Available online at http://www.advancedscientificjournal.com
http://www.krishmapublication.com
IJMASRI, Vol. 1, issue 10, pp. 316 -322, Dec. -2021
https://doi.org/10.53633/ijmasri.2021.1.10.011

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY ADVANCED SCIENTIFIC RESEARCH AND INNOVATION (IJMASRI)

ISSN: 2582-9130 **IBI IMPACT FACTOR 1.5** **DOI: 10.53633/IJMASRI**

RESEARCH ARTICLE

A REAL – TIME SYSTEM FOR RECOGNITION OF AMERICAN SIGN LANGUAGE BY USING DEEP LEARINING

[1]Mohit Panwar, [1]Rohit Pandey, [1]Rohan Singla and [1]Kavita Saxena

[1]Computer Science and Engineering, Maharaja Agrasen Institute of Technology, Rohini, Delhi

Abstract

Every day we see many people, who are facing illness like deaf, dumb etc. There are not as many technologies which help them to interact with each other. They face difficulty in interacting with others. Sign language is used by deaf and hard hearing people to exchange information between their own community and with other people. Computer recognition of sign language deals from sign gesture acquisition and continues till text/speech generation. Sign gestures can be classified as static and dynamic. However static gesture recognition is simpler than dynamic gesture recognition but both recognition systems are important to the human community. The ASL American sign language recognition steps are described in this survey. There are not as many technologies which help them to interact with each other. They face difficulty in interacting with others. Image classification and machine learning can be used to help computers recognize sign language, which could then be interpreted by other people. Earlier we have Glove-based method in which the person has to wear a hardware glove, while the hand movements are getting captured. It seems a bit uncomfortable for practical use. Here we use visual based method. Convolutional neural networks and mobile ssd model have been employed in this paper to recognize sign language gestures. Preprocessing was performed on the images, which then served as the cleaned input. Tensor flow is used for training of images. A system will be developed which serves as a tool for sign language detection. Tensor flow is used for training of images.

Keywords: ASL recognition system, convolutional neural network (CNNs), classification, real time, tensor flow

Introduction

Sign language is a language that provides visual communication and allows individuals with hearing or speech impairments to communicate with each other or with other individuals in the community. According to the World Health Organization, the number of hearing-impaired individuals has recently reached 400 million. For this reason, recent studies have been accelerated to make disabled people communicate more easily. If studies in the literature are examined, Basic Component Analysis is used in the extraction of the feature vector in work done by Mahmoud Zaki and Samir Shaheen in 2011 and Hidden-Markov Model is used as the classifier . In the study conducted by Ching Hua et al. In 2014, motion sensors were used for feature extraction and k-NN and Support Vector Machine (SVM) for classification . In 2015, Cao et al. used the depth comparison feature of Microsoft Kinect to obtain feature vectors and used the random forest and constrained link angle algorithm to classify the obtained vectors . K-Nearest Neighbors (k-NN) Classifier was used in American Sign language recognition by Dewinta Aryanie and Yaya Heriadi in 2015. In the study made by Jin et al. In 2016, Speeded Up Robust Feature (SURF) algorithm for feature extraction and SVM as classifier were used .

There are many different sign languages for every language in the world. There are over two hundred languages in utilize around the world today such as the Chinese, Spanish, Irish, British, and American Sign Language (ASL), which is the most common sign language in the world. of thousands of words, including very alike hands poses, in addition to similarities between some signs. It is necessary to find another way to make communication possible between the majority of hearing communities and the deaf community. Automatic ASL recognition system is a new way of understanding the meaning of deaf signs without needing the help of expert.

This technique can be used to translate signs into sounds or texts based on the users‟ needs. Signs recognition system is still a difficult problem. Many researchers have tried hard to solve this problem because it requires the detection and recognition of the required information hands poses and hands movements.

The static gestures-based system is still difficult due to visual similarities in different signs. For example, the letters N and M appear to be identical, and are just distinguishable by the situation of the thumb. Likewise, there are enormous variations depending on the viewpoint of the camera. The advantages of a deep learning with CNNs were employed to solve this problem and achieve a real time and accurate sign fingerspelling recognition model.

In this article, an attempt was made to design a real-time and high performance translator for those who do not know the sign language. This study proposes a CNN (Convolutional Neural Network) structure for feature extraction and classifier, and then the hand locating process was applied to construct the real-time system.

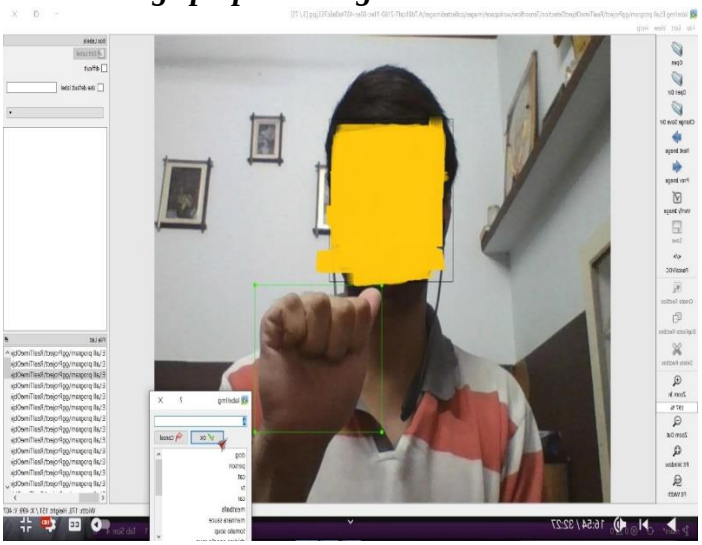
Image processing

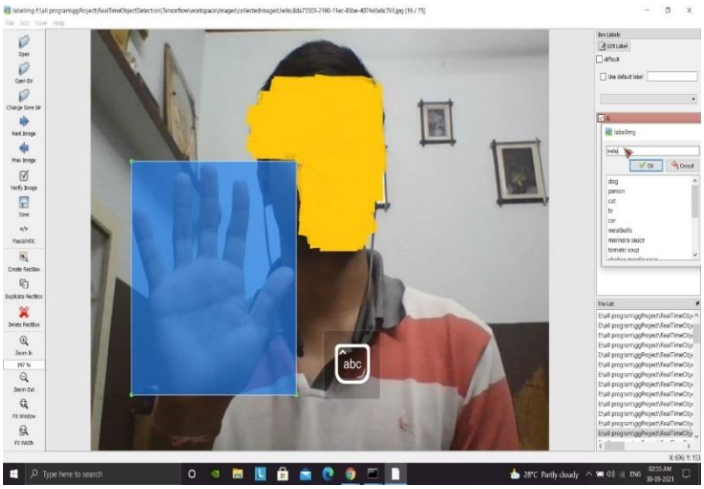
A. Introduction

Image processing is a method for performing various operations on a picture, to create an enhanced image and/or to extract some useful information from input information. It is a kind of signal during which input is a picture and output could be a picture or characteristics/features

related to that picture.

B. Image preprocessing





LabelImg tool

LabelImg is a graphical image annotation tool and label object bounding boxes in images. It is a free, open source tool for graphically labeling images. It's written in Python and uses QT for its graphical interface. It's an easy, free way to label a few hundred images to try out your next object detection project.

It is written in Python and uses Qt for its graphical interface. Annotations are saved as XML files in PASCAL VOC format, the format used by ImageNet. Besides, it also supports YOLO and CreateML formats. This tool is used here for image processing and training of model

Methodology

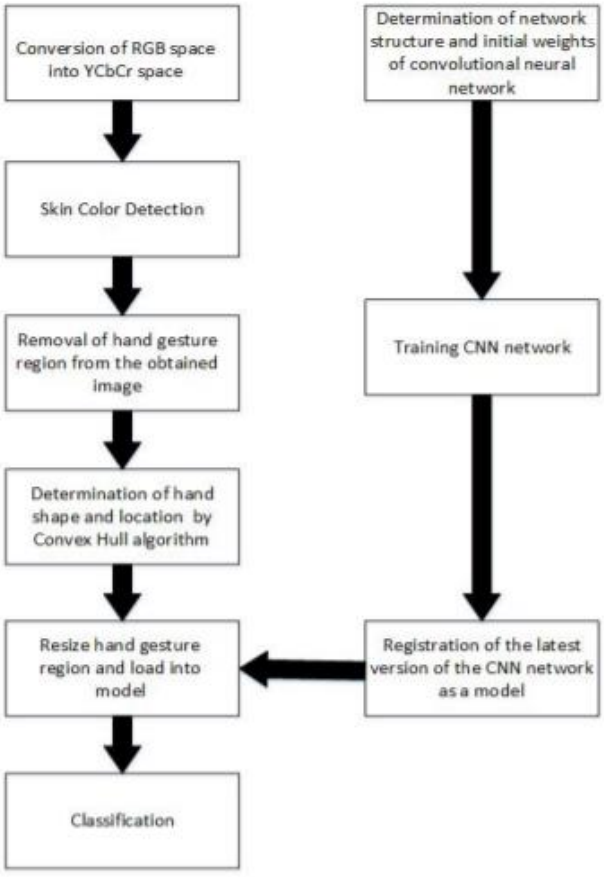
System architecture

In this research, a real-time ASL fingerspelling recognition was built with CNNs algorithm using real colouring images. It comprised a total of 26 alphabets, including J and Z, in addition to two classes for space and delete. This system was divided into three phases, and the first phase represents the collection of data. The methods of Hand-Gesture recognition explored by this research required a large dataset for training, so it has been decided to build new datasets that included a wider variety of features such as different lightings, different skin tones, different backgrounds, and a wide variety of situations. The second phase was a multi-class recognition with CNN, while the last phase was the writing system, which represented the communication between the computer and the user. This system facilitates the communication between the majority of hearing communities and the deaf community. It is an input system that uses a PC camera.

Convolutional Neural Network

Traditionally in recognition system, a classifier structure follows image processing and morphological process units. With the use of deep learning methods, it has become possible to classify in high performance by the minimum number of image processing and morphological processing steps.In this paper, convolutional neural network is used as a fine classifier with tensorflow and Keras libraries in Python. These libraries work efficiently on powerful modern GPUs (Graphics Processing Units) that allows doing much faster computation and training. In recent years, CNN based classifications and researches are very popular and have proven to be successful in areas like image classification and

recognition. Rectified Linear Unit (ReLu) is used as activation function, which makes converge much faster while still presents good quality.



- Dataset will be augmented to include a greater number of images for our training. In this step of data augmentation, we will rotate and flip each of the images in our dataset.
- Visualization of total images in the data set will be done.
- We divide our data into the training set which will contain the images on which the CNN model will be trained, and the test set with the images on which our model will be tested. And then we will build our Sequential CNN model with various layers such as Conv2D, MaxPooling2D, Flatten, Dropout and Dense.
- We will fit images in the training set using Kera"s library.
- We will be using the Haar Feature-based Cascade Classifiers for detecting the features of the face. This cascade classifier is designed by OpenCV to detect the frontal face by training thousands of images for a better accuracy and good results.
- In the last step, we will use the OpenCV library to run an infinite loop to use our web camera in which we detect the face using the Cascade Classifier. The resulting model will predict the possibility of each of the two classes (without mask, and with mask). If the probability of a masked face is higher, the model will highlight it with green box and if probability of unmasked face is detected higher, a red box will be used to highlight the results in the live camera.

Algorithm used and CNN

- Convolutional neural network (CNN) is a type of deep learning model for processing data that has a grid pattern, such as images, which is then further used in images recognition, images classifications, objects detections, recognition faces etc.
- During this proposed method, the sign language detection model is made using the TensorFlow and Keras libraries. this enables the user to form the new layers for the model step by step. The different layers used for our CNN model is described below.
- The primary layer is that the Conv2D layer with 100 filters and therefore the filter size or the kernel size of 3X3. ReLu function stands for Rectified long measure which may output the input directly if is positive, otherwise, it'll output zero.
- Subsequent layer is again a Conv2D layer with another 100 filters of an equivalent filter size 3X3.

- And in the next step, we use the Flatten layer to flatten all the layers into one 1D layer.
- After the Flatten layer, the Dropout (0.5) layer is used to stop the model from overfitting.

The SoftMax function outputs a vector which can represent the probability distributions of every of the input units.

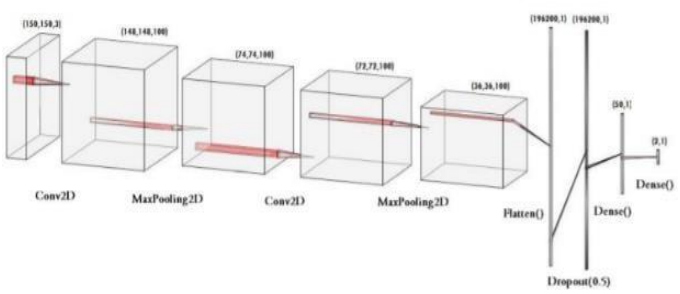


Fig. CNN working diagram

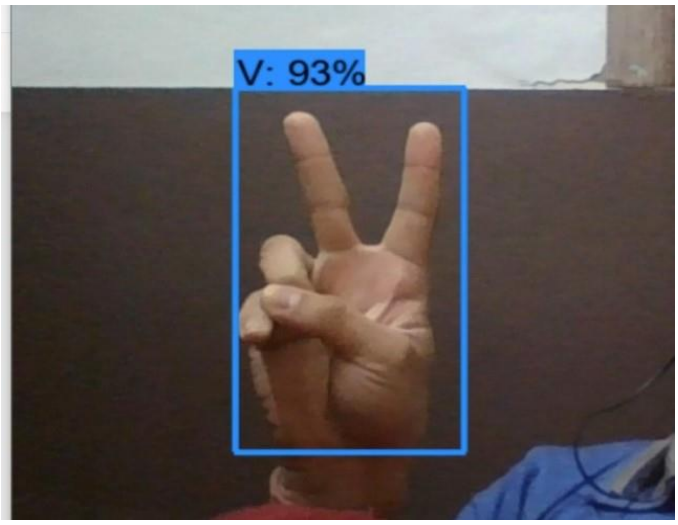
Result

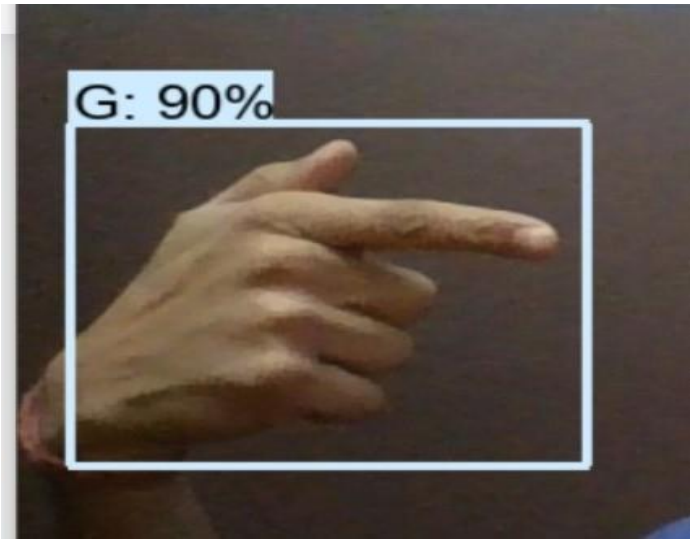
This model aims to demonstrate the future uses of the system, so tests are administered using the recall metrics, Precision, and therefore the corresponding macro average and weighted average the target of using these metrics is to gauge the model under various conditions. Recall and precision indicate the power of the model to properly detect true positives. Recall also considers the false negatives detected, and therefore the precision of the false positives detected by the model

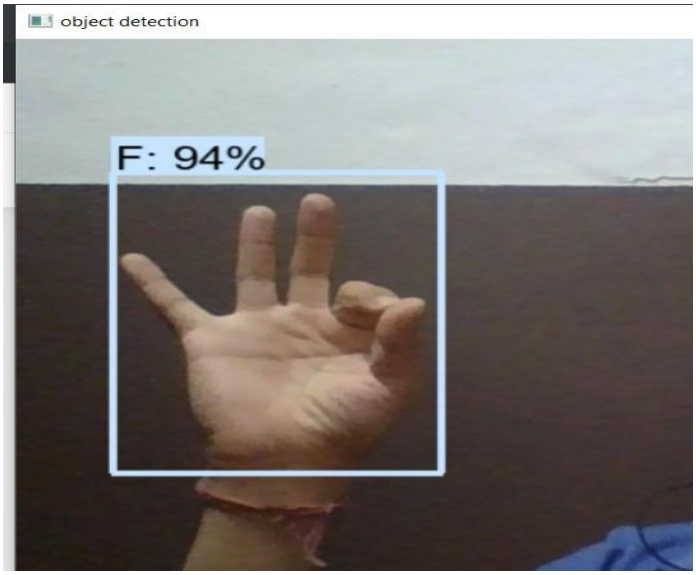
A real-time ASL fingerspelling recognition with a CNN algorithm using real colouring images from a PC camera was introduced. In this paper, deaf signs are translated into text statements to help creating a writing system that can be used as an input system for a computer using any computer camera. This system showed good results by taking advantage of a deep learning technique. This section discusses all the results that were obtained from the experiment. CNNs were used as the recognition system, in which each ASL sign was represented as an individual category.

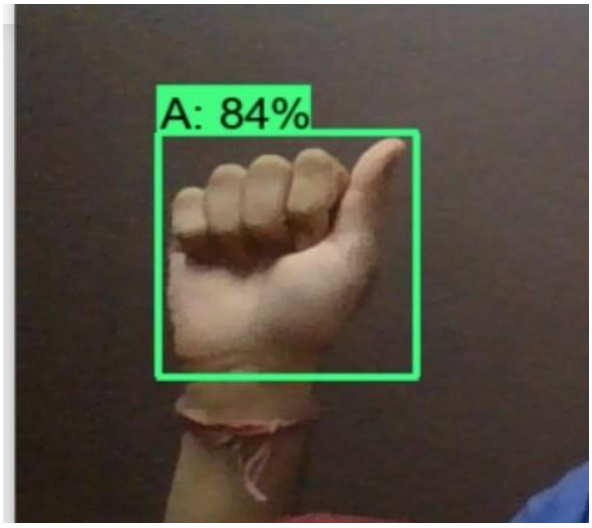
During a successful execution of the program, an accuracy of 99.88% was obtained using Deep Supervised Learning CNN. The model calculated the accuracy percentage and plotted the data in the accuracy vs epoch plot, and results were tested and found accurate as well. During the successful run, expected class matched with the correct class, showing that the model worked.

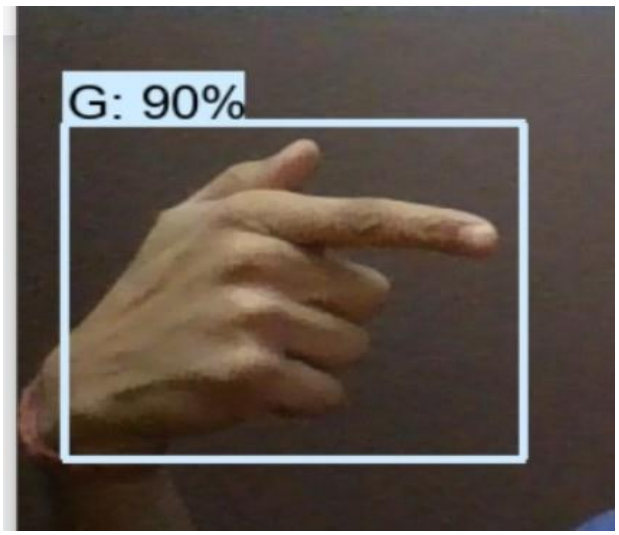
Given images shows the results given by our model











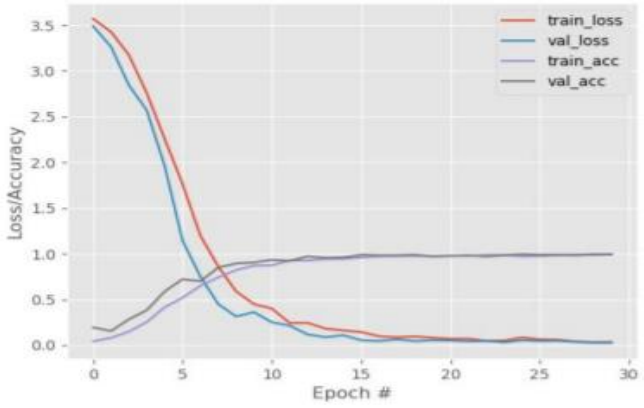


Fig. Training loss and accuracy on American Sign Language

During a successful execution of the program, an accuracy of 99.88% was obtained using Deep Supervised Learning CNN. The model calculated the accuracy percentage and plotted the data in the accuracy vs epoch plot, and results were tested and found accurate as well. During the successful run, expected class matched with the correct class, showing that the model worked.

The accuracy of this model when is run at 75 epochs, comes out to be 99.88%, which is one of the highest among recent work done in the existing models. Our model has the following numbers as the results, it has accuracy: 99.88%, specificity: 99.6%, sensitivity: 99.77%.

Conclusion

The comparison between the studies in the literature and the proposed method and given Proposed system has advantage in terms of test accuracy according to these similar studies. Also, the applicability as real-time system is validated. System achieved high accuracy even with the letters that have similar gestures.

The signs for all the alphabets from A to Z are being recognized using the combinational neural networks architecture. The advantage of using the algorithm is high processing speed which

can produce results in real-time manner. The speed of processing is increased due to the neural network architecture.

Future Perspective

This research work can be extended to recognize the rotation and distance invariant ASL Alphabets gestures, numbers gestures and other complex gestures in different background (plain and complex), location (indoor and outdoor), lighting conditions (day and night light) in real time environment. This research work can also be extended to recognize English words and sentences which need video processing.

For the future upgrades, this technique could be used on smart sensors where Cloud services could be used collect multimedia data. There could also be work done in developing mobile phone-based application which would be very handy.

For the future upgrades, this technique could be used on smart sensors where Cloud services could be used collect multimedia data. There could also be work done in developing mobile phone-based application which would be very handy.

Reference

1. Dong, C., Leu, M.C. Yin, Z. (2015). “American Sign Language alphabet recognition using Microsoft Kinect,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, USA, pp. 44–52.
2. Pansare, J.R. and Ingle, M. (2016). “Vision-based approach for American Sign Language recognition using Edge Orientation Histogram,” in Proc. International Conference on Image, Vision and Computing (ICIVC), Portsmouth, UK, 2016, pp. 86–90.
3. Joshi, A., Sierra, H and Arzuaga, E. (2017). “American sign language translation using edge detection and cross correlation,” in Proc. IEEE Colombian Conference on Communications and Computing (COLCOM), Cartagena, Colombia, 2017, pp. 1–6.
4. Raheja, J.L., Mishra, A and Chaudary, A. (2016). Indian Sign Language Recognition Using SVM 1. Pattern Recognition and Image Analysis. September; 26(2)
5. Liao, B., Li, J. Ju, H and Ouyang, G (218). "Hand Gesture Recognition with Generalized Hough Transform and DC-CNN Using Realsense," 2018 Eighth International Conference on Information Science and Technology (ICIST) , Cordoba,pp. 84-90.
6. Asha Thalange, Dr. S. K. Dixit, (2016). “COHST and Wavelet Features Based Static ASL Numbers Recognition”, 2 nd International Conference on Intelligent Computing, Communication & Convergence (Elsevier), pp.455-460.
7. Pei Xum (2017). “A real time hand gesture recognition and human computer interaction,” Dept. of Electrical and Computer Engineering, University of Minnesota, 2017, pp. 1-8.

*****