Real-time Sign Language Detection

Prof. Kapil Mundada, Varad N. Somawar.

Department of Instrumentation and Control

Vishwakarma Institute of Technology, Pune.

Abstract: **There is a need of a method or an application that can recognize sign language gestures so that the communication is possible even if someone does not understand sign language. With this work, we intend to take a basic step in bridging this communication gap using Sign Language Recognition. In this project Deep-Learning approach was used for model training to recognise the signs used in real time. By using TensorFlow and OpenCV we can do the detections in real time. We make use of Convolution Neural Network (CNN) for training and to classify the images of the 5 gestures for the phrases used in American sign language. And the average detection rate in real-time is above 80%.**

Keywords: *Sign language, Deep-learning, TensorFlow, OpenCV, Convolution Neural Network (CNN).*

1. Introduction

Sign language has been constantly used to communicate with deaf people. But there's still a communication gap is present because of lack of knowledge of sign language. So in this project we tried to recognize 5 signs of American Sign Language (ASL) using Machine learning. ASL is globally adopted sign language and it has a variety of signs for the phrases we use in daily life. And it will be easier to communicate if there is a model which can recognize these signs in real time and this will greatly reduce the communication gap. In this project we have used Jupyter notebook for coding. So, in this project we have used TensorFlow for the object detection purpose and OpenCV for the capturing and processing images. So, in proposed system the dataset was created and using CNN trained the model to classify those images.

2. Existing Literature

There are many projects available for this problem statement, but most of them are for recognizing the sign language alphabets and digits or some projects are recognizing their local sign language.

First paper we studied was by Siming He[1]. He proposed a system having a dataset of 40 common words and 10,000 sign language images. To locate the hand regions in the video frame, Faster R-CNN with an embedded RPN module is used. It improves performance in terms of accuracy. Detection and template classification can be done at a higher speed as compared to single stage target detection algorithm such as YOLO. The detection accuracy of Faster R-CNN in the paper increases from 89.0% to 91.7% as compared to Fast-RCNN.

Then there is research doneby Rekha, J[2]. which made use of YCbCr skin model to detect and fragment the skin region of the hand gestures. They used principle of curvature and extracted image features and classified with Multi class SVM, DTW and non-linear KNN. For training dataset of 23 Indian Sign Language static alphabet signs was used and 25 videos for testing were used. The experimental result obtained were 94.4%for static and 86.4% for dynamic.

Then another one is Real-Time Sign Language Gesture (Word) Recognition from Video Sequences Using CNN and RNN by Sarfaraz Masood, Adhyan Srivastava, Harish Chandra Thuwal and Musheer Ahmad[3]. To train the model on spatial features, they used inception model which is a deep convolutional

neural network (CNN) and recurrent neural network (RNN) to train the model on temporal features. Our dataset consists of Argentinean Sign Language (LSA) gestures, belonging to 46 gesture categories. The proposed model was able to achieve a high accuracy of 95.2% over a large set of images.

Then let's take a look at work done by by Rung-Huei Liang, Ming Ouhyoung[4]. They implemented a prototype system with a lexicon of 250 vocabularies in Taiwanese Sign Language (TWL). This system uses HMMs (Hidden Markov Models) for 51 fundamental postures, 6 orientations, and 8 motion primitives. Their average rate of detection was 80.4%.

Then another one is done by Mehreen Hurroo , Mohammad Elham[5]. Their work was about Sign Language Recognition System using Convolutional Neural Network and Computer Vision they were able to detect 10 American sign gestures with the accuracy of 90%.

3. Methodology

The first part in the proposed methodology is to collect the data. Many researchers have been using different type of sensors and camera for data collection, in this system web cam was used to capture images. So, there were 15 images captured for each sign and they were further divided into 80%for testing and 20% for training. Further feature extraction was done using labelling the gesture from the image using labelImg package in python. Then CNN was applied for training and classification. Then the model would be able to predict the signs after evaluation.

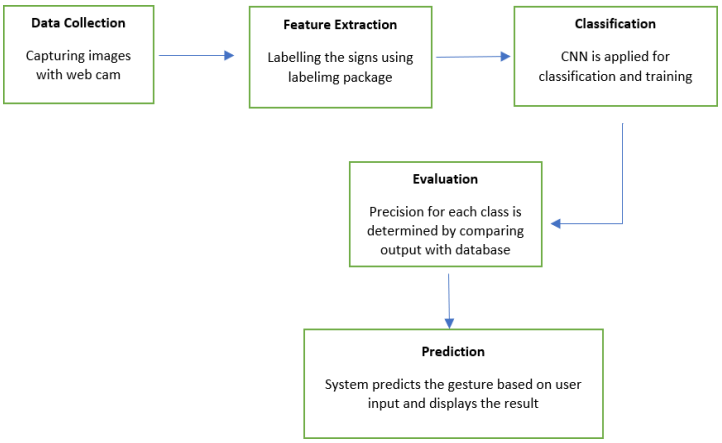

Fig1. Block diagram of system architecture

1) Data collection: Data collection is an important part of any system as the performance of the system is depended on it. For this model the images were captured using webcam a total of 50 images were captured out of which 80% was for training and 20% was for testing. The images were collected using the same background and were written in jpg format.

2) Feature extraction: One of the crucial parts of data processing is to extract the features we need from an image data. There is only a specific part of the image that we need for this model. So, that part was done by labelling that part using the LabelImg package. So this package generates the file in xml format which has the specific part we need.

3) Classification: In this proposed method, a 2-D CNN is applied with TensorFlow library. The layers of the CNN scan the images with filters. This particular step extracts important features and pass them on further.

4) Software Structure of System

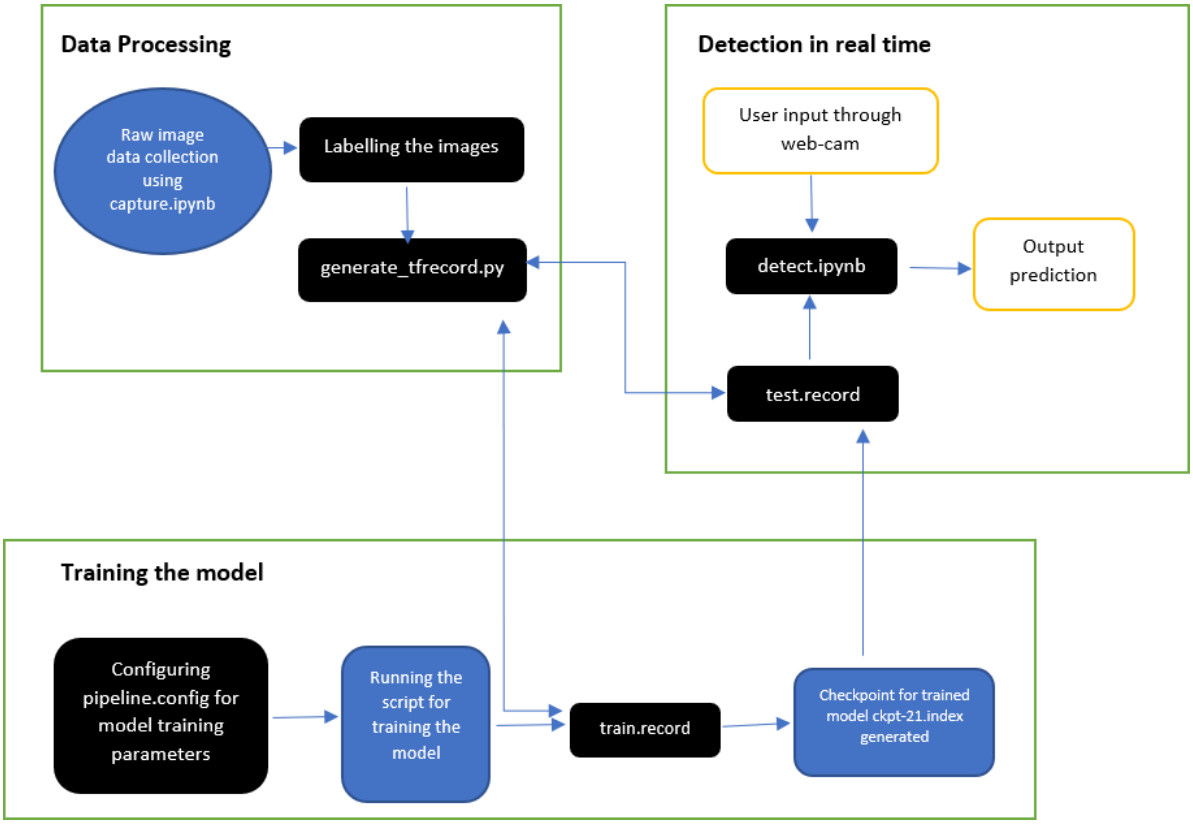

Fig2. Block diagram of software Architecture

4. Experimentations and Results

After running this project the detection was successful. Model was able to detect all the 5 signs precisely with speed.

Fig3. Image labelling

Fig4. Detection in real-time

5. Conclusion

It can be concluded that this project will help in closing the communication gap. Also, this will remove the need of mediator or translator as this project will serve as the mediator while communicating. This model was able to detect all the signs with average accuracy of 80% and the highest precision was 100% which was relatively better than the previous work.

6. Future scope

Right now this model works on image sequencing only in future we plan to make this model for video sequencing as well and also, we will try to give speech as output of the detected sign so that blind people can take benefit of this system.

References

1. He, Siming. (2019). Research of a Sign Language Translation System Based on Deep Learning. 392-396. 10.1109/AIAM48774.2019.00083.
2. International Conference on Trendz in Information Sciences and Computing (TISC). : 30-35, 2012.
3. sign-language-gesture-recognition by Sarfaraz Masood, Adhyan Srivastava, Harish Chandra Thuwal and Musheer Ahmad. doi:10.1007/978-981-10-7566-7_63
4. A real-time continuous Gesture recognition system for sign language by Rung-Huei Liang, Ming Ouhyoung. DOI:10.1109/AFGR.1998.671007
5. Sign Language Recognition System using Convolutional Neural Network and Computer Vision. IJERTV9IS120029