

Sign Language Recognition System for Communicating to People With Disabilities

Yulius Obi

Faculty of Computer Science

BINA NUSANTARA
UNIVERSITY

Tangerang, Indonesia

yulius.obi@binus.ac.id

Kent Samuel Claudio

Faculty of Computer Science

BINA NUSANTARA
UNIVERSITY

Tangerang, Indonesia

kent.claudio@binus.ac.id

Vetri Marvel Budiman

Faculty of Computer Science

BINA NUSANTARA
UNIVERSITY

Tangerang, Indonesia

vetri.budiman@binus.ac.id

Abstract—Sign language is one of the most reliable ways of communication with disabled people, as it can be done anywhere. However, most people don't understand sign language. Therefore, we have come up with an idea to make an application that can recognize sign language and can convert it to text in real-time. In this research, we use American Sign Language (ASL) datasets and the Convolutional Neural Networks (CNN) classification system. In the classification, the hand image is first passed through a filter and after the filter is applied, the hand is passed through a classifier which predicts the class of the hand gestures. This research focuses on the accuracy of the recognition. Our Application resulted in 96,3% accuracy for the 26 letters of the alphabet.

Keywords—Computer Vision, Convolutional Neural Networks (CNN), American Sign Language (ASL), real-time, sign language recognition

I. INTRODUCTION

AI or Artificial Intelligence is one of the fields of computer science that studies human intelligence to make artificial intelligence capable of solving problems.

Computer vision is a subcategory of Artificial Intelligence (AI). The goal of computer vision is to extract useful information from images. However, it is very difficult to implement. Computer vision has been used to manufacture robots and photo scans and is also used in the automotive, medical, mathematical, and industrial fields [18][25].

Deaf people have problems communicating with normal people in their daily lives. One of the

reasons for that is there are not many people that understand American Sign Language (ASL) [10][12]. As such, the purpose of this research is the recognition of hand gesture or ASL, which the system will change into text that can be read in real-time, making communication with people that have special needs easier. Hand gesture recognition is also in Human-Computer Interaction (HCI) because it interacts with the user directly. Human-Computer Interaction (HCI) is a study, planning, or design of interaction between users and computers. One of the interactions that are usable for a hand gesture recognition system is displaying text that is composed of alphabets read by the system [1][9].

In this research, we will utilize Computer Vision and Pattern Recognition technology to detect hand movements in real-time using a webcam/live camera. Afterward, we will use American Sign Language (ASL) datasets and use the Convolutional Neural Networks (CNN) classification system. This research focuses on the accuracy of recognizing letters of the alphabet and provides the results in a text in real-time.

II. LITERATURE REVIEW

Currently, Human-computer interaction (HCI) is generally done using a mouse, keyboard, remote control, or touch screen. However, interpersonal communication is done naturally, such as through voice and physical movement, which is generally considered to be more flexible and efficient.[1] However, according to Zhi-hua Chen et al (2014), due to the rapid development of software and hardware, new types of HCI are needed. In

particular, speech recognition and gesture recognition have received great attention in the field of HCI [2]. To perform gesture recognition, Artificial Intelligence (AI) technology is also needed, specifically computer vision. In computer vision technology, there are several things that can be researched, one of which is real-time motion-based recognition. There are a variety of different ways that can be used to create recognition systems. The aim of research in this field is usually to increase the accuracy of the recognition performed and to perform several gestures recognition, such as hand movements, sign language, and body movements.

In general, recognition technology can recognize many things such as patterns, faces, body movements, or hand movements for different purposes. In 2013, there was a study on gesture recognition in real-time which already had a success rate of more than 68% for each gesture. This study uses an optical flow feature and is combined with a face detector [11]. However, Tarek Frikha and Abir Presentche were more interested in making hand gesture recognition [10]. To support research on hand gesture recognition, there was a study in 2018 that introduced a dataset and a benchmark called EgoGesture [18]. EgoGesture uses Hierarchical Hidden Markov Model and Classification methods, as well as using the Cambridge Hand Gesture dataset, and is able to produce a dataset that is very useful for research. One application of recognition technology is Automotive Human-Machine technology. In its field, there is hand gesture recognition to control applications on mobile tablets. This research uses ToF sensors, PCA-based preprocessing, and Convolutional Neural Network which gives satisfactory results for the riders [17].

Although many applications are made with recognition technology, the focus of this research is on the field of communication, especially sign language. Communication is the most important thing for every human being to be able to share their thoughts and ideas. Communication can be said to be successful if the communication partner receives and understands the message [19]. However, for people who have hearing difficulties or cannot speak, communication will be difficult so other forms of communication are needed [20], such as by writing or using body gestures. However, written communication is less practical because people with hearing problems are generally less proficient in writing spoken language. What's more, this type of communication is impersonal and slow in face-to-face conversation. For example, when there is an accident, agile communication with the doctor is often required, and written communication is not always possible [4]. Of these various forms of communication, sign language is the most effective tool of communicating [21].

Although sign language is the most effective, communication is sometimes difficult because not everyone understands sign language [3] [22]. Sign Language has different forms and hand gestures in each country, such as American Sign Language (ASL) that is used in a study by Shruti Chavan, Xinrui Yu and Jafar Saniie (2021) [14] and Indonesian Sign Language (ISL) that is used in a study conducted in 2021 [12]. Therefore, we need a tool that can recognize and convert sign language into understandable text with a high level of accuracy as well as easy to use [23].

To be able to recognize and change the sign language into understandable text, the device must be able to receive a photo or video input. The photo or video input is then processed before finally being classified. This photo or video processing can be done in several ways or steps, such as background blurring or removal [24] [25], edge detection [23] [25] [26], point cloud processing [27], skin-color detection [6] [27], converting Region of Interest (ROI) to grayscale image and blurring ROI with Gaussian blur, Contour-Extraction [6] [26], Data augmentation such as re-scaling, zooming, shearing, rotation, width and height-shifting also into the dataset [22][7], HSV colorspace, Global Threshold, Adaptive Gaussian Threshold [20], KAZE feature detection, KMeans algorithm [21], as well as converting BRG to RGB and converting RGB to thresholding [19]. After doing the image processing using these various methods, the results of the image processing are then classified to get the results. The classification can also be done in various ways, such as with Histogram matching algorithm [23], Nearest Neighbor [21] [27], SVM, NBC [21], Hidden Markov model [27], Extreme Learning Method (ELM)[9] and the most frequently used method is the Convolution neural network or CNN [19] [20] [22] [25] [8] [5].

Therefore, we are looking for references regarding suitable methods and algorithms to use. In 2019, there was research that discussed the best out of these 3 methods, the methods being Wavelet Transforms method, Empirical Mode Decomposition method and Convolutional Neural Networks method which gave the result that the CNN method was more accurate but had relatively high memory usage [13]. Furthermore, in 2019 there was also research about Indonesian Language Recognition (ISL) using the CNN method and the YOLOv3 architecture which also uses the ISL dataset. The research gave an accuracy of 100% for images and 73% for videos [15]. Then in the following year, the Machine Learning method using OpenCV and TensorFlow gave positive results and was supported with a different dataset from other research [16].

The conclusion that we can convey is that to facilitate communication with people who have

hearing difficulties or people who cannot speak, a Human-computer interaction (HCI) media is needed. So, at our current times, a sign language recognition system is needed to make the communication between deaf people and normal people easier. Research on sign language recognition has been carried out by researchers using different methods. In our opinion, the topic is very interesting and the results make a positive impact on society. Therefore, this research will focus on making a sign language recognition system that uses the American Sign Languages (ASL) dataset with the Convolutional Neural Network (CNN) method which is easy to understand and has a high level of accuracy.

III. METHODOLOGY

The steps required to run the language recognition model are generally divided into several sections. First, the application will enter input from the user's live-camera, then the input will be read and the results of the photo will be displayed as characters which will finally be assigned to a word. We have included our flowchart for our research process as shown below.

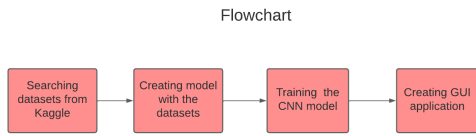


Figure 1. Research process

A. Datasets

The dataset that we will use is a dataset that we took from the Kaggle website entitled “ASL Hand Sign Dataset (Grayscaled & Thresholded) [28]” which contains 24 classes that have been applied a gaussian blur filter. However, because the source code we use requires 27 classes, we use the 3 dataset classes taken from Nikhil Gupta, namely the datasets for classes J, Z, and 0 (blank), to fill the gaps in the datasets in Kaggle.

Compared to the datasets from the code we used [29], our datasets have a higher number of images. The images that we use are 30526 images which are divided into 27 classes for training data and 8958 images which are divided into 27 classes for testing data, while the datasets from the source code that we use only have 12845 images which are divided into 27 classes for training data and 4268 images which are divided into 27 classes for testing data.

B. Implementing Algorithm

This research uses a dataset obtained from Kaggle [28] because it has a larger database than Nikhil Gupta's. By using Kaggle's dataset, we are able to analyze the differences from Nikhil Gupta's research caused by having a different dataset. After that, we used that data set for training and testing on the model created by Nikhil Gupta.

This model uses the Convolutional Neural Network (CNN) method for Gesture Classification which consists of 2 layers. At Layer 1, this model performs image processing to predict and detect the number of frames of user input. Layer 1 consists of the CNN model in which there are 7 layers (1st Convolution Layer, 1st Pooling Layer, 2nd Convolution Layer, 2nd Pooling Layer, 1st Densely Connected Layer, 2nd Densely Connected Layer, and Final Layer), Activation Function, Pooling Layer, Dropout Layer, and Optimizer. In Layer 2, there are 2 layers of algorithms that check and predict symbols or letters that look similar to each other, so that they can detect and display an accurate letter. In addition, our model uses the Hunspell library to implement the autocorrect feature, where the user can click on one of the application's 3 suggestions to form a word. We have included our flowchart for our gesture classification process as shown below.

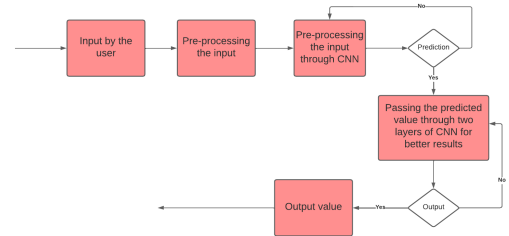


Figure 2. Gesture classification process

After the model was modified and trained using the new datasets, we used the Sign Language To Text Converter application which was created by Nikhil Gupta. In this application, the hand image captured by the camera is directly converted to grayscale and then processed using gaussian blur and also adaptive threshold. Then, the image will be processed by the model that has been made to make predictions and display the output character. Next, the predicted result or the letter must remain the same for a few seconds to be combined into a word. We also modified the GUI of the application by adding an image containing a list of ASL symbols to make it easier for the user. The modified GUI (Graphical User Interface) of the application can be seen in the images below.

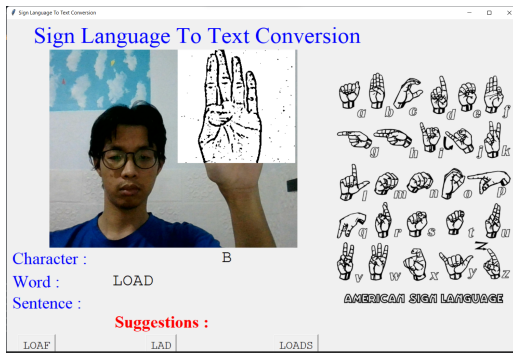


Figure 3. Application GUI

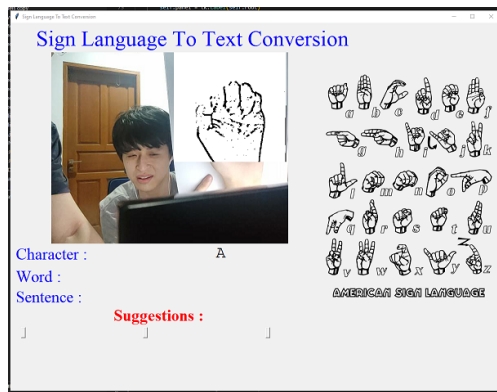


Figure 4. Application GUI

C. Evaluate

Performance and accuracy measurements will be carried out by looking at the results of the train data and test data using epochs 20 times. The iteration produces accuracy values and loss values for train data and test data so it can find out the values that indicate the level of success and loss values.

In conducting trials and evaluations of the application that has been made, we use a confusion matrix to calculate the model's performance and accuracy. The data used for the confusion matrix is a dataset created by Nikhil Gupta in his research.

Certain parameters or variables that may affect application performance when executing the application are as such.

- Device 1's specification used for testing the application has an i5-9300H processor, 1650Ti VGA, 8GB RAM, 1TB HDD and 256GB SSD, Windows 11 Operating System, and HD WEBCAM LAPTOP
- Device 2's specification used for testing the application has an i7-4790H Processor, NVidia GeForce GTX 970 VGA, 16GB RAM, 1TB HDD and 128GB SSD,

Windows 10 Operating System, and a POCOPHONE F1 smartphone camera

- Background complexity in testing the application.

IV. RESULT AND DISCUSSION

By using a different training and testing dataset and number of epochs from the research by Nikhil Gupta, we produced a model with training and validation accuracy of 89,1% and 98,6% respectively, as well as training and validation errors of 35,5% and 5,3% respectively. We have visualized the accuracy and the loss from our training and validation model of our research algorithm in the form of graphs as shown below.

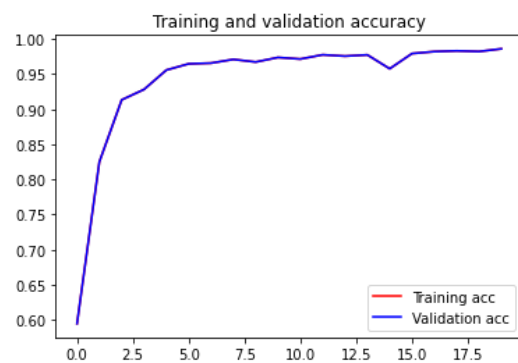


Figure 5. Training dan validation accuracy

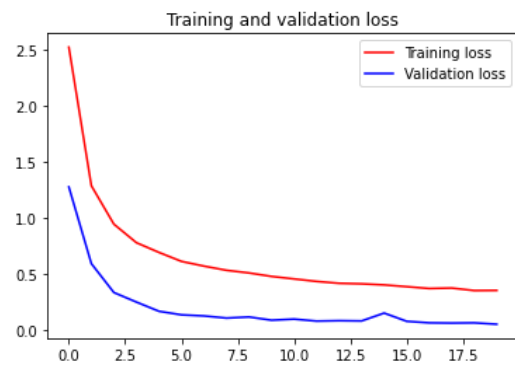


Figure 6. Training dan validation loss

The results of our model compared with the one by Nikhil Gupta are very different, namely, our model has a greater loss value in training and validation, having 35,5% loss value compared to Nikhil Gupta's 3.7% in training, and 5,3% loss value compared Nikhil Gupta's 0.18% on validation. Our model also has a lower accuracy value in training and validation than Nikhil Gupta's model, having 89,1% accuracy compared to Nikhil Gupta's 99% in training and 98,6% accuracy compared to Nikhil Gupta's 99.9% in validation. This difference in results could be influenced by the number of epochs performed for different

models and the difference in the amount of data in the dataset.

In testing and evaluating the model, we use a confusion matrix created by testing using a dataset created by Nikhil Gupta's research because it has the same number of classes, the same type of data, and a smaller amount of data. The accuracy resulting from the confusion matrix is 96,3%. Just like the research conducted by Nikhil Gupta, we did not implement a background remover, so when running the application we need a background that does not have objects on the box that reads hand gestures. In addition, good lighting is required when running the application. In testing the application, to test the "blank" gesture (a box to read hand gestures must be empty with nothing in it), our application requires a completely empty background condition. We have provided an image of the confusion matrix for our results below.

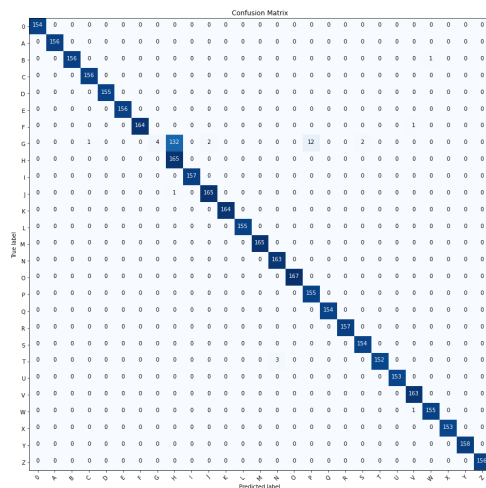


Figure 7. Confusion matrix

V. CONCLUSION

In this study, the steps we took were finding a dataset from the Kaggle website, creating a new model using the 2 layers Convolutional Neural Network (CNN) method using the dataset we got, training our CNN model, and creating a GUI for our application. Our application can read hand gestures in real time and combine them into a word with a final accuracy rate of 96,3%. However, the process of forming letters into a word requires the letters to remain the same for a few seconds. Therefore, for future research, we suggest implementing background removal methods, searching for a way to speed up the process of forming letters into a word so the wait time will be shortened, and multiplying the layers in the CNN model so the accuracy of the model will be greater. One might also reconsider using other methods other than CNN, as it might lead to better results.

REFERENCES

- [1] Q. Zhang, S. Xiao, Z. Yu, H. Zheng, and P. Wang, "Hand gesture recognition algorithm combining hand-type adaptive algorithm and effective-area ratio for efficient edge computing," 2021, doi: 10.1117/1.JEI.30.
- [2] Z. H. Chen, J. T. Kim, J. Liang, J. Zhang, and Y. B. Yuan, "Real-time hand gesture recognition using finger segmentation," *Scientific World Journal*, vol. 2014, 2014, doi: 10.1155/2014/267872.
- [3] R. A. Kadhim and M. Khamees, "A real-time american sign language recognition system using convolutional neural network for real datasets," *TEM Journal*, vol. 9, no. 3, pp. 937–943, Aug. 2020, doi: 10.18421/TEM93-14.
- [4] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign Language Recognition using Convolutional Neural Networks."
- [5] Vazquez Lopez, "HAND GESTURE RECOGNITION FOR SIGN LANGUAGE TRANSCRIPTION," 2017.
- [6] H. Huang, Y. Chong, C. Nie, and S. Pan, "Hand Gesture Recognition with Skin Detection and Deep Learning Method," in *Journal of Physics: Conference Series*, Jun. 2019, vol. 1213, no. 2, doi: 10.1088/1742-6596/1213/2/022001.
- [7] M. Z. Islam, M. S. Hossain, R. U. Islam, and K. Andersson, "Static Hand Gesture Recognition using Convolutional Neural Network with Data Augmentation."
- [8] E. Padmalatha, S. Sailekya, R. Ravinder Reddy, C. Anil Krishna, and K. Divyarsha, "Sign language recognition," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 2128–2137, Sep. 2019, doi: 10.35940/ijrte.C4565.098319.
- [9] A. Miao and F. Liu, "Application of human motion recognition technology in extreme learning machine," *International Journal of Advanced Robotic Systems*, vol. 18, no. 1, 2021, doi: 10.1177/1729881420983219.
- [10] A. Zeineb, C. Chaala, T. Frikha, and A. Hadriche, "HAND GESTURE RECOGNITION SYSTEM." [Online]. Available: <https://sites.google.com/site/ijcsis/>
- [11] M. Bayazit, A. Couture-Beil, and G. Mori, "Real-time Motion-based Gesture Recognition using the GPU." *MVA*. 2013.
- [12] Mulyanti, B., Widiaty, I., & Yuni Wulandari, I. (2021). INDONESIAN SIGN LANGUAGE CONVERTER INTO TEXT AND VOICE AS SOCIAL INTERACTION TOOL FOR INCLUSION STUDENT IN VOCATIONAL HIGH SCHOOLS. In *Journal of Engineering Science and Technology Special Issue on AASSEECC2021*.
- [13] N. Alnaim and M. Abbod, "Hand gesture detection using neural networks algorithms," *International Journal of Machine Learning and Computing*, vol. 9, no. 6, pp. 782–787, 2019, doi: 10.18178/ijmlc.2019.9.6.873.

- [14] S. Chavan, X. Yu, and J. Saniie, "Convolutional Neural Network Hand Gesture Recognition for American Sign Language." [Online]. Available: <http://ecasp.ece.iit.edu/>
- [15] S. Daniels, N. Suciati, and C. Fathichah, "Indonesian Sign Language Recognition using YOLO Method," IOP Conference Series: Materials Science and Engineering, vol. 1077, no. 1, p. 012029, Feb. 2021, doi: 10.1088/1757-899x/1077/1/012029.
- [16] A. B., K. Krishi, M. M., M. Daaniyaal, and A. H. S., "Hand gesture recognition using machine learning algorithms," Computer Science and Information Technologies, vol. 1, no. 3, pp. 116–120, Nov. 2020, doi: 10.11591/csit.v1i3.p116-120.
- [17] N. Zengeler, T. Kopinski, and U. Handmann, "Hand gesture recognition in automotive human-machine interaction using depth cameras," Sensors (Switzerland), vol. 19, no. 1. MDPI AG, Jan. 01, 2019. doi: 10.3390/s19010059.
- [18] Y. Zhang, C. Cao, J. Cheng, and H. Lu, "EgoGesture: A New Dataset and Benchmark for Egocentric Hand Gesture Recognition," IEEE Transactions on Multimedia, vol. 20, no. 5, pp. 1038–1050, May 2018, doi: 10.1109/TMM.2018.2808769.
- [19] S. Khadar Sharif, V. Vjiet, C. Sri Varshini, G. Sreekanth, G. Hruday, and M. Chandu, "Sign Language Recognition." [Online]. Available: www.ijert.org
- [20] I Putu Iduar Perdana, I Ketut Gede Darma Putra, I Putu Arya Dharmadi, "Classification of Sign Language Numbers Using the CNN Method."
- [21] A. Bain, S. Birajdar, and S. Manonmani, "Sign language Recognition System using Machine Learning," International Research Journal of Engineering and Technology, 2021, [Online]. Available: www.irjet.net
- [22] A. H. Al-Obodi, A. M. Al-Hanine, K. N. Al-Harbi, M. S. Al-Dawas, and A. A. Al-Shargabi, "A saudi sign language recognition system based on convolutional neural networks," International Journal of Engineering Research and Technology, vol. 13, no. 11, pp. 3328–3334, 2020, doi: 10.37624/ijert/13.11.2020.3328-3334.
- [23] S. Rajan, R. Nagarajan, A. Kumar Sahoo, M. Gowtham Sethupati, "Interpretation and Translation of American Sign Language for Hearing Impaired Individuals using Image Processing," International Journal of Recent Technology and Engineering, vol. 8, no. 4, pp. 415–420, Nov. 2019, doi: 10.35940/ijrte.d6966.118419.
- [24] D. Zhang, D. J. Lee, and Y. P. Chang, "A new profile shape matching stereovision algorithm for real-time human pose and hand gesture recognition," International Journal of Advanced Robotic Systems, vol. 11, no. 1, Feb. 2014, doi: 10.5772/57515.
- [25] S. Juneja, A. Juneja, G. Dhiman, S. Jain, A. Dhankhar, and S. Kautish, "Computer Vision-Enabled Character Recognition of Hand Gestures for Patients with Hearing and Speaking Disability," Mobile Information Systems, vol. 2021, 2021, doi: 10.1155/2021/4912486.
- [26] A. K. Hamed AlSaedi and A. H. Hassin AlAsadi, "A new hand gestures recognition system," Indonesian Journal of Electrical Engineering and Computer Science, vol. 18, no. 1, pp. 49–55, 2019, doi: 10.11591/ijeecs.v18.i1.pp49-55.
- [27] T. Kapuscinski, M. Oszust, M. Wysocki, and D. Warchol, "Recognition of hand gestures observed by depth cameras," International Journal of Advanced Robotic Systems, vol. 12, Apr. 2015, doi: 10.5772/60091.
- [28] ASL Handsign Dataset (Grayscaled & Thresholded). (2022). Furkan Akdeniz. <https://www.kaggle.com/datasets/furkanakdeniz/asl-hand-sign-dataset-grayscaled-thresholded?resource=download-directory&select=asl-dataset>
- [29] Sign Language To Text Conversion.(2021). Nikhil Gupta <https://github.com/emnikhil/Sign-Language-To-Text-Conversion>