Hand Gesture Recognition System for Arabic letters and numbers Using Deep Learning

Mogeeb A. Saeed [1], Mohammed Hashem Almourish [2], Ahmed Y. A. Saeed[3]

[1]Department of Computer Networks & Distributed Systems, Faculty of Engineering & IT -Taiz University- Yemen.
[2]Department of Communication & Computer Engineering, Faculty of Engineering & IT -Taiz University- Yemen.
[3]Department of Software Engineering, Faculty of Engineering & IT -Taiz University- Yemen.
mogeeb1982@gmail.com[1],Mohmedalmourish@gmail.com[2],ahmed12.12f@gmail.com [3]

Abstract. Recognizing hand gestures is a key to overcome many of the difficulties people with a physical disability have in communicating with the general public. Therefore, it was necessary to develop a technology that solves this problem and enables people with disabilities to communicate with people without problems. This study presents a hand gesture recognition system to recognize Arabic numbers and letters using artificial intelligence that uses deep learning technology. The convolution neural network (CNN) was used as a deep learning model to train data sets on hand gestures, where the gesture images were displayed on the network entrance and changed Scale the images to the same size to extract the image features and then categorize them into text. The results showed that the CNN model achieved an accuracy of 99% in the testing phase.

Keywords: hand gesture recognition, deep learning, convolution neural network.

1. Introduction

In recent years, many researchers have become interested in solving problems facing people with physical, visual or hearing disabilities. The main problem is how can people with disabilities integrate into society able to communicate and express without difficulties. The importance of the research lies in finding a method of communication between the deaf and the community by finding an easy way to communicate between them without difficulties or obstacles by converting the hand gestures of the deaf into a language that society understands. A gesture can be defined as a special physical movement of the hand, head, or body that has its own sign language to communicate with others. Several studies have been conducted to identify hand gestures by automatically recognizing sign language using a computer that helps

deaf people communicate with their environment [1]. Krueger (1991) was the first who proposed computer recognition of gesture in the mid-seventies [2].

After that, many techniques were developed to recognize gestures by computer, but recent studies show that the use of artificial intelligence systems shows a higher accuracy in recognizing the images of hand gestures. In this paper, artificial intelligence was proposed to recognize images of hand gestures for deaf people. Where it showed that the use of deep learning and neural networks as models of artificial intelligence is effective in identifying images of gestures, whether letters or numbers in real time. The CNN model was proposed in deep learning technique to recognize images of hand gestures, and the results showed that the CNN model achieved an accuracy of 99%. This paper is organized as follows: section 2 is introduced about techniques that used, Section 3 is presented the methodology of the system which including data collection and training phase that used, section 4 is introduced the result that gotten from implementation and discussion them and the last section is presented the conclusions of the system.

Related work

In this part, we will review the studies that focused on recognizing hand gestures using modern technologies, which provide an intelligent method for human-computer interaction. Methods for recognizing hand gestures have evolved with the development of computer techniques.

Reference [3] presented a paper for automatic electromyography based on hand gesture recognition system which used a fast and accurate method for hand gesture categories recognition. The accuracy of this system achieves to 99.78%. And the study [4] presented a model for real-time Arabic number recognition using Hidden Markov Model which can recognize isolated and continuous gesture. The accuracy of this research was 98.94% for isolated and 95.7% for continuous gestures. Nguyen et al. [5] presented real-time hand gesture recognition system which used for recognition gestures in unconstrained environments that the system has been tested. In this system, the American Sign Language has been recognized which can recognize (digits, letter spelling and alphabets). While reference [6] presented a paper for recognition the American letters (A to Z) which used different HMM topologies with different states. Kim et al. [7] introduced a paper for finger movement recognition using EMG sensors. In this paper, combination of Bayes and KNN classifier have been used to classify 20 classes. The accuracy of this system reached to 94%. Reference [8] introduced the ANN implementation of 13 gesture recognition. The architecture of ANN has 45 inputs, 2 hidden layers and 14 outputs. The accuracy of this neural network was 96.02%. Rafiqul et al. [9] presented studies related to gesture recognition using PCA techniques, HMMs and ANNs, and also presented a hand gesture recognition system by detecting an image of the hand and then extracting its features and then recognizing it. The authors of [10] presented a method for identifying sign language with data gloves, and two different neural network systems were used to recognize sign language. The study [11] proposed a system for recognizing hand gestures using hand gloves that sends signals via a wireless system. Sensors are placed inside

the hand gloves and then the gloves are moved with specific gestures and the signals are sent using Bluetooth to be recognized with an accuracy of up to 94%. The authors of [12] demonstrate a method for hand gesture detection using image processing where the image is segmented by color and its features are extracted, then machine learning algorithms are applied to recognize hand signals. In Study [13] a method was proposed to convert ASL hand gestures into voice using a Convolutional Neural Network (CNN). They created a desktop application that captures American sign language hand gestures via a computer camera, translates them into text, and then converts them into voice.

2. Methodology

A. Data collection

The dataset was collected from the Arabic alphabet sign language dataset [14]. A dataset of gestures pictures, which consist of 54049 images (28 letters and 10 digits), was compiled and categorized into 38 classes on the output. The figure 1 shows some images of hand gestures.

Figure 1. ARABIC SIGN LANGUAGE ALPHABETS

B. Pre-processing

In this step, all images in the dataset have been resized into (200x200) to fit the input layer for CNN. Also converted them to binary level.

C. Training Phase: In this phase, the sign language dataset is split into 70% for the training phase and 30% for the test. Training is done by the CNN model, which is used to extract features from the dataset and then classify it. The final weights of the trained data are saved in an expert form, which is then used to test the data.

D. Testing Phase: In this phase, 30% of the dataset is tested, and all images in the dataset are resized to (200 x 200) and converted to the binary level to be equal with the input data.

1. **Convolutional Neural Network**: The CNN model is a deep learning algorithm consisting of input, hidden and output layers. It is designed manually according to the data to be processed. The number of layers and filters is not fixed and can be increased or decreased. In the research,

three convolutional layers were used to train the sign language dataset [15]. The activation function is usually the RELU layer which is used to determine the output to the network. Figure 2 shows the structure of a particular CNN model.

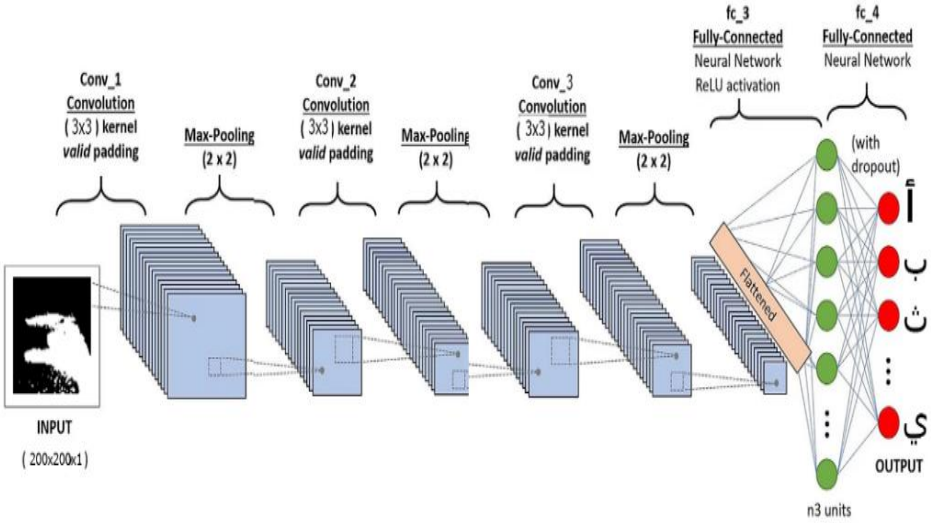

Figure 2. CNN architecture.

3. Results

In this part, we will present the results obtained during the implementation of a convolutional neural network. Where, we divided the dataset into two parts, 70% for training and 30% for testing data. The convolutional neural network was trained in 10 periods and the size of 64 batches as shown in figure 3, where it achieved an accuracy of 98.7% in the training phase and 99% in the test phase as shown in the table 1.

Table 1. results of training by CNN model.

Training loss	0.0428
Validation loss	0.0106
Training accuracy	98.7%
Validation accuracy	99%

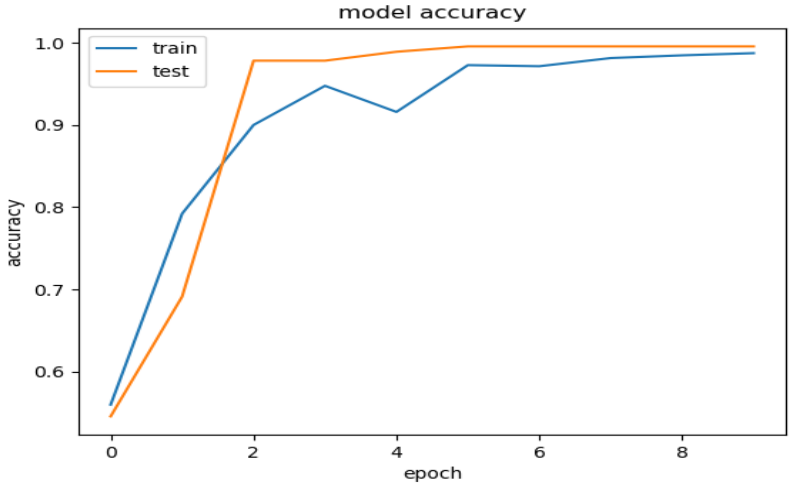

Figure 3. Training and testing accuracy of CNN model.

4. Conclusion

This paper proposed a system to recognize hand gestures for Arabic letters and numbers using artificial intelligence. CNN was used and it showed effective in solving this problem. The data set was divided into 70% for training and 30% for testing and the model was trained with batch size 64 and 10 epoch. The system achieved 99% accuracy in the testing phase.

References

[1] M. M. Hasan and P. K. Misra, "Brightness factor matching for gesture recognition system using scaled normalization," International Journal of Computer Science & Information Technology, vol. 3, no. 2, pp. 35–46, 2011.

[2] Krueger, M.W.: Artificial Reality II, vol. 10. Addison-Wesley, Reading (1991).

[3] Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., & Acharya, U. R. (2020). A deep learning approach for Parkinson's disease diagnosis from EEG signals. Neural Computing and Applications. https://doi.org/10.1007/s00521-018-3689-5.

[4] M Elmezain, A Al-Hamadi, J Appenrodt, B Michaelis. "A hidden markov model-based continuous gesture recognition system for hand motion trajectory." 2008 19th International Conference on Pattern Recognition. IEEE, 2008.

[5] D. B. Nguyen, S. Enokida, and E. Toshiaki. RealTime Hand Tracking and Gesture Recognition System. IGVIP05 Conference, CICC, pp. 362-368, 2005.

[6] L. Nianjun, C. L. Brian, J. K. Peter, and A. D. Richard. Model Structure Selection & Training Algorithms for a HMM Gesture Recognition System. In International IWFHR, pp. 100-106, 2004.

[7] Kim J, Mastnik S, André E (2008) EMG-based hand gesture recognition for realtime biosignal interfacing. In: Proceedings of the 13th international conference on Intelligent user interfaces, ACM, pp 30–39.

[8] Akmeliawati R, Dadgostar F, Demidenko S, Gamage N, Kuang YC, Messom C, Ooi M, Sarrafzadeh A, SenGupta G (2009) Towards real-time sign language analysis via markerless gesture tracking. In: Instrumentation and measurement technology conference, I2MTC'09, IEEE, pp 1200–1204.

[9] Rafiqul Zaman Khan & Noor Adnan Ibraheem "Survey on Gesture Recognition for Hand Image Postures" Vol. 5, No. 3; May 2012

[10] K. Murakami and H. Taguchi, "Gesture recognition using recurrent neural networks," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching through Technology, pp. 237–242, 1999.

[11] L. Kau, W. Su, P. Yu and S. Wei, "A real-time portable sign language translation system," 2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS), Fort Collins, CO, 2015, pp. 1-4, doi: 10.1109/MWSCAS.2015.7282137.

[12] Shahriar S, Siddiquee A, Islam T, Ghosh A, Chakraborty R, Khan AI, Shahnaz C, Fattah S,"Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning," TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 1168-1171, doi:10.1109/TENCON.2018.8650524.

[13] Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur, Dr. Dayananda P, 2020, Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCAIT – 2020 (Volume 8 – Issue 15),

[14] Latif, Ghazanfar; Alghazo, Jaafar; Mohammad, Nazeeruddin; AlKhalaf, Roaa; AlKhalaf, Rawan (2018), "Arabic Alphabets Sign Language Dataset (ArASL)", Mendeley Data, V1, doi: 10.17632/y7pckrw6z2.1.

[15] Convolutional Neural Network. [online] Available at: https://en.wikipedia.org/wiki/Convolutional_neural_network.