Survey on Translation of Gesture-Based Sign Language

Shivangi Deshmukh[1], Anujkumar Yadav[2], Tapasya Kite[3], Vaishnavi Gurav[4], Prof. Ketaki Katre[5]
[1, 2, 3, 4, 5] Dept of Information Technology
[1, 2, 3, 4, 5] G.S. Moze College of Engineering, Pune, India

Abstract- *Most of us communicate our thoughts through voice and facial gestures, but according to the most recent poll, around 1% of the population in India is deaf and mute. These folks interact with others using hand gestures and facial expressions. However, most individuals find it difficult to comprehend gestures. To close this gap, we provide static gesture categorization based on sign language norms, which we subsequently transform to text and speech of a specific local dialect. Approaches for identifying hands and recognizing sign language may be classified into two categories: standard methods and deep learning methods. With the great successes of deep learning in the field of computer vision in recent years, it has been demonstrated that the deep learning approach has many benefits, such as rich feature extraction, powerful modelling capacity, and intuitive training. As a result, this article investigates hand locating and sign language recognition of common sign language using a neural network.*

Keywords- Gesture Recognition, Convolutional Neural Network, Deep Learning, Sign Language Recognition, Human Computer Interaction (HCI).

I. INTRODUCTION

Many deaf children struggle to get a suitable education in India. According to the 2011 Indian census, over 1.3 million persons have "hearing impairment." In comparison, the National Association of the Deaf in India estimates that 18 million persons nearly 1% of the Indian population are deaf.

Dumb individuals communicate via hand signals, therefore normal people have difficulty comprehending their language based on the signs they make. As a result, systems that identify various signs and deliver information to ordinary architecture.

This paper demonstrates how CNN results in extremely high levels of accuracy in solving —computer vision problems. We find a fingerspelling sign language translator with a 95% curacy rate. When we are conducting the project, there are a few finer elements that must be considered. The thresh must be managed to avoid distorted grayscales in the frames. If we run into this problem, we'll have to either reset the histogram or hunt for regions with good lighting. We may also wear gloves to avoid the issue of the signee's different skin tones. We were able to make accurate predictions in this study after we began testing using a glove.

The basic method to developing a system with the necessary capabilities comprises both hardware-based systems and software-based systems such as computer vision. A hardware-based system necessitates various sorts of accoutrements, like as:

- Hard Disk minimum of 40 GB.
- 2 GB minimum RAM
- Dual Core and up ,15” Monitor.
- Integrated webcam or external webcam and software-based systems like:
- Python
- TensorFlow
- Keras
- pip
- OpenCV

We are working on the model for converting signs to text and speech so dumb and deaf people easily communicate with normal or people like them.

1.1 SIGN LANGUAGE-

Sign language is the most comfortable and natural form of communicating between deaf and mute people, and it is also the primary instrument for special education institutions to educate and express concepts. Sign language is a natural language that communicates meaning through the shape, position, movement, and facial expressions of the hands. Sign language, like other natural languages, has a regulated grammar and a comprehensive vocabulary system.
The alphabets in Indian Sign Language (ISL) are shown in the figure 1.

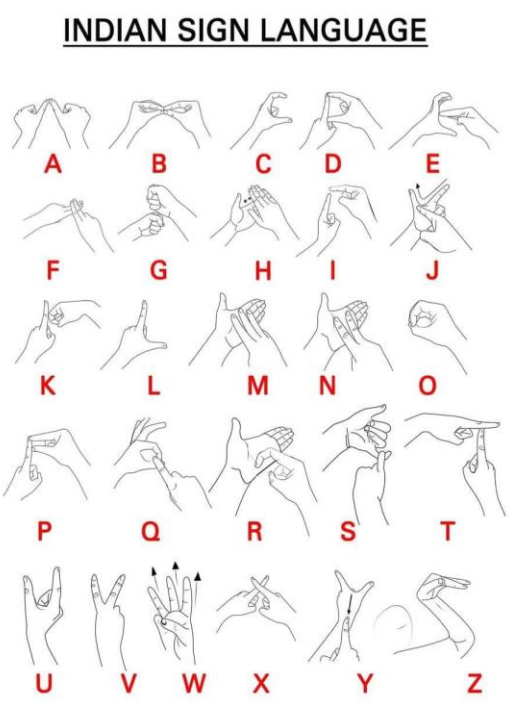


Figure 1. Alphabets in ISL

II. LITRATURE SURVEY

One of the two techniques can be used to identify sign language. The first one is through the Hardware-based system. In this approach, the user is needed to wear gloves. The second approach is the vision-based approach, in this the gesture is recognized using the concept of computer vision. In term of the hand locating, the current methods can be divided into traditional hand locating algorithm and hand locating algorithm based on deep learning. In the detection algorithm based on traditional methods, the researchers use prior knowledge to extract features, and combine the method of classifier to detect gestures. As the visual colour feature of human hand, skin colour feature can distinguish most objects from human hands. Some scholars also use Bayesian model to train skin colour model with skin colour pixel points as sample data, and use region growth to obtain relatively complete human hand region, making this method not easily disturbed by the shape change of human hand, scale change and rotation. With the successful application of Convolutional Neural Network (CNN) and target detection algorithm, a new research direction has been brought to hand Locating.

In the paper presented by Siming He (Ridley College, St. Catharines, Canada) studies hand locating and sign language recognition of common sign language based on neural network, and the main research contents include:

1. A hand locating network based on the Faster R-CNN is established to recognize the sign language the part of the hand in the picture or video, and the result of recognition is handed over to successive processing.
2. A sign-language recognition framework and a 3D CNN feature extraction network based on LSTM (long- and short-time memory) coding and decoding network are constructed for the sign language images of sequences.
3. To solve the problem of RGB sign language video or image recognition in practical problems, this paper combines 3D CNN feature extraction network LSTM encoding and decoding and hand locating network, to build the recognition algorithm

Lihong Zheng, Bin Liang, Ailian Jiang (School of Computing and Mathematics Charles Sturt University) present a complete overview of deep learning-based methodologies for sign language recognition. They discuss various types of such approaches designed for the recognition from viewpoints of available modalities provided by depth sensors, feature extraction and classification. In addition, they summaries the currently available datasets of sign language, including gestures of finger spelling and vocabulary words.

III. CONVOLUTONAL NEURAL NETWORK

A convolutional neural network (CNN) is a form of artificial neural network that is especially built to analyze pixel input in image recognition and processing. CNNs are strong image processing, artificial intelligence (AI) systems that employ deep learning to do both generative and descriptive tasks, frequently utilizing machine vision, which includes image and video identification, recommender systems, and natural language processing (NLP).

A convolutional neural network has four layers: the convolutional layer, the pooling layer, the ReLU correction layer, and the fully connected layer.

3.1. CONVOLUTION LAYER:

Convolutional neural networks rely heavily on the convolutional layer, which is always at least the first layer.
Its goal is to identify the existence of a collection of characteristics in the pictures that are fed into it. Convolution filtering is used to do this: the basic idea is to "drag" a window representing the feature on the picture and compute the convolution product between the feature and each area of the scanned image. In this sense, a feature is analogous to a filter: the two ideas are interchangeable. Therefore, the convolutional layer receives several images as input and computes the convolution of each with each filter. The filters are properly correlated to the features in the images that we

want to find.We obtain a feature map for each pair (image, filter) that informs us where the features are in the picture: the greater the value, the more the corresponding location in the image matches the feature.

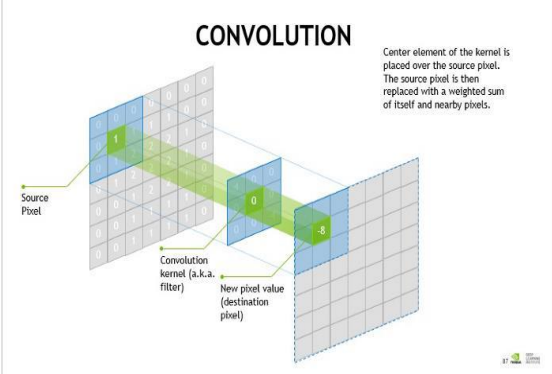


Fig. 2 Convolution Layer

3.2. POOLING LAYER:

This sort of layer is frequently put between two convolution layers: it receives numerous feature maps and applies the pooling operation to each of them.

The pooling method reduces the size of the photos while keeping their vital properties.

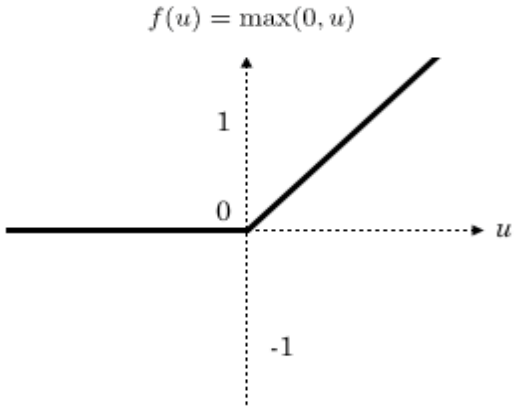
To do this, we divide the image into regular cells and keep the maximum value within each cell. In practice, small square cells are typically used to avoid losing too much information.

The most popular options are 2x2 adjacent cells that do not overlap or 3x3 cells separated by a 2-pixel step (thus overlapping).

The pooling layer minimizes the number of parameters and calculations in the network. This enhances network efficiency and prevents over-learning.

3.3 THE ReLU CORRECTION LAYER:

The real non-linear function defined by ReLU(x)=max is referred to as ReLU (Rectified Linear Units) (0, x). Visually, it seems to be as follows:



All negative values received as inputs are replaced by zeros by the ReLU correction layer. It performs the function of activation.

3.4 FULLY CONNECTED LAYER:

Because the fully connected layer is often the last layer of any neural network, whether convolutional or not, it is not unique to CNNs

.

This layer accepts an input vector and outputs a new vector. It accomplishes this by applying a linear combination and, maybe, an activation function to the receiving input values.

As an input to the network, the last fully connected layer 0classifies the picture: it outputs a vector of size N, where N is the number of classes in our image classification task. Each member of the vector represents the likelihood that the supplied image belongs to a class.

IV. CONCLUSION

In our paper, we try to improve the ability of people with speech and hearing disabilities to communicate with those who are not familiar with Indian Sign Language. It greatly limits their ability to communicate because of this. In order to have such technology closer to home and make it possible for our disabled community to communicate with other people, we useISL.

REFERENCES

[1] Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton, "ImageNet Classification with deep convolutional neural networks," 2012, Proceedings of the 25th International Conference on Neural Information Processing Systems, pp-1097-1105.

[2] Das, A., Gawde, S., Suratwala, K., & Kalbande, Dr. Dhananjay Kalbande, "Sign Language Recognition Using Deep Learning on Custom Processed Static Gesture Images," 2018 International Conference on Smart City and Emerging Technology (ICSCET). doi:10.1109/icscet.2018.8537248

[3] Kshitij Bantupalli and Ying Xie, American Sign Language Recognition using Deep Learning and Computer Vision 2018 IEEE International Conference on Big Data (Big Data)

[4] Siming He, Research of a Sign Language Translation System Based on Deep Learning, 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), ©2019 IEEE

[5] Lihong Zheng, Bin Liang and Ailian Jiang Recent Advances of Deep Learning for Sign Language Recognition, 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)

[6] IJTRS-V2-I7-005 Volume 2 Issue VII, August 2017 survey on hand gesture techniques for sign language recognition.

[7] Neel Kamal Bhagat, Vishnusai Y, Rathna G N, "Indian Sign Language Gesture Recognition using Image Processing and Deep Learning," 2019 Digital Image Computing: Techniques and Applications (DICTA). doi:10.1109/dicta47822.2019.8945850.

[8] Citra Suardi, Rosa Andrie Asmara, Anik Nur Handayani, Aji Prasetya Wibawa, Huzain Azis, Lilis Nur Hayati, Design of Sign Language Recognition Using E-CNN, 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT) @IEEE