

**NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY**

**School of Electrical Engineering and Computer Sciences**

MUJTABA SHAHID FAIZI

BSCS-5A

#131818

Assignment 1

Computer Vision

**Submitted to**: Sir Saad Bashir

Dated: 01-05-2019

**Single Shot Detector (SSD) Comparisons:**

This technique uses a unified detection approach (i.e. gives the required output in a single forward pass of the network) that speeds up the process of object detection, by eliminating the need for the region proposal network, unlike the RFCN series. It utilizes a good CNN architecture like VGG-16 to extract feature maps and apply small convolution filters to these feature maps to detect objects. SSD uses an input image and ground truth boxes for each object during training. Similarly, evaluation is done on a small set of default boxes of different aspect ratios & scales at each location in the several feature maps at different layers, by matching with the ground truth boxes. These default boxes can be considered similar to those of the anchor boxes used in faster RCNN. Shape offsets and all object scores are predicted, for all of the default boxes. Non-maximum suppression is also followed to choose only the confident boxes among many overlapping bounding boxes, to make the final predictions.

SSD allows more aspect ratios and flexibility than YOLO, that has its predefined grid cells’ aspect ratio fixed. This makes SSD boxes to wrap around the objects in a more accurate and tighter way. SSD has better capability to detect objects in multiple scales than YOLO. Also, SSD adds more convolutional layers than YOLO, for scale invariance purposes. In addition, SSD surpasses YOLO in terms of speed, which is why it is considered best for real-time object detection, and can be trained for better accuracy as well.

**Region-based Fully Convolutional Network (R-FCN) Comparisons:**

In the RCNN series approaches, region proposals are generated, then ROI pooling is done, and then finally FC layers are used for classification and bounding box regression.

Like the faster RCNN, RFCN also utilizes the research proposed network to get the region proposals (i.e. where convolutions and different anchor boxes are used at various locations inside the feature maps). The difference is that RFCN removes all complexity after the ROI pooling by removing all the FC layers after ROI Pooling. A set of position-sensitive score maps are constructed that incorporates information w.r.t a relative spatial position (say, to the right of an object), that is also used to inject translation invariance in classification. There is also a position-sensitive ROI pooling layer on top of this FCN, that conducts selective pooling to aggregate the outputs of the last convolutional layer, and finally generates scores for each region of interest.

So, basically region proposals makes use of the same set of score maps to perform average voting, that is far from a complex calculation. Also, no learnable layer is needed after the ROI layer, making RFCN more faster & less complex than faster RCNN. In addition, the mean average precision (mAP) metric has also been observed to be more for RFCN than faster RCNN.

**SSD vs RCFN:**

SSD doesn’t work well for small objects, since the multiple convolutions can make the small objects disappear. This is why, RFCN is considered for its accuracy, if speed is not an issue for small objects. However, SSD can outperform all in accuracy for large objects. Also, in general SSD outperforms RFCN in speed with light and fast feature extractors.