Indian sign language recognition system

Varshitha P[1], Penujuru Pranitha[2], R Nithin[3], Mylarapu Vineetha[4], Dr.Jalaja G[5]

[1,2,3,4] *Student, Dept. of Computer Science and Engineering, BNM Institute of Technology, Karnataka, India*
[5]*Associate Professor, Dept. of Computer Science and Engineering, BNM Institute of Technology, Karnataka, India*

---***---

Abstract - This article briefs about the development of Indian sign language recognition system using machine leaning. Two hand Gestures are generally used in Indian sign language to communicate. Creating model of machine learning that will be able to identify the different hand gestures used in sign language for finger spelling is the aim of this paper. On the datasets, Convolutional Neural Network (CNN) algorithms are applied.

*Key Words***:** Sign Language, Machine Learning, Convolutional Neural Network

1. INTRODUCTION

Many individuals who are not familiar with sign language find it difficult to communicate without an interpreter. Therefore, a device that transcribes sign language symbols into plain text can assist with real-time communication, and can also provide people with immersive training to learn sign language. Hand gestures are used in sign language to communicate meaning. A lot of research has been done in the corresponding American Sign Language (ASL) area, but the same cannot be said for ISL, unfortunately. The key challenges that have led to little research being done in ISL have been the lack of common datasets, occluded features and variance in the language with locality. This paper is an attempt to research the challenges of Indian Sign Language character classification (ISL) and to gather a dataset and then use different features.

2. DEVELOPMENT OF INDIAN SIGN LANGUAGE RECOGNITION

Several researchers have focused on American sign language(ASL) recognition but ISL(Indian sign language) is different from ASL. A very small amount of research has been done in this area, hence it is viable to explore further and achieve feasible results which motivated us to take up this topic. Two hands are used by ISL for communicating whereas a single hand is used by ASL. Due to overlapping of hands, the use of both hands also contributes to obscurity of features. In addition to this, lack of datasets and variation in sign language with locality have resulted in restricted attempts to detect ISL gestures. This paper can be used as communication bridge between normal people and deaf and dumb people using Indian sign language.

In this paper we will be using a custom dataset for real time prediction of more words related to ISL and on sentence formation. We are adding a feature to send Emergency text messages using gestures. We would be training the model based on single and both hands to improve accuracy. Both text and speech prediction of gesture would be included.

3. THEORETICAL BACKGROUND OF DEEP LEARNING METHODS USED FOR INDIAN SIGN LANGUAGE RECOGNITION

During the last 10 years, deep learning has demonstrated to be a wonderful technique within the field of Artificial Intelligence. Deep learning methods are used to solve various problems like image processing, speech recognition, and linguistic communication processing . As deep learning can learn robust and effective feature representation through layer-by-layer feature transformation of the initial signal automatically, it is an honest capability to deal with some challenges.

3.1 Convolutional Neural Networks

CNNs (Convolutional Neural Networks) are similar to conventional ANNs in that they are composed of neurons that achieves self-optimization through learning. An input will still be received by each neuron. CNNs consist of three types of layers. The three types of layers are convolutional layers, pooling layers and fully-connected layers. The other classification algorithms is much higher as compared to pre-processing required in a ConvNet.

The ConvNet architecture is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Each individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. To cover the entire visual area a collection of such fields overlap.
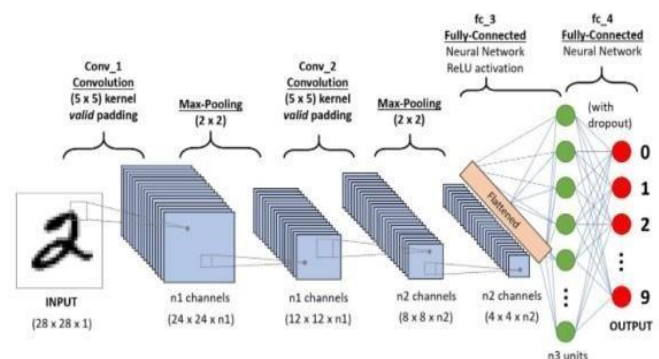
3.1 Artificial Neural Networks

Artificial neural networks are computer structures loosely inspired by the biological neural networks that make up animal brains, typically simply called neural networks. An ANN is based on a set of artificial neurons called linked units or nodes that loosely model the neurons in the biological brain. Artificial Neural network is typically organized in layers. Layers are being made up of many interconnected 'nodes'. A neural network may contain the following 3 layers:

a. Input layer

The purpose of the input layer is to receive the values of the explanatory attributes for each observation as input. Usually, the number of explanatory variables is equal to the number of input nodes in an input layer. The 'input layer' presents the network patterns, which communicates to one or more 'hidden layers'. The input layer nodes do not change the data. From the input layer, it duplicates each valueand send to all the hidden nodes.

b. Hidden layer

The transformations to the input values inside the network are performed by Hidden layers. Here the coming arcs that go from other hidden nodes or from input nodes connected to each node. Outgoing arcs are connected to output nodes or to other hidden nodes in hidden layer. The actual processing in hidden layer, is done via a system of weighted 'connections'. The hidden layers can be more than one. Here the values that are entering a hidden node which is multiplied by weights, is a set of predetermined numbers stored in the program. A single number is produced when the weighted inputs get added.

c. Output layer

The hidden layers here link to an output layer. It receives connections from hidden layers or from input layer.

The output layer then returns an output value which corresponds to the prediction of the response variable. There is usually only one output node in classification problems. The output values are produced when active nodes in the output layer combine and change the data. The proper selection of the weights provides useful data manipulation in neural network.

Conventional information processing is different.

3.2 Rule-Based Classifier:

Rule-based classifiers are just another form of classifier which, depending on the use of different "if...else" rules, makes the class decision. These rules are easily interpretable and are therefore commonly used to create descriptive models for these classifiers.

The method of background subtraction is used to detect the hand area from the background. Then the palm and the fingertips are segmented. On the basis of segmentation, the fingers in the hand picture are discovered and recognized.

4. ABSTRACT SPECIFICATION OF SUBSYSTEMS

Input Set (Hand Gestures):

The image acquisition of the person transmitting in the signal language can be obtained by using a camera. It is necessary to initiate the acquisition manually. A camera sensor is necessary to capture the signer's movements.

An algorithm for image recognition takes an image as an input and outputs what the image is made up of. In other words, the output is a class mark (for example, "'A Letter' Hand Sign Gesture," B" Letter' Hand Sign Gesture," "'C Letter' Hand Sign Gesture," etc.). In other words, a class mark is the output. The performance is, in other words, a class mark. The algorithm needs to be trained to understand the differences between different classes and to get to know the content of the image.If you want to find 'A letter' in pictures, you need to train an image recognition algorithm with thousands of 'A letter' images and thousands of background images that do not contain 'A letter'. Needless to mention, this algorithm can only understand objects / classes that it has learned. We concentrate on two-class classifiers (binary). Under the hood, several popular object detectors have a binary classifier, Inside a sign language detector is an image classifier that says whether a patch of an image is a hand gesture or a meaning.

Pre Processing

To normalize contrast and brightness effects, the input image is pre-processed. An input image or patch of an image is often cropped and resized to a fixed size as part of the pre-processing. This is necessary because on a fixed sized image, the next step, feature extraction, is performed.

Filtering is the first preprocessing phase. From the acquired image, unnecessary noise is removed. Background subtraction forms the next major stage. This processing results in a binary image in which white is colored with the pixels that form the hand and all the others are black. This processing includes the classification as part of human skin or not of each pixel of the image.

Classification

The form of the gesture being recognized is based on the camera location, the distance from the camera of the signer, etc. During real time execution, these methods must maintain a balance between precision and computational complexity. Image recognition is a method of marking and sorting objects by certain groups in the image. We need to educate it by showing thousands of examples and backgrounds before implementing a classification algorithm. Various learning algorithms view things differently, but the general concept is that learning algorithms treat function vectors as points in higher dimensional space, and try to find planes/surfaces that divide the higher dimensional space in such a way that all examples relating to a certain class are on one side of theplane/surface. For classification purposes, we use the Convolutional Neural Network algorithm as it is the best and most reliable.

CNN Algorithm:

Convolution is a phase through which a new product is combined and generated by two functions. We have to think of an image as a matrix of pixels when it comes to pictures. Each pixel has its own value, but it is combined with other pixels, and an image produces a result.

To detect certain features in the image, CNN adds filters. The manner in which the convolutional neural network operates entirely depends on the type of filter applied. So, we can provide the network with as many different features as possible when applying machine learning solutions to image classification. Then upon preparation, it will evaluate their principles. CNNs are composed of three types of layers. When such layers are piled, a CNN architecture has been created. The fundamental functionality of the above example of CNN can be broken down into four main areas.

5. CONCLUSIONS

Machine translation is a very hot research subject in the field of natural language processing at present. Machine learning helps to train a human brain-like translation system. In text processing, the CNN Translator API offers better performance. With the aid of a well-trained neural network, word alignment, reordering and language modeling can be done. Open CV, CNN and translate API are capable of recognizing and translating sign language and unknown text language into known language. Sign language recognition is a crucial communication aid for visually disabled, deaf and dumb individuals. This instrument can help bridge the gap between individuals who are natural and deaf/dumb. We use the Convolution Neural Network to classify sign language characters, including alphabets and numerals, with exceptional precision. Instead of only characters, we construct a real-time application that can define the sign language, including terms and phrases.

REFERENCES

[1] Richmond Alake, "Understanding and Implementing LeNet-5 CNN Architecture", June 2020.

[2] Madeline Schiappa, "Understanding the Backbone of Video Classification: The I3D Architecture", June 2020, https://towardsdatascience.com/understanding-the backbone-of-video-classification-the-i3d-   architecture-d4011391692

[3] Muttu Mariappan H , Dr. Gomathi V . "Real Time Recognition of Indian Sign Language" Conference: 2019 International Conference on Computational Intelligence in Data Science (ICCIDS).

[4] Neel Kamal Bhagat, Vishnusai Y, Rathna G N ," Indian Sign Language Gesture Recognition using Image Processing and Deep Learning,2019".

[5] T. Bohra, S. Sompura, K. Parekh and P. Raut, "Real-Time Two Way Communication System for Speech and Hearing Impaired Using Computer Vision and Deep Learning," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 734-739.

[6] Priyankur Sarkar, "What is LDA: Linear Discriminant Analysis for Machine Learning", September 2019.

[7] Kumar Mahesh, "Conversion of Sign Language into Text," International Journal of Applied Engineering Research ISSN 0973- 4562 Volume 13, Number 9 (2018)pp. 7154-7161.

[8] Suharjito, Suharjito & Gunawan, Herman & Thiracitta, Narada & Nugroho, Ariadi. (2018). Sign Language Recognition Using Modified Convolutional Neural Network Model. 1-5.

[9] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141.

[10] Sik-Ho Tsang, "Review: Inception-v3-1 st Runner Up (Image Classification) in ILSVRC 2015", September 2018, https://shtsang.medium.com/review-inception- v3- 1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c

[11] Sign Language Recognition Using Convolutional Neural Networks, International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 5 Issue: 6,2017.