HAND GESTURE RECOGNITION USING CONVOLUATIONAL NEURAL NETWORK

Harikrishnan H
Dept. of Computer Science & Eng.
Amal Jyothi College of Engineering
Kottayam, India
harikrishnanh2023@cs.ajce.in

Joel Johny
Dept. of Computer Science & Eng.
Amal Jyothi College of Engineering
Kottayam, India
joeljohny2023@cs.ajce.in

Marshel C Mathews
Dept. of Computer Science & Eng.
Amal Jyothi College of Engineering
Kottayam, India
marshelcmathews2023@cs.ajce.in

Meghana Suresh
Dept. of Computer Science & Eng.
Amal Jyothi College of Engineering
Kottayam, India
meghanasuresh2023@cs.ajce.in

Honey Joseph
Dept. of Computer Science & Eng.
Amal Jyothi College of Engineering
Kottayam, India
honeyjoseph@amaljyothi.ac.in

***Abstract*—In this paper, a system for enhancing communication between sign language users and non-users is presented. The system recognizes and converts sign language into text and voice using picture data collected by a computer webcam. A masking approach is used to preprocess the picture data, then a Convolutional Neural Network (CNN) algorithm is used for feature extraction and classification. The system was trained using a dataset of masked images representing the 26 letters of the English alphabet, totaling 45500 images for training and 6500 images for testing. The system's goal is to make it simpler for those who use sign language and those who do not to communicate with one another. Keywords: Hand Gesture Recognition, Feature Extraction, Pre-Processing, Classification, CNN.**

Keywords: Feature Extraction, Pre Processing, Classification, CNN

I. INTRODUCTION

The advancement of technology has brought a great deal of convenience to our lives, and it has the potential to help individuals with speech impairments communicate more effectively as well. The system we are proposing converts sign language into the English alphabet, making it easier for those who do not understand sign language to comprehend the message being conveyed.

Sign language interpreters can provide a valuable service to the deaf and hard-of-hearing community, but they are not always available, and even when they are, there is always a risk of miscommunication. Computerized interpreters can provide a more reliable alternative, ensuring accuracy in communication at a much lower cost. This system, which converts signs into text or audio, can be very helpful in real-time applications, such as business meetings or social gatherings.

The video capturing aspect of the system is processed using a Convolutional Neural Network (CNN) algorithm, which maps and summarizes features through its internal three layers. The algorithm then forms classes based on these features, and predicts the image, giving the output. This internal function of the algorithm ensures accuracy and reliability in the system, providing a better overall communication experience for people with speech impairments.red output file.

II. LITERATURE SURVEY

In the paper Finger spelling in sign language is recognized using convolutional neural network (CNN) architecture in the paper Automated sign language finger spelling using convolution neural network by Beena, M. V., Namboodiri, M. A., Dean, P. The authors explain the importance of finger spelling in sign language communication and the need for an efficient finger spelling recognition system.

The paper presents an approach to automatically recognize finger spelling in sign language using convolutional neural network (CNN) architecture. The authors explain the importance of finger spelling in sign language communication and the need for an efficient finger spelling recognition system.

The proposed approach involves preprocessing the input image to obtain the region of interest (ROI) containing the hand, and then extracting the hand shape and finger tip locations using contour analysis. The extracted features are fed into a CNN architecture consisting of convolutional layers, pooling layers, and fully connected layers to classify the input as a specific finger spelling. The authors conducted experiments on two datasets - an Indian sign language dataset and a publicly available American sign language dataset. The proposed approach achieved an accuracy of 98.5 percent and 99.7 percent on the Indian and American sign language datasets, respectively.

The authors conclude that the proposed approach using CNN architecture is an effective method for automatic finger spelling recognition in sign language. They also discuss the potential applications of the proposed approach, such as in aiding communication for people with hearing and speech impairments.

In the paper, Sign language converter by Arsan, T., Ülgen, O. The authors introduce the concept of a sign language converter, which is a system that converts sign language into text or speech to facilitate communication between people with hearing and speech impairments and the general population. The paper presents an overview of existing sign language recognition and conversion systems and proposes a new approach to improve the accuracy and efficiency of such systems.

The authors discuss various techniques used for sign language recognition, such as data glove-based systems, camera-based systems, and marker-based systems. They also highlight the challenges involved in sign language recognition, such as variations in sign language across different regions and individuals, and the need for robust feature extraction and classification methods.

The authors propose a new approach that combines both camera-based and data glove-based systems for sign language recognition. The system uses a camera to capture the hand gestures and a data glove to capture the finger movements, which are then combined to improve the accuracy of recognition. The authors also propose a new feature extraction method based on histogram of oriented gradients (HOG) to extract relevant features from the hand gestures.

The authors discuss various techniques used for sign language conversion, such as speech synthesis and text generation. They also propose a new approach that uses a rule-based method to generate text from the recognized sign language gestures. The proposed approach uses a hierarchical set of rules to translate the sign language gestures into corresponding text phrases.

The authors conducted experiments on a dataset of 20 sign language gestures and achieved an average recognition accuracy of 97.5 percent. They also evaluated the text generation performance of the proposed approach and found that it achieved an accuracy of 92.5 percent.

The authors conclude that the proposed sign language converter system is an effective method for improving communication between people with hearing and speech impairments and the general population. They also discuss the potential applications of the proposed system, such as in education and healthcare.

Works of Pigou, L., Dieleman, S., Kindermans, P. J., Schrauw, presents a novel approach to sign language recognition using convolutional neural networks (CNNs). The authors highlight the importance of sign language recognition for facilitating communication between people with hearing and speech impairments and the general population. They also discuss the limitations of existing sign language recognition systems and propose a new approach to improve their accuracy and efficiency.

The proposed approach involves preprocessing the input image to obtain the region of interest (ROI) containing the hand, and then extracting relevant features from the hand using CNNs. The CNN architecture consists of convolutional layers, pooling layers, and fully connected layers. The authors also propose a new method for data augmentation to increase the size of the training dataset.

In the experiments conducted on two publicly available datasets - the RWTH-BOSTON-104 dataset and the RWTH-BOSTON-50 dataset. The proposed approach achieved an accuracy of 96.3 percent and 98.0 percent on the two datasets, respectively. The authors also compared the performance of their approach with other state-of-the-art methods and found that it outperformed them in terms of accuracy and efficiency.

In the paper The authors conclude that the proposed approach using CNNs is an effective method for sign language recognition. They also discuss the potential applications of the proposed approach, such as in improving communication for people with hearing and speech impairments. The authors highlight the limitations of the proposed approach, such as the need for a large training dataset and the sensitivity of the method to hand pose and lighting conditions. They suggest future research directions to address these limitations.

The paper presents a study on hand gesture recognition using deep convolutional neural networks. The authors explain the importance of hand gesture recognition and its potential applications in various fields such as human-computer interaction, sign language recognition, and robotics.

The authors review the literature on hand gesture recognition and explain the limitations of previous approaches. They also describe the advantages of using deep convolutional neural networks for this task.

The authors collected a dataset of hand gestures using a Microsoft Kinect sensor. The dataset contains 11 hand gesture classes, and each class has 1,000 samples. The authors also explain the preprocessing steps they applied to the data.

The authors used a deep convolutional neural network (CNN) to classify hand gestures. They explain the architecture of their CNN, which consists of several convolutional and pooling layers followed by fully connected layers. They also describe the training process and the hyperparameters they used.

The authors report the performance of their CNN on the hand gesture recognition task. They achieved an accuracy of 99.1 percent on their test set, which outperforms previous approaches on this dataset. They also provide a confusion matrix to show the classification results for each gesture class.

The authors discuss the strengths and limitations of their approach. They also compare their results with previous studies and highlight the advantages of using deep CNNs for hand gesture recognition.

The authors conclude that their deep CNN approach is effective for hand gesture recognition and can be used in various applications. They also suggest some directions for future research, such as exploring different CNN architectures and investigating the transfer learning approach.

III. Proposed Method

This section describes the details of the proposed method. The proposed system can translate sign language to text and then to audio,thus can improve communication with sign language.

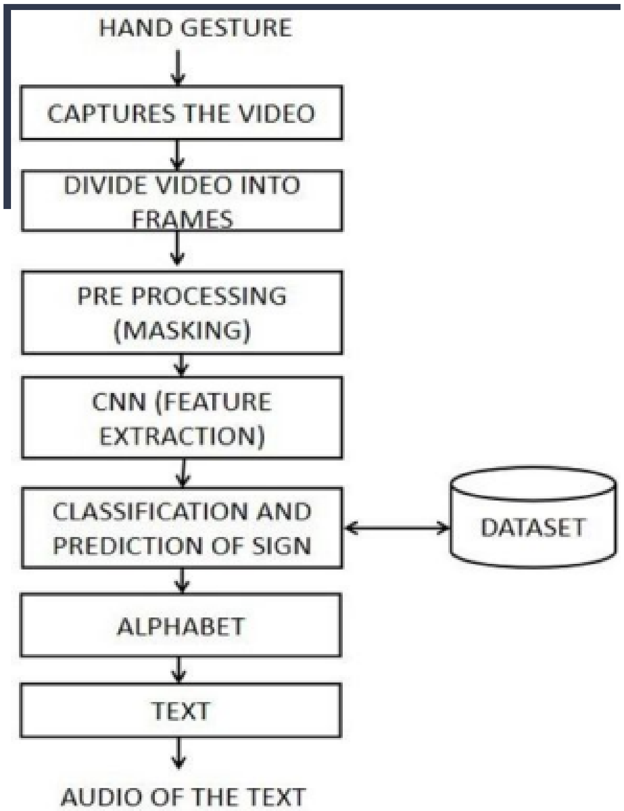

Fig. 1. BLOCK DIAGRAM OF PROPOSED METHOD

A. *Dataset*

The dataset in this study is made up entirely of the English alphabet, and the sign language used is American Sign Language (ASL). The dataset is divided into training and testing sets before the model is trained, with 45500 images used for training and 6500 images used for testing. Each alphabet has 1750 photos in the testing set. The model is trained using the supervised learning method, and then tested using the testing data set to determine its accuracy. To create masked photos, the images are pre-processed using the masking technique.

B. *Functional Components*

1) Hand Gesture: Hand gestures are non-verbal actions performed with the hands and fingers that convey meaning or emotion. They are often used to complement or emphasize spoken language, or to communicate when words are not enough or not possible.

Hand gestures can be intentional or unconscious, and they vary in meaning and significance across cultures. Some gestures are universal, such as a wave goodbye, while others are more specific to certain cultures or contexts.

2) Pre-processing and Masking: Pre-processing involves preparing the input data to be suitable for feeding into the CNN. This may include tasks such as normalization, resizing, and data augmentation. Normalization involves scaling the input pixel values to a fixed range, typically between 0 and 1, to help the model converge faster during training. Resizing involves resizing the input images to a fixed size, which is important to ensure that all images have the same dimensions and to reduce the computational load. Data augmentation involves generating new training examples by applying transformations such as rotation, scaling, and flipping to the original images. This helps the model learn to be more robust to variations in the input data.

Masking involves selectively ignoring certain parts of the input data during training or inference. This is particularly useful in image segmentation tasks, where the goal is to identify the boundaries of objects in an image. Masking can be used to focus the model's attention on the regions of the image that are relevant for the task, while ignoring irrelevant regions such as the background. Masking can also be used to handle missing or corrupted data in the input, by selectively ignoring the affected regions and focusing on the remaining data.

3) Feature Extraction: Convolutional Neural Networks (CNNs) are commonly used for feature extraction in computer vision applications, including image classification, object detection, and hand gesture recognition.

CNN can be used to extract features from images of hands making different gestures. The CNN architecture typically consists of multiple layers, including convolutional layers, pooling layers, and fully connected layers.

The features of the input images are learned by the convolutional layers. They extract features like edges, corners, and textures from the input image by applying a set of teachable filters. The pooling layers preserve the most crucial data while reducing the spatial dimensions of the feature maps created by the convolutional layers. The flattened feature maps from the pooling layers are input to the fully connected layers, which then do the final classification into various hand movements.

The process of using a pre-trained CNN for feature extraction involves removing the final classification layer from the network and using the output of the last pooling layer as the feature representation for the input image. This approach is known as transfer learning, and it allows the features learned by the CNN on a large dataset to be reused for a different task with a smaller dataset, such as hand gesture recognition.

IV. CONCLUSION

This project's primary goal is to bridge the communication gap between smart and regular people. Here, we use the American sign languages to train the convolutional neural network. In this Hand Gesture Recognition paper, we'll use Python and OpenCV to build a hand gesture recognizer. For the purposes of gesture identification and detection, we employ the MediaPipe and Tensorflow frameworks, respectively. The Convolutional Neural Network is then used to implement and train it using American Sign Language. This system enables us to make a gesture, which it can then recognize, and it can also construct and voice out a sentence. Getting rid of the undesirable image backdrop and noise that is captured in the area of interest is the most challenging task.

References

[1] Beena, M. V., Namboodiri, M. A., Dean, P. G. (2017)., "Automatic sign language finger spelling using convolution neural network." IEEE TRANSACTIONS ON Nanalysis.

[2] Arsan, T., Ülgen, O. (2015), " Sign language converter. International Journal of Computer Science Engineering Survey (IJCSES), 6(4), 39-51.

[3] N. Muthukumaran, "Analyzing Throughput of MANET with Reduced Packet Loss', Wireless Personal Communications, Vol. 97, No. 1, pp. 565-578, November 2017."

[4] N. Muthukumaran, R. Aiswarya, S. Anna Sankari K. Divya Bharathi, "Deep Learning Neural Network Based Human Emotion Classification with ANFIS Algorithm, Irish Interdisciplinary Journal of Science Research, Vol.4, Iss.3, Pages 105-111, JulySeptember 2020."

[5] Grandhi, C., Liu, S., Rahoria 'American Sign Language Recognition using Deep Learning."

[6] VP. Anubala, N. Muthukumaran and R. Nikitha "Performance Analysis of Hookworm Detection using Deep Convolutional Neural Network', 2018 International Conference on Smart Systems and Inventive Technology, pp." J. Springer-Verlag London Ltd., part of Springer Nature 2020
García-Hernández, Y. Ledeneva and C. E. Millán-Hernández, "Extractive Automatic Text Summarization Based on Lexical-Semantic Keywords," in IEEE Access, vol. 8, pp. 49896-49907, 2020.

[7] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan and M. M. Kabir, "A Survey of Automatic Text Summarization: Progress, Process and Challenges," in IEEE Access, vol. 9, pp. 156043-156070, 2021.