IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105110

AMERICAN SIGN LANGUAGE RECOGNITION WITH CONVOLUTIONAL NEURAL NETWORKS

Reva Sirdeshpande, Sakshi Khese, Veena Bhosale, Pratiksha Keswani

Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Karvenagar, Pune, India.

Abstract: The only way speech and hearing-impaired people can communicate is by using the sign language. The main problem with this kind of communication is that the non-impaired people, who cannot understand the sign language, would not be able to communicate with these people or vice versa. This thesis is about intentionally researching a algorithm which can to allow deaf and dumb communities to communicate effectively. Thus, this study is also aimed to extract features from finger and hand motions. This paper shows the American sign language which recognizes 26 hand gestures. The system contains four modules such as: pre-processing and hand segmentation, feature extraction, sign recognition, and sign to text and audio. The project uses an image processing system to identify, especially English alphabetic sign language, and convert it into text. The proposed model takes image sequences and extracts temporal and spatial features from them. We then use a CNN (Convolution Neural Network) for recognizing spatial features.

Keywords: Convolution Neural Networks (CNN), Machine learning, Sign Recognition, Rectified Linear Unit (ReLU)

I. INTRODUCTION

American Sign Language (ASL) substantially facilitates communication in the deaf community. There are only 250,000-500,000 people who can understand American Sign Language (ASL). This significantly limits the number of people that they can comfortably communicate with. The alternative of written communication is cumbersome, impersonal, and even impractical when an emergency occurs. To diminish this obstacle and to enable communication, we present an ASL recognition system that uses Convolution Neural Networks (CNN) in real-time to translate a gesture of a user as ASL signs into text. Sign language is a type of language that uses manual communication to convey meaningful messages to other people.

American Sign Language Recognition can be performed using two approaches in the domain of Computer Vision, i.e using image processing, and video processing. From the known set of alphabets and numbers, 34 gestures (A-Z, except 'J' and 'Z', and 0-9) are static and can be easily recognized using image processing.

Deep Neural Network is prominently used for the classification and recognition of gestures in video processing as it is useful to solve classification obstacles having complex data, and the work involves unlabeled/unstructured data. Based on the classification result achieved, the text associated with the input gesture is displayed.

Our approach first converts the video into frames and then pre-processes the frames to convert them into grey-scale images. Then Convolution Neural Network (CNN) classifier is used to build the classification model which then classifies the frames into 26 different classes representing 26 English alphabets. Lastly, based on the properties calculated during the earlier phase, transliteration of signed gestures into text is carried out. Consequently, gestures will be recognized through training on various images of different signers signing each sign.

Consequently, gestures will be recognized through training on various images of different signers signing each sign.

II. RELATED WORK

Two methods of classification for prediction in the final layer can be used, i.e. the pool layer, and the soft-max layer[2]. The study also discussed concepts and use of RNN, and CNN for image classification. It discusses an efficient approach for the sign gestures of the English alphabets' set. The letters 'J' and 'Z' have been excluded, and the transliteration of the rest of the 24 alphabets, along with the numbers are performed[3]. It summarizes the use of Deep Learning in recognizing the sign language gestures. The dataset collected by Massey University, Institute of Information and Mathematical Sciences, in 2012, has been used. The dataset has been given to the Convolutional Neural Network for training[4].

The difference between the average recognition rate of the proposed technique against multiple existing techniques has been graphically demonstrated[3]. The impact of various background factors on the average recognition rate of the proposed system are demonstrated[5].In this paper, continuous sign recognition framework and LS-HAN have been proposed. Both these novels are used to eliminate pre-processing of temporal segmentation. Also, the proposed LS-

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105110

HAN is divided into three components which are two-stream CNN which is used for video feature representation generation, a hierarchical attention network used for latent space-based recognition and lastly the Latent space beneficial for semantic gap bridging[6].A real time HGR system based on American Sign Language (ASL) recognition is proposed which has better accuracy. Required gesture images are obtained from mobile's camera. The processing phase includes extraction of five features which are eccentricity, fingertip finder, elongated ness, rotation, and segmentation[7]. Multi-Modality American Sign Language Recognition proposes a system that captures ASL videos using a Kinect depth sensor. These captured ASL videos are used for the process of learning. The prediction of the ASL components is done using new videos of input[8]. Lenet achieved a great result and showed the potential of CNN, the development in this area due to limited computing power and the amount of the data. CNN can only solve some easy tasks such as digit recognition, but for more complex features like faces and objects, a HarrCascade or SIFT feature, SVM classifier was a more preferred approach[9].

III. PROPOSED METHOD

In this proposed work, an effort has been placed to recognize ASL Alphabets, which mainly depends on one hand and fingers. The process of identifying ASL alphabets is distributed in phases such as preprocessing the input image, computation of the region properties of the preprocessed image, and transliteration from treated image to text.

The four basic phases of the functioning of the neural network are convolutional, activation, pooling, and fully connected layers.

1. Image feature extraction.

The convolutional layer is made up of neurons with learnable weights and biases. Each neuron receives several inputs, takes a weighted sum over them, and passes it through an activation function and responds with an output. In the first layer, basic features such as borders or edges are filtered by the neurons. Each convolution kernel can extract features across the entire input plane, but usually the neurons are used to handle different regions of the image plane to create the feature patterns over the identical size of receptive field.

2. Neuron activation using ReLU, and image stack reduction using Max Pooling. Relu also known as rectified Linear Unit is an activation function. Relu transforms functions, along with that it is also used for smoothening the inputs from the convolution layer. The smoothened inputs reduce the sensitivity of the filters to variations and noises. Relu is beneficial for the system as it does not activate all the neurons together. The reason why Relu is used in CNN is because it makes the nonlinear images in the data set linear. The softmax function is another type of Activation Function used in neural networks to compute probability distribution from a vector of real numbers. This function generates an output that ranges between values 0 and 1 and with the sum of the probabilities being equal to 1. The softmax function is represented as follows:

This function is largely used in multi-class models. It returns the probability of each class. The target class has the highest probability. It appears in almost all the output layers of the

deep learning architecture where they are used. The primary difference between the sigmoid and softmax Activation Function is that while the former is used in binary classification, the latter is used for multivariate classification.

3. Classification of the Input Gesture.

In most neural network models, the final layers are fully connected layers. These layers compile the input from the previous layers i.e., convolutional or pooling layers. This input is flattened before passing it to the fully connected layer. It is the final layer where the classification actually happens. Here, we take the filtered images and reduce them to form one single list.

IV. CONCLUSION

It is an approach targeted to solve the issues encountered by people with hearing and speech disabilities. It comprises two major components, i.e. analyzing the gestures from the images, and classifying the gestures from the images. We have investigated two methods for classification: using the Max pooling layer and using the SoftMax layer for the final predictions.

REFERENCES

[1] [1]. Shivashankara S, Srinath S, American Sign Language Video Hand Gestures Recognition using Deep Neural Networks, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249- 8958, Volume-8 Issue-5, June 2019.

[2] [2]. Bantupalli, Kshitij and Xie, Ying, “American Sign Language Recognition Using Machine Learning and Computer Vision” (2019). Master of Science in Computer Science Theses.

[3] [3]. Shivashankara S, Srinath S, “ American Sign Language Recognition System: An Optimal Approach ”, International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.10, No.8, pp. 18-30, 2018.DOI: 10.5815/ijigsp.2018.08.03.

[4] [4]. Taskiran, Murat KAallA¨ aoA¨ lu, Mehmet Kahraman, Nihan. (2018).A Real-Time System For Recognition of American Sign Language by Using Deep Learning. 10.1109/TSP.2018.8441304.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105110

[5] [5]. S. Shivashankara, S. Srinath , aAn American Sign Language Recog-nition System using Bounding Box and Palm FEATURES Extraction Techniquesˆa, International Journal of Recent Technology and Engineer- ing (IJRTE) ISSN: 2277-3878, Volume-7 Issue-4S, November 2018.

[6] [6]. Md. Mohiminul Islam, Sarah Siddiqua, and Jawata Afnan, "Real Time Hand Gesture Recognition Using Different Algorithms Based on American Sign Language", IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR), 2017.

[7] [7]. Chenyang Zhang, Yingli Tian, Matt Huenerfauth, "Multi-Modality American Sign Language Recognition", IEEE International Conference on Image Processing (ICIP), pp.2881-2885, 2016.

[8] [8]. Srinath S, Ganesh Krishna Sharma, "Classification approach for Sign Language Recognition", International Conference on Signal, Image Processing, Communication & Automation, 2017.

[9] [9]. Linqin, C. Shuangjie and X. Min, "Dynamic hand gesture recognition using RGB-D data for natural human-computer interaction," Journal of Intelligent and Fuzzy Systems, 2017.

[10] [10]. F. Ronchetti, Q. Facundo and A. E. Cesar, "Handshake recognition for argentinian sign language using probsom," Journal of Computer Science & Technology, 2016.

[11] [11]. Ashok K Sahoo, Gouri Shankar Mishra and Kiran Kumar Ravulakollu (2014), "Sign Language Recognition: State of the Art." ARPN Journal of Engineering and Applied Sciences 9(2).

[12] [12]. Alice Caplier, Sebastien Stillittano, Oya Aran, Lale Akarun, Gerard Bailly, Denis Beautemps, Nouredine Aboutabit and Thomas Burger (2007)."'Image and Video for Hearing Impaired People."' EURASIP Journal on Image and Video Processing.

[13] [13]. Siddharth S. Rautaray and Anupam Agrawal (2015), "Vision based Hand Gesture Recognition for Human Computer Interaction: A Survey." Artificial Intelligence Review, Springer 43 : 1–54.

[14] [14]. Pramod Kumar Pisharady and Martin Saer Beck (2015), "Recent Methods and Databases in Vision based Hand Gesture Recognition: AReview." Computer Vision and Image Understanding 141 : 152–165.

[15]. Jiangtao Cao, Siquan Yu, Honghai Liu and Ping Li (2016), "Hand Posture Recognition based on Heterogeneous Features Fusion of MultipleKernels Learning." Multimedia Tools Applications, Springer 75 : 11909–11928.