3 Results and Analysis

3.1 For SSD Mobilenet version-2:

The Mobile Net SSD model is a single-shot multibox detection (SSD) network that examines the pixels of an image that are within the bounding box coordinates and class probabilities to perform object detection. The model's architecture is based on the concept of an inverted residual structure, as opposed to traditional residual models. In this scenario, the input and output layers of the residual block are bottlenecked. Additionally, nonlinearities in intermediate layers are minimized, and a lightweight depthwise convolution is employed. This model is included into the TensorFlow object detection API. [11] There was a range of accuracy of 88% to 93% for this real-time sign language detection model.

|  |  |  |  |
| --- | --- | --- | --- |
| Number of Image Used for Training | True Result | False Result | Accuracy (%) |
| 50 | 24 | 26 | 48 |
| 100 | 54 | 46 | 54 |
| 200 | 144 | 56 | 72 |
| 500 | 435 | 65 | 87 |

Table 3.1:- The Accuracy Table for SSD Mobilenet v-2



Figure 3.1: Real time Sing Language Detection Using SSD Mobilenet Version-2

3.1 For YOLO version-5:

There was a range of accuracy of 88% to 93% for this real-time sign language detection model. YOLO is an abbreviation for the term 'You Only Look Once'. This is an algorithm that identifies and recognizes different objects in a photograph (in real-time). YOLO performs object detection as a regression problem and returns the class probabilities of the recognized images. To detect objects in real-time, the YOLO algorithm employs convolutional neural networks (CNN). To detect objects, the algorithm requires only one forward propagation through a neural network, as the name implies. This means that the entire image is predicted in a single algorithm run. CNN is used to predict multiple class probabilities and bounding boxes at the same time [12]. As a result of this real-time sign language detection model, there was a range of accuracy of 90\% to 98\% using the YOLOv5 algorithm.

|  |  |  |  |
| --- | --- | --- | --- |
| Number of Image Used for Training | True Result | False Result | Accuracy (%) |
| 50 | 24 | 26 | 48 |
| 100 | 56 | 43 | 56 |
| 200 | 152 | 48 | 76 |
| 500 | 446 | 54 | 89.2 |

Table 3.1:- The Accuracy Table for YOLO Algorithm Version-5

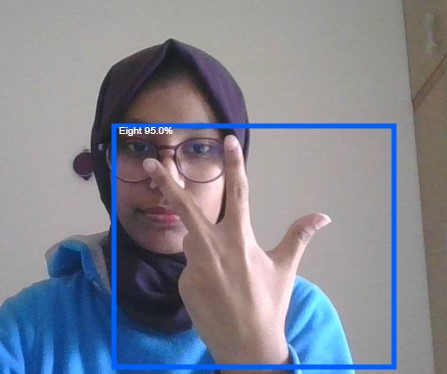


Figure 3.1: Real time Sing Language Detection Using YOLO Algorithm Version-5

4 Conclusion

4.1 Discussion and conclusion

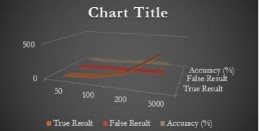


Figure 3.1: Accuracy Graph for SSD Mobile net

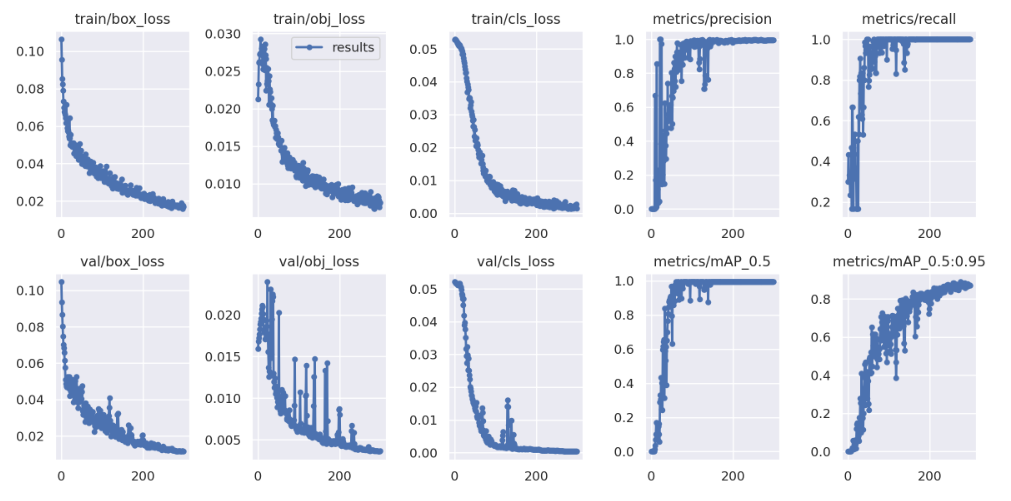


Figure 3.2: Accuracy and object loss Graph for YOLOv5

The primary objective of a sign language detecting system is to facilitate communication between able-bodied and deaf individuals through the use of hand gestures. The suggested system is accessible through webcam or any other built-in camera that detects and processes the indicators for recognition. We may conclude, based on the model's outcome, that the suggested system can provide precise results under controlled light and intensity. Additionally, it is simple to add custom gestures, and photographs captured from a variety of angles and frames will increase the model's precision. Increasing the dataset makes it simple to scale up the model to a huge size. The model is limited by environmental constraints such as low light intensity and an unmanaged background, which reduce the detection's accuracy. Consequently, we will next endeavor to eliminate these defects and expand the dataset for more precise results.

Abstract

A Real Time Sign Language detector is a huge step towards creating a bridge between the deaf, dumb and the general population. Computer recognition of sign language begins with the acquisition of sign gestures and progresses to the generation of text or speech. There are two types of sign gestures: static and dynamic. Although static gesture recognition is easier than dynamic gesture recognition, both systems are useful to the human community. It is a pleasure and honor to demonstrate and implement such a system that can detect sign language using a convolutional neural network. To implement Transfer learning, we utilized a Pre-Trained SSD Mobile net V2 architecture, and YOLO (You only look once) version-5 trained on our own dataset. We created a robust classification system that consistently classifies sign language in the majority of instances. Furthermore, this strategy will be extremely beneficial to sign language learners in terms of sign language practice. Throughout the project, various human-computer interface methodologies for posture recognition were investigated and evaluated. The most effective approach was determined to be a series of image processing techniques with human movement classification. Dataset consists of 50 different classes, including Bangla Alphabets, Numbers, Bangladesh, Dhaka, and Salam. Without a controlled background and low light, the system can recognize selected Sign Language signs with an accuracy of 88-93% using SSD Mobilenet V2 model and 90%--98% using YOLO V5 Algorithm.

iii