Sign Language Recognition Using Convolution Neural Network

Varshitha Sannareddy, Mounika Barlapudi, Venkata Koti Reddy Koppula, Gali Reddy Vuduthuri, and Nagarjuna Reddy Seelam

Abstract Sign language is one of the medias to communicate with deaf and dumb people; usually, it is not known to normal people. So, it becomes a challenging task to establish a communication between normal people and hearing impaired person. Many tools are developed to help them, but unfortunately not produce accurate results. To interact with them, we are using various fingers gestures; then, designed model will convert those gesture into words or alphabets into specific language. The predicted model is helpful to reduce the gap between the normal people and hearing impaired person. In our proposed sign language recognition algorithm, we focused on deep convolution neural network to produce better accuracy.

Keywords Sign language recognition · Convolution neural network · Support vector machine

1 Introduction

In the world to interact with any person, we require some language in the form of textual, vocal or visual representation. But in case of visually challenged persons means those who have hearing impaired problems communication is very tedious task. To communicate with them require a suitable media through visual, i.e., sign language [1]. It is very useful to persons those who have difficulties with speaking or hearing. Sign language is one of the popular communication media it used various ways like hand motions, facial expressions and body movements to express something. In the world, already some of the popular sign languages are exists with various functionalities and limitations [2, 3]. Some of the popular sign languages are Indian Sign Language (ISL), Polish Sign Language, American Sign Language, etc., based on the geographical conditions languages are changed like spoken languages. Because of these every sign language have some limitations.

V. Sannareddy (✉) · M. Barlapudi · V. K. R. Koppula · G. R. Vuduthuri · N. R. Seelam
Computer Science and Engineering, Lakireddy Balireddy College of Engineering, Mylavaram, Andhra Pradesh 521230, India

V. S. Reddy et al. (eds.), *Soft Computing and Signal Processing*, Advances in Intelligent Systems and Computing 1413, https://doi.org/10.1007/978-981-16-7088-6_59

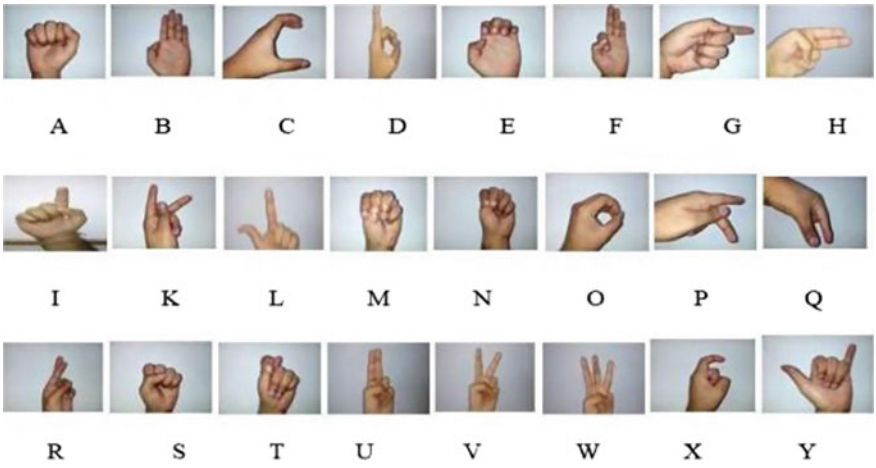


Fig. 1 Hand gestures of ISL

Many software tools and packages are also developed teach and understand the sign language, but the usage is limited because of not producing accurate results. To overcome the limitations, we proposed an algorithm to interact with a hearing impaired person accurately by recognizing and understanding the sign correctly. We considered Indian sign Language (ISL) to check the efficiency of the proposed algorithm.

The following Fig. 1 conveys the hand gestures of ISL.

To understand ASL, so many algorithms are already proposed, but ISL is completely different from ASL. In ASL, they used only one hand to give the signs, and in ISL they used two hands to provide the signs.

2 Related Work

Ansari and Harit [4] have produced significant research contributions to categorize the Indian signal language gestures accurately. In that they used different alphabets, numbers and different movements all together of 140 different samples they categorized. To detect the various parts in the body like hand they used traditional unsupervised learning technique, i.e., K-Means clustering algorithm. They used Gaussian distribution also to extract the features required for train the data set with the accuracy reaches 90%.

Deora and Bajaj [5] adopted PCA algorithm to recognize the sign language gestures. In this, they also used artificial neural networks to recognize in efficient manner. They considered very small data set and produced results are also not satisfactory. But, compared to neural networks, PCA produces accurate results.

Zhang et al. [6] adopted convolution neural networks to perform sign language recognition. To perform the process in efficient manner, they established two-step process—first one is extract features then next followed by classification process. To extract the features, they used CNN and artificial neural network is used to classify them. In proposed algorithm, they used Italian sign gestures of 27 members. CNN adopted max pooling technique to extract the features and forwarded to ANN. This model produces the highest accuracy of 91.7%.

3 Methodology

In the proposed algorithm, we adopted the approach described Fig. 2. In the figure, it contains the steps to recognize the various sigs worked on the ISL.

In data set, we are going to analyze total of 4800 sign images of ISL of the English alphabets, and it consists of 26 different class labels in the data set. The data set is as shown in Fig. 1. At pre-processing stage, it resize the images into 640 × 480 pixels. Both feature extraction and classification techniques will be performed after performing the normalization on images. The following Fig. 3 provides information about workflow diagram of the proposed model. The proposed model can perform in various steps.

1. Image Acquisition
2. Image pre-processing
3. Feature extraction
4. Apply CNN
5. Classification.

Convolution Neural Network Architecture is shown in Fig. 4.

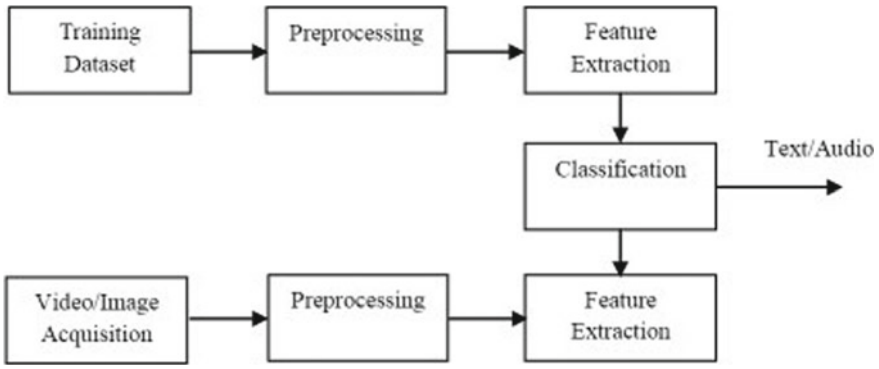


Fig. 2 Block diagram of proposed model

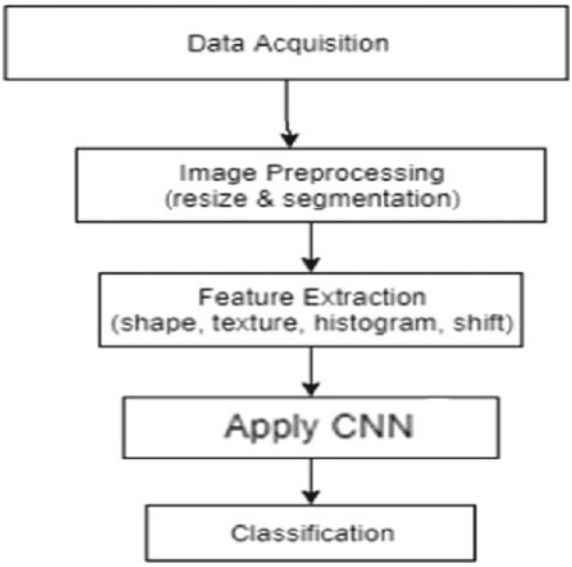


Fig. 3 Work flow diagram of proposed model

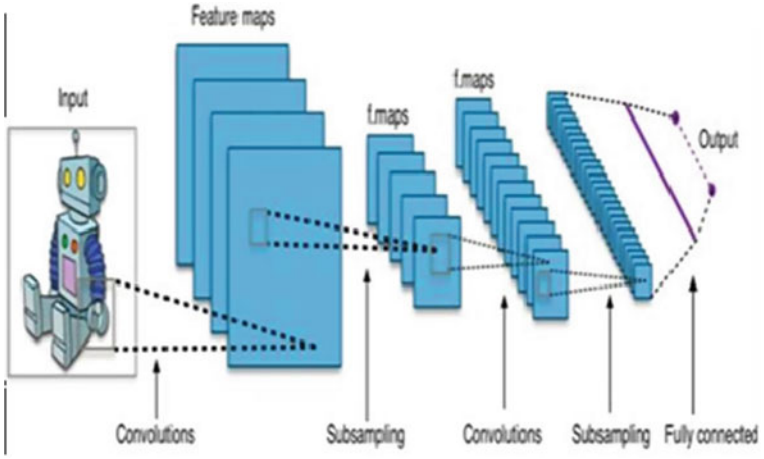


Fig. 4 Convolution neural network architecture

4 Implementation

4.1 Image Acquisition

Image acquisition is the process of creating the photographic images, such as the interior structure of an object. The term is often assumed to include the compression, storage, printing, and display of such images. To acquire frames in real time we use

Fig. 5 Acquiring frames in real time

various built in functions in the open CV library. The following python code used to capture the images in real time.

cap = cv2.VideoCapture(0)

ret, img = cap.read()

The following Fig. 5 provides an example used to represent acquiring frames in real time.

4.2 Image Pre-processing

The main objective of image pre-processing to remove unwanted pixels by applying various techniques like image crop, applying filtering techniques to improve the brightness of the images and also to remove the unwanted noise or outliers detected in the images. During this, all acquired images are in the form of RGB. But by applying various techniques, they will be converted to binary images. In this step, only all images are resized into specific and size, by using image segmentation we are going to detect the image boundaries. The following Fig. 6 shows about the frames after applying the pre-processing steps.

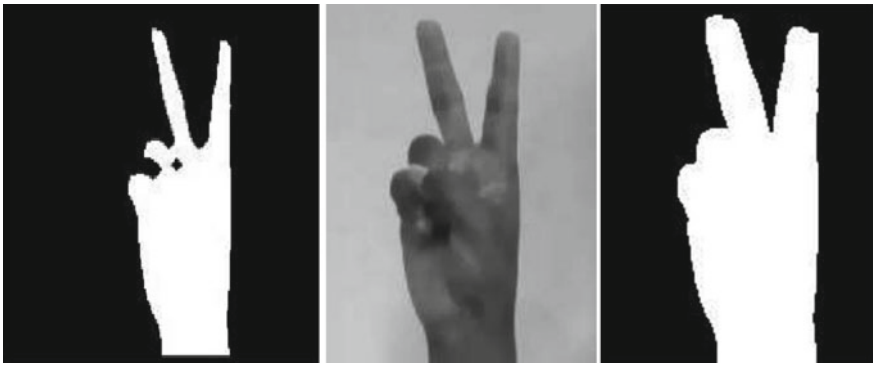


Fig. 6 Images after applying various filtering techniques

4.3 Feature Extraction and Classification

It is the most important step by using this, we are going to create a model for sign recognition. To extract relevant features from the images, we adopted CNN technique. It contain more than one convolution layer to extract relevant features. To extract the topological properties from an image, we adopted feed-forward network.

On the selected features apply popular classification technique like SVM, Random Forest and KNN to design the model using trained data set. After fitting, the model predicts the values for test data set.

Program Code

The following python function is used to predict the hand gestures of ISL. In this, we used opencv and tensorflow frame work to design the proposed model. Initially for every alphabet, some integer is assigned.

```
  gestures = {     1:'A',     2:'B',     3:'C',     4:'D',
5:'E',   6:'F',     7:'G',     8:'H',     9:'I',     10:'K',
11:'L',      12:'M',      13:'N',      14:'O',      15:'P',
16:'Q',          17:'R',          18:'S',          19:'T',
20:'U',21:'V',    22:'W',    23:'X',    24:'Y'}
  def predict(gesture):
      img = cv2.resize(gesture, (50,50))
      img = img.reshape(1,50,50,1)
      img = img/255.0
      prd = model.predict(img)
      index = prd.argmax()
      return gestures[index]
```

5 Results and Discussions

The results of this proposed method will give alphabets as output corresponding to the captured hand gestures. We can use various deep learning algorithms to predict this output, but accuracy differs from one algorithm to other. Here, we used deep convolution neural networks in order to predict more accurate output. Here, the output will be alphabets that relate to the hand gestures captured. As shown in Fig. 7, if there is no hand in front of camera then there will be no output and if any symbol that matches with the data set is there in front of camera as shown in Fig. 8 then the corresponding alphabet is shown.

In our proposed work, Convolution Neural Network and Adam optimizer with learning rate 0.01 and dropout 0.25 is used. Accuracy of proposed system is as shown in Fig. 9.

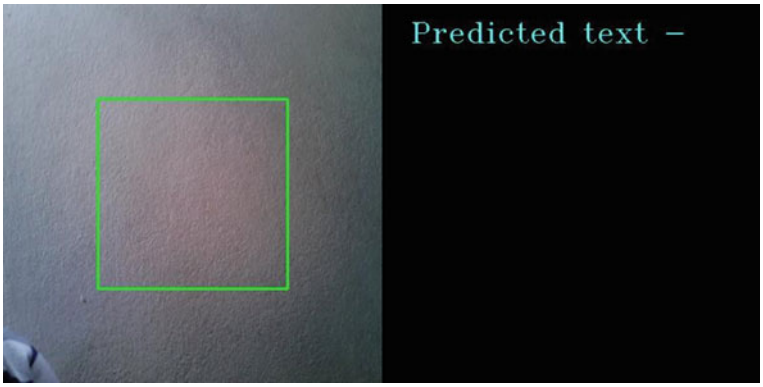


Fig. 7 Proposed method predicted result

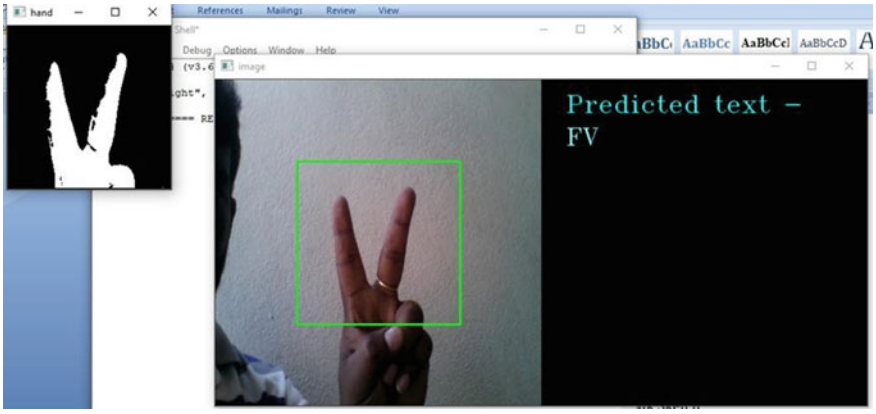
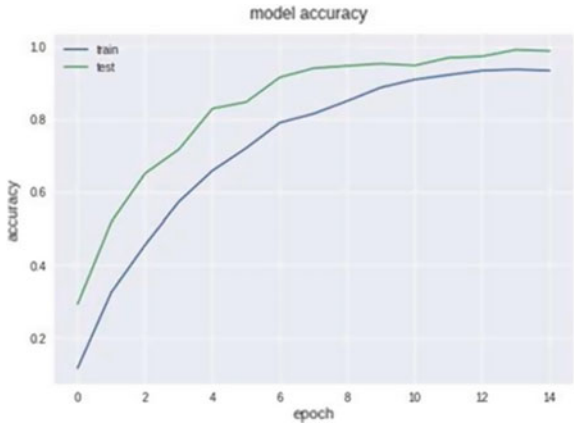


Fig. 8 Proposed model predicts as letter V

Fig. 9 Accuracy of proposed model



6 Conclusion and Future Work

In this paper, the sign language system is proposed using the convolution neural network algorithm after researching about various algorithms. Although our proposed task is to recognize only the alphabets, it is 98.74% accurate which is higher than existing systems. The proposed system is now only suitable for ISL signs for alphabets only. It is not suitable for numbers, body movements, sentences and facial expressions. In future, it will be extended to work with different forms of signs. And also to improve the accuracy, we can adopt mixed breed clustering or classification techniques in future.

References

1. A.K. Sahoo, G.S. Mishra, K.K. Ravulakollu, Sign language recognition: state of the art. ARPN J. Eng. Appl. Sci. (2014)
2. J. Singha, K. Das, Indian sign language recognition using Eigen value weighted Euclidean distance based classification technique. Int. J. Adv. Comput. Sci. Appl. **4**, 2 (2013); N.V. Tavari, A.V. Deorankar, Indian sign language recognition based on histograms of oriented gradient. Int. J. Comput. Sci. Inf. Technol. **5**(3), 3657–3660 (2014)
3. NMANIVAS. Gesture recognition system. https://github.com/nmanivas/Gesture-Recognition-System
4. Z. A. Ansari, G. Harit, Nearest neighbour classification of Indian sign language gestures using kinect camera. Sadhana **41**(2), 161–182 (2016)
5. D. Deora, N. Bajaj, Indian sign language recognition, in *2012 1st International Conference Emerging Technology Trends in Electronics, Communication and Networking (ET2ECN)* (2012). https://doi.org/10.1109/ET2ECN.2012.6470093
6. C. Zhang et al., Multi-modality American sign language recognition, in *2016 IEEE International Conference on Image Processing (ICIP)* (2016). https://doi.org/10.1109/ICIP.2016.7532886