Journal of Engineering and Applied Sciences 14 (19): 6982-6990, 2019
ISSN: 1816-949X

Design of arabic recognition application using Convolutional neural network

Salma Shofia Rosyda, Budhi Irawan and Anggunmeka Luhur Prasasti
Department of Computer Engineering, Faculty of Electrical Engineering, Telkom University, Bandung, Indonesia, anggunmeka@telkomuniversity.ac.id

Abstract: Arabic is one of the languages which attracted, needed and used by a lot of people in the world. Many countries have used Arabic language in courses which related to the international world. Arabic also used by Muslim people, beacause Arabic is the main language of the hooly book of Muslim (Quran). The importance to know and to understand the basic of Arabic language for the temporary needs like when we travel to a country which use Arabic as their main language include during Hajj and Umrah. At this time the smartphone has become a major need for humans, based on that then made Arabic writing recognition application which aims to help pilgrims of Hajj and Umrah who can not speak Arabic. In this study, the method used is the method of Convolutional Neural Network (CNN). Test results from Arabic handwritten image classification using CNN resulted in an average accuracy of 60%. It can be concluded that the CNN method used in this application is able to do a good classification.

Key words: Image processing, Arabic writing recognition application, Convolutional Neural Network (CNN), Muslim people, smartphone, classification

INTRODUCTION

According to Indonesia dictionary (KBBI) language is the words used by tribes that contain a certain meaning. Arabic is a tool commonly used by the Arabs to convey intent and purpose. Arabic itself is the official language of 25 countries in the world and is also one of the international languages.

Generally, not many Indonesians understand Arabic while the majority of Indonesians are Muslim, of course, every year the country sends Umrah and Hajj pilgrims. In 2016, Indonesia sent 154,421 pilgrims to perform Hajj (Statistik, 2017). Because in the majority location of Hajj and Umrah using Arabic, then Indonesian must understand the Arabic language, especially, Arabic used in daily.

Therefore, the Arabic writing recognition application is made because it can detect the Arabic writing through the input image. That way, pilgrims can take pictures of Arabic writing to know the translation and its latin. This app uses image processing, image processing is processing with the input of the image. The method used for the classification is Convolutional Neural Network (CNN). This app is designed on android operating system.

MATERIALS AND METHODS

In this study, we will explain the system overview, pre-processing and classification. There are three processes on preprocessing, i.e., resize, grayscale and thresholding. In the classification will explain how the classification process for this application using CNN method.

System overview: The system designed on this Android operating system aims to read Arabic handwriting that has punctuation (syakal), this system can also translate and make latin from Arabic writing. The recognition process begins with image taking, image capture can be done from smartphone camera android or from the image in the smartphone gallery. Arabic writing in the form of the image will go through the process of pre-processing. Preprocessing consists of three processes, i.e., resize, grayscale and thresholding. Secondly, the preprocessing result will enter into the classification process using CNN. This process aims to categorize images into predefined classes. After that, the classification results will be latin and translation. Here's an overview of the system.

System design: In designing the system will be discussed about the process on the system that has been described in Fig. 1.

Corresponding Author: Anggunmeka Luhur Prasasti, Department of Computer Engineering, Faculty of Electrical Engineering, Telkom University, Bandung, anggunmeka@telkomuniversity.ac.id

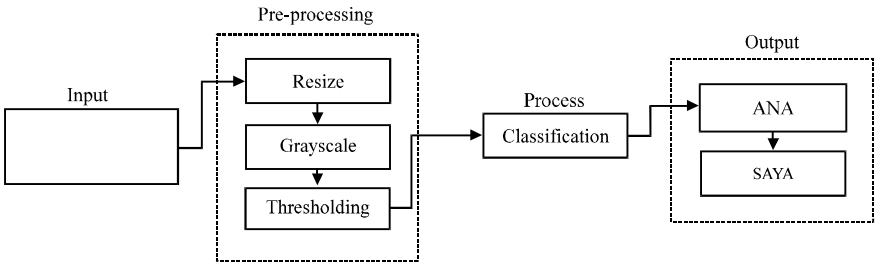

Fig. 1: General system overview

Pre-processing: Pre-processing is the process before the image is converted into text. This process aims to determine the important information contained in the images entered into the system. In pre-processing there are three processes:

Resize: Is a process of change in a discrete image defined on a set of coordinates to a new set of coordinates. This process can be done by resampling used to increase the number of points on the image to improve the quality of the image (Wang *et al*., 2008). In this application, the image is resized to 30×30.

Grayscale: In this process, the image becomes gray level. The image used as input is a combination image of Red, Green and Blue (RGB). The image inputted in this process is represented to degrees of gray. In grayscale, there is only dark (black) and light (white). The black color in decimal is 0 while the white in decimal is 255. In black bit is 0 while white is 11111111. The values of a-c represent the color channels, Red = (0.299), Green = (0.587) and B = (0.144). The numbers that represent the colors are in accordance with the sensitivity of the human eye, if they meet the colors. Here, is the formula of grayscale (Lim and Isa, 2011; Zhao and Tian, 2012):

$$J(x, y = [a \times R(x, y)] + [b \times G(x, y) + [c \times B(x, y)] \quad (1)$$

Thresholding: Thresholding is the process of changing the grayscale image to binary (Black-White) to know clearly where the object and where the background in the image. In this process, if the input image pixel of the grayscale process has a value greater than a predetermined value and constant (the treshold value) then the pixel of the pixel output is 1 and if the pixel has a value smaller than the Threshold value (T) then the pixel output will be assigned a value of 0 (Roy *et al*., 2014; Sivakumar and Murugesh, 2014):

$$T(X, Y) = \begin{cases} 1\ F(X, Y) \geq T \\ 0\ F(X, Y) < T \end{cases} \quad (2)$$

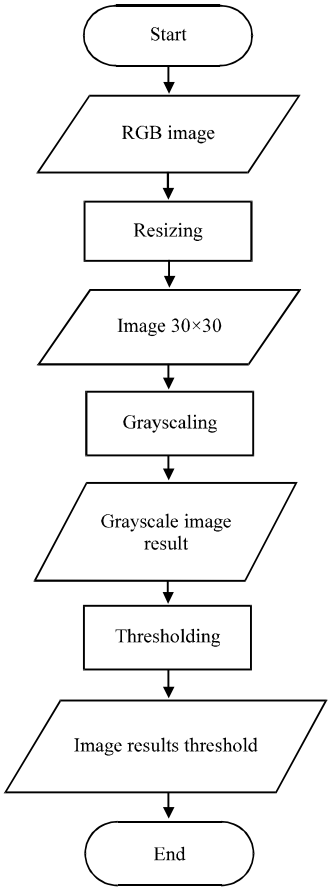

Fig. 2: Pre-processing

Where:
T = Threshold value
F (X, Y) = Value before being processed

The preprocessing process is illustrated in the following flowchart (Fig. 2).

Classification: Classification is a grouping process based on similarities and differences. The images obtained will be classified using layers present on CNN. CNN consists of three main layers.

Convolution layer: This layer is the base layer that builds a CNN which aims to extract features from the input image. In this layer, the feature representation of the image in the input is studied (Song *et al*., 2016; Rawat and Wang, 2017; Akhand, *et al*., 2015). The first kernel will move from the top left corner to the top right corner and from the top right corner the kernel moves to the lower right corner (Maitra *et al*., 2015). The kernel is moved per element, so that, the move process will be repeated until the kernel is completely moved to the lower right corner. The first convolution layer extracts low-level features such as edges, lines and corners while the layers are higher. The formula for operating the convolution process as:

$$f_c = \sum_m \sum_n I_{(m, n)} K_{(i, j)} \quad (3)$$

Where:
f_c = The result of the convolution process
I = Input image matrix
K = Filter/Kernal

Subsampling layer: This is a layer for the image reduction process to increase the invariance of feature positions. There are two types of pooling commonly used for image processing, namely max pooling and averange pooling (Murtiwiyati, 2013). Max pooling has a way of working by taking the maximum value of a non-overlapping sub-region (Putra, 2016) while mean collection works by calculating the mean value of the merging area (Ramadhan *et al*., 2016):

$$pool(i, j) = \frac{1}{M} \sum_m X_{(i, j)} \quad (4)$$

The fully connected layer is the last layer after going through the convolution layer and repeating itself repeatedly in accordance with the designed architecture (Reza and Mazumder, 2012). Acting as a class in the convolutional neural network, this layer is included in the neural network convolutional architecture which consists of input, hidden layer and output layer (Ramadhan *et al*., 2016). Every neuron has full connections to all activations in the previous layer as seen in regular neural networks. The last fully-connected layer is called the "output layer" and in classification settings it represents the class scores (Fig. 3).

Classification is a grouping process based on similarities and differences. The images obtained will be classified using layers that are on CNN. Here's how CNN works on the Arabic writing introduction application: the characteristic vector of the Arabic script obtained from the pre-processing stage used as the initial input of the classification process. Data is divided into two: data training and data testing.

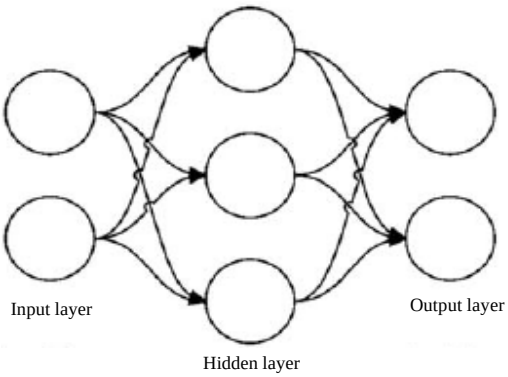

Fig. 3: Fully connected layer (Culjak *et al*., 2012)

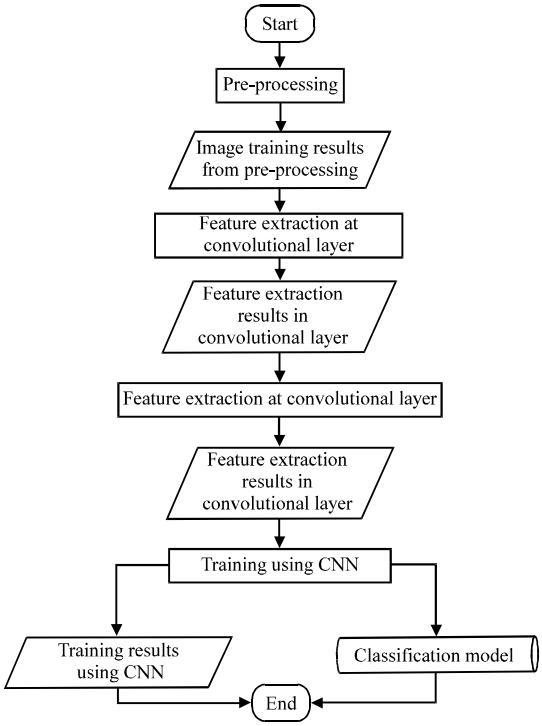

Fig. 4: Classification flow chart using CNN

Training data entered in sentence form and words amounted to 20 and 50 copies of each sentence and word. In the CNN process training process, the first process is drawing for training into the pre-processing process, after which the image enters the convolution layer to extract trained properties using CNN. Testing data is an Arabic word image. Imagery will go through a pre-processing process and then classified using the Convolutional Neural Network (CNN) method. In the process of pre-processing the image will be resized to 30×30 (Fig. 4).

In the CNN classification there are three main layers of the convolution layer, the subsampling layer and the

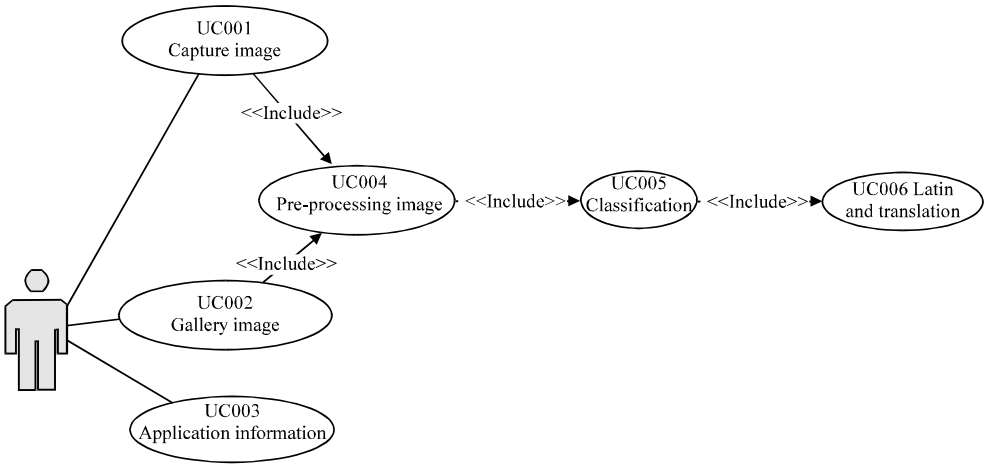

Fig. 5: Use case

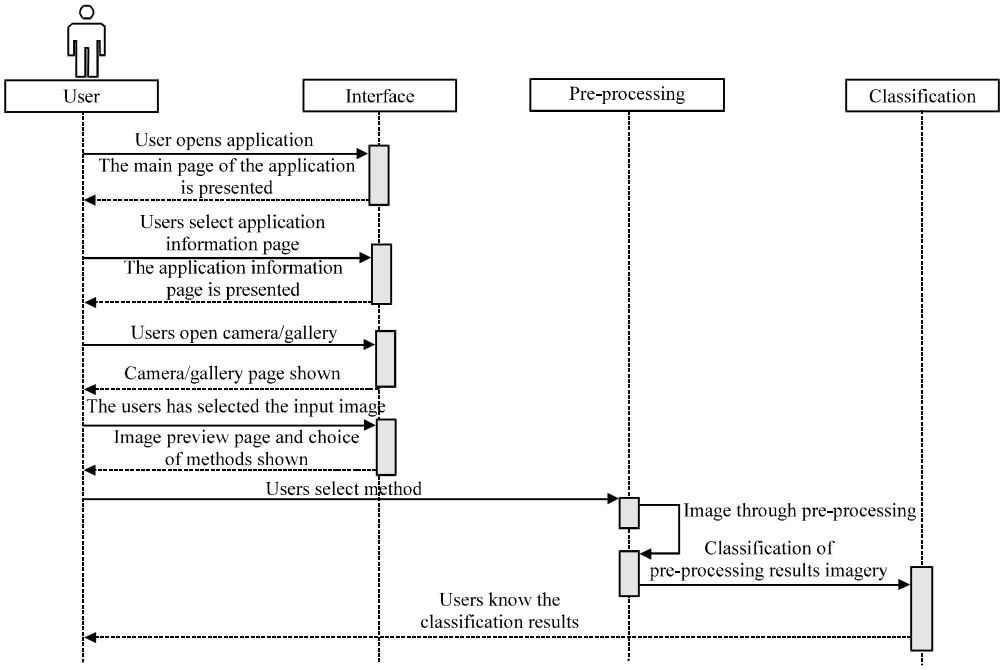

Fig. 6: Sequence diagram

fully connected layer. The additional layer applied to CNN in this application is the softmax layer. In the convolution layer, the image to the input is a 30×30 matrix. Once processed in the shock layer, the image is processed in the subsampling layer using an average pooling with the image input of a convolution process with a size of 28x28. After that the image matrix will be processed in a layer fully connected to the 14×14 matrix size. The 14×14 matrix size comes from the subsampling process. In the last layer of CNN is the output layer that is fully connected to the output of this layer is 20, 20 comes from the number of classes that contain Arabic writing applications. For the softmax layer, the process that occurs, here is a process that aims to minimize noise on the image. Here is a classification diagram.

Use case diagram: Use case diagram is used to provide a functional overview of the system. It consists of two main elements of actors and relationships, both the relationship between actor and use case and the relationship between use case and use case. Here's a use case diagram of the app you created (Fig. 5 and 6).

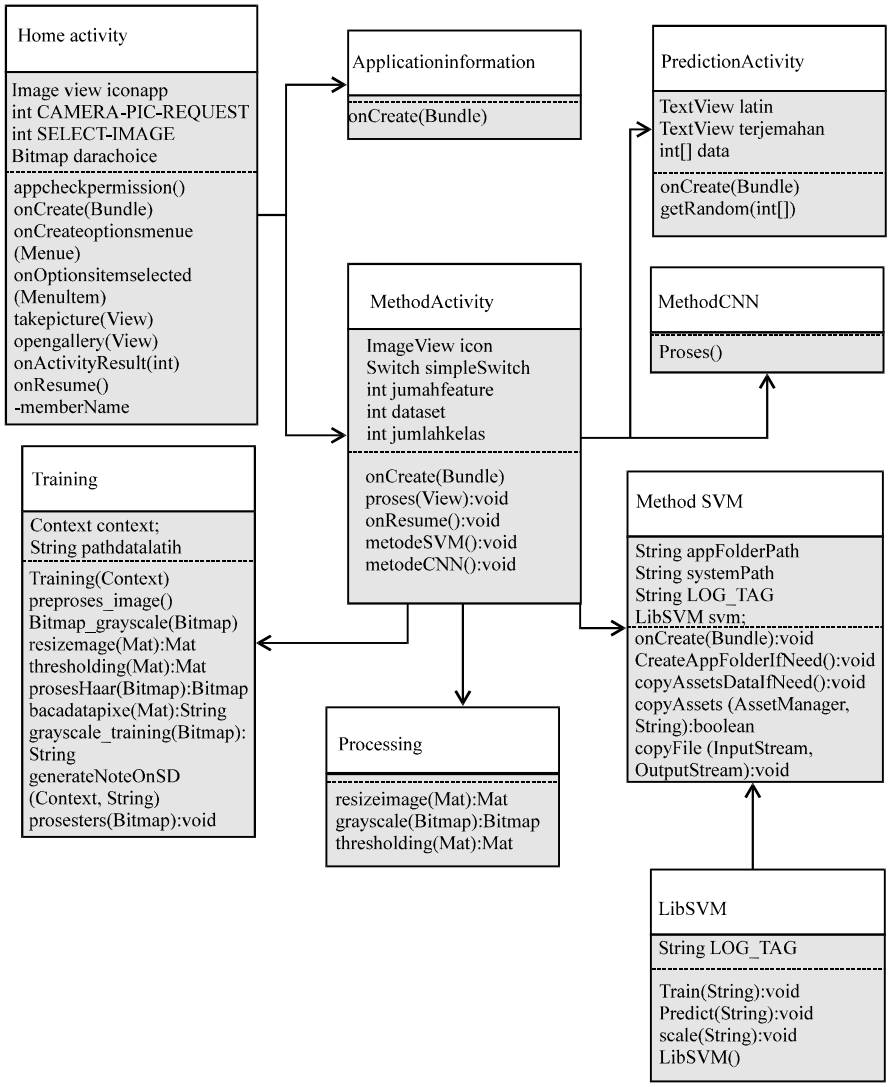

Fig. 7: Class diagram

Sequence diagram: Is a diagram that serves to provide an overview of interaction objects that exist in the system at certain times. This diagram illustrates the interaction between the user, the view, the pre-processing and the classification. First the user opens the app after which the view from the main page of the app is displayed (Fig. 6). After that the user can select the application information menu to be able to find information related to the application. After that the user clicks the camera/gallery menu on the main page of the app, then the user can view and use the camera/gallery to insert the image to be processed in pre-processing. Once, the image is selected, the image preview and method selection will be displayed. Once, processed in the pre-processing the image will go into the classification stage to be able to detect the results of its arabic detection. The last user will know the result of the classification. The following is a sequence diagram introduction to Arabic writing application.

Class diagram: The class diagram is a diagram illustrating the relationship between the classes contained in the application. Here is a class diagram of Arabic writing recognition application (Fig. 7).

Application display implementation: This study will describe the appearance of Arabic writing recognition applications. The following is an explanation of the Arabic writing recognition application appearance.

Fig. 8: Home

Fig. 9: Application information page

Home: On this page there is a logo and text "SmartlationS", the image below is the main page view of the application. On this page there are several menus that can be operated. The following is an explanation of the menus on the main page (Fig. 8):

- Camera menu, to make image acquisition via. camera
- Gallery menu, to select an image from the gallery
- Application information menu, to display information about the application

Application information page: On this page, there is text application information. In the picture, below is a page that shows information related to the application. This page can be accessed through the about menu on the main page (Fig. 9).

Select page method: On this page there are two menu options that work to select a method to process an input image. This page is accessible after the user selects the image you want to process on the main page. Here is a view of the page selection method (Fig. 10 and 11).

Translation results page: On this page there is a result of the detection of Arabic writing in latin as well as in

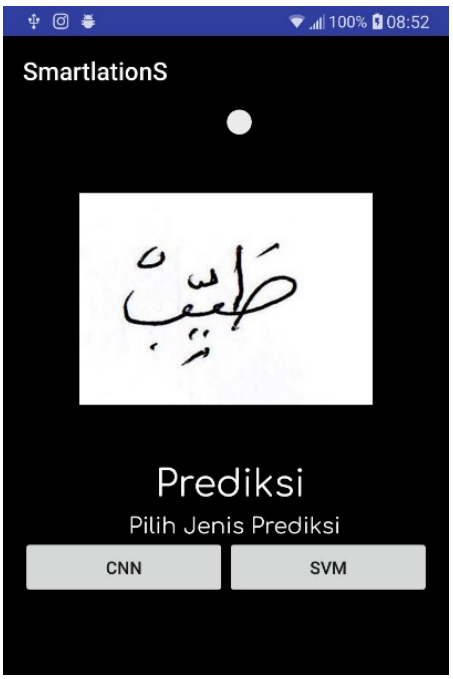

Fig. 10: Select page method

translation. This page can be accessed after the user selects the method you want to use to detect images. Here is a translated page view (Fig. 11).

Fig. 11: Translation results page

RESULTS AND DISCUSSION

In this study will be discussed about the test results and analysis of each test that has been done on Arabic tlisa recognition applications. The purpose of testing is to ensure that the built system can perform the task correctly. The following is a test conducted on Arabic writing application.

Testing Arabic writing accuracy: The purpose of this test is to determine the level of accuracy of the system by making variations in the number of models used in the system. In the system there are 1,000 models with a size of 30×30 pixels. Testing will be done on every word and sentence specified. This system is able to translate Arabic languange to Bahasa (Indonesian language) with 60% average accuracy. The following is an accuracy test Table 1.

Table 1: Accuracy testing

Image testing	Word/sentence	Translation (Indonesian language)	Accuracy (%)
كَيْفَ حَالُكَ؟	Khaifa Haluka	Apa kabar (Laki-Laki)	60
كَيْفَ حَالُكِ؟	Khaifa Haluki	Apa kabar (Perempuan)	60
طَيِّبٌ	Toyyib	Baik	50
جِهَةٌ	Jihatun	Arah	90
فِي	Fii	Dalam	40
أَهْلاً وَسَهْلاً	Ahlan Wa Sahlan	Halo/Hai (Respon)	70
مَرْحَبًا	Marhaban	Halo/Hai (Sapa)	70
نَعْلٌ	Na'lun	Sandal	60
مَا سْمُكَ؟	Masmuka	Siapa Namamu (Laki-laki)	50
مَا سْمُكِ؟	Masmuki	Siapa Namamu (Perempuan)	50
أُدْخُلْ	Udkhul	Masuk	60
إِسْمِي	Ismi	Nama Saya	70
جَوَّالٌ	Jawaalun	Ponsel	70
أَمَامَ	Amaama	Di Depan	50
فَوْقَ	Fauqo	Di Atas	70
تَحْتَ	Tahta	Di Bawah	80
لَمْ	Lam	Belum	50
يَمِينٌ	Yaminun	Kanan	50
حَمَّامٌ	Hamam	Kamar Mandi	40
خَلَاصٌ	Kholasun	Sudah	60

Table 2: Responsetime testing

Devices	Processor	Android version	Respons time (sec)
1	1.8 GHz quad-core	Marshmellow (6.0.0)	76
2	1.9 GHz octa-core	Nougat (8.0.0)	73

Table 3: The result of testing the acquisition distance effect

Distance (cm)	Accuracy (%)
10	60
15	30
20	10

Table 4: Testing the effect of image viewing angle

Angles	Accuracy (%)
Leaning to the left	40
Leaning to the right	40

Table 5: Testing CNN method

Epoch	Learning rate (%)	Accuracy (%)
1000	75	70
1500	75	80
1000	100	90
1500	100	90

From Table 1, it can be seen that the highest accuracy value is found in the test word Jihatun with 90% accuracy. The things that affect the accuracy level are clear writing and adequate lighting. In some words like the fii word in the system has little difficulty in detecting because the example, for the word there is a black line caused by the lack of scanning process in the printer.

Response time testing: This test aims to determine the effect of the processor clock speed on the response time speed of the application system. This test is done by running applications on two types of smartphones with different processors. In device 1, the system response time for CNN is measured, the same is done on device 2. Table 2 is the test result of the system response time.

From the results of testing the response time of the system obtained a response time of 76 sec while using the device with quad-core processor speed of 1.8 Ghz. While on devices with octa-core 1.9 Ghz processor, the response time obtained for the CNN method is 73 sec.

Testing the acquisition distance effect: This test is performed to determine the accuracy of the system based on the distance specified in the image acquisition process. In this test, a distance of 10, 15 and 20 cm has been set. The following is a Table 3 of results of acquisition distance testing effects.

Based on the test results in table above it can be seen that the highest accuracy value is achieved with a distance of 10 cm. It can be concluded that the ideal distance for image acquisition is 10 cm with the accuracy reaching 60% while at a distance of 15 and 20 cm there are some letters that are difficult to detect.

Testing the effect of image viewing angle: This test is carried out with the aim to determine the effect of shooting angle (shooting position) on the level of accuracy at detection. The following is a test from the point of view of taking pictures. Table 4 and 5 shows an accuracy of 40% for each corner of the acquisition.

Metode CNN (learning rate and epoch): Testing the level of learning rate and epoch is done with the aim to find out how the influence of the two parameters on the detection accuracy. The following are levels of learning rate and epoch.

After doing this test, it can be seen that the highest accuracy is obtained when the epoch is 1000 and 1500 with the learning level is 100%. The lowest accuracy is obtained when the epoch is 1000 and the learning level is 75%.

CONCLUSION

Based on the implementation and testing of Arabic writing recognition applications referring to the analysis and design that has been made can be summarized as follows: detection of Arabic tulsions using the Convolutional Neural Network (CNN) classification was successfully carried out. The parameters contained in the test have an effect on the level of accuracy. The highest accuracy in the process of detecting Arabic text using the CNN method is obtained in the word "Jihatun" by 90% using training data of 50 copies of each word or sentence. This value is still not optimal and is still below the average, due to the limitations of image training data contained in the database.

REFERENCES

Akhand, M.A.H., M.M. Rahman, P.C. Shill, S. Islam and M.H. Rahman, 2015. Bangla handwritten numeral recognition using convolutional neural network. Proceedings of the International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), May 21-23, 2015, IEEE, Dhaka, Bangladesh, ISBN:978-1-4673-6675-5, pp: 1-5.

Culjak, I., D. Abram, T. Pribanic, H. Dzapo and M. Cifrek, 2012. A brief introduction to OpenCV. Proceedings of the 35th International Conference on Convention MIPRO, May 21-25, 2012, IEEE, Opatija, Croatia, ISBN:978-1-4673-2577-6, pp: 1725-1730.

Lim, W.H. and N.A.M. Isa, 2011. Color to grayscale conversion based on neighborhood pixels effect approach for digital image. Proceedings of the 7th International Conference on Electrical and Electronics Engineering (ELECO 2011), December, 1-4, 2011, Bursa, Turkey, pp: 1-4.

Maitra, D.S., U. Bhattacharya and S.K. Parui, 2015. CNN based common approach to handwritten character recognition of multiple scripts. Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR), August 23-26, 2015, IEEE, Tunis, Tunisia, ISBN:978-1-4799-1805-8, pp: 1021-1025.

Murtiwiyati, G.L., 2013. Designing Indonesian culture learning application for elementary school children based on android. Comput. Sci. J. Inf. Syst., 12: 1-10.

Putra, I.W.S.E., 2016. Image classification using Convolutional Neural Network (CNN) on Caltech 101. Ph.D Thesis, Sepuluh Nopember Institute of Technolo, Surabaya, Indonesia.

Ramadhan, I., B. Purnama and S. Al Faraby, 2016. Convolutional neural networks applied to handwritten mathematical symbols classification. Proceedings of the 4th International Conference on Information and Communication Technology (ICoICT), May 25-27, 2016, IEEE, Bandung, Indonesia, ISBN:978-1-4673-9880-0, pp: 1-4.

Rawat, W. and Z. Wang, 2017. Deep convolutional neural networks for image classification: A comprehensive review. Neural Comput., 29: 2352-2449.

Reza, H. and N. Mazumder, 2012. A secure software architecture for mobile computing. Proceedings of the 2012 9th International Conference on Information Technology-New Generations, April 16-18, 2012, IEEE, Las Vegas, Nevada, USA., ISBN:978-1-4673-0798-7, pp: 566-571.

Roy, P., S. Dutta, N. Dey, G. Dey and S. Chakraborty *et al.*, 2014. Adaptive thresholding: A comparative study. Proceedings of the International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), July 10-11, 2014, IEEE, Kanyakumari, India, ISBN:978-1-4799-4191-9, pp: 1182-1186.

Sivakumar, V. and V. Murugesh, 2014. A brief study of image segmentation using thresholding technique on a noisy image. Proceedings of the International Conference on Information Communication and Embedded Systems (ICICES 2014), February 27-28, 2014, IEEE, Chennai, India, ISBN:978-1-4799-3834-6, pp: 1-6.

Song, X., X. Gao, Y. Ding and Z. Wang, 2016. A handwritten Chinese characters recognition method based on sample set expansion and CNN. Proceedings of the 3rd International Conference on Systems and Informatics (ICSAI), November 19-21, 2016, IEEE, Shanghai, China, ISBN:978-1-5090-5522-7, pp: 843-849.

Statistik, B.P., 2017. [Indonesian statistics in 2017]. Statistics Indonesia Government office, Central Jakarta, Indonesia. (In Indonesian)

Wang, Y.S., C.L. Tai, O. Sorkine and T.Y. Lee, 2008. Optimized scale-and-stretch for image resizing. ACM. Trans. Graphics, 27: 1-8.

Zhao, X. and D. Tian, 2012. The architecture design of streaming media applications for Android OS. Proceedings of the IEEE International Conference on Computer Science and Automation Engineering, June 22-24, 2012, IEEE, Beijing, China, ISBN:978-1-4673-2007-8, pp: 280-283.