

Arabic Handwritten Digit Recognition Using Convolutional Neural Network

Dewy Yuliana

Informatics Engineering Department

Faculty of Computer Science,
Universitas Sriwijaya
Palembang, Indonesia
09121002056@student.unsri.ac.id

Abstract— Arabic is the most widely used language in the world, especially the Arab League Country. Of course, in those countries often use Arabic numeral in bank and business application, postal zip code and data entry application. This research have focused on handwriting recognition of Arabic numeral that have unlimited variation in human handwriting such as style and shape. The propose methods on the deep learning technique is Convolutional Neural Network. LeNet-5 architect also used in training and recognizing handwritten image of Arabic numeral as much as 70000 images derived from MADbase dataset. The experimental result on 10000 images of database used is by comparing the number of epoch in training process yields, and the average accuracy is 97.67%

Keywords: Handwritten Digit Recognition, Arabic Numeral, Deep Learning, Convolutional Neural Network











1. INTRODUCTION

Optical Character Recognition (OCR) system development is considered one of the most important in the field of pattern recognition research. OCR allows the machine to recognize characters automatically through optical mechanisms. In other words, OCR is an electronic translator for handwritten drawings in computer format [1]. Research in the field of OCR is still very needed for various languages in the world [2]. The main problem on this research focused on handwritten digit recognition that has become one of the OCR branches.

Each country has an officially used language. This official language is usually used as official communications such as legislation, official correspondence and as a means of interaction associated with the implementation of functions of a position and language that must be studied in the country's education. One of the five most widely used languages in the world is Arabic. Evident from reports made by the CIA that there are at least twenty countries that use Arabic as an official language or may be called a national language[3]. While in research conducted by [4] said that there are twenty-four countries that make it a national language. In addition, Arabic is also the liturgical language of Muslims which can be seen from the book of guidance written in Arabic. Many other languages are adapted from Arabic such as Persian, Urdu and Hindi. Nevertheless, OCR research on Arabic both letters and numbers still has not received

sufficient attention [2], [4]. Sample handwritten digit in Arabic language are shown in Table 1.

Table 1. Samples of Arabic Handwritten Digit

Latin Digit	Arabic Digit	Arabic Handwritten Digit
0	٠	
1	١	
2	٢	
3	٣	
4	٤	
5	٥	
6	٦	
7	٧	
8	٨	
9	٩	

Research on handwriting recognition in both letters and numbers has been done in various languages. Arabic is still mostly an unexplored field of study, with the maximum work done previously on this language being based on printed character. In [4] conducted a study entitled Handwritten Arabic Numeral Recognition using a Multi-Layer Perceptron. This study has an accuracy of 94.93% of 3000 handwritten Arabic handwriting data obtained from three hundred participants of different ages and gender. The accuracy is derived from experiments using fifty four hidden networks on Multi-Layer Perceptron. Further research was conducted by [5] in his research by using Gabor Filter as the basic feature and SVM to classify Arabic numerals by using sample data of 21120 samples from 44 participants and produce 97.94% accuracy with scale 4 and orientation 6 on implementing Gabor Filter.

In the study of [6] with the title A novel hybrid CNN-SVM classifier for recognizing handwritten digits yield accuracy of 94.40%. The SVM method is used on the output layer in Convolutional Neural Network. The data used in this research are MNIST data with 60000 data train and 10000 test data. Later in [7] conducted research to recognize Arabic numbers using backpropagation method. In this study conducted Selvi & Meyyappan produce accuracy of 96%.

The next research was conducted by [8] using HMM in classifying the Latin number with accuracy of 97.2% of

10000 test data and train data of 60000 data sourced from MNIST data. In the same year, [9] conducted research by combining SVM, Fuzzy C-Means and Unique Pixel methods to classify Arabic numerals. They used a public dataset of 3510 samples. The accuracy is 88% from 40% sample data as test data. Further research was conducted by [10] using Dynamic Bayesian Network with to classify Arabic numerals as much as 10000 test data yielding 85.26% accuracy.

In this research, the proposed method is Deep Learning, which has recently stirred the world of machine learning which can increase the accuracy in the classification. Deep learning has a lot of architecture like CNN. CNN is a multi-layer feed-forward neural network that extracts properties from input data and is trained with a neural network back-propagation algorithm.

2. THE PROPOSED METHOD

1) Preprocessing

Before the images are trained and tested, firstly the images must be converted from RGB image to binary image. In this pre-processing stage as shown in fig. 1, used OpenCV with python [11].



Figure 2. Preprocessing Result.
(a) Original Image [12]. (b) Binary Image.

2) Recognition

As mention in the previous subsection, in order to recognize Arabic handwritten digit, this research have used convolutional neural network as identifier.

Convolutional Neural Network (CNN) is one method of deep learning commonly used in image data. CNN can be used to detect and recognize objects in an image. Frankly speaking, CNN is not much different from the usual neural network. CNN consists of neurons that have weight, bias and activation function also CNN are trained with backpropagation algorithm. But architecturally, CNN is different from other deep learning architectures because each layer on the CNN architecture has a different purposes. The illustration layer in CNN shown in fig. 2.

CNN has a special architecture of the Artificial Neural Network, which has three additional architectural ideas that comprise local receptive fields, weight sharing, and sub-sampling [13]. Local receptive fields are extracting the basic features of an insert, for example the edge of the image. Then the feature becomes inserted for the next layer and extracted again, so obtained more specific features.

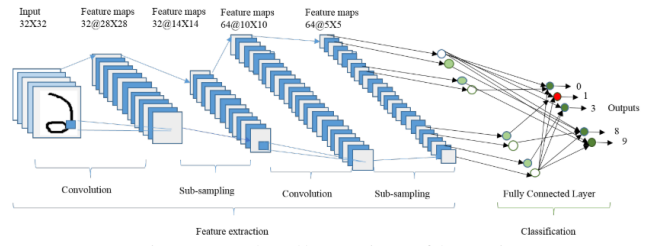


Figure 2. The Illustration of layer in CNN
(Source:[14])

Sharing parameters serves to reduce the number of usage parameters without reducing the ability of CNN, so the memory usage on the computer becomes more efficient. Sub-sampling serves to reduce the number of hidden units in the hidden layer and the sensitivity to shifting and distortion of the input feature

a) Convolution Layer

Convolutional layer is the basic layer of CNN architecture. This layer usually works for convolution calculations that effect the image. The aim of this layer is to extract the feature from image. The equation according to [15] as follows;

$$y_j^{(l)}(x, y) = \phi^{(l)} \left(\sum_{i \in I} \sum_{(u, v) \in K} w_{ji}^{(l)}(u, v) y_i^{(l-1)}(x + u, y + v) + b_j^{(l)} \right)$$

Where $w_{ji}^{(l)}$ denotes the weight on from neuron i to neuron j in layer l , $y_i^{(l-1)}$ denotes output layer on $l - 1$ or input image for the first convolution layer and $b_j^{(l)}$ is the bias on neuron j on layer l . So, that formula is the activation of respective convolution maps is then simply the sum of all convolution result and the bias. The illustration of calculations on convolutional layer shown in Fig.3.

7	2	3	3	8
4	5	3	8	4
3	3	2	8	4
2	8	7	2	7
5	4	4	5	4

 \ast

1	0	-1
1	0	-1
1	0	-1

 $=$

6	-9	

$2 \times 1 + 5 \times 1 + 3 \times 1 + 3 \times 0 + 3 \times 0 + 2 \times 0 + 3 \times -1 + 8 \times -1 + 8 \times -1 = -9$

Figure 3. The illustration of calculations on convolutional layer

b) Subsampling Layer

Pooling layer or subsampling is useful to change the input feature to represent the statistical results of the surrounding features, so the resulting feature size will be much smaller than the previous features [16]. The representation is also useful for reducing the sensitivity to shifting and distortion of features. For example, in the case of face detection in the image, where the face position can vary, then with pooling, convolutional neural network can detect faces in the image without the need to segment first. There are two types of pooling methods that are often used that shown in Fig.4.



Figure 4. Max Pooling(a) and Mean Pooling(b)
(source : [17])

Pooling method that will be used is max-pooling method because max pooling shows better result in [18]. The equation according to Sugumori (2016) as follows;

$$y_{ij}^{(l)} = \max(a_{(l_1 i + s)(l_2 j + t)}^{(k)})$$

Here, l_1 and l_2 are the size of pooling filter and $s \in [0,1]$, $t \in [0,1]$.

c) Fully Connected

Fully Connected Layer is the culmination of a process on CNN that functions as a classifier. In multilayer perceptron, this layer is called hidden layer. The equation for this layer as follows:

$$y^l = \phi (w \cdot y^{l-1} + b)$$

Where, w denotes the weight, b denotes the bias and y^{l-1} denotes output from layer $l - 1$. The illustration of this layer shown in fig.5.

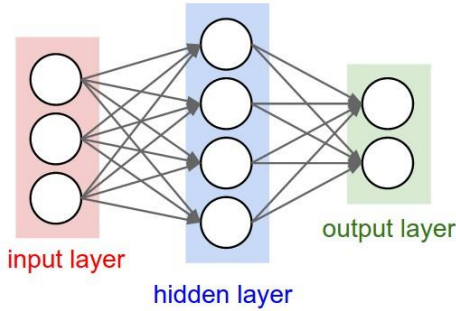


Figure 6. Fully Connected Layer

d) Architecture Network

In this study, the method is used the LeNet-5 architecture with 8 layers including 1 output layer, 3 convolution layers, 2 sub-sampling layers and 2 fully-connected layers. The CNN architecture is shown in fig. 7. The scale of weights and biases in every convolution layers are 5×5 and 1×1 which is initiated with random weight and then propagated overfitting or overlapping.

In the output layer, this research used softmax activation function.

$$P(o_i) = \text{softmax}(H^{out})_i = \frac{e^{H_i^{out}}}{\sum_j e^{H_j^{out}}}$$

3. DATASET

As explain in the previous subsection, CNN could learn internal features from data. In order to do this, data must be put into the input layer in a proper way. The pre-processing was applied in this study is converting image to binary image. Which the value of pixel between 0 or 255. Dataset was used from MADBase are composed 70000 digits written by 700 writer [12]. This dataset has the same format as MNIST dataset.

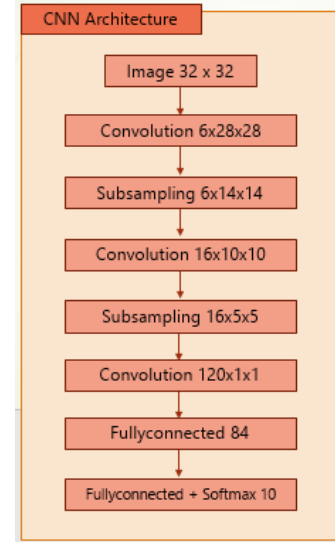


Figure 7. CNN architecture

The dataset is partitioned into two set, 60000 digit as training set and 10000 digits as testing set, as exemplified in fig. 8.



Figure 8. Samples of MADBase dataset

4. SYSTEM ARCHITECTURE

Figure 9 shows the proposed system architecture. The system incorporates two process.

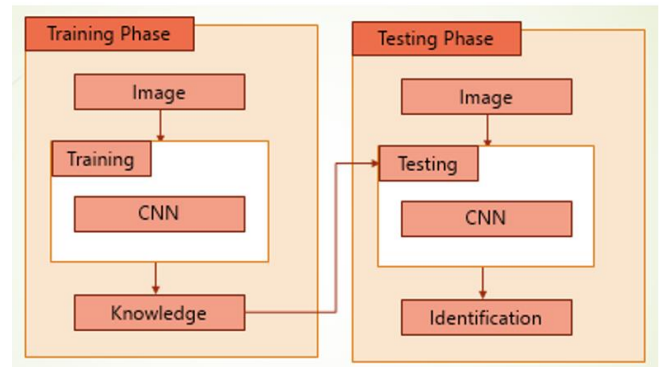


Figure 9. System Architecture

The first process trained image data that has been manipulated by pre-processing process with CNN. When training process done, the model or knowledge has been

saved. The next process is identification process which is testing the knowledge that get from training process.

5. EXPERIMENT RESULT

The performance of CNN was done in training and recognition process. The propose architecture that used in training is also used in testing process. Training and testing process are implemented using java language in Netbeans 8.2 without using the existing library with hardware Intel i5 @ 1.70GHz, 8 GB RAM.

In order to evaluate CNN method, the recognition system was trained based on 60000 images with backpropagation algorithm for 250 epochs. The test is comparing four outcomes from scenarios with differences in the number of different epochs. The result of our experiment shown in table 2, 3, 5 and fig. 10. Average recognition rates in our experiment is 97.67% from 10000 images.

6. CONCLUSION

From the result of test conducted, it can be concluded that the system can recognize object with an average value of 97.67%. This result shows the proposed method is suitable for the MADBase dataset.

TABLE 2. Testing With 100 Epoch

Epoch	Sum Of Data	Unrecognized	Recognized	Percentage (%)
100	10000	275	9725	97.25
150	10000	211	9789	97.89
200	10000	219	9781	97.81
250	10000	228	9772	97.72
Average				97.6675

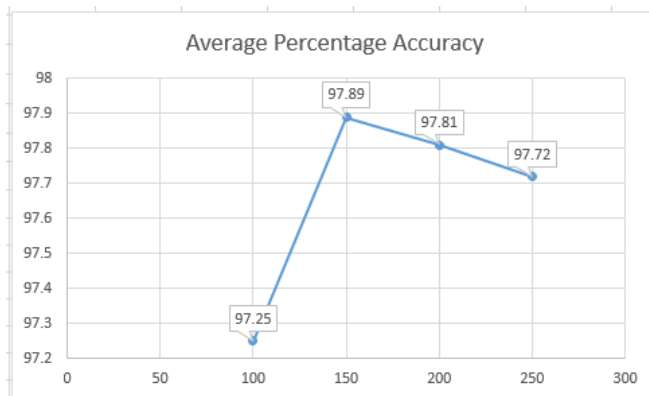


Figure 10. Comparison of Test Results

REFERENCES

- [1] R. I. Zaghloul, D. M. K. B. Enas, and F. AlRawashdeh, "Recognition of Hindi (Arabic) Handwritten Numerals," *Am. J. Eng. Appl. Sci.*, vol. 5, no. 2, pp. 132–135, 2012.
- [2] A. Ashiquzzaman and A. K. Tushar, "Handwritten Arabic Numeral Recognition using Deep Learning Neural Networks," pp. 3–6, 2017.
- [3] "The World Factbook — Central Intelligence Agency." [Online]. Available: <https://www.cia.gov/library/publications/resources/the-world-factbook/fields/2098.html>. [Accessed: 20-Jul-2018].
- [4] N. Das, A. Mollah, S. Saha, and S. Haque, "Handwritten Arabic Numeral Recognition using a Multi Layer Perceptron," *arXiv Prepr. arXiv1003.1891*, pp. 200–203, 2010.
- [5] S. Mahmoud, "Recognition of writer-independent off-line handwritten Arabic (Indian) numerals using hidden Markov models," *Signal Processing*, vol. 88, no. 4, pp. 844–857, 2008.
- [6] X. X. Niu and C. Y. Suen, "A novel hybrid CNN-SVM classifier for recognizing handwritten digits," *Pattern Recognit.*, vol. 45, no. 4, pp. 1318–1325, 2012.
- [7] P. P. Selvi and T. Meyyappan, "Recognition of Arabic numerals with grouping and ungrouping using back propagation neural network," *Proc. 2013 Int. Conf. Pattern Recognition, Informatics Mob. Eng. PRIME 2013*, pp. 322–327, 2013.
- [8] S. S. Ali and M. U. Ghani, "Handwritten Digit Recognition Using DCT and HMMs," *2014 12th Int. Conf. Front. Inf. Technol.*, pp. 303–306, 2014.
- [9] M. Takruri, R. Al-Hmouz, and A. Al-Hmouz, "A three-level classifier: Fuzzy C Means, Support Vector Machine and unique pixels for Arabic handwritten digits," *2014 World Symp. Comput. Appl. Res. WSCAR 2014*, pp. 1–5, 2014.
- [10] J. H. Alkhateeb and M. Alseid, "DBN - Based learning for Arabic handwritten digit recognition using DCT features," *2014 6th Int. Conf. Comput. Sci. Inf. Technol. CSIT 2014 - Proc.*, pp. 222–226, 2014.
- [11] M. Fachrurrozi, "Multi-Object Face Recognition Using Content Based Image Retrieval (CBIR)," no. x, pp. 193–197, 2017.
- [12] "AHDBase." [Online]. Available: <http://datacenter.aucegypt.edu/shazeem/>. [Accessed: 20-Jul-2018].
- [13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.
- [14] Md Zahangir Alom, Paheding Sidike, T. M. Taha, and Vijayan K. Asari, "Handwritten Bangla Digit Recognition Using Deep Learning," 2017.
- [15] S. Duffner, "Face Image Analysis With Convolutional Neural Networks," *Thesis*, pp. 1–191, 2007.

- [16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. 2016.
- [17] Y. Kim, S. Member, and T. Moon, "Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks," vol. 13, no. 1, pp. 1–5, 2015.
- [18] D. Scherer, Andreas Muller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6354 LNCS, no. PART 3, pp. 92–101, 2010.
- [19] Y. Sugumori, *JAVA DEEP LEARNING ESSENTIALs*. 2016.