# Single Shot Detector

| Name | ID |
| --- | --- |
| Md. Zahidul Haque | 18-36130-1 |
| Ahmed, Faisal | 18-36639-1 |
| Protiva Haque | 18-37648-1 |
| Nishat, Asif Mostofa Md | 18-39259-3 |

## Abstract:

This report represents an implementation of Single Shot Detector (SSD) model. Deep learning is emerging as a powerful tool and has become a leading machine learning tool in computer vision and image analysis. SSD is designed for object detection in real-time. Faster R-CNN uses a region proposal network to create boundary boxes and utilizes those boxes to classify objects. The implementation and working principle have discussed in this report.



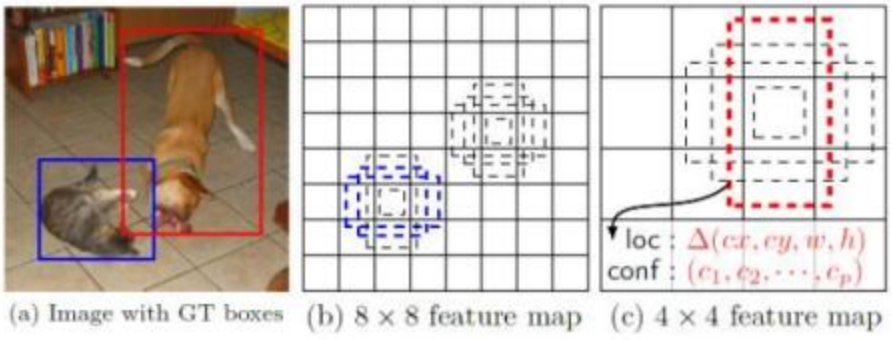
**Keywords:** Deep learning, Computer Vision, Object detection, SSD


## Introduction:

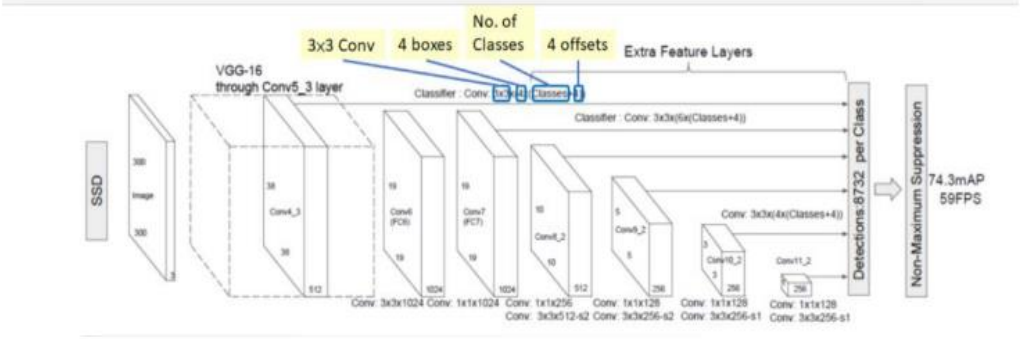
SSD is designed for object detection in real-time. Faster R-CNN uses a region proposal network to create boundary boxes and utilizes those boxes to classify object. While it is considered the start of-the-art in accuracy, the whole process runs at 10 frames per second. Real-time processing needs SSD speeds up the process by eliminating the need for the region proposal network. To recover the drop in accuracy SSD applies a few improvements including multi-scale features and default boxes. These improvements allow SSD to match the Faster R-CNN accuracy using lower resolution images, which further pushes the speed higher. According to the following comparison, it achieves the real-time processing speed and even beats the accuracy of the Faster R-CNN.

## Related Works:

1. Single Shot Detector is a name like YOLO providing results through a single shot. Proposed regions are not chosen in the CNN. A countable number of features extraction is processed in the way and a particular area is obtained in some feature layers. In the next step, CNN layer is applied in the extracted feature layer.
2. At first in our image processes through some convolutions to feature extraction. Through this step, we get a feature layer which is m*n*p our Image will be divided into different size scaled boxes with different aspect ratio.
3. Each of the bounding boxes are computed c class scores and offset relative with the original default bounding box shapes and then we get output.
4. Now every training image is sampled randomly with original input image and randomly sample a patch. Now each patch is of [0,1,1] or original size.
5. Now SSD may not be an efferent model but it sure gives us more accuracy and truth. If we compare with YOLO, it gives us efficiency but fails to be more accurate.




(a) Image with GT boxes (b) $8 \times 8$ feature map (c) $4 \times 4$ feature map




## Proposed Model:

This section presents our proposed approach for detecting the objects in real-time from images by using convolutional neural network deep learning process. The previous algorithms such as CNN, faster CNN, faster RCNN, YOLO, and SSD are only suitable for highly powerful computing machines and they require a large amount of time to train. In this paper, we have tried to overcome the limitations of the SSD algorithm by introducing an improved SSD algorithm with some improvement. The proposed scheme uses improved SSD algorithm for higher detection precision with real-time speed. However, SSD algorithm is not appropriate to detect tiny objects, since it overlooks the context from the outside of the boxes. To address this issue, the proposed algorithm uses depth-wise separable convolution and spatial separable convolutions in their convolutional layers. Specifically, our proposed approach uses a new architecture as a combination of multilayer of convolutional neural network. The algorithm comprises of two phases. Firstly, it reduces the feature maps extraction of spatial dimensions by using resolution. Second, it is designed with the application of small convolutional filters for detecting objects by using the best aspect ratio values. The major objective during the training is to get a high-class confidence score by matching the default boxes with the ground truth boxes. The advantage of having multiple layers leads to significant results in detection. Single shot detector was discharged at the tip of Gregorian calendar month 2016 and thus arrived at a new set of records on customary knowledge sets like Pascal VOC and COCO. The major problem with the previous methods was how to recover the fall in precision, for which SSD applies some improvements that include multi-scale feature map and default boxes. For detecting a small object with higher resolutions, feature maps are used. The training set of improved SSD

algorithm depends upon three main sections, i.e., selecting the size of box, matching of boxes, and loss function.

## Results and Discussion:

The accuracy of the proposed model is more than 79.8%. The training time for this model is about 5–6 h. These convolutional neural networks extract feature information from the image and then perform feature mapping to classify the class label. The prime objective of our algorithm is to use the best aspect ratios values for selecting the default boxes so that we can improve SSD algorithm for detecting objects.

## Conclusion:

This study develops an object detector algorithm using deep learning neural networks for detecting the objects from the images. The research uses am improved SSD algorithm along with multilayer convolutional network to achieve high accuracy in real time for the detection of the objects. The performance of our algorithm is good in still images .