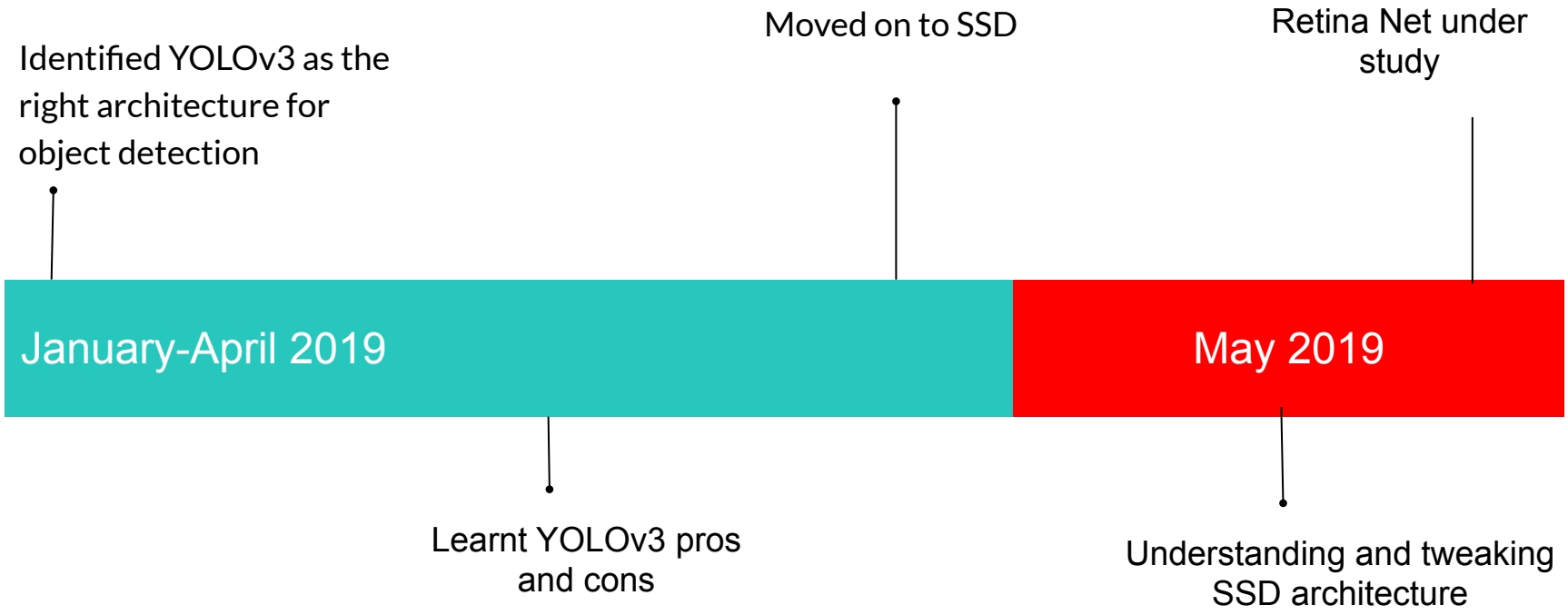
Computer Vision
Object & Pedestrian Detection

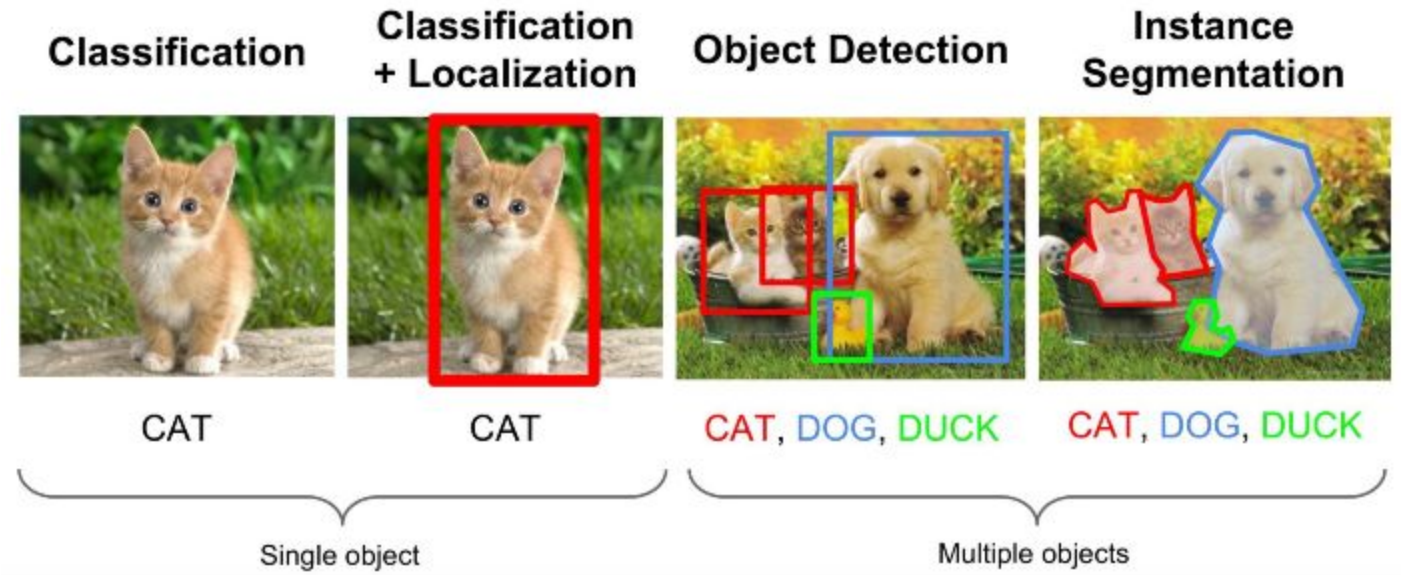
(For Internal Evaluation)

R&D Team - Analytics Labs

Milestones


Identified YOLOv3 as the right architecture for object detection
Moved on to SSD
Retina Net under study
January-April 2019
May 2019
Learnt YOLOv3 pros and cons
Understanding and tweaking SSD architecture

Identifying our need



We need **Object Detection,** at the same time, a regression and a classification task

Evaluation of any model :

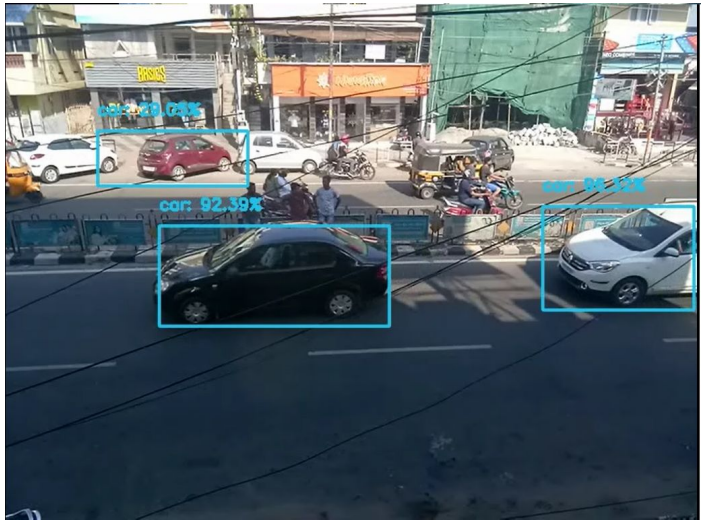
- Several datasets have been released for object detection challenges.
- Researchers publish results of their algorithms applied to these challenges.
- Specific performance metrics have been developed to take into account the spatial position of the detected object and the accuracy of the predicted categories.

Available Datasets

PASCAL2 Pattern Analysis, Statistical Modelling and Computational Learning	COCO Common Objects in Context	
• 10 000 images • 20 categories • 2012	• 120 000 images • 80 categories • 2012	• 500 000 images • 200 categories • 2013
Although, the PASCAL VOC dataset contains only 20 categories, it is still considered as a reference dataset in the object detection problem.	Both associated labeled data are not publicly available to avoid overfitting on the test dataset.	It is rarely used because the size of the dataset requires an important computational power for training. Also, the high number of classes complicates the object recognition task.

Created Dataset : AL-image dataset

- 10 x 5 x 60 x 30 = 90,000 images
- 0 labels
- Highlights:
  - Indian Traffic
  - Product developed on it will be more robust
  - Product developed can be easily sold to Indian market
  - Data is not cleaned



Performance Metrics

- **Intersection over Union (IoU)** : It corresponds to the overlapping area between the predicted box and the ground-truth box

- **mean Average Precision (mAP)** : It is simply the mean of the Average Precisions computed over all the classes of the challenge. The mAP metric avoids to have extreme specialization in few classes and thus weak performances in others.

- The COCO challenge has developed an **official metric** to avoid an over generation of boxes. It computes a mean of the mAP scores for variable IoU values in order to penalize high number of bounding boxes with wrong classifications.

Generally used Architectures

- **R-CNN (Region-based Convolutional Network)**
- **Fast-RCNN (Fast Region-based Convolutional Network)**
- **Faster R-CNN (Faster Region-based Convolutional Network)**
- **You Only Look Once (YOLO)**
- **Single Shot Detector (SSD)**

R-CNN

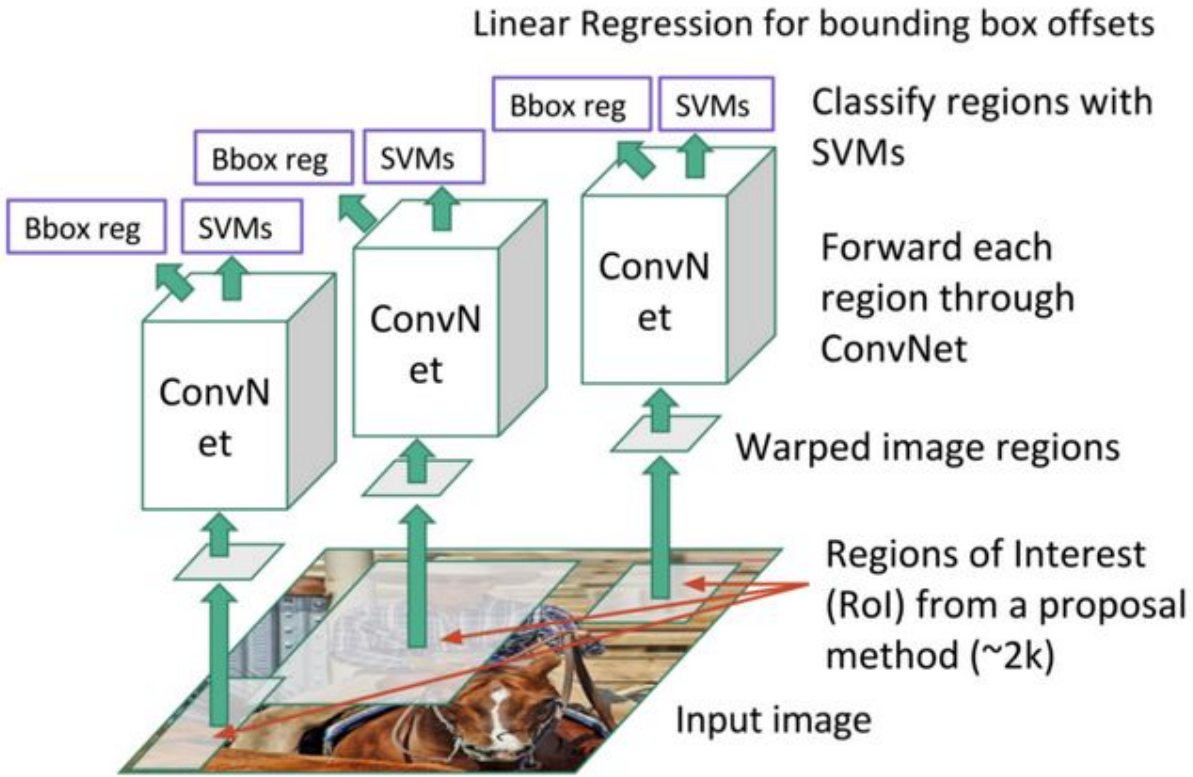
- Intuitively begin with the region search and then perform the classification. Modifications to improve processing time includes usage of selective search over exhaustive search.

- It initializes small regions in an image and merges them with a hierarchical grouping. Thus the final group is a box containing the entire image.

- The detected regions are merged according to a variety of color spaces and similarity metrics.

- The output is a few number of region proposals which could contain an object by merging small regions.

R-CNN Architecture



R-CNN Results

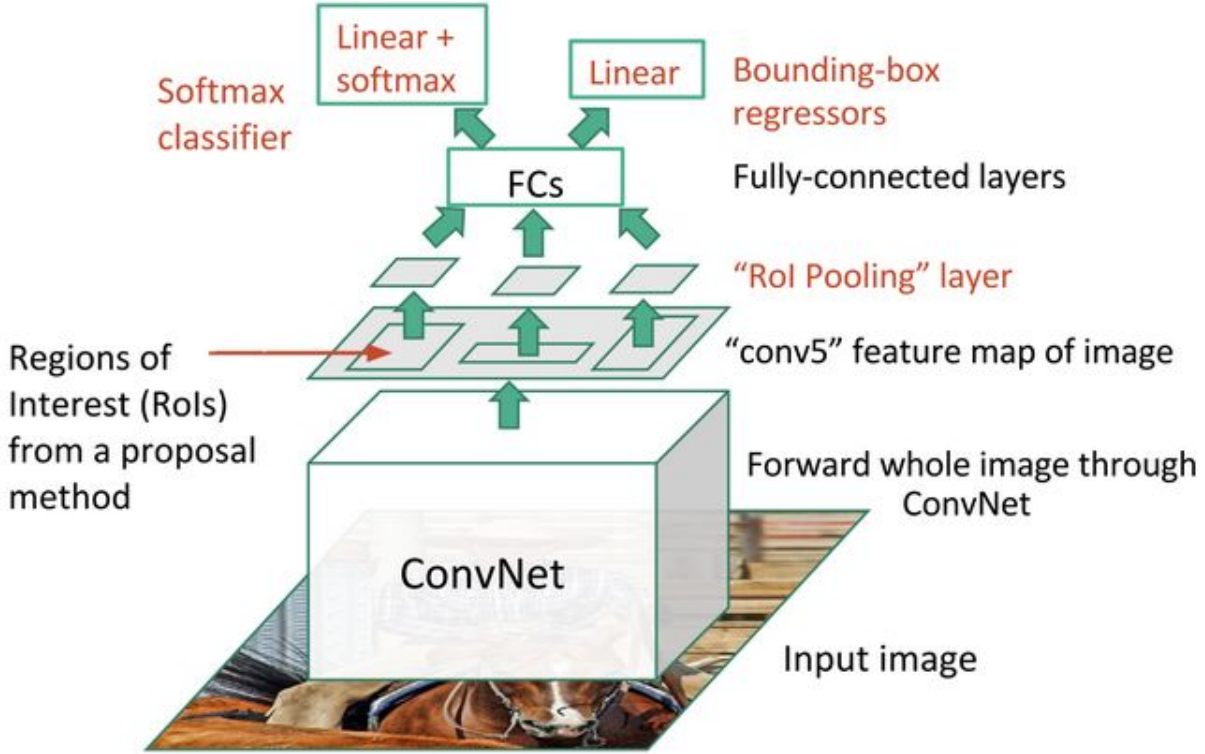
The best R-CNNs models have achieved a 62.4% mAP score over the PASCAL VOC 2012 test dataset (22.0 points increase w.r.t. the second best result on the leader board) and a 31.4% mAP score over the 2013 ImageNet dataset (7.1 points increase w.r.t. the second best result on the leader board).

Model	PASCAL VOC 2007	PASCAL VOC 2010	PASCAL VOC 2012	COCO 2015 (IoU=0.5)	COCO 2015 (IoU=0.75)	COCO 2015 (Official Metric)	COCO 2016 (IoU=0.5)	COCO 2016 (IoU=0.75)	COCO 2016 (Official Metric)	Real Time Speed
R-CNN	x	62.4%	x	x	x	x	x	x	x	No

Fast R-CNN

- A main CNN with multiple convolutional layers is taking the entire image as input instead of using a CNN for each region proposals (R-CNN).
- **Region of Interests** (RoIs) are detected with the selective search method applied on the produced feature maps. Formally, the feature maps size is reduced using a RoI pooling layer to get valid Region of Interests with fixed height and width as hyperparameters.
- Each RoI layer feeds fully-connected layers[1] creating a features vector.
- The vector is used to predict the observed object with a softmax classifier and to adapt bounding box localizations with a linear regressor.

Fast R-CNN Architecture



Fast R-CNN Results

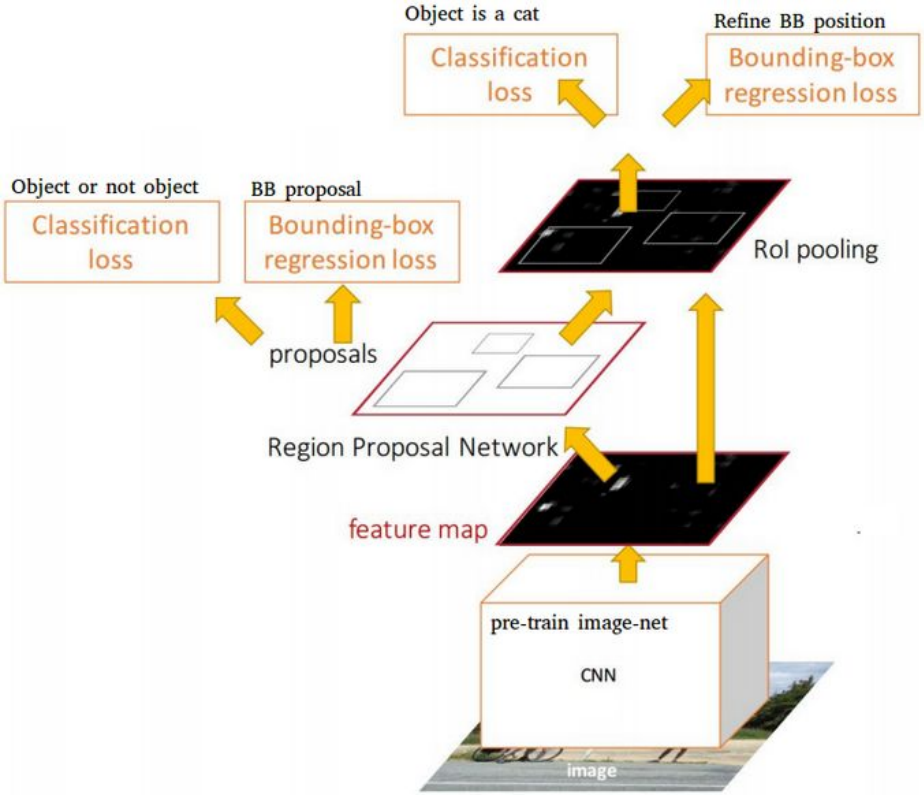
The best Fast R-CNNs have reached mAp scores of 70.0% for the 2007 PASCAL VOC test dataset, 68.8% for the 2010 PASCAL VOC test dataset and 68.4% for the 2012 PASCAL VOC test dataset.

Model	PASCAL VOC 2007	PASCAL VOC 2010	PASCAL VOC 2012	COCO 2015 (IoU=0.5)	COCO 2015 (IoU=0.75)	COCO 2015 (Official Metric)	COCO 2016 (IoU=0.5)	COCO 2016 (IoU=0.75)	COCO 2016 (Official Metric)	Real Time Speed
Fast R-CNN	70.0%	68.8%	68.4%	x	x	x	x	x	x	No

Faster R-CNN

- The previous model is computationally expensive.
- The Faster Region-based Convolutional Network (Faster R-CNN) is a combination between the RPN and the Fast R-CNN model.
- Faster R-CNN uses RPN to avoid the selective search method, it accelerates the training and testing processes, and improve the performances.
- The RPN uses a pre-trained model over the ImageNet dataset for classification and it is fine-tuned on the PASCAL VOC dataset.
- Then the generated region proposals with anchor boxes are used to train the Fast R-CNN.
- This process is iterative.

Faster R-CNN Architecture



Faster R-CNN Results

The best Faster R-CNNs have obtained mAP scores of 78.8% over the 2007 PASCAL VOC test dataset and 75.9% over the 2012 PASCAL VOC test dataset. They have been trained with PASCAL VOC and COCO datasets. One of these models[2] is 34 times faster than the Fast R-CNN using the selective search method.

Model	PASCAL VOC 2007	PASCAL VOC 2010	PASCAL VOC 2012	COCO 2015 (IoU=0.5)	COCO 2015 (IoU=0.75)	COCO 2015 (Official Metric)	COCO 2016 (IoU=0.5)	COCO 2016 (IoU=0.75)	COCO 2016 (Official Metric)	Real Time Speed
Faster R-CNN	78.8%	x	75.9%	x	x	x	x	x	x	No

Why YOLO over CNN?

- YOLOv3 is extremely **fast** and **accurate**
- Can easily **trade off** between **speed** and **accuracy** simply by changing the size of the model, **no retraining** required
- YOLO trains on full images and directly optimizes detection performance where as CNNs simultaneously predict bounding boxes and class accuracy

YOLO

1. Reasons globally about the image when making predictions.
2. Sees the entire image during training and test time.
3. Implicitly encodes contextual information about classes as well as their appearance.

YOLO

Pratabidya’s slideshow begins

Single Shot Detector (SSD)

- Predicts all at once: the bounding boxes and the class probabilities with a end-to-end CNN architecture

- The model takes an image as input which passes through multiple convolutional layers with different sizes of filter (10x10, 5x5 and 3x3).

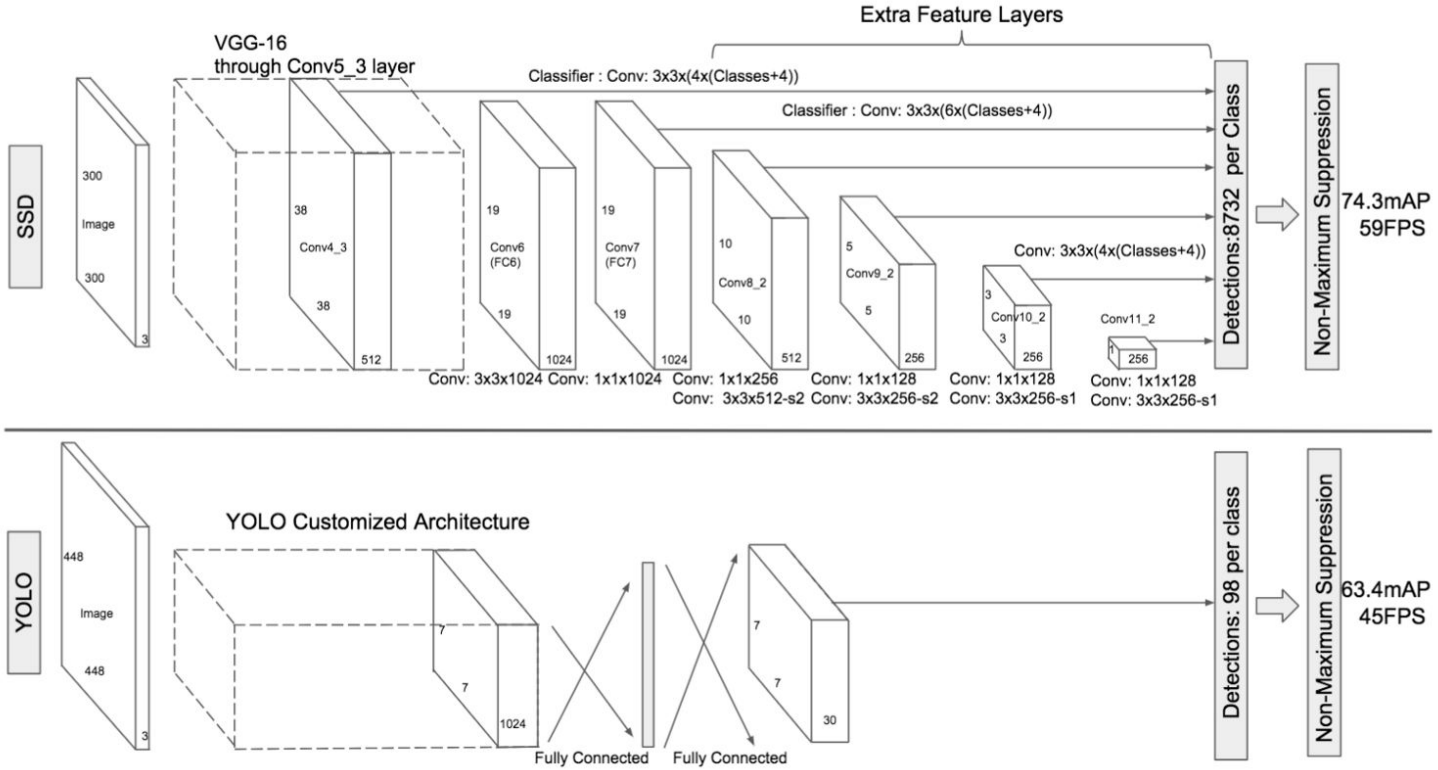
- Feature maps from convolutional layers at different position of the network are used to predict the bounding boxes.

- They are processed by a specific convolutional layers with 3x3 filters called extra feature layers to produce a set of bounding boxes.

- **Non-Maximum Suppression** method is also used at the end of the SSD model to keep the most relevant bounding boxes

- SSD512 model which is the SSD300 with an extra convolutional layer for prediction to improve performances.

SSD - Architecture



SSD - Results

They have obtained mAP scores of 83.2% over the 2007 PASCAL VOC test dataset and 82.2% over the 2012 PASCAL VOC test dataset. Over the test-dev dataset of the 2015 COCO challenge, they have had a score of 48.5% for an IoU = 0.5, 30.3% for an IoU = 0.75 and 31.5% for the official mAP metric.

Model	PASCAL VOC 2007	PASCAL VOC 2010	PASCAL VOC 2012	COCO 2015 (IoU=0.5)	COCO 2015 (IoU=0.75)	COCO 2015 (Official Metric)	COCO 2016 (IoU=0.5)	COCO 2016 (IoU=0.75)	COCO 2016 (Official Metric)	Real Time Speed
SSD	83.2%	x	82.2%	48.5%	30.3%	31.5%	x	x	x	No

Comparative Analysis

Model	PASCAL VOC 2007	PASCAL VOC 2010	PASCAL VOC 2012	COCO 2015 (IoU=0.5)	COCO 2015 (IoU=0.75)	COCO 2015 (Official Metric)	COCO 2016 (IoU=0.5)	COCO 2016 (IoU=0.75)	COCO 2016 (Official Metric)	Real Time Speed
R-CNN	x	62.4%	x	x	x	x	x	x	x	No
Fast R-CNN	70.0%	68.8%	68.4%	x	x	x	x	x	x	No
Faster R-CNN	78.8%	x	75.9%	x	x	x	x	x	x	No
YOLO	63.7%	x	57.9%	x	x	x	x	x	x	Yes
SSD	83.2%	x	82.2%	48.5%	30.3%	31.5%	x	x	x	Yes

Conclusion

In this short period of time, amongst all the algorithms we have seen that SSD outperforms the other algorithms in terms of mAP however, the runtime optimization and more availability of YOLO algorithm has rendered it more usable in Industry.

In light of experimentation, we have created our very own dataset, and run the input samples through two YOLO architectures and a SSD architecture. Owing to lack of ground-truth data availability of the project, we have not been able to test our implemented algorithms.

Thank You