TEST Engineering & Management

Sign Language Translation into Text and Speech using CNN and open CV

Aman Kumar, University Institute of Engineering, Chandigarh University, Chandigarh, India
AkashYadav, University Institute of Engineering, Chandigarh University, Chandigarh, India
Ruchika Gupta, University Institute of Engineering, Chandigarh University, Chandigarh, India

Article Info
Volume 82
Page Number: 16228 – 16234
Publication Issue:
January-February 2020

Abstract:

There are approximately 700,000 deaf and dumb people in this world. Approximately 60% of these people are born deaf and dumb. The language of deaf and dumb that uses body components to convey the message is thought as sign language. Because of the comparative lack of prevalent sign language consumption among the society, deaf and different challenged individuals tend to face problem on a daily basis. Inability to speak is considered to be true disability. People with this incapacity use completely different modes to speak with others, there are number of ways out there for his or her communication, one such common technique of communication is sign language. Developing sign language application for deaf individuals may be vital, as they‟ll be able to communicate easily with even those who don‟t recognize sign language actions. Here the job targets to take the fundamental actions in maintaining the verbal gap among traditional individuals and verbally challenged individuals.For them, means of communication becomes very difficult and the only way they can communicate is by means of sign language. Although, a deaf and dumb person might know sign language but the person he wants to communicate with may not know. We want to bridge this gap and provide a way to process sign language and convert it into text and speech with the help of machine learning and neural networks

Article History
***Article Received**: 18 May 2019*
***Revised**: 14 July 2019*
***Accepted**: 22 December 2019*
***Publication**: 28 February 2020*

***Keywords:** Sign Language Recognition, Convolutional Neural Networks, Gesture Recognition, OpenCV, Image Processing with CNN, Video Processing With CNN*

I. INTRODUCTION

A person who knows sign language can communicate effectively with another person who also understands sign language. The problem arises when a person wants to communicate with the rest of the world. The worlds population is around 7.7 billion and only 700,000 people are deaf and dumb. Hence, having a system that allows the disabled to communicate easily is very important. Our work will enable the disabled person to put forward his ideas in front of the world and allow the rest of the world to understand them effectively. This can be a breakthrough in the lives of the disabled people as they will be able to communicate with everyone without the dilemma whether others will understand them or not. When a person communicates, he can communicate by framing long or small sentences. Small sentences can be easily conveyed with sign language but the retention of long sentences is difficult. Our work will translate the sign language used by the user to both text and speech. As a result, retention of the sentences will become very easy. Sign language translated into text and speech will also ignite the interest of others and it will help them to respond effectively without any errors. We have trained the network under different configurations, also analyzed and tabulated the obtained results. The observation displays the correctness improves as more data is added to totally several subjects during the time of training. Another crucial applications of hand motion identification is to check the language based on sign which is a vital tool for communication for physically challenged, . people who are dumb and deaf. As they are suffering a lot and many difficulties in their life so this will help in the their normal day to day life.

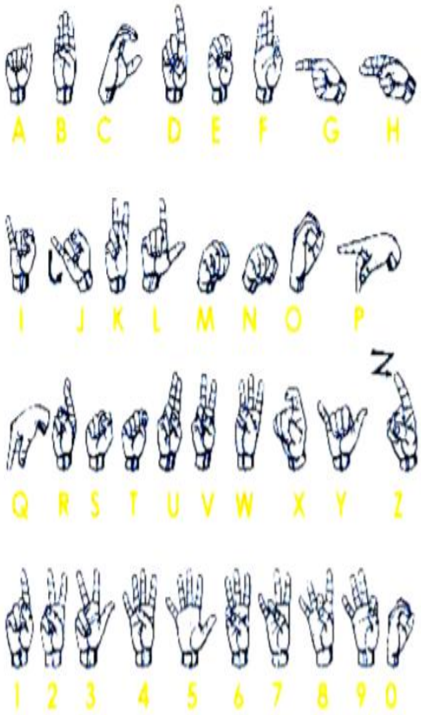


Fig. 1: Hand Signs For Alphabets

A. Motivation: Our work will tremendously make the lives of thousands of people very easy. We are providing a new method of communication which doesn't require everyone to learn sign language at the same time doesn't places any extra burden on the disabled.

B. Problem Statement: Classify the English alphabets from a live video stream with the help of Convolutional Neural Networks.

C. Objectives: This work focuses on taking the certain actions in maintaining the miscommunication among normal people and verbally challenged people using sign action language.The main objective of our project is on contributing to the sector of automatic sign language recognition. We focus on the recognition of the static sign language gestures. This work focused on deep learning approach to recognize 24 alphabets and 0-9 numbers. We created a convolutional neural network classifier that can recognize static sign language gestures with high accuracy. We have trained the network under different configurations, also analyzed and tabulated the obtained results. The observation displays the correctness improves as more data is added to totally several subjects during the time of training. Another crucial applications of hand motion identification is to check the language based on sign which is a vital tool for communication for physically challenged, . people who are dumb and deaf. As they are suffering a lot and many difficulties in their life so this will help in the their normal day to day life.

II. Related work

Lionel Pigou [1] made use of two CNN neural networks, one for evaluating the hand gestures and the other to evaluate the body language. Their dataset consisted of 20 Italian gestures made by different people and in various environments. Hey were able to identify 20 gestures with high accuracy irrespective of the surrounding. They have used dropout layer to reduce the chances of model overfitting. R Cui [2] and their team used Recurrent Convolutional Neural Networks for continuous sign language detection. They have made use of temporal convolution, bidirectional LSTM and pooling for temporal feature extraction. They have evaluated their model on a publicly available dataset RWTH-PHOENIX-Weather. It contains sentences in the German sign language. J Huang [3] and their team have used 3D CNN to get information rich features. They have used open source SLR datasets. They have compared various models such as LSTM, S2VT, LSTM-A, LSTM-V and HAN. The were able to get accuracy of 0.736, 0.745, 0.757, 0.768, 0.793 respectively for all the above models. SruthiUpendran [4] and their team introduced “American Sign Language Interpreter System for Deaf and Dumb Individuals”. The talk about methods could validate 20 among the 24 ASL. The characters like S also N, A and M can not be verified because of the similarities in their hand signs.Bhumika Gupta, PushkarShukla, Ankush

Mittal [5].They extract image. Features are then put in a separate matrix. Correlation is computed for these and is fed to a K-Nearest Neighbour Classifier. 179 gestures were recognized correctly among the 200 gestures. They worked on the dataset of Indian Sign Language.

III. Methodology

We have used two technologies, namely, Convolution Neural Networks and OpenCV. Convolutional Neural Networks is a state of the art technique for image data. We have used two 2D ConvolutionalNeural Networks layers of filter size 5 and 15 respectively. We have set the kernel size = 5 for both the layers. We have used ReLU activation function.

$$\mathrm{Re}\,Lu: f(x) = x+ = \max(0, x)$$

We have used ReLU activation function as it avoids the problem of vanishing gradient. For the output, we have used a Dense layer with Softmax activation. We have taken a dataset of 550 hand signs images. The dataset consists of 11 distinct alphabets: A, B, C, D, E, G, H, I, L, O, V. For each alphabet, we have 50 hand sign image of that alphabet. Thee hand sign image of the alphabets are of dimension 64 x 64 with RGB channel. Though the image are black and white, we have still used RGB channel. We have trained our Convolutional Neural Networks model on this dataset. We have used OpenCV for live classification of hand signs. In this, we capture a video frame form the live video. We only capture a region of interest and not the whole video frame. The next step is to convert it into an array and resize it to 64 x 64. Then we use our pre trained model to predict the class of the captured video frame. For multiclass classification, we have trained another Convolutional Neural Networks model on all 26 alphabets. The dataset consisted of 1872 images of dimension 64 x 64.

IV. Proposed Work

1. Binary Classification: Binary or binomial classification is that the task of classifying the objects of a given set into two teams (predicting which cluster each one belongs to) on the idea of a classification rule.In this, we have made a prediction between any two alphabets.We have used a CNN model to classify the alphabets.The input to the model is the images of alphabets, each of size 64 x 64 x 3. The image is of RGB format. Hence, the depth of the images is 3. The model consists of two convolution layers with activation of ReLu, two layer of max pooling and one layer which is fully connected with Sigmoid activation.We have used Sigmoid here because the ouput of the model is a binary value.

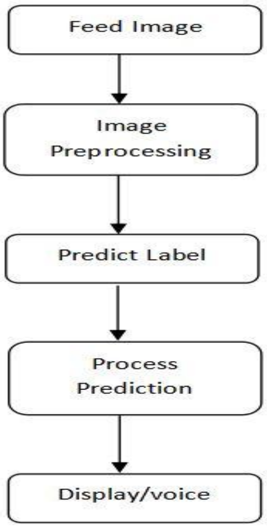


Fig. 2: Flowchart of the Proposed Work

2. Multiclass Classification: In this, we have trained the model on the complete dataset (A-Z) and classified all the alphabets.Here, instead of using Sigmoid activation in the fully connected layer, we have used the Softmax activation for multiclass classification.The optimizer we have used in this model is Adam as it combines the advantages of both AdaGrad and AdaDelta optimizers and works well for computer vision problems.The output given by the model is a 26 dimensional array. According to the prediction made by the model, only one value in the array is 1 and the rest are 0.

3. This sub-module will capture hand signs and classify them. A user will be able to construct words and sentences in a real time environment with the

help of different signs. Here, we capture an image using OpenCV. The captured image is converted to an image array. This array is given as input to a CNN model and the model predicts the output. The output of the model is a 26 dimensional array.

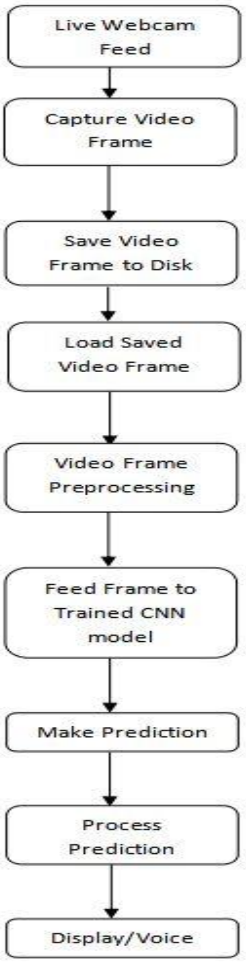


Fig. 3: Detailed Flowchart of the Proposed Work

EXPERIMENTAL RESULTS

1. Multiclass Classification: For multiclass classification the model achieves an accuracy of 99.73% after 10 epochs. The model is not able to perform well in the case of alphabets like A and E as they have extremely similar hand signs. The model was run for 10 epochs. The Convolutional Neural Networks model began to meet convergence at 6 epochs. The output of the model is a vector of size 26. The position at which the value is 1 denotes the alphabet at that position.

The graph below shows the accuracy of the model for multiclass classification with respect to epochs. As the epoch increases, the accuracy of the model also increases.

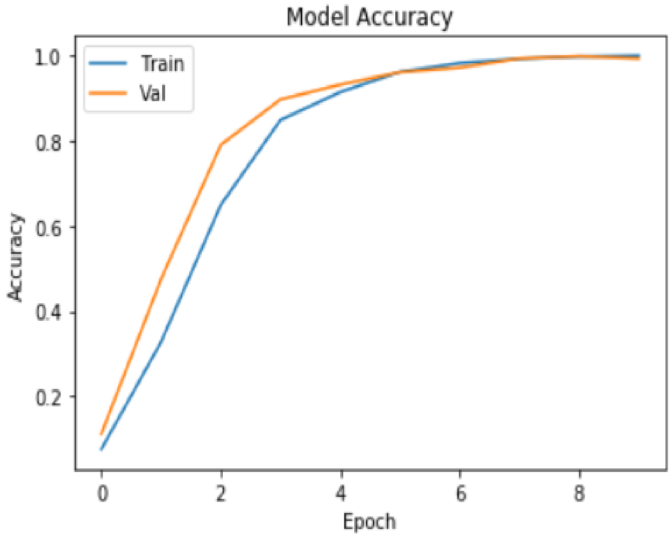


Fig. 4: Model Accuracy - Multiclass Classification

Fig. 5, Fig. 6, Fig, 7 and Fig. 8 shows the hand signs for alphabets "D", "H", "M" and "C" respectively. Some of the predictions made by this model are:

i. Case 1: When hand sign is of of alphabet "D"

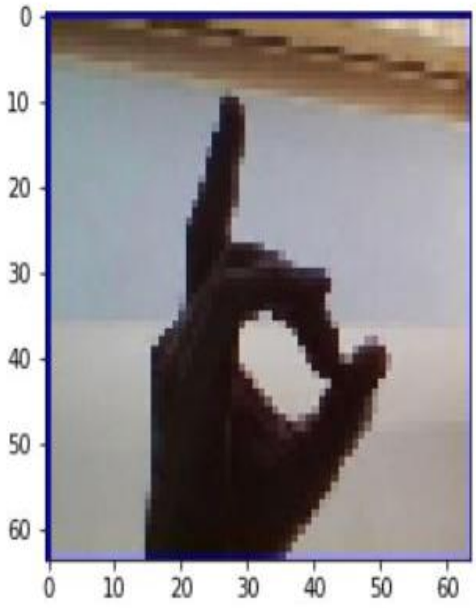


Fig. 5: Hand Sign of alphabet "D"

```
[[0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
  0. 0.]]
```

ii. Case 2: When hand sign is of of alphabet "H"

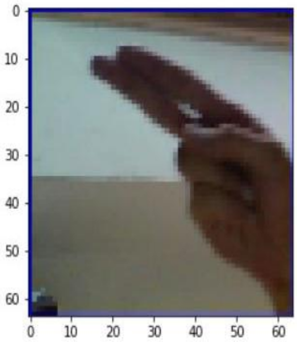


Fig. 6: Hand sign of alphabet “H”

```
[[0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
  0. 0.]]
```

iii. Case 3: When hand sign is of alphabet “M”

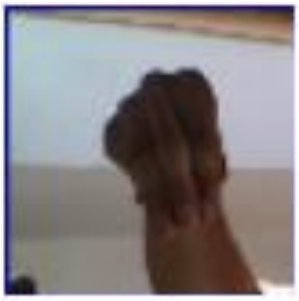


Fig. 7: Hand sign of alphabet “M”

```
[[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
  0. 0.]]
```

iv. Case 4: When hand sign is of alphabet “C”

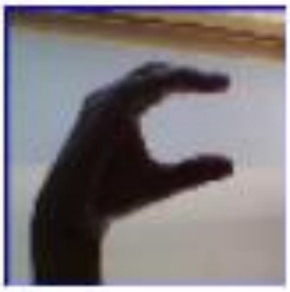
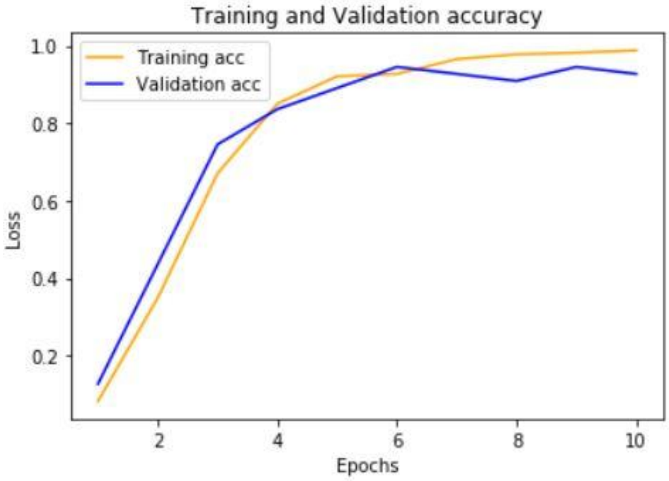


Fig. 8: Hand sign of alphabet “C”

```
[[0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
  0. 0.]]
```

2. Live Hand Sign Classifier: The below two graphs Fig. 9 and Fig. 10 shows the training and validation accuracy along with training and validation loss respectively. The model performed very well on the dataset. After 10 epochs, the Convolution Neural Networks model achieved training accuracy of 98.79% and validation accuracy of 92.73%.



3. Fig. 9: Training and Validation Accuracy

4. Fig. 10 shows that the model loss was highest at 1 epoch and lowest at 10 epochs. The highest and lowest loss encountered is 2.4253 and 0.0588 respectively.

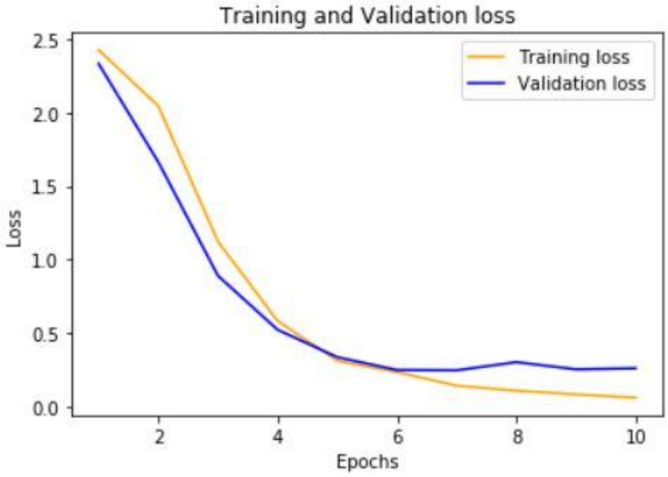


Fig. 10: Training and Validation Accuracy

Fig. 11, Fig. 12, Fig, 13 and Fig. 14 shows the hand signs for alphabets “L”, “H”, “D” and “G” respectively.

Some of the predictions made by this model are:

i. Case 1: When hand sign is of alphabet “L”

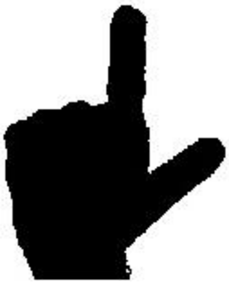
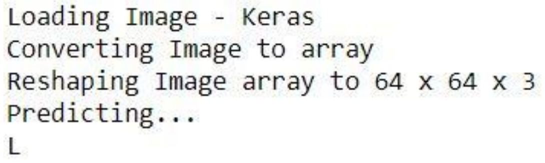


Fig. 11: Live case for alphabet “L”

Output:



```
Loading Image - Keras
Converting Image to array
Reshaping Image array to 64 x 64 x 3
Predicting...
L
```

ii. Case 2: When hand sign is of alphabet “H”

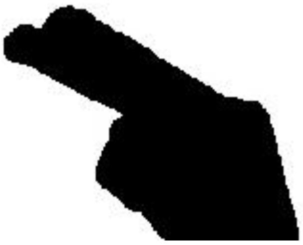
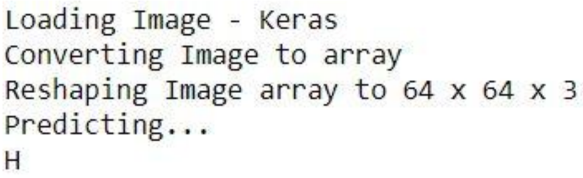


Fig. 12: Live Case for alphabet “H”

Output :



```
Loading Image - Keras
Converting Image to array
Reshaping Image array to 64 x 64 x 3
Predicting...
H
```

iii. Case 3: When hand sign is of alphabet “D”

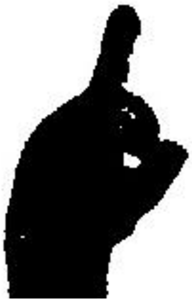
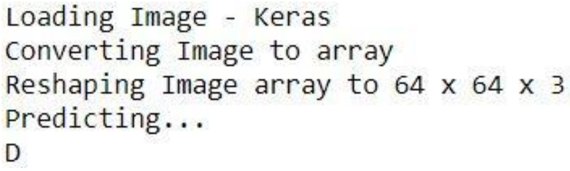


Fig. 13: Live case for alphabet “D”

Output:



```
Loading Image - Keras
Converting Image to array
Reshaping Image array to 64 x 64 x 3
Predicting...
D
```

iv. Case 4: When hand sign is of alphabet “G”

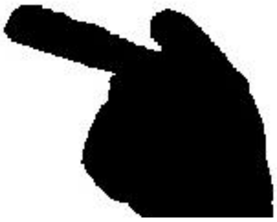
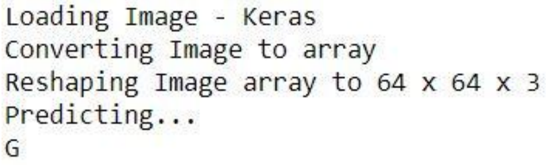


Fig. 14: Live case for alphabet “G”

Output:



```
Loading Image - Keras
Converting Image to array
Reshaping Image array to 64 x 64 x 3
Predicting...
G
```

V. Conclusion

This idea will help to bridge the gap between 700,000 deaf and dumb people and the rest of the people. Our model has shown great results in the classification of hand signs and faced ambiguity in only two alphabets, namely, A & E. The model can be further improved by increasing the size of the dataset and adding instances with different environments, people, colors etc. The limited size of the dataset has affected the accuracy of the model.Our idea is going to help maintain the huge gap among normal and physically challenged people. With the observations and results we get above one can finish that CNN gives a supereb and markable correctness in recognizing the language based on sign also characters which are having alphabets and numbers. Our work can now also extended for purpose of making a time application based on real time, that will recognize the language based on sign and also the words, complete structure of sentences to identify not by only characters.

FUTURE SCOPE

One more important feature that can be added to this project is Gesture Recognition. We have only classified alphabets till now. There is a huge possibility of training our model to recognize gestures. Gesture recognition will make this project even more robust and effective. It will allow the user to communicate like never before. Gesture recognition will require a huge number of volunteers in order to prepare the dataset. Also, training the model on such a big dataset will require lots of computing power.

REFERENCES

1. Pigou, Lionel, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen. "Sign language recognition using convolutional neural networks." In European Conference on Computer Vision, pp. 572-578. Springer, Cham, 2014.
2. Cui, Runpeng, Hu Liu, and Changshui Zhang. "Recurrent convolutional neural networks forcontinuous sign language recognition by staged optimization." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7361-7369. 2017.
3. Huang, Jie, Wengang Zhou, Qilin Zhang, Houqiang Li, and Weiping Li. "Video-based sign language recognition without temporal segmentation." In Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
4. Upendran, Sruthi, and A. Thamizharasi. "American Sign Language interpreter system for deaf and dumb individuals." In 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), pp. 1477-1481. IEEE, 2014.
5. Gupta, Bhumika, PushkarShukla, and Ankush Mittal. "K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion." In 2016 International Conference on Computer Communication and Informatics (ICCCI), pp. 1-5. IEEE, 2016.