IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

Real Time Indian Sign Language Detection System

Nikhil Salvi[1], Qureshi Zaid[2], Nikhil Bhodke[3], Prof. Sanjay Jadhav[4]
Department of Computer Engineering[1,2,3,4]
Mahatma Gandhi Mission College of Engineering and Technology, Navi Mumbai, Maharashtra, India

Abstract: *Sign language, as a different form of the communication language, is important to large groups of people in society. There are different signs in each sign language with variability in hand shape, motion profile, and position of the hand, face, and body parts contributing to each sign. A challenging area of computer vision research is the recognition of visual sign languages. In recent years, deep learning techniques have significantly improved the many models that have been suggested by various academics. We examine the deep learning-based vision-based models for sign language recognition. An important advancement in enhancing communication between the deaf and the overall populace is a real-time sign language detector. We are glad to present the development and application of a model for recognizing sign language based on a Convolutional neural network (CNN). We utilized a pre-trained SSD mobile net V2 architecture trained on our own dataset in Order to apply transfer learning to the task.*

Keywords: Machine Learning, Python, OpenCV, Sign Language Recognition, TensorFlow..

I. INTRODUCTION

Numerous technological developments and much study have been made to benefit the deaf and dumb. Deep learning and computer vision are both tools that can be used to further the cause.

This can be extended to constructing automatic editors, where a person can easily write by using just their hand gestures, which can be very helpful for the deaf and dumb in interacting with others as understanding sign language is not something that is common to all.

Sign language is largely used by the disabled, and there are few others who understand it, such as relatives, activists, and teachers. Sign language detection is a project implementation for designing a model in which web camera is used for capturing images of hand gestures which is done by open cv. After capturing images, labelling of images are required and then pre trained model SSD mobile net v2 is used for sign recognition. Thus, an effective path of communication can be developed between deaf and normal audience.

Researchers are receiving increasing attention these days for creating commercially viable sign language recognition technology. Research is conducted in a variety of methods. It begins with the data collection techniques. The cost of a suitable device necessitates a variety of data collecting techniques, but a low-cost technique is required for the commercialization of a sign language recognition system.

Worldwide, deaf and hard-of-hearing persons mostly communicate using sign languages (Izzah &Suciati, 2014). It is the strongest and most successful method for bridging the social interaction and communication gap between them and the able-bodied persons. By converting sign language into spoken words and the other way around, sign language interpreters help close the communication gap with the hearing impaired. Though, the limitations of using interpreters are the variable structure of sign languages and the scarcity of skilled sign language interpreters worldwide (Kudrinko et al., 2021). Researchers have recently focused more on developing sign language.

Worldwide, deaf and hard-of-hearing persons mostly communicate using sign languages (Izzah &Suciati, 2014). It is the strongest and most successful method for bridging the social interaction and communication gap between them and the able-bodied persons. By converting sign language into spoken words and the other way around, sign language interpreters help close the communication gap with the hearing impaired. The flexible nature of sign languages and the scarcity of skilled sign language interpreters worldwide provide difficulties for the use of interpreters, however (Kudrinko et al., 2021).

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

In this article, we presented systems using the Python OpenCV and Keras libraries. In this project, we develop a sign detector that recognizes alphabets and can be readily expanded to recognize a huge variety of other signs and hand gestures, such as the numerals.

II. RELATED WORK

Sign languages are defined as an organized collection of hand gestures having specific meanings which are employed from the hearing impaired people to communicate in everyday life [3]. Being visual languages, they use the movements of hands, face, and body as communication mediums. There are over 300 different sign languages available all around the world [5]. Though there are so many different sign languages, the percentage of population knowing any of them is low which makes it difficult for the specially-abled people to communicate freely with everyone SLR provides a means to communicate in sign language without knowing it.

It recognizes a gesture and translates it into a commonly spoken language like English. SLR is a very vast topic for research where a lot of work has been done but still various things need to be addressed. The machine learning techniques allow the electronic systems to take decisions based on experience i.e. data. The classification algorithms need two datasets ‗ training dataset and testing dataset. The training set provides experiences to the classifier and the model is tested using the testing set [6]. Many authors have developed Indian Sign Language is a visual spatial efficient data acquisition and classification methods [3][7]. Based on data acquisition method, previous work can be categorized into language that was developed in India. Indian Sign Language is a natural language with its own two approaches: the direct measurement methods and the vision based approaches [3]. The direct measurement methods are based on motion data gloves, motion capturing systems, or sensors.

video source that contains signing gestures is converted into gray-scaled frames, from which features are extracted by applying a directional histogram. Lastly, clustering is used to classify the signs, according to their features, into one of the predefined classes. The authors achieved a 100% sign recognition rate in their study and concluded that the 36-bin histogram method was more accurate than the 18-bin histogram method.

The existing SLR system involves - Hardware modules, Flex Sensors, Immobile sensory equipment and kits. The implemented canteen SLR system is a real time recognition system, end to end communication, easy to interface, flexible and has minimal software specifications and requirements. The different tech stacks used by similar Sign Language recognitions systems are Python, TensorFlow, OpenCV, Keras, NumPy and various machine learning algorithms.

III. PROPOSED SYSTEM

In this article, we presented systems using the Python OpenCV and Keras libraries. In this project, we develop a sign detector that recognizes alphabets and can be readily expanded to recognize a huge variety of other signs and hand gestures, such as the numerals. This project divided into 3 parts:

3.1 Creating the dataset

The dataset we require is reasonably easy to find online, however for this project, we will be building the dataset ourselves.

Every frame that recognizes a hand in the created ROI (region of interest) will be saved in a directory (here gesture directory) that has two folders, train and test each of which has ten folders containing images takenusing the create_gesture_data.py program

The directory structure, phonology, morphology, and grammar must first be developed. It employs body and head movements, hand gestures, face expressions, and arm motions to produce semantic information that conveys words and emotions.

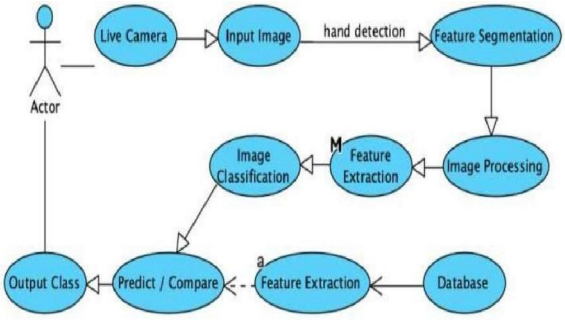
Indian Sign Language motions can be detected and recognized from grayscale photos using the method Nandy et al. provided. A train and a test have the same directory in their approach. Now, in order to construct the dataset, we use OpenCV to obtain the live camera feed and create a ROI, which is just the area of the frame where we want to identify hands for gesture detection. The ROI is indicated by the red box, and this window is used to access the webcam's live video.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

We compute the background's accumulated weighted average and subtract it from the frames that have a distinguishable foreground object in front of the background in order to distinguish between the background. This is accomplished by calculating the accumulated average for the backdrop after accumulating the weight for a subset of frames (in this case, 60 frames).

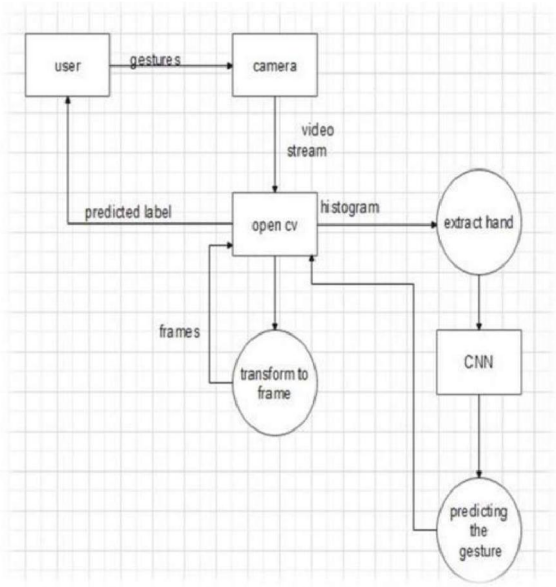
When we know the backdrop's cumulative average, we can locate any object that obscures the background by deducting it from each frame we read after the first 60 frames.

(We put up a text using cv2.putText to display to wait and not place any item or hand in the ROI while detecting the background).



3.2 Calculate threshold value :

Now, we use cv2 to identify the contours for each frame and calculate the threshold value.Utilizing the function segment, findContours returns the max contours (the object's outermost contours). We can tell if there is a hand in the ROI by looking at the contours to see if any foreground objects are being picked up there. For the letter or number we are detecting it for, we begin to save the picture of the ROI in the train and test sets, respectively, when contours are recognized (or a hand is present in the ROI).



For the train dataset, we save 30 images for each alphabets to be detected, and for the test dataset, we do the same and create 20 images for each alphabets.

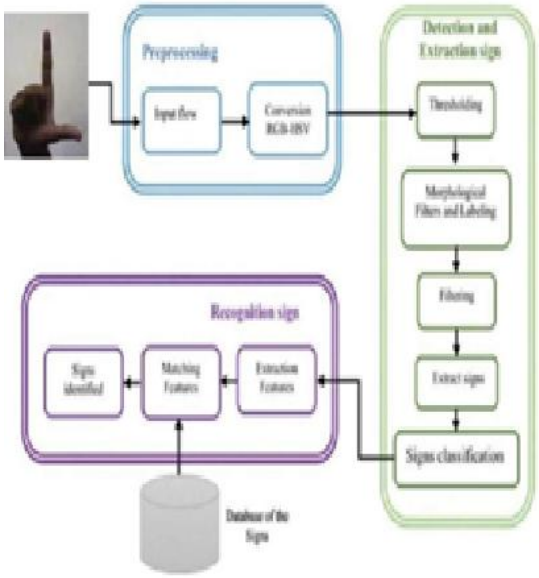
Below are the example of data we have created fully trained and test after calculating region of interest and threshold value.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

3.3 Training CNN

We now train a CNN using the newly produced data set. In order to load the train and test sets of data, we first load the data using the Image Data Generator of Keras. Each of the names of the number folders will be the class names for the loaded images. The Plot Images function is used to plot images from the imported dataset. Next, we fit the model and save it so that it can be used in the final module (model_for_gesture.py). Alternate hyperparameters may be used, depending on some trial and error. Call backs of Reduce LR on plateau and early halting are employed in training, and both of them depend on one another on the loss in the validation dataset.

The validation dataset is used to calculate the accuracy and loss for each epoch. If the validation loss is not decreasing, the model's LR is reduced using the Reduce LR function to prevent the model from overshooting the loss minima. Additionally, we are using the early stopping algorithm so that the training is stopped if the validation accuracy continues to decline for a number of epochs. The example includes the callbacks as well as the two alternative optimization methods used, Adam (a combination of Adagrad and RMSProp) and stochastic gradient descent, which updates the weights at every training instance.



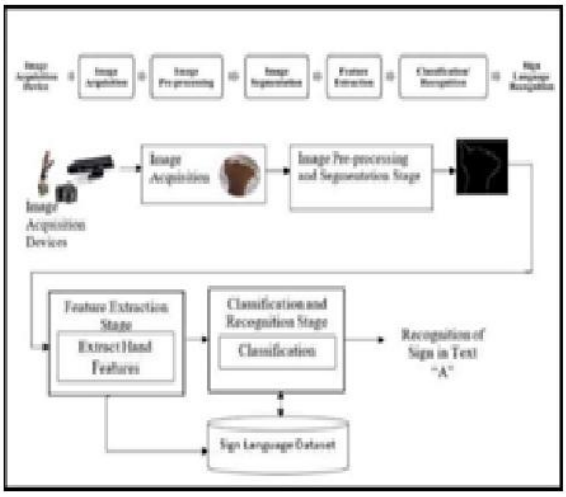
We discovered that the SGD model appeared to have improved accuracy. As you can see, the model had a training accuracy of 100% and a validation accuracy of roughly 81%.While detecting on the live cam feed, we are visualizing and testing the model in this instance to see if everything is operating as expected. The label names for the various expected labels are contained in the word dictionary. The test and train folders are taken into account by the Image Data Generator based on the order of the folders inside the test and training datasets.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

3.4 Predict the data/gestures

In doing so, we establish a bounding box for the ROI and determine the accumulated average. Using the call backs stated above, we fit the model using the train batches after compiling it for 10 epochs (the number of epochs may vary depending on the user's parameter selection). Using the call backs stated above, we fit the model using the train batches after compiling it for 10 epochs (the number of epochs may vary depending on the user's parameter selection).

Now that the model has been imported, several of the relevant variables have been initialized, including the backdrop variable and the dimensions of the ROI. This function computes the background accumulated weighted average of the dataset. Segmenting the hand is the process of detecting the hand's maximal contours and threshold image.



The effectiveness of models for sign recognition based on the work we did to create the dataset. To recognize any foreground object, this is done.

Now that the maximum contour has been determined and a hand has been identified, the ROI's threshold is handled as a test image. Using keras, we load the model that was previously saved. Models are loaded, and the threshold picture of the hand-containing ROI is sent into the model as an input for prediction.

Table 5 lists the imports required for the background photos in model_for_gesture.py. Although ResNet18 and EfficientNet_B1 had more trainable parameters, MobileNet_V2 outscored them in this technique. The ResNet18 and MobileNet_V2 had above 99% accuracy, precision, sensitivity, and F1_score combined. The specificity of the models ResNet18, Mobile Net V2, and Efficient Net B1 is 100%, 100%, and 99.98%, respectively, demonstrating that they have a very low false alarm rate. Efficient Net B1, while having more parameters than Mobile Net V2, performed the worst out of the three CNNs utilized to solve this sign identification challenge. Efficient Net, however, has an overall accuracy precision, sensitivity, and F1 score of above 98%, indicating that it is not the model with the best performance despite being the deepest of the three networks for sign recognition.

IV. METHODOLOGY

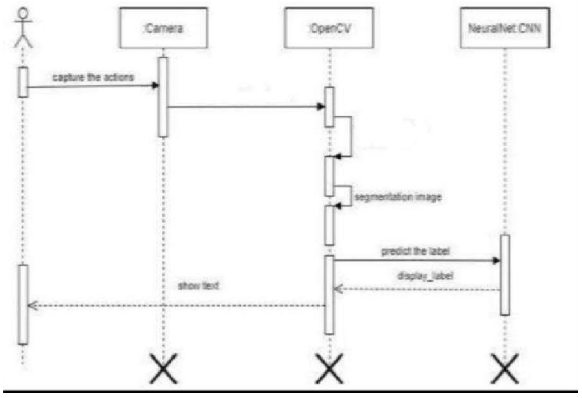
In order to provide text and audio output for the illiterate, our initiative seeks to record sign language while it is being used by signers in real-time. Due to the ease of portability and movement that the camera-based method offers over other techniques, a camera-based approach will be used for this. A camera-capable device will first record a video of the signer. Then, our application will process this video. The video would be split up into several frames, turning it into a raw visual sequence. The boundaries will then be initially determined by processing of this image sequence. This will help to divide the many body components being photographed by the camera into two main subdivisions. hands and head. The head component will be further divided into the categories of position, motion, and facial expression. The movement of the hands will be used to deduce postures and gestures. The WLASL Dataset will then be used to classify all of the data after it has been matched against it. Words will be created as a result of the categorization.

The proposed model can be used in a variety of ways, but its primary application is as a captioning tool for calls including video communication, like Facetime. The model would have to be running frame-by-frame, predicting what

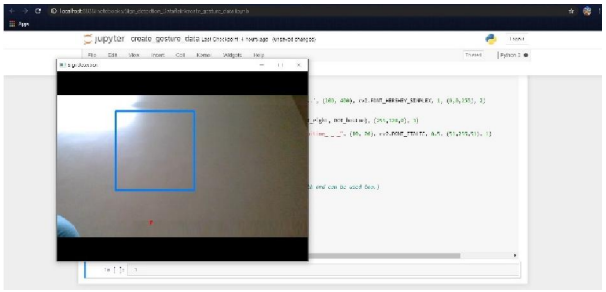
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

sign is being presented at all times, in order to construct such an application. In order to more properly assess the words being presented using ASL, we can also recognize whether a person is not exhibiting a sign or is switching between signs using other techniques. A completely functional sign language to text translator might be created using this technique to connect the letters that are displayed into words and even sentences. For people with hearing impairments, a gadget like this would make it much easier for them to take use of virtual communication's advantages. As demonstrated, the model is precise. forecasts the character that the camera is currently showing. The program additionally shows the CNN Keras model's classification confidence along with the predicted character.

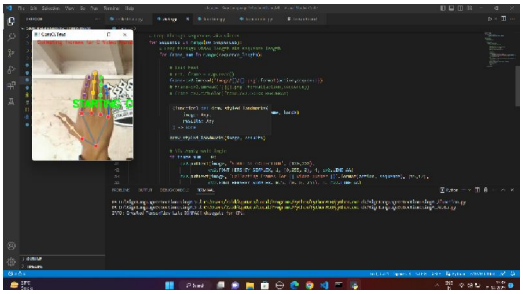
This project's ability to transform phrases with a 5 millisecond time delay between each word is an additional crucial feature.



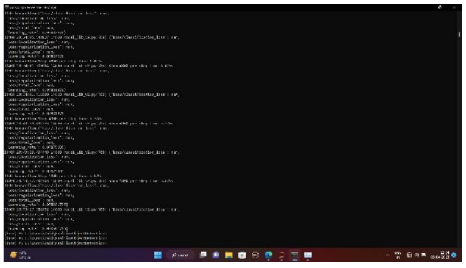
V. RESULTS



Creating ROI.



Detecting hand points.



Training Model.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

VI. CONCLUSION

Intelligent systems for sign language recognition continue to garner academic interest in light of current developments in machine learning and computational intelligence techniques. Thus, it can be said that more attention needs to be paid to the uncontrolled environment setting in order to make the vision-based gesture recognition system ready for real-life application. This is because it can give researchers the chance to improve the system's ability to recognise hand gestures in any type of environment.

VII. FUTURE SCOPE

For ISL word and sentence level recognition, we can create a model. A system that can recognise changes in the temporal space will be needed for this. By creating a comprehensive offering, we can bridge the communication gap for those who are deaf or hard of hearing. The deaf community in India uses Indian Sign Language (ISL). However, ISL is not utilised in deaf schools to instruct deaf students. ISL-focused teaching strategies are not emphasised in teacher training programmes. There is no curriculum that includes sign language

REFERENCES

[1]. R.H. Abide, M. Arslan, J.B. Iroko: Sign language translation using deep convolutional neural networks KSII Transactions on Internet and Information .

[2]. Systems, 14 (2) (2020).R. Sharma et al. Recognition of Single Handed Sign Language Gestures using Contour Tracing descriptor. Proceedings of the World Congress on Engineering 2013 Vol. II, WCE 2013, July 3 - 5, 2013, London, U.K ChandraKarmokar, B.; Alam, K.M.R.; Siddique, M.K. Bangladeshi Sign Language Recognition Employing Neural Network Ensemble. Int. J. Comput. Appl. 2012, 58, 43 46. [CrossRef].

[3]. Kang, Byeongkeun, Subarna Tripathi, and Truong Q. Nguyen. "Real-time sign language fingerspelling recognition using convolutional neural networks from depth map." Pattern Recognition (ACPR), 2015 3rd IAPR Asian Conference on. IEEE, 2015.

[4]. LOKHANDE, Priyanka; PRAJAPATI, Riya; PANSARE, Sandeep. Data gloves for sign language recognition system. International Journal of Computer Applications, 2015, 975: 8887.

[5]. HORE, Sirshendu, et al. Indian sign language recognition using optimized neural networks. In: Information technology..