Review on Sign Language Detection Using Convolutional Neural Networks

[1]Sharan Lional Pais, [2]Dikshit Kotian, [3]T K Harshith Prasad, [4]Srihari B, [5]Anjali CJ

[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student
Information Science and Engineering
Alva's Institute of Engineering and Technology

***Abstract*: Sign language is an indispensable communication means for deaf-mute people because of their hearing impairment. At present, sign language is not popular communications method among hearing people, so that most of the hearing are not willing to have a talk with the deaf-mute, or they must spend much time and energy trying to figure out what the correct meaning is. There has been various research work been done to find an optimal solution to the sign language recognition. This paper reviews one of such works for the sign language recognition using convolutional neural networks.**

***Keywords*: Sign Language, CNN, Image**

I. INTRODUCTION

Communication can comprehensively be characterized as trade of thoughts, messages and data between at least two people, through a medium, in a way that the sender and the recipient communicate the message in good judgment, that is, they create basic comprehension of the message. We convey through discourse, signals, non-verbal communication, perusing, composing or through visual guides, discourse being quite possibly the most usually utilized among them. However, unfortunately, for the speaking and hearing-impaired minority, there is a communication gap.

Sign Language is the common language for the deaf and dumb, something that works out easily as a type of non-verbal communication between signers. Not many individuals communicate using sign-based communication. Additionally, in spite of mainstream thinking, it's anything but a worldwide language. The alternative of written communication is cumbersome, because the deaf community is generally less skilled in writing a spoken language. This type of communication is impersonal and slow in face-to-face conversations. The limitation, combined with the absence of information about communication via sign by verbal speakers, makes a detachment where the two people can't productively convey.

Such an issue increments under a particular setting, for example, crisis circumstances, where first-reaction groups, for example, emergency services, cops may be not able to appropriately go to a crisis given that collaborations between the involved parties become a hindrance for dynamic when time is scant. Developing a cognitive-capable tool, that serves to perceive gesture-based communication in a one-of-a-kind way, is an absolute necessity to decrease obstructions between the deaf and dumb and emergency individuals under this unique circumstance.

II. RELATED WORKS

Xinyun Jiang proposed a model for hand gesture recognition using a Support Vector Machine. In this study, the proposed system aimed to recognize hand gestures for sign language. The process of gesture recognition can be categorized into four stages, namely data acquisition, pre-processing, features extraction and classification. The input data of the system was acquired from the camera, and then processed to find where the hand gesture was, for further processing. Feature extraction is a stage to complete the transformation of a pre-processed image into sets of feature vectors. Afterwards, for classification, the Support Vector Machine (SVM) is used to model the selected features, mapping them to a specific hand gesture. The model was trained to recognize only five alphabets.

A database containing five gestures was built, of which 1,400 pictures played by 4 people with different ages and sexes, were mapped for each gesture. Samples were also collected in various environments, grouped into six categories according to the lighting conditions, including most of situations taken in daily life. A method that adaptively determines the limits of skin color range was developed. Two methods of noise removal are proposed, which could effectively work for obtaining cropped images only containing ROI. When dealing with the orientation problem, the rotation angle was found accurately and a method was proposed to solve the problem that useful information may be cut off when rotation. By using PCA, 8 features were extracted to represent the original data, which reduces the computational complexity. The one-versus-one SVM classifier was established, which could complete classification effectively in a relatively short time.

Juan Zamora-Mora worked on a Costa – Rican sign language detection model that used EgoHands dataset separated into a training-set with twenty-seven thousand two hundred sixty-two images, and a test-set with two thousand five hundred five images. MobileNet-V1 model was able to detect hands in LESCO videos successfully until the two hundred thousand training iteration. The same algorithm under fifty thousand iterations was not able to properly label hands in the five test videos used. The results showed a small percentage of frames where the algorithm was not able to detect hands.

To detect and track hands, a two-step process was used. The first step trained a convolutional neural network using the EgoHands dataset. This first step will test its accuracy using the mean average precision metric. The second step tested the ability of the trained model to recognize hands in video recordings of deaf people signing phrases in Costa Rican sign language. Haar detection and Convolutional-based detection were used for extracting the features. Haar detection is based on a set of low-level features selected from several rectangular regions across the image which produces a high number of features filtered by Adaboost and Convolutional-based detection had many sub sampling layers capable of performing object detection with high-level accuracy. Two checkpoints of the model were obtained at 50 thousand (50k) and 200 thousand (200k) iterations. The 50k model obtained a value of (ninety-two dot three percent) 92.3% for the mAP, and the second model with 200k iterations a value of (ninety-six dot one percent) 96.1%.

III. METHODOLOGY

In this research, a computer vision-based learning strategy was proposed to see hand signs in different designs. To recognize different sign signals, AI technique which contains preparing classifiers with HOG structures was applied. CNN algorithm for forecasting directed signs was selected.

A.Data cleaning and data pre-processing

In this process, information was cleaned, revamped, and adjusted into readable for the algorithms. Pre-processing, reshaped and rescaled crude information to find a way into the neural network. Especially cleaning diminishes commotion in the crude information.

B.Feature extraction

Unnecessary information was standardized from the raw pictures to extract the principle highlights. Feature extraction intends to decrease the quantity of features in a dataset and can be executed in Python utilizing Keras application programming interface. An API is utilized to coordinate connection among third party stages for access to its highlights and administrations. Keras is a neural network based API and equipped for running on top of TensorFlow. Those features which represent images were extracted. Important features from the image data was found. Then the correlation between the data was found.

C.Train test split

In this progression, the model prepared was trained. 25000 pictures were taken and the outcome was anticipated. The pictures were split into 80% for the train information and 20% for the test information. The purpose behind parting the dataset into 80% and 20% was to train the model in a neural network. The preparation and test information contained pictures and the objectives. The pictures were converted into pixels called a feature matrix. The target was isolated from the feature matrix. The neural networks can ascertain a tremendous measure of highlights. Data for train purposes passed through eight layers in the training phase.

D.Convolutional Neural Network (CNN)

Deep learning models are abstrusely stimulated by data dealing out and outlines in genetic nervous structure, yet have various variances from the physical and functional properties of the genomic brain which make them mismatched with neuroscience signs. It was a typical term for multi-layer NN. CNN is a kind of feed-forward Artificial Neural Networks that was known to be incomprehensibly incredible inside the field of picture sorting and accreditations.

Different layers of CNN are:

(I) Convolution: - The most extreme utilization of the convolution measure in the circumstance of a CNN is to properly recognize the data. They are separating the features here. Those include on or after the picture which is an association for the primary layer. Three-dimensional between connections of pixels is safeguarded by this. It is finished by utilizing little squares of picture by achievement of picture highlights.

(II) Non-linearity: - ReLu is a Nonlinear strategy known as Rectified Linear Unit. ReLU is a method that is finished per pixel which at that point surpasses all the negative figures of each pixel by zero out of a feature map.

(III) Max-Pooling: - It is a discretization technique which is a consideration of many samples. The point was to down-sample an input representation. It can lessen dimensionality and consider assumptions to be made about features restricted in the subregions binned. It is furthermore named down an experimental group that helps in decreasing the size of each element map. When pooling is done, in the end, the 3D feature map will be restored to the Minimum 1D featured vector.

(IV) Fully Connected layer (Classification): - In these layers. The entirety of the input sources come from One layer additionally associated with Each and every other actuation unit of the Following layer. In the greatest current AI model, only a few layers are completely associated layers which gather the information mined by past layers, so it can make a last output. This is the second most Time taking layer after the Convolution.

E. Images of the sign language

For the work, a training dataset was created. The raw pictures taken for this specific research were gathered from some mute and deaf people, to know how they speak with people like them.The sign was collected as a picture.

IV. RESULTS

Bangla Sign language dataset which comprises 24168 examples (fundamental characters: 18745 and numerals: 5423) was used for training. The proposed model generated 100% testing accuracy on digits, and 97.5% testing accuracy for the letters in alphabets of Bangla. The accuracy reached at 98.75% while the two kinds of tests were used. The dataset contained only hand images hence, the

performance of the model when the rest of the human body is captured cannot be guaranteed. Images used for training were not captured under different lighting and background conditions. The number of epochs required for reaching the higher training and testing accuracy was more than 140. The total accuracy of detection becomes slightly lower when both accuracy were used.

V. CONCLUSION

In this paper, a model was developed to detect Bangla Sign Language using the Convolution Neural Network (CNN). The dataset was split into 80:20 ratio where 80% was used for training purpose and rest of them was test purpose. CNN was able to detect and classify the signs in the images into Bangla digits and numbers. The accuracy achieved was 98.75%. This work shows that convolutional neural networks can be used to accurately recognize different signs of a sign language. One advantage of CNNs is that it requires no feature engineering, as each layer in the network is able to automatically extract the target object features, by encoding low-level attributes in the initial hidden layers such as shapes and edges, to more complex objects such as body parts in the following layers. CNN also requires a great number of samples for training that ranges from hundreds to thousands for each class which might take some time for a dataset to get prepared, due to the fact that labeling each image is usually a manual task performed with a tool able to select the target object on each image.

REFERENCES

[1]. Sohrab Hossain, Bengali Hand Sign Gestures Recognition using Convolutional Neural Network, ICIRCA-2020
[2]. Xinyun Jiang , Hand Gesture Detection based Real-time American Sign Language Letters Recognition using Support Vector Machine, IEEE, 2019
[3]. Juan Zamora-Mora, Real-Time Hand Detection using Convolutional Neural Networks for Costa Rican Sign Language Recognition, CONTIE, 2019
[4]. M.F. Tolba, A.S.Elons, Recent Developments in Sign Language Recognition Systems