

# Arabic Handwritten Recognition Using Deep Capsule Networks

Adwaa Bukhari

Artificial Intelligence, Midocean University

Assignment 2, Problem Statement and Objective

## 1. PROBLEM STATEMENT

Manual work of documenting and digitalizing the paper work is consuming time and cost. Therefore, handwritten recognition system can be employed to read the extract the document data very efficiently however, the accuracy of the output still a biggest challenge specially in Arabic Language. Conventional Neural Network (CNN) is used commonly which has multiple connected layers to recognize pattern of the images due to the ability by using conventional filters and non-linearity units. According to the recent researches, the most gap and limitation of the model is the cursive nature of Arabic, and the limited dataset, which still might lead to manual work to clarify the output work, therefore, this needs to enhance the model to achieve the minimum percentage of errors.

## 2. PROJECT OBJECTIVE

To improve and analyze the accuracy of the system, we need to develop a more robustness Arabic handwriting recognition. Capsule Networks (CapsNets) have been introduced as an alternative to Convolutional Neural Networks (CNN). CapsNets are group of capsules composed of neurons representing the presence of an entity and the vector indicating the relationship between features by encoding instantiation parameters [7]. Unlike the CNN, CapsNets needs less data to train, and faster training time. This makes the accuracy of the network recognition higher.

## 3. LITRATURE REVIEW

Handwriting recognition has gained a lot of attention in the field of pattern recognition and machine learning. Many researchers have proposed different technologies. In this review section we present previous works on Arabic handwritten recognition methods.

Elsawy et al., n.d. [1] claim that the performance of Arabic recognition is affected by the limited dataset. He uses Convolution Neural Networks (CNN) which has the best performance in images and big data. A CNN consists of a number of layers (convolutional layers, pooling layers, fully connected layers). Dataset collected and made by 60 participants between 19 to 40 years. Each participants wrote each character ten times on different form. This dataset composed approximately 16K. The accuracy of the experimental result is 94.9%. Therefore, Ahmad et al., 2020 [2] presented couple of the methods using Multi-Dimensional Long Short-Term Memory Networks (MDLSTM) and Connectionist Temporal Classification (CTC) for recognition Arabic characters. They investigated the capability of data augmentation methods to improve the performance of deep learning models on the KHATT dataset. They implemented data-augmentation

techniques for increasing data images in pre-processing step (blurring, text-line contouring, edge enhancing, bleeding-through effect). The model achieved 80.02% recognition rate on KHATT dataset.

Alaasam et al., 2017 [3] trained a Convolution Neural Networks (CNN) to recognize historical Arabic handwritten text. To avoid the limitation of historical handwritten text that might have loosely layout format and overlapping components in a historical text. Deep convolutional networks have been successfully applied to problems such as text recognition, image recognition, and character recognition. Convolutional Neural Networks (CNN) consist of convolutional layers and pooling layers followed by fully connected layers. This architecture uses the 2D structure of an image as direct input without the need to hand crafted features for representing image data. They studied two ideas to overcome the large dataset which are: extend a small labeled dataset of sub-words by automatic synthesis of these sub words, then extend these labeled datasets by generating various forms of sub-words using NASKH fonts. Experiments were conducted on three datasets: historical handwritten text images, synthesized text image dataset, and the printed text images dataset which used the same CNN architecture and same optimization algorithm. The best accuracy is 85%, that was obtained for combination of the handwritten text image dataset and synthesized text image dataset.

In this paper Elleuch et al., 2016 [4] proposed a new design based-Support Vector Machine (SVM) of Convolution Neural Network (CNN) classifier architecture with offline Arabic handwritten text recognition which CNN as an automatic feature extractor from raw images and it let SVM do classification by analyzing the error classification rate on the Arabic handwritten character. The target is to mix the CNN respective capacities and the SVM to obtain a new effectual recognition system inspired by the two formalisms and using HACDB database and IFN/ENIT database. The result of the experimental achieve a classification error rate of 6.95%. So, in order to improve this architecture, they added a dropout technique on the CNN based-SVM model to the only fully connected hidden layers, reduces the error to 5.83% with HACDB database.

Amrouch et al., 2018 [5] Deployed system to integrate Convolution Neural Networks (CNN) and Hidden Markov Models HMM classifier, to use HMM model the dynamic of Arabic handwriting and CNN to extract salient features. The normalized input images are provided to the first convolutional layer and the designed CNN is trained by stochastic gradient descent (SGD) with momentum. The HMM baseline is trained by a new features vector obtained from the outputs of the hidden layer (FCL). Once the HMM classifier has been well trained, it performs the recognition task and makes new decisions on testing images with such automatically extracted features. They use the IFN/ENIT database that consists of 946 handwritten Tunisian city names and their corresponding post codes, which achieve a rate of 88.95%.

El-Melegy et al., 2019 [6] proposed high performance of the Convolution Neural Networks (CNN) recognition system in the recognition of Arabic handwritten literal amounts that are used in bank cheques. They use AHDB standard dataset which consists 50 different Arabic literal amounts. They used 80% of the dataset for training and 20% for testing. The model achieved 96.58%, Therefore, they improved the model performance by applying the data augmentation that achieved 97.8% accuracy.

The table [1] summarize the literature reviewed on Arabic handwriting recognition. From the literature review, I observe that CNN is the most widely used method to analyzing and processing images because it can learn highly abstract features and its very efficiently to identify object. However, CNNs require large datasets for training to achieve good results and this is what they all acknowledged especially in Arabic language due to its complicated overlapping alphabet compared to other languages.

| Paper                          | Method                     | Dataset                         | Result                               |
|--------------------------------|----------------------------|---------------------------------|--------------------------------------|
| (Elsawy et al., n.d. 2017) [1] | CNN                        | Collect and made by participant | Accuracy 94.9%                       |
| (Ahmad et al., 2020) [2]       | MDLSTM + CTC               | KHATT                           | Accuracy 80.02%                      |
| (Alaasam et al., 2017) [3]     | CNN                        | HAHPT                           | Accuracy 85%                         |
| (Elleuch et al., 2016) [4]     | CNN based-SVM with Dropout | HACDB + IFN/ENIT                | Error rate 5.83%<br>Error rate 7.05% |
| (Amrouch et al., 2018) [5]     | CNN Based HMM              | IFN/ENIT                        | Accuracy 88.95%                      |
| (El-Melegy et al., 2019) [6]   | CNN                        | AHDB                            | Accuracy 97.8%                       |

#### REFERENCES:

- [1] Elsayy, A., Loey, M., El-Bakry, H. M., El-Sawy, A., & El-Bakry, H. (n.d.). *Arabic Handwritten Characters Recognition using Convolutional Neural Network*.  
<https://www.researchgate.net/publication/313891953>
- [2] Ahmad, R., Naz, S., Afzal, M., Rashid, S., Liwicki, M., & Dengel, A. (2020). A deep learning based Arabic script recognition system: Benchmark on khat. *International Arab Journal of Information Technology*, 17(3), 299–305. <https://doi.org/10.34028/iajit/17/3/3>.
- [3] Alaasam, R., Kurar, B., Kassis, M., & El-Sana, J. (2017). *Experiment study on utilizing convolutional neural networks to recognize historical Arabic handwritten text*.  
<https://beratkurar.github.io/papers/Asar2017Experiment.pdf>.
- [4] Elleuch, M., Maalej, R., & Kherallah, M. (2016). A New design based-SVM of the CNN classifier architecture with dropout for offline Arabic handwritten recognition. *Procedia Computer Science*, 80, 1712–1723. <https://doi.org/10.1016/j.procs.2016.05.512>.
- [5] Amrouch, M., Rabi, M., & Es-Saady, Y. (2018). Convolutional feature learning and CNN based HMM for Arabic handwriting recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10884 LNCS, 265–274.  
[https://doi.org/10.1007/978-3-319-94211-7\\_29](https://doi.org/10.1007/978-3-319-94211-7_29).
- [6] El-Melegy, M., Abdelbaset, A., Abdel-Hakim, A., & El-Sayed, G. (2019). Recognition of Arabic Handwritten Literal Amounts Using Deep Convolutional Neural Networks. *Lecture Notes in*

*Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11868 LNCS, 169–176. [https://doi.org/10.1007/978-3-030-31321-0\\_15](https://doi.org/10.1007/978-3-030-31321-0_15).

- [7] Goceri, E. (2021). *ANALYSIS OF CAPSULE NETWORKS FOR IMAGE CLASSIFICATION*. Biomedical Engineering Department, Engineering Faculty, [https://www.ihci-conf.org/wp-content/uploads/2021/07/04\\_202107L007\\_Goceri.pdf](https://www.ihci-conf.org/wp-content/uploads/2021/07/04_202107L007_Goceri.pdf).