# Review of Object Detection Algorithms using CNN

**Presentation** · March 2019

DOI: 10.13140/RG.2.2.26529.25447

CITATIONS
0

READS
658

**3 authors**, including:

Farhana Sultana
University of Gour Banga, India
**9** PUBLICATIONS **7** CITATIONS
SEE PROFILE

A. Sufian
University of Gour Banga, West Bengal, India
**31** PUBLICATIONS **47** CITATIONS
SEE PROFILE

# A Review of Object Detection Models based on Convolutional Neural Network

**Farhana Sultana[1], Abu Sufian[1], Paramartha Dutta[2]**

[1]University of Gour Banga

[2]Visva-Bharati University

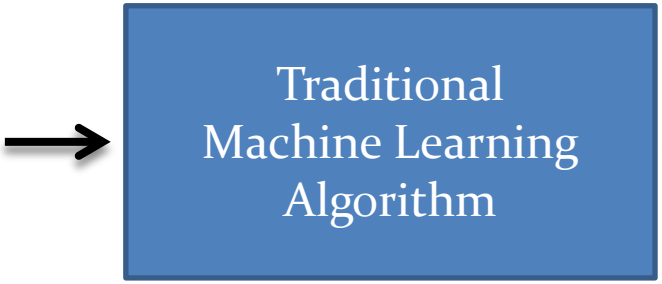
**2nd ICCDC-2019, HIT, Haldia**

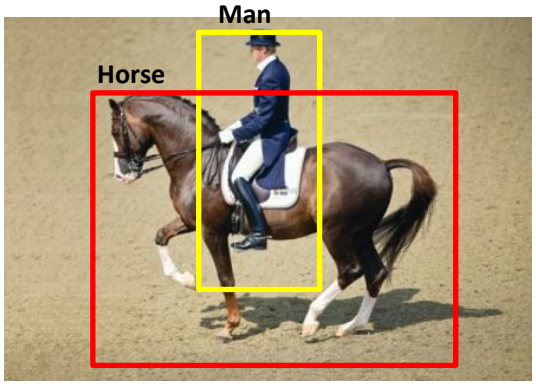
# Contents

# Introduction

- Object Detection
  - Fundamental problem of Computer Vision
  - the problem of estimating the class and location of objects contained within an image

Input Image



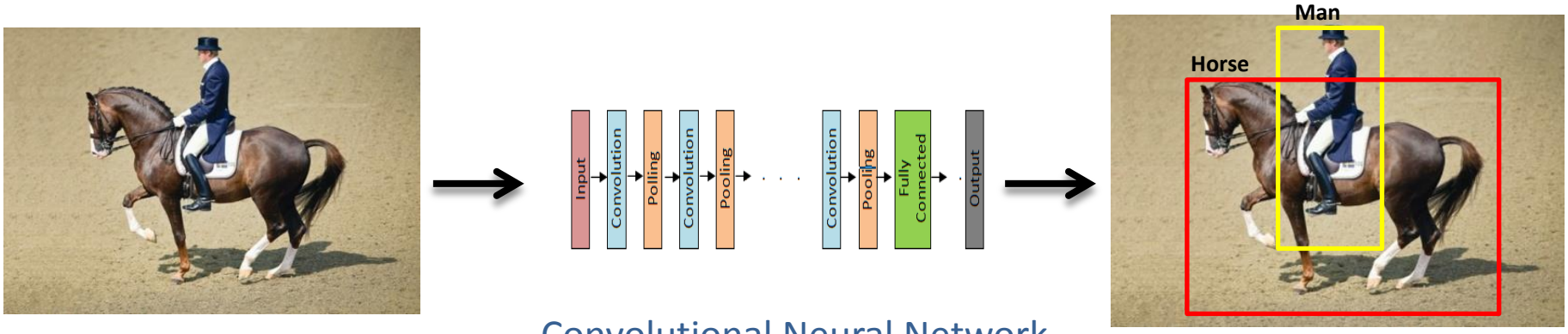




Detected objects with class and location

- Convolutional Neural Network shows state-of-the-art performance in image classification [10] and object detection task.

# Introduction

- Object Detection
  - Fundamental problem of Computer Vision
  - the problem of estimating the class and location of objects contained within an image




Input Image

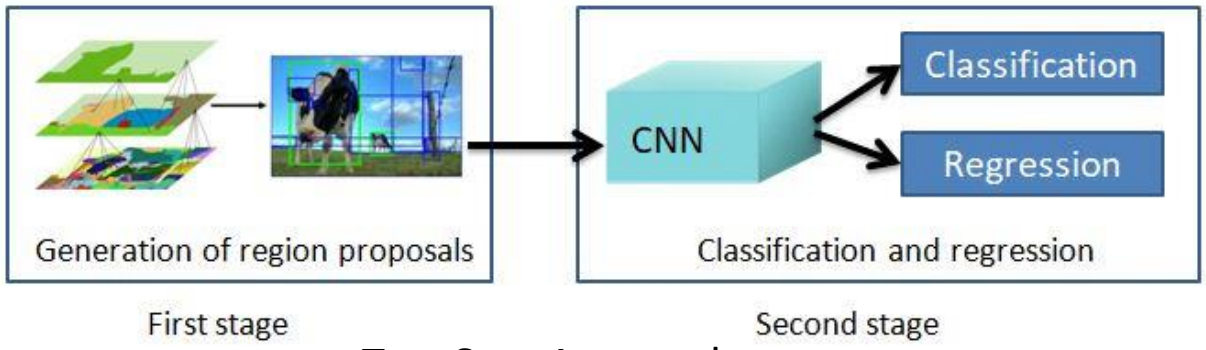
Convolutional Neural Network

Detected objects with class and location

- Convolutional Neural Network shows state-of-the-art performance in object detection task.

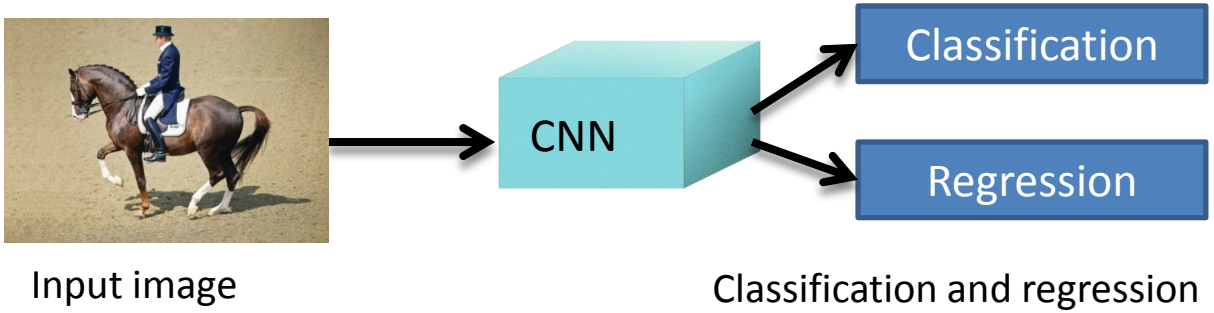
# Architectural Approach

- Two stage approach
  - R-CNN
  - SPP-net
  - Fast R-CNN
  - Faster R-CNN
  - Mask R-CNN




Two Stag Approach

- One Stage approach
  - YOLO
  - SSD
  - YOLO9000
  - RetinaNet
  - RefineDet

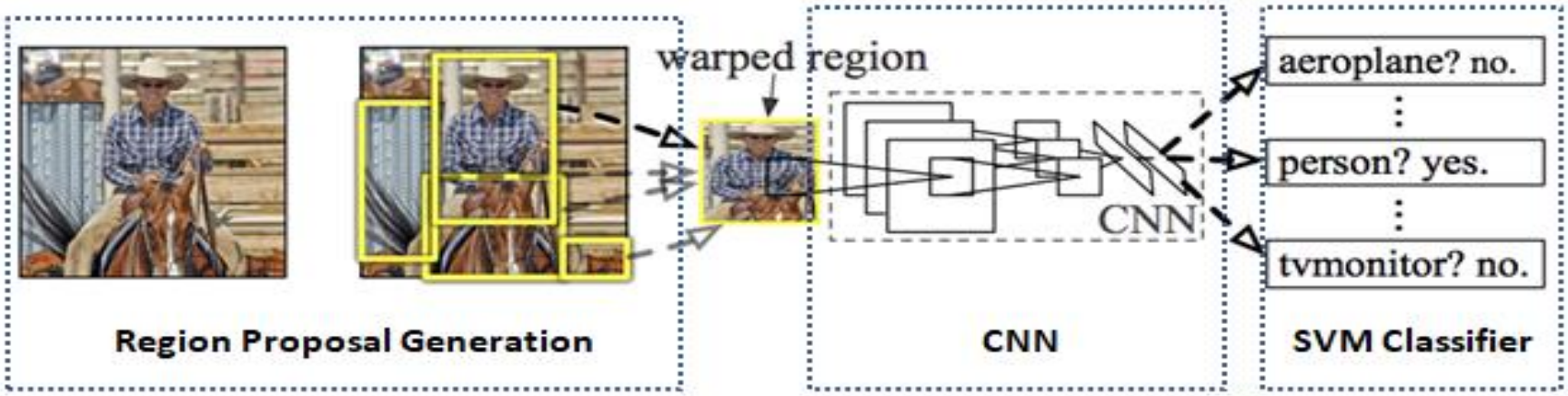



One Stag Approach

# Different Object Detection Models based on Two Stage Approach

# R-CNN(2014)

- First Convolutional Neural Network based object detection model
- R-CNN composed of three blocks
  - Region Proposal Generation (Selective Search)
  - Convolutional Neural Network (AlexNet)
  - SVM Classifier (20 classes)
- Achieves more than 30% mAP relative to the previous best result on PASCAL VOC-2012.

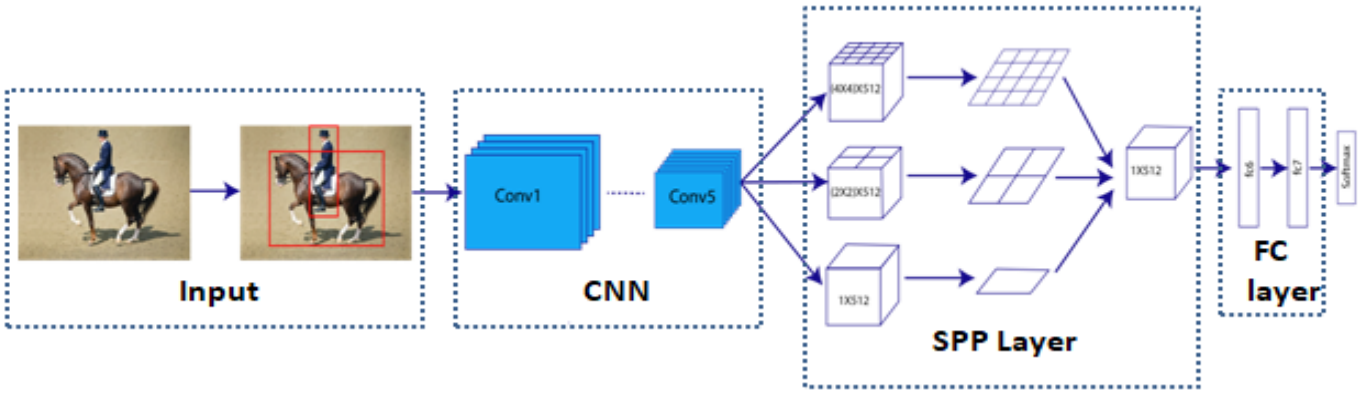



Rich feature hierarchies for accurate object detection and semantic segmentation : [R. B. Girshick], CVPR, 2014

Contd.

# SPP-net(2014)

- R-CNN limitation
  - CNNs need fixed sized input image but region proposals are arbitrary in size
  - For each region proposal – a CNN is needed
- Spatial pyramid pooling (SPP) can generate fixed sized output regardless of input size.
- SPP-net
  - Included SPP layer in between  conv layers and FC layers of R-CNN
  - Increases scale invariance and reduces over fitting
  - Much faster than R-CNN – one time feature map calculation

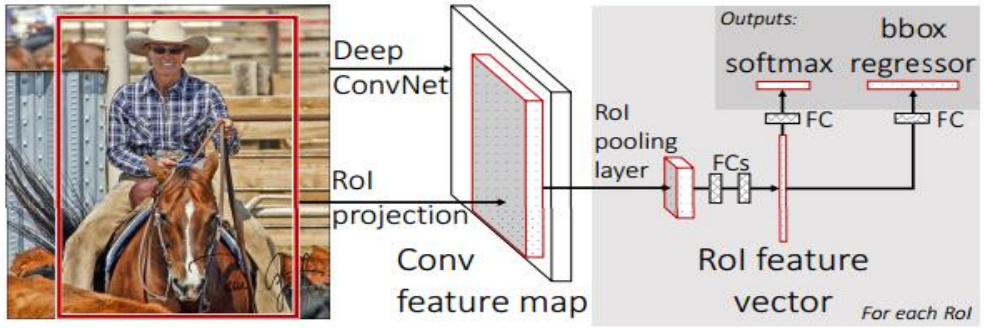



Spatial pyramid pooling in deep convolutional networks for visual recognition: [He et al.], ECCV, 2014

Contd.

# Fast R-CNN(2015)

- R-CNN and SPP-net suffers from
  - Multi-stage pipeline training
  - Expensive training in terms of space and time and
  - Slow object detection.
- Fast R-CNN
  - Single-stage training algorithm – fast training and testing
  - Conv layers takes entire image and a set of object proposals as input
  - Region of Interests (RoI) are generated form feature map
  - RoI pooling layer reshaped RoIs into a fixed length feature vector
  - Output layers - classification and bounding box regression
  - Much Faster

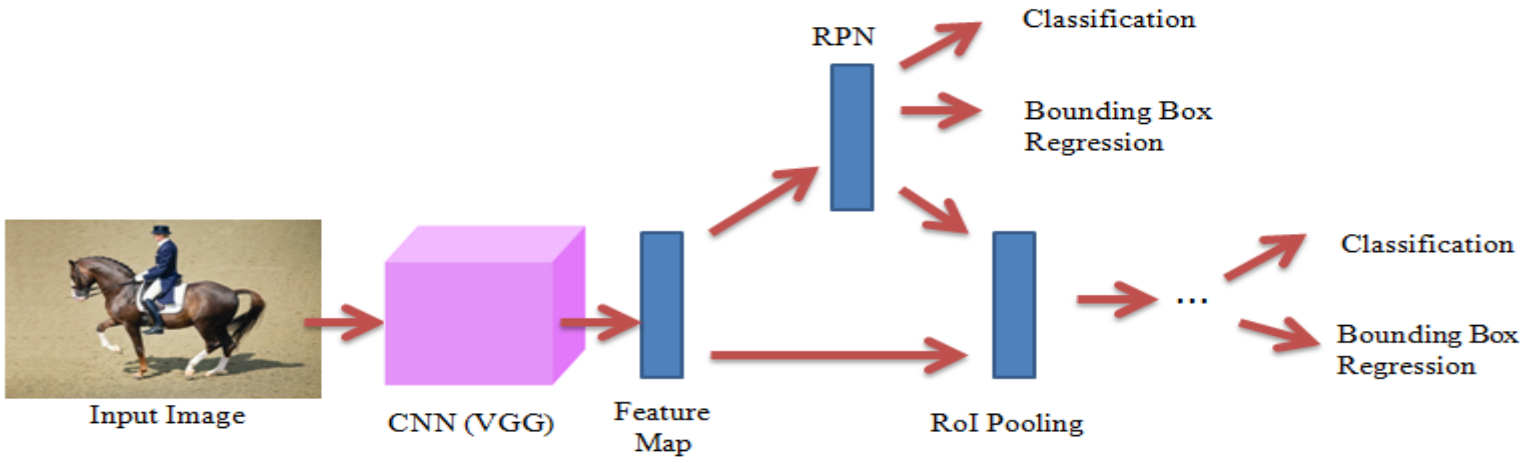



Fast R-CNN : [R. B. Girshick], ICCV, 2015

Contd.

# Faster R-CNN(2015)

- R-CNN, SPP-net and Fast R-CNN depend on
  - Slow and time consuming region proposal algorithm
- Faster R-CNN
  - Introduced Region Proposal Network (RPN)
  - RPN and Fast R-CNN share convolutional features
  - Used Anchors of three different scales and aspect ratio
  - Used RoI pooling layer like Fast R-CNN

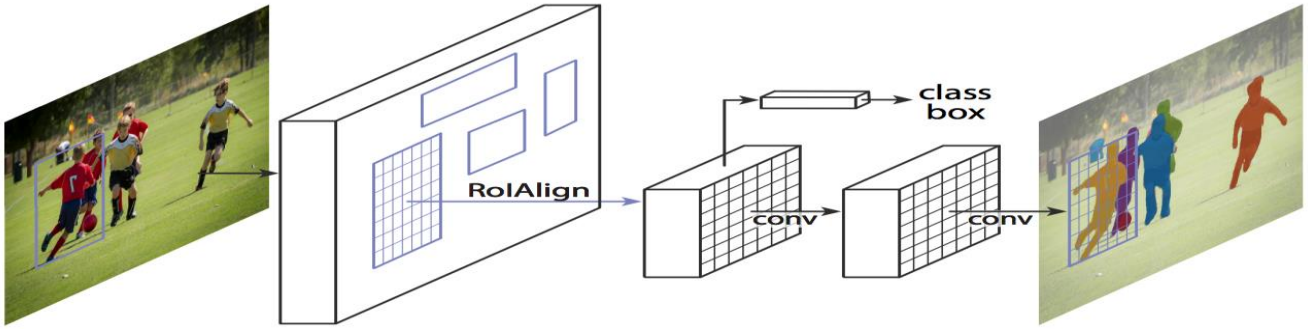



Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks : [Ren et al.], TPRMI, 2015

Contd.

# Mask R-CNN(2017)

- Previous Models
  - Detects objects with bounding box
- Mask R-CNN
  - Introduced instance segmentation
  - Output of Mask R-CNN
    - Class label
    - Bounding box offset and
    - Binary object mask
  - RoIAlign layer instead of RoI pooling layer
  - Easy to generalize to other task



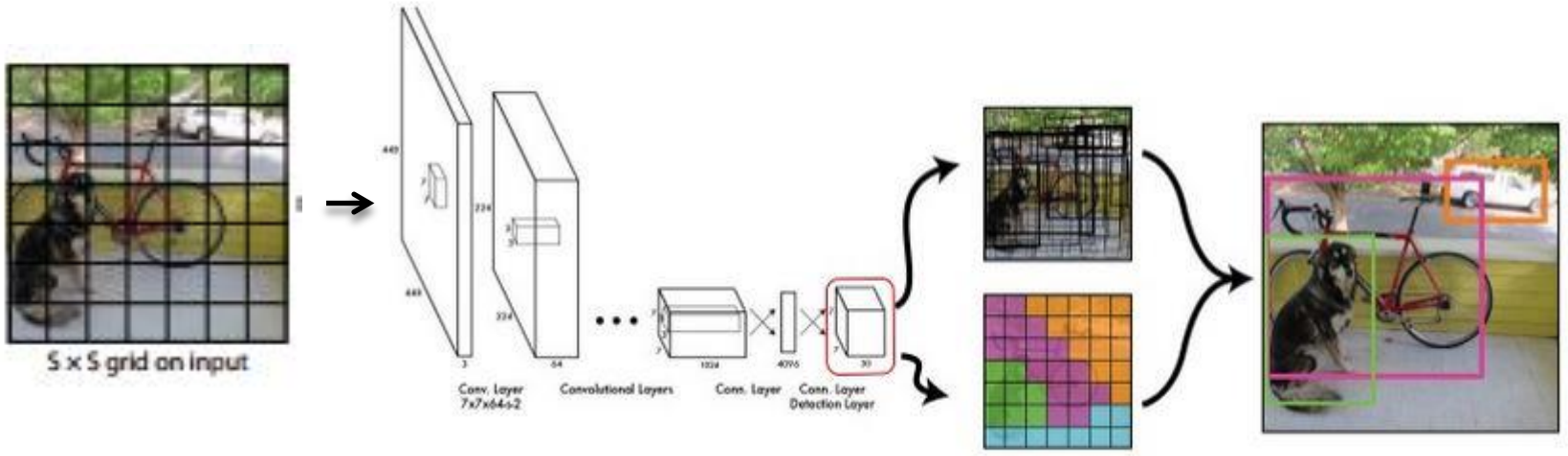

Mask R-CNN: [He et al.], ICCV, 2017

Contd.

# Different Object Detection Models based on One Stage Approach

# YOLO(2016)

- Previous Models
  - Classifier-based and region proposal based
- YOLO
  - First unified  neural network
  - Predicts class probabilities and bounding boxes directly from full input image using a simple CNN in one evaluation
  - Extremely fast and processes image in real time

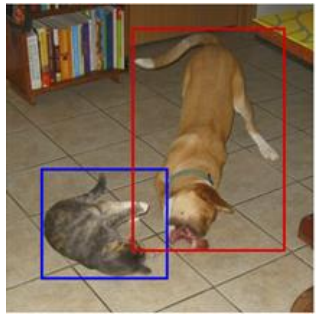


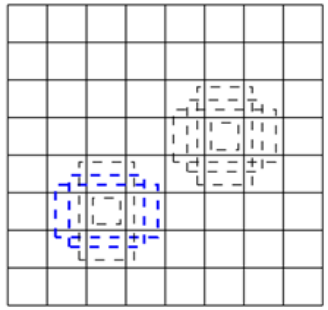

You only look once: Unified, real-time object detection : [Redmon et al.], CVPR,, 2016

Contd.

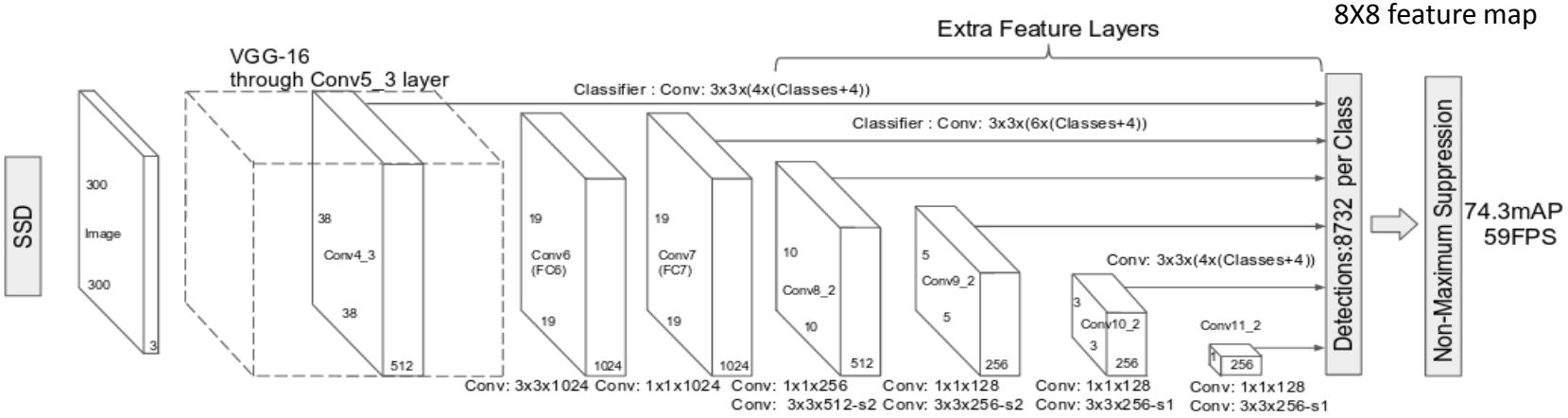
# SSD(2016)

- Previous Models
  - Prediction depends on only single scale feature map
- SSD
  - Eliminates region proposal generation
  - Multi-scale feature maps for detection
  - Convolutional predictors for detection
  - Default boxes and aspect ratios
  - Achieves a good balance between speed and accuracy.





8X8 feature map




SSD: Single shot multibox detector : [Liu et al.], ECCV, 2016

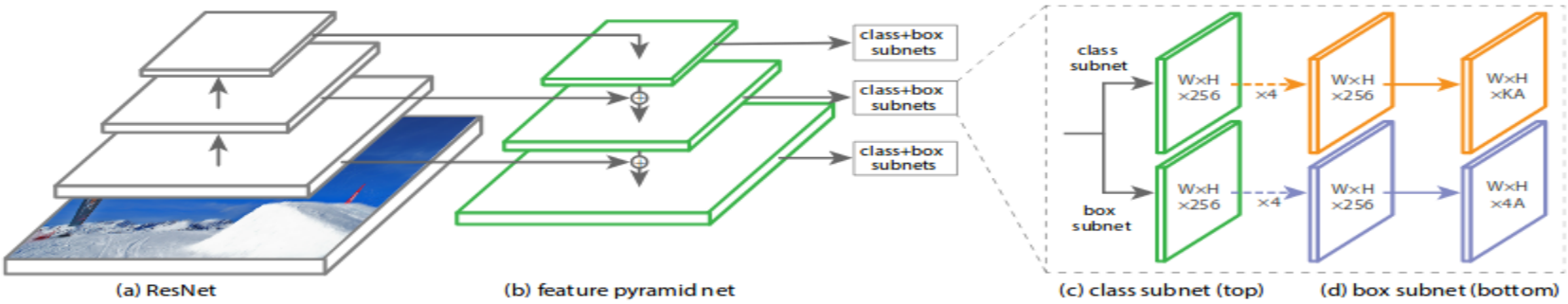
Contd.

# YOLO9000(2017)

- YOLO
  - Makes a significant number of localization errors
  - Has relatively lower recall than region proposal based method
- YOLO9000
  - Focuses on improving recall and localization
  - Real-time object detector
  - High resolution classifier
  - Capable of detecting more than 9000 object categories
  - Jointly optimizes detection and classification.
  - Used WordTree  hierarchy to combine data from various sources
  - Enhances the performances of YOLO without decreasing its speed.

# RetinaNet (2017)

- One stage models are faster than two stage models but slower -
  - Extreme class imbalance problem during training
- RetinaNet
  - Introduced Focal Loss for dense object detection
  - Fully convolutional object detector –
    - Feature Pyramid Network (FPN) as backbone network
    - Two task specific subnetworks for classification and detection.
  - Matched the speed of previous one stage detectors
  - Surpassed the accuracy of previous two stage detectors

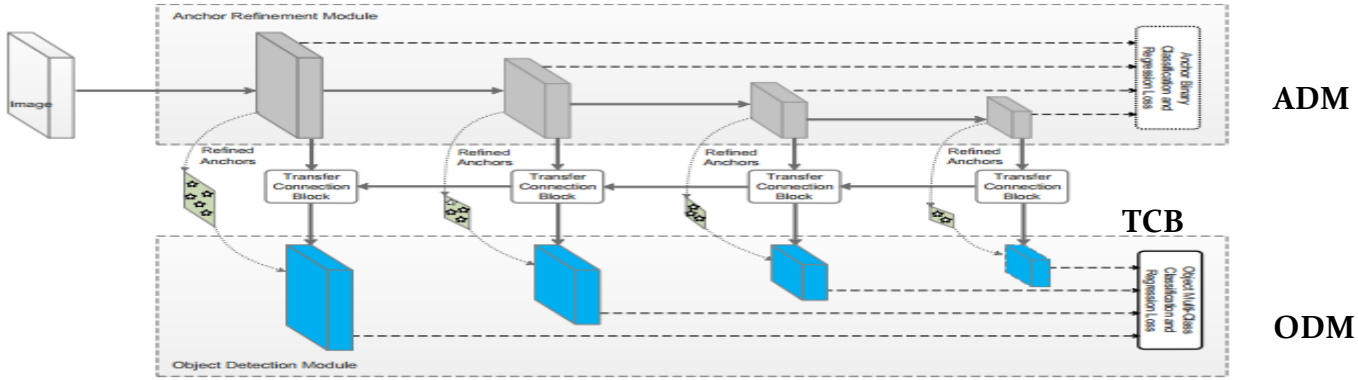



Focal loss for dense object detection : [Lin et al.], ICCV, 2017

Contd.

# RefineDet (2018)

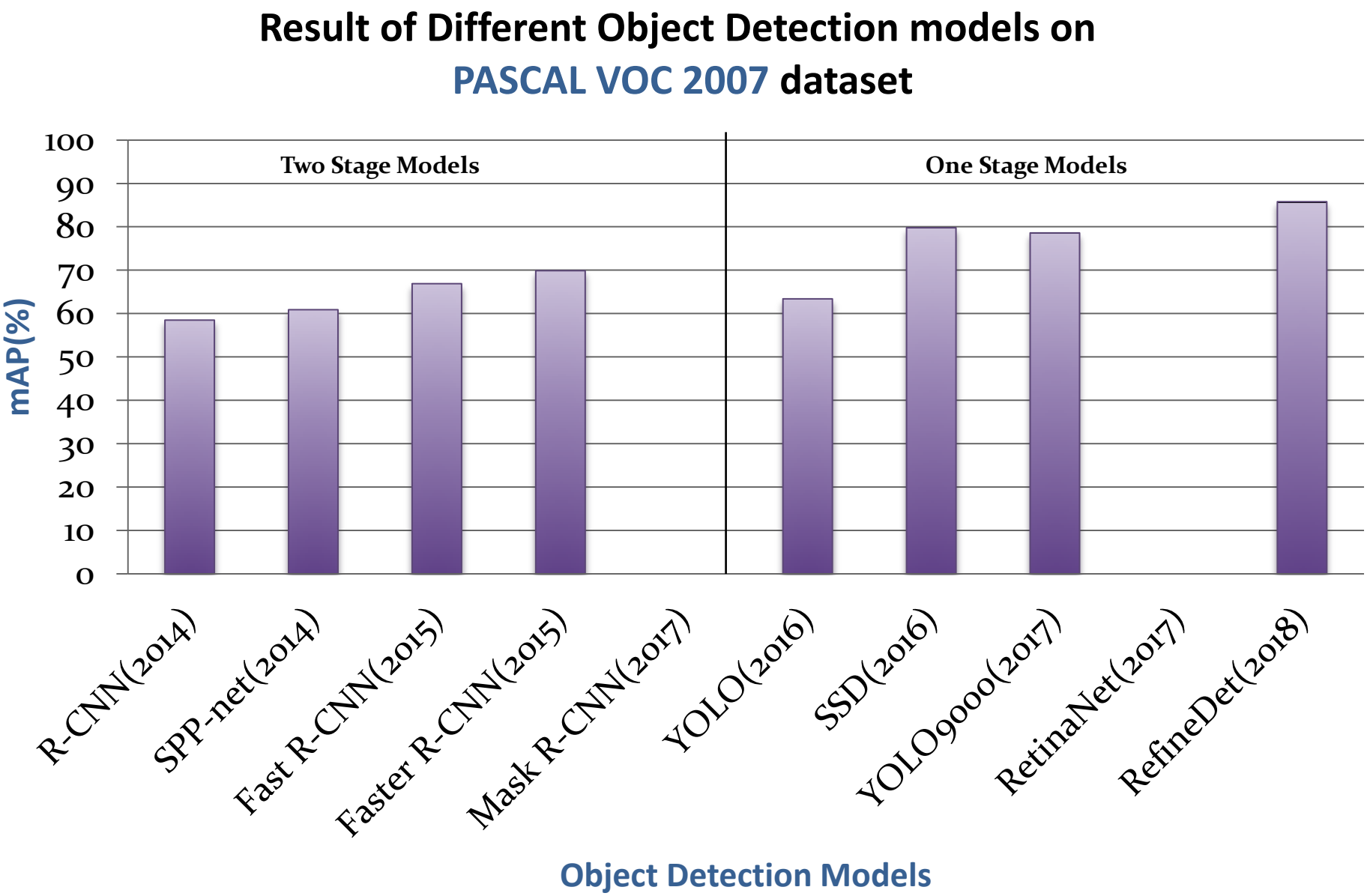
- Single-shot object detector based on feed forward convolutional network.
- Consists of two interconnected modules
  - The anchor refinement module (ARM)
  - The object detection module (ODM).
- The transfer connection block (TCB) transfer the features in the ARM to predict locations, sizes and class labels of object in the ODM.



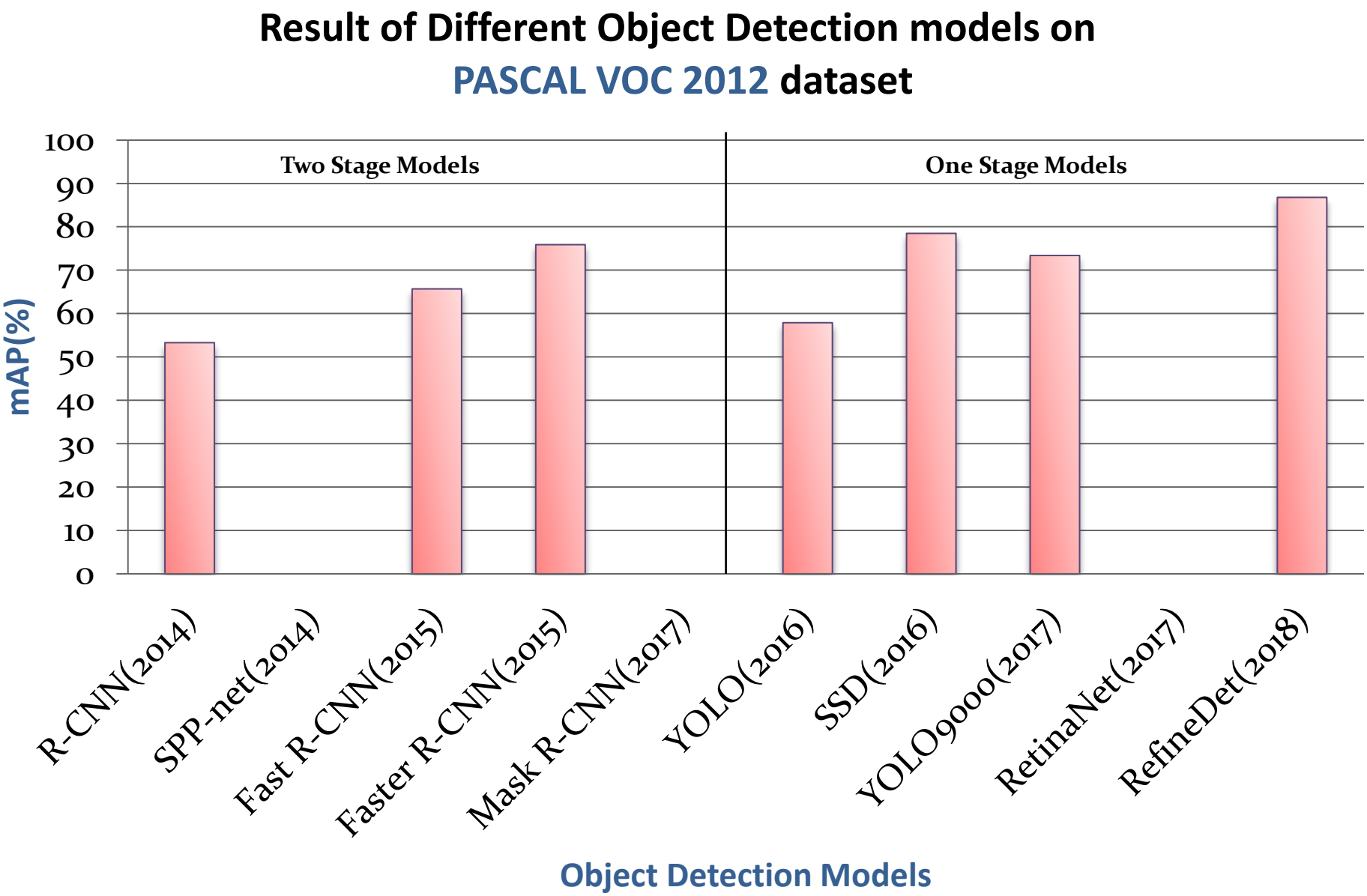

Single-shot refinement neural network for object detection : [Zhang et al.], CVPR, 2018

Contd.

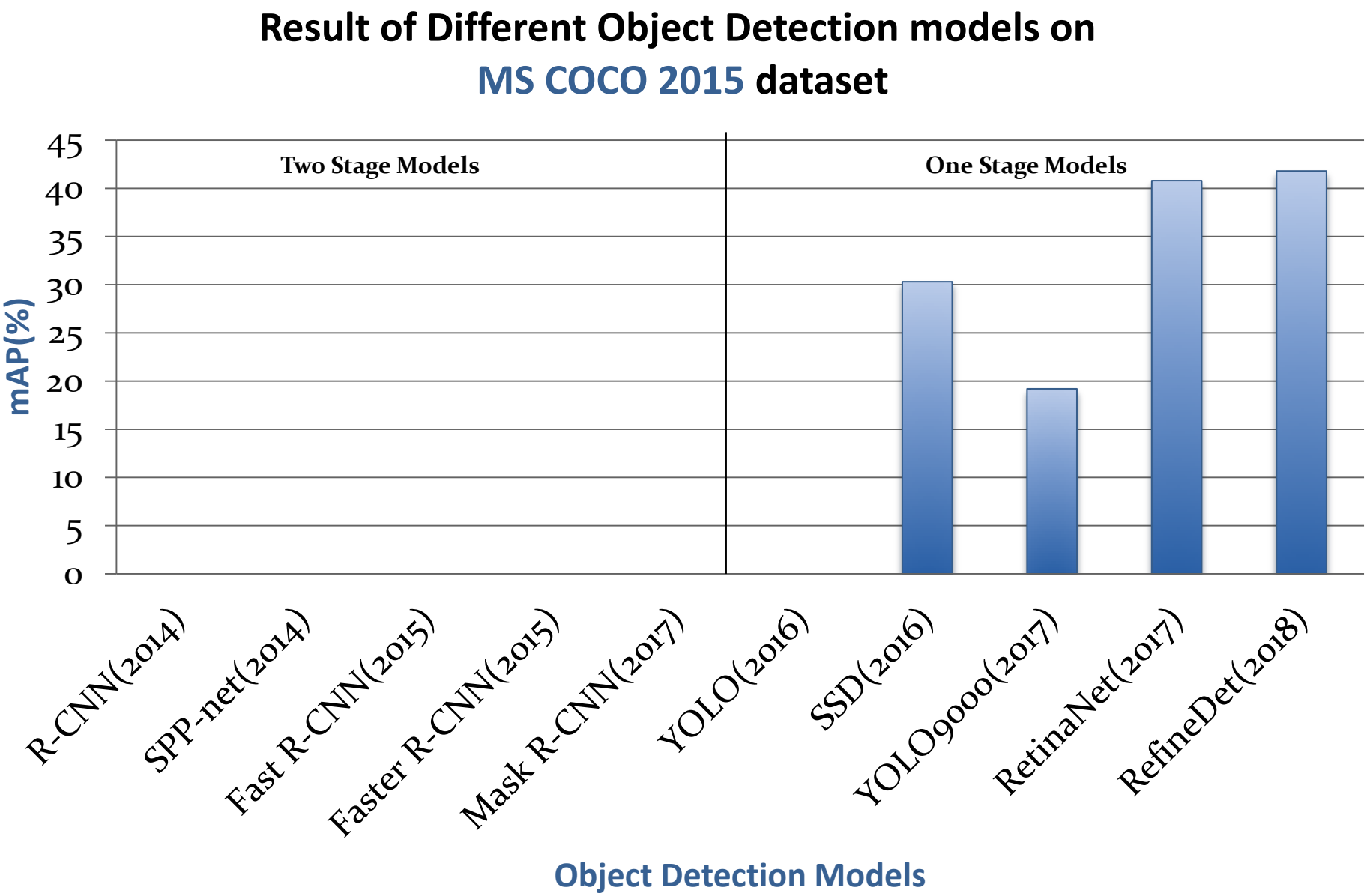
# Comparative Result



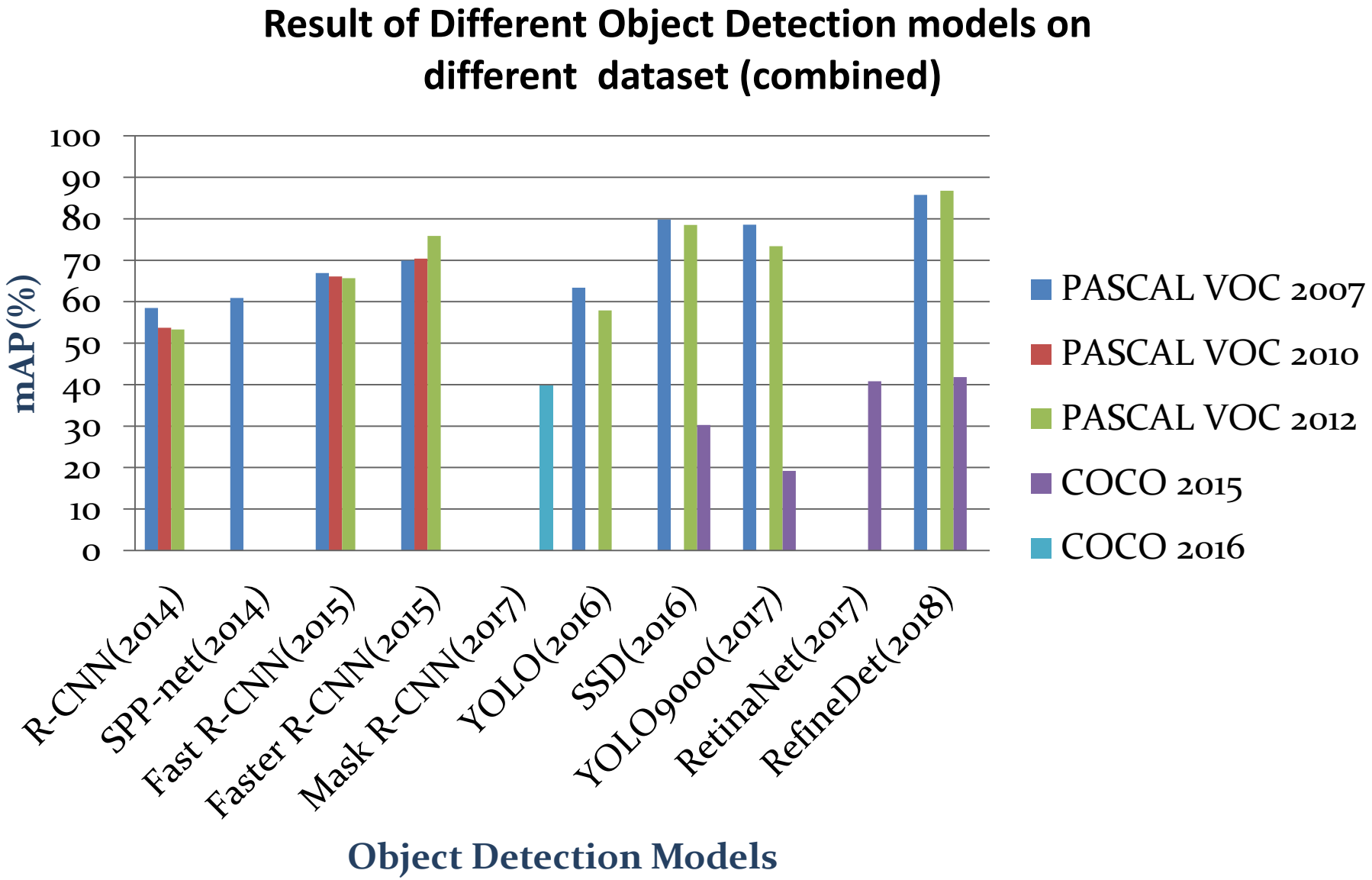
# Comparative Result



# Comparative Result



# Comparative Result



# Conclusion

- State-of-the-art object detection models can be categorized into two different approaches: two-stage and one-stage.
- Two-stage models gave higher accuracy than one-stage models in object detection but they are slower.
- R-CNN, SPP-net and Fast R-CNN were slow because of external region proposal Network.
- Faster R-CNN overcome that problem using RPN.
- Mask R-CNN added instance segmentation to the architecture of previous model.
- YOLO, SSD gave us a way for fast and robust object detection.

Contd.

# Conclusion

- RetinaNet focuses on improving loss function for better detection.
- RefineDet combined the merit of both two-stage and one-stage approach.
- Progress of various models are mainly because of better CNN models, new detection architecture, different pooling method, novel loss design etc.
- The improvements of different models give us the hope for more accurate and faster real time object detection.

# References

## References


1. R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation”, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580{587, 2014.
2. R. B. Girshick, “Fast R-CNN”*CoRR*, vol. abs/1504.08083, 2015. [Online]. Available: http://arxiv.org/abs/1504.08083
3. K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition”, *CoRR*, vol. abs/1406.4729, 2014. [Online]. Available: http://arxiv.org/abs/1406.4729
4. K. He, G. Gkioxari, P. Doll´ar, and R. B. Girshick, “Mask r-cnn,” *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980{2988, 2017.
5. J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection”, *CoRR*, vol. abs/1506.02640, 2015. [Online]. Available: http://arxiv.org/abs/1506.0264
6. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single shot multibox detector”, in *ECCV*, 2016.
7. J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517{6525, 2017.
8. T. Lin, P. Goyal, R. B. Girshick, K. He, and P. Doll´ar, “Focal loss for dense object detection”, *CoRR*, vol. abs/1708.02002, 2017. [Online]. Available: http://arxiv.org/abs/1708.02002
9. S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, “Single-shot refinement neural network for object detection”, *CoRR*, vol. abs/1711.06897, 2017. [Online].Available: http://arxiv.org/abs/
10. F. Sultana, A. Sufian, and P. Dutta, Advancements in image classification using convolutional neural network, in proceeding of ICRCICN, 2018.

# Any Question?

# Thank you