field of Arabic manuscripts. Even if, the use of these methods based on deep learning to predict the class of characters, such as the Handwritten numbers, has achieved a great result. Hence, the idea of using methods based on deep learning techniques to classify words and characters in images of Arabic manuscripts.

In this paper, we propose two classification methods to predict the class of each word, using the HADARA80P dataset. The first one uses a simple neural network and the last one uses a convolutional neural network. The experimental results obtained by these two methods are very interesting

Keywords

Classification, Computer vision, Deep convolutional neural network, HADARA80P Dataset, Handwritten Arabic, Image processing, Neural network.

0.1 Introduction

Statistically speaking, Arabic is one of the top five most widely spoken language in the present world, where it is used by more than 267 million people, it is an official language in more than 16 countries and spoken widely in a number of other countries as well [1]. In addition to that, Arabic is the liturgical language of the religion of Islam, hence a large number of religious texts and writings are in Arabic. Numerous languages including Persian, Hindi, and Urdu have taken inspiration from Arabic.

Arabic handwritten technologies have been developed for many years in different tasks. Despite this development that has been shown by the researchers, these methods still not accurate enough for practical applications. The difficulty of handwritten is mainly caused by a multitude of confusing characters and excessive cursiveness in Arabic handwritings. In addition, Handwritten documents, either historical or modern, always suffer from variability in writing style, not only among different authors but also for documents of the same writer [2]. Figure 1 shows some examples of Arabic words that are very similar and are often confusing. Arabic contains a lot of such confusing characters.

In recent years, Deep Convolutional Neural Networks (DCNN) have achieved a good results in different computer vision tasks like classification, image recognition and detection. Composed of many layers, DNNs can model much more complicated functions than shallow networks. The availability of large-scale training data and advances in computing technologies have made the training of such deep networks possible, leading to a widespread adoption of DNNs in many problem domains. For example, deep convolutional neural networks (DCNNs) have shown outstanding performances in many image recognition field. 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, none of existing methods for word recognition are based on deep convolutional neural network. In this context, we built two methods that predict the class of each word based on deep convolutional neural networks. In this research, we review some projects that have been worked on different handwritten kinds, we use the famous HADARA80P dataset to evaluate ours proposed methods.

The rest of this document is organized as following: in section 0.2, we talk briefly on some projects have been done in different handwritten kinds, and describe the convolutional neural network and HADARA80P. In section 0.3, we present the experimental results of our work. finally, conclusions are provided in section 0.4.

Figure 1: Examples of Arabic manuscripts "Source ' HADARA80P dataset ""

0.2 Related Work

0.2.1 Methods

Regarding handwritten problems, there are several challenges to solve these problems, since many documents are in defective editions, and since the low quality of old manuscripts, the complexity of Arabic script and the different writing styles. 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 SIFT (Scale Invariant Feature Transform), geometric features and HOG-based descriptors. In contrast the modern methods are based on deep convolutional neural networks such as [7],[8][9]. In this section, we mention some projects that have worked on the different problems in different handwritten kinds such as Handwritten Arabic, Hangul, Chinese.

Using SIFT (Scale Invariant Feature Transform) algorithm as a feature extractor for interest words points [3], the authors could propose method for semi-automatic transcription based on two steps: the first one is to segment the digitized manuscripts on words, while the second searches in a database of image words and their equivalent in text mode the corresponding text words of each image word if it exists.

inspired by [10], 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], [9] also have worked on character recognition by using convolutional neural network, the first one for Handwritten Hangul and last one for Handwritten Chinese.

0.2.2 Convolutional neural networks

A- Neural Networks

A multilayer perceptron "MLP" (or neural network) is a class of feedforward artificial neural networks. A MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer (See Figure 2). The output layer can in principle contain several perceptrons (neurons or units). 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.

B- 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:

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

A convolutional neural network is generally composed of three kinds of layers, a convolutional layer, pooling layer, and fully connected layer. See figure 4.

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 in reducing 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: After several convolution and pooling layers, convolutional neural networks generally ends up with fully connected layers. 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.

0.2.3 Arabic Handwritten

In this work, we based on HADARA80P dataset published in [11]. This dataset is based on the historical handwritten book ""*badalu almaaun fi fadlu altaaun*"", which was written by the author, EL Hafid Ibn Hajr El Askalani, and published in 06.833 AH (Islamic calendar) corresponds to Feb. 1430 AD. The book contains about 250 text pages grouped into five chapters.

The HADARA80P datset contains about 80 manuscript images consisting of one inlay cover and the first 79 of the complete book. For this work, we could obtain 300 word samples from [12], this database includes 150 samples of each word which are "Allah" and "Tauun". The images are scaled to the pixel size of 100x100. For the first proposed method, we defined the images as a first step with gray scale pixel values then we flatten them to a vectors, which gives us a vectors of size 10000. For the second method, we provide the images as is it.

0.3 Methods and Experiments

Simple Neural Network

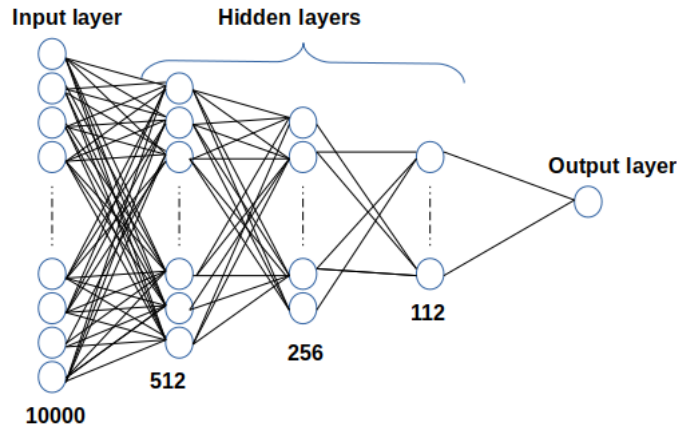

Figure 2: Our architecture.

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 segmoid 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 3 shows the evolution of the precision of our model

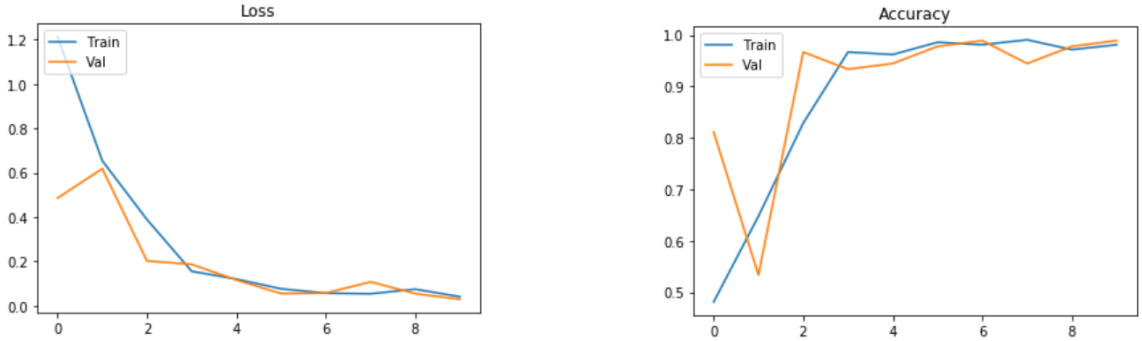

Figure 3: Accuracy and loss of our neural network

Deep Convolutional Neural Network

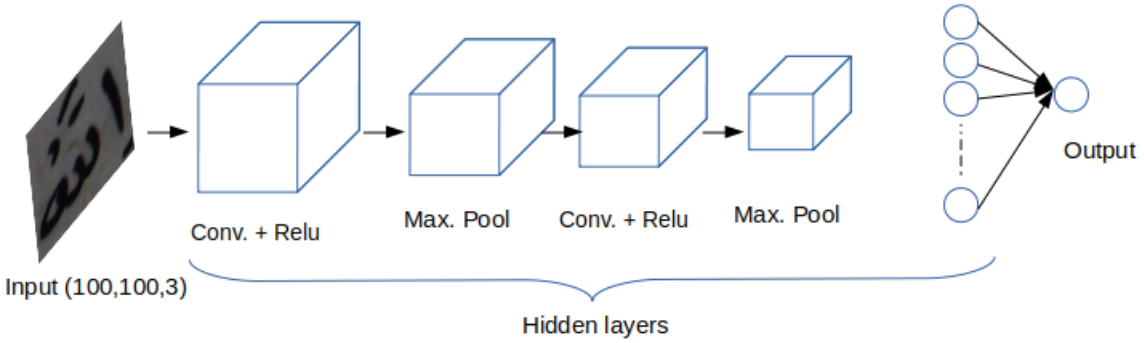

Figure 4: Our architecture.

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 4).

The figure 5 shows the evolution of our model.

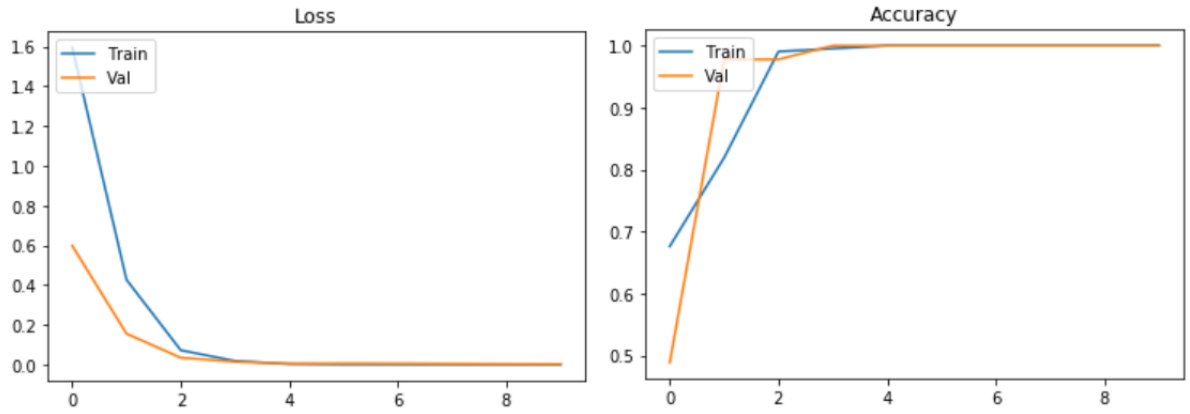

Figure 5: Accuracy and loss of our neural network using 10 epochs

0.4 Conclusion

In this paper, we proposed two methods for word classification. The first one based on deep neural network and the last one on deep convolutional neural network. These proposed methods help to predict the class of each word either Allah or Atauun which are obtained from the HADARA80P dataset. We show empirically that our proposed methods are able to achieve state-of-the-art performance of 100% on the HADARA80P datasets.

Bibliography

[1] A. Ashiquzzaman and A. K. Tushar,"*Handwritten Arabic Numeral Recognition using Deep Learning Neural Networks*" pp. 3–6.

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

[3] N. El Makhfi and R. Benslimane, "*Feature extraction in segmented words for semi-automatic transcription of handwritten arabic documents*", J. Theor. Appl. Inf. Technol., vol. 70, no. 1, pp. 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* " Eur. J. Sci. Res., vol. 28, no. 1, pp. 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, pp. 239–244, 2016.

[6] T. Faisal and S. AlMaadeed, "*Enabling indexing and retrieval of historical Arabic manuscripts through template matching based word spotting*" vol. 1, pp. 57–63, 2017.

[7] S. Sudholt, G. A. Fink, and A. W. Spotting, "*PHOCNet: A Deep Convolutional Neural Network for Word Spotting in Handwritten Documents*" pp. 277–282, 2016.

[8] I. K. X. Xie, "*Handwritten Hangul recognition using deep convolutional neural networks*" pp. 1–13, 2015.

[9] Weixin Yang, Lianwen Jin+, Zecheng Xie, Ziyong Feng"*Improved Deep Convolutional Neural Networl(For Online Handwritten Chinese Character Recognition using Domain-Specific Knowledge*"

[10] J. Almazán, A. Gordo, A. Fornés, and E. Valveny "*Adapting off-the-shelf CNNs for Word Spotting and Recognition*"

[11] W. Pantke, M. Dennhardt, D. Fecker, V. Märgner, and T. Fingscheidt, " *An Historical Handwritten Arabic Dataset for Segmentation-Free Spotting - HADARA80P*". Proc. Int. Conf. Front. Handwrit. Recognition, ICFHR, vol. 2014-Decem, pp. 15–20, 2014.

[12] Technische Universität Braunschweig, HADARA80P Dataset https://www.ifn.ing.tu-bs.de/research/data/HADARA80P