Object Localization with Classification in Real Time

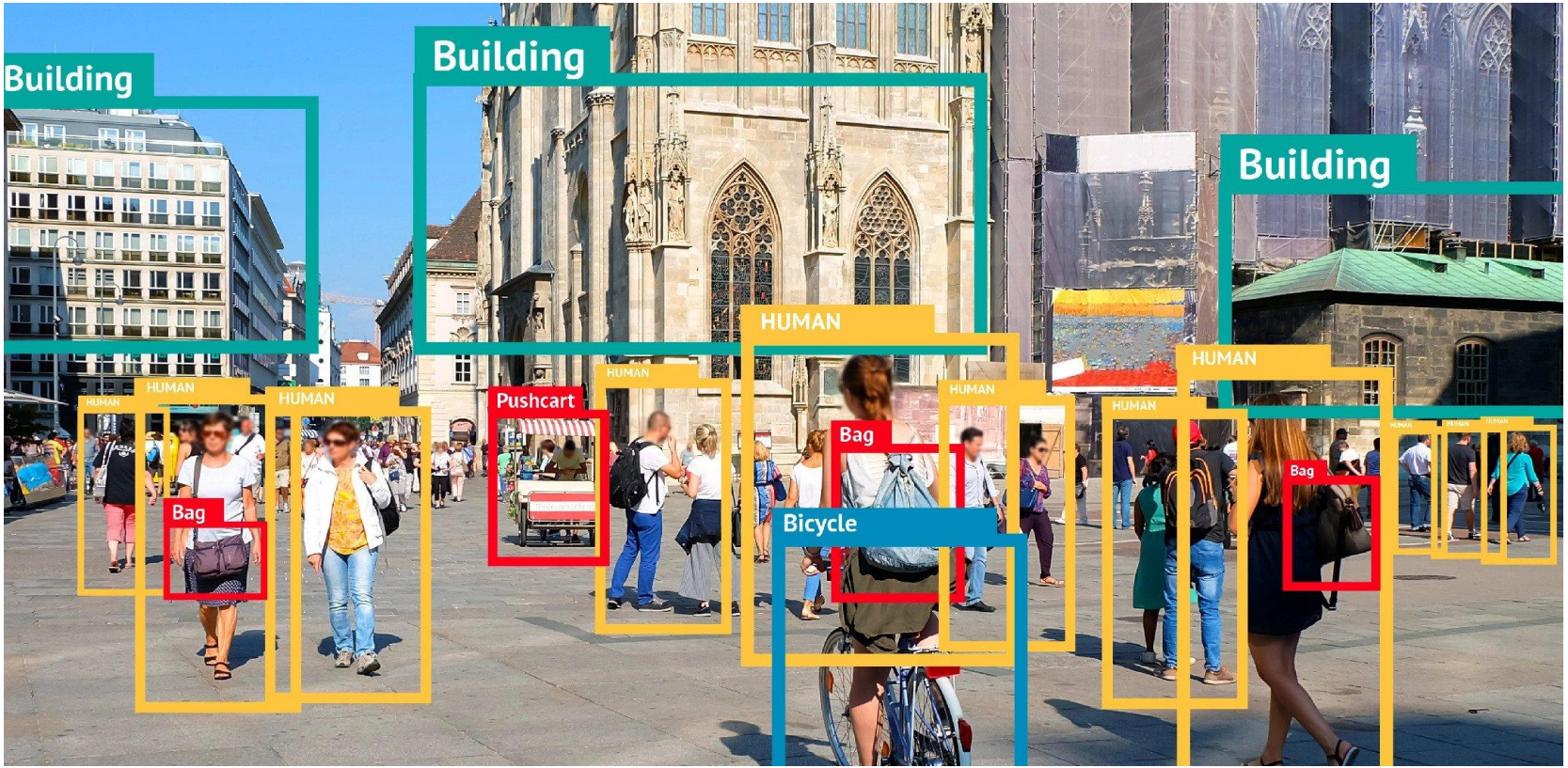
SHOBHIT
19CS06008
M.TECH 2nd YEAR
COMPUTER SCIENCE AND ENGINEERING

Under the guidance of :
Dr. Debi Prosad Dogra

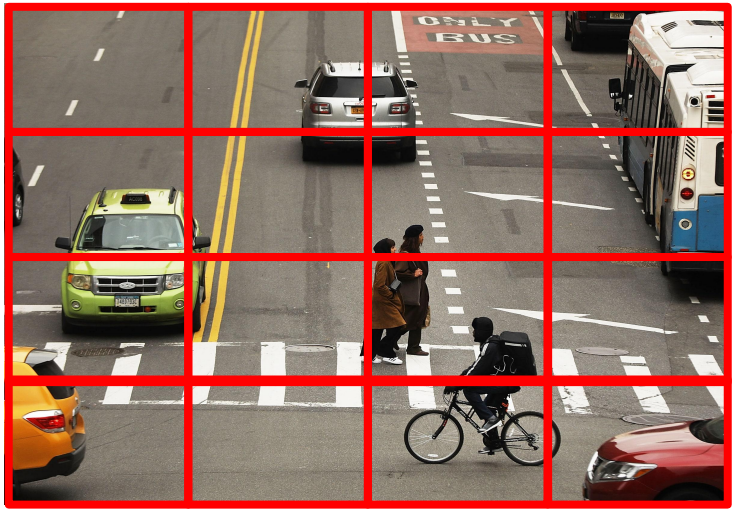
Content:

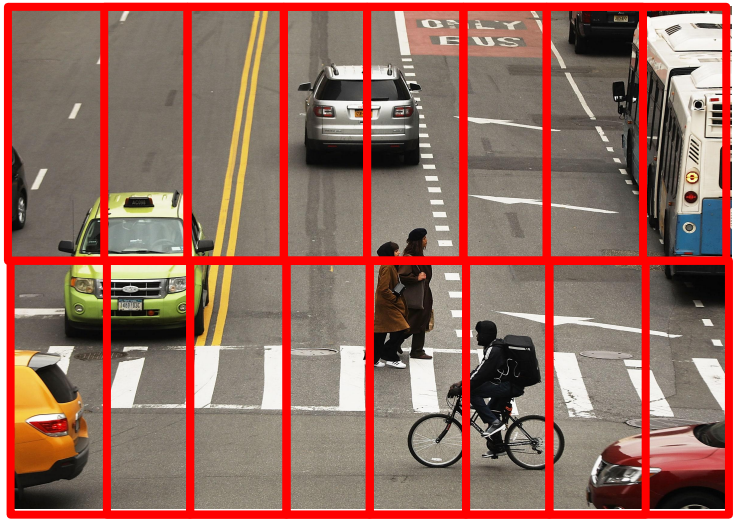
1. Introduction
2. Sliding Window Detection
3. Convolutional Implementation of Sliding Window
4. Region Based Convolutional Neural Networks(RCNN)
5. YOLO
6. Implementation

Introduction:



Sliding Window Detection:

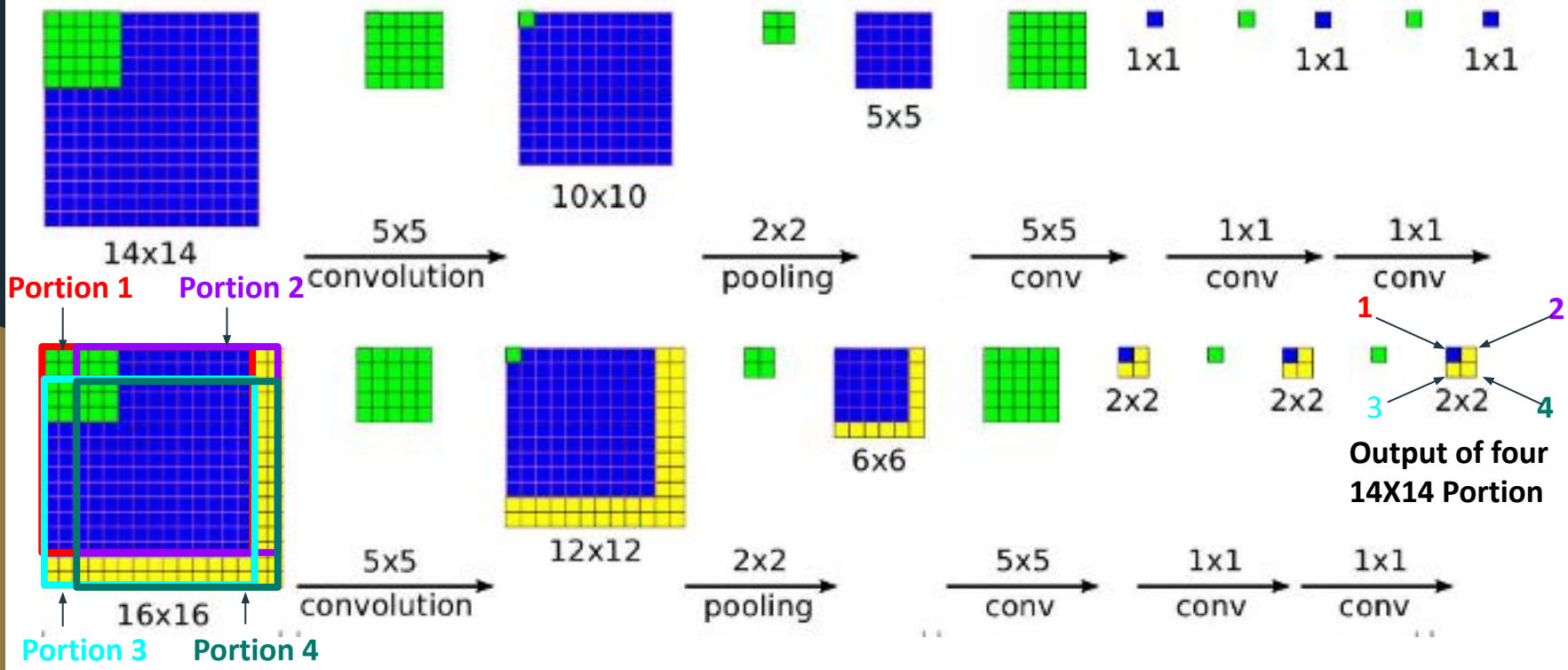




Medium Window Size
with high stride

Large Window Size
with low stride

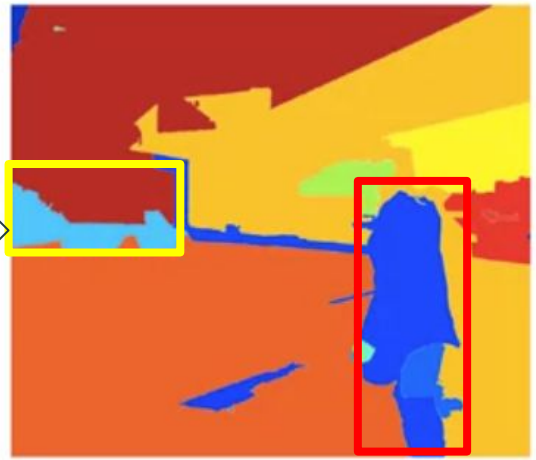
Convolutional Implementation of Sliding Window



[Sermanet et al.,2014, OverFeat:Integrated recognition, localization and detection using convolutional networks]

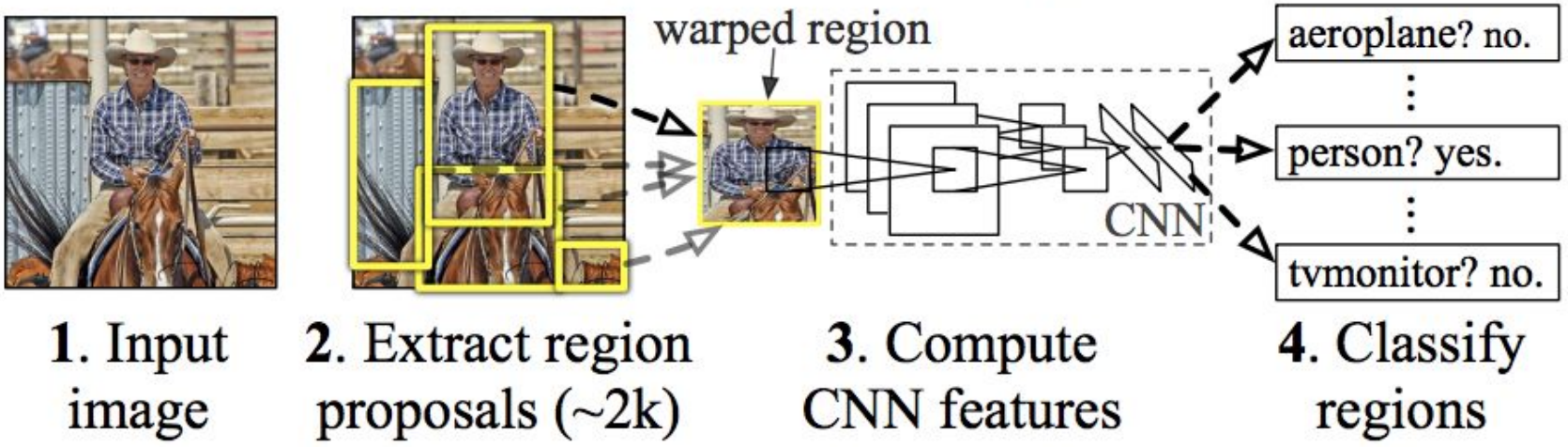
Region Based Convolutional Neural Networks (RCNN)

Segmentation Algorithm



Original Image

Segmentation Image



[Girshik et. al. 2013 Rich feature hierarchies for accurate object detection and semantic segmentation]

Other Versions:

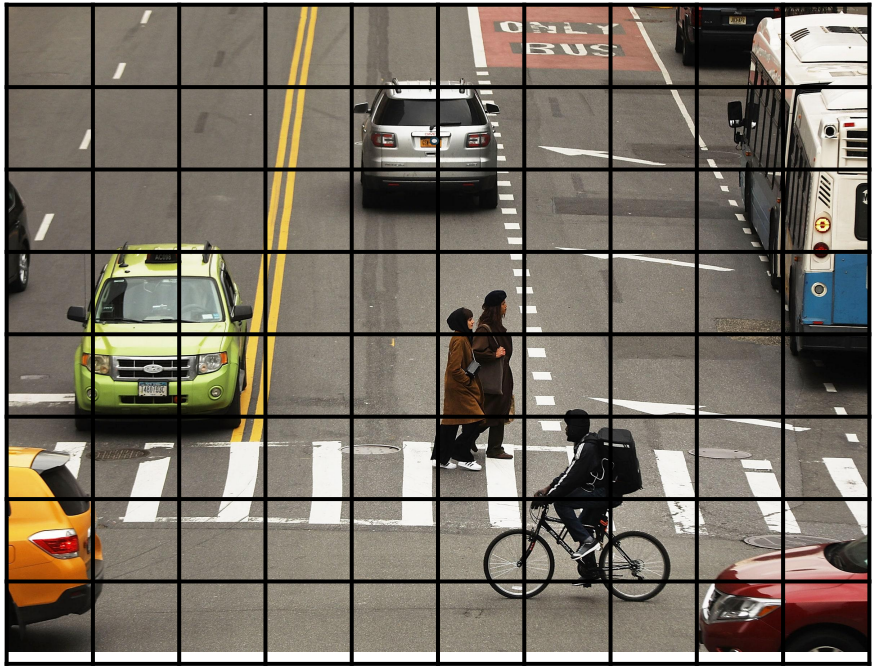
1. Fast RCNN
2. Faster RCNN

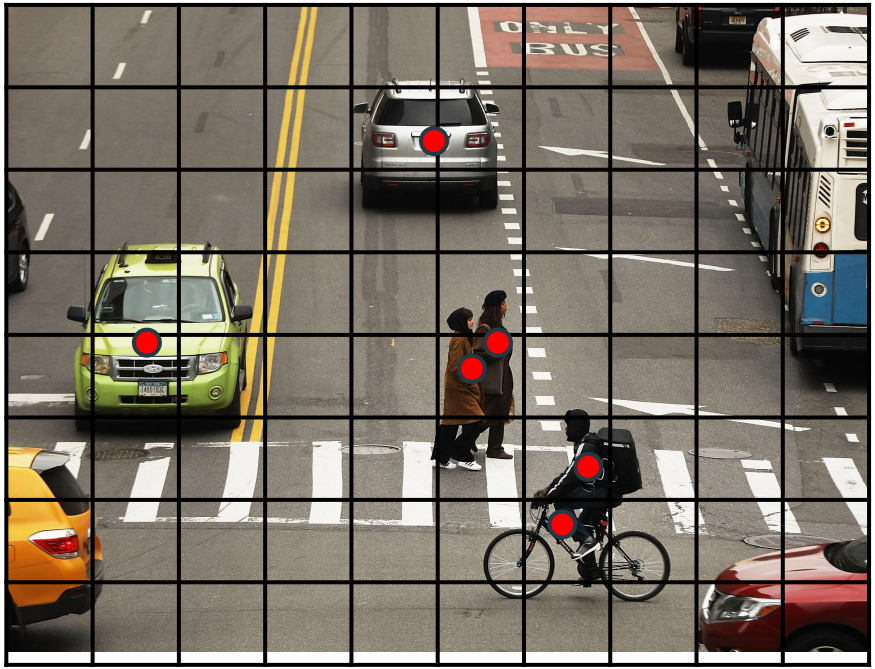
Drawbacks:

1. Training is expensive in space and time.

2. Test-time detection is slow.

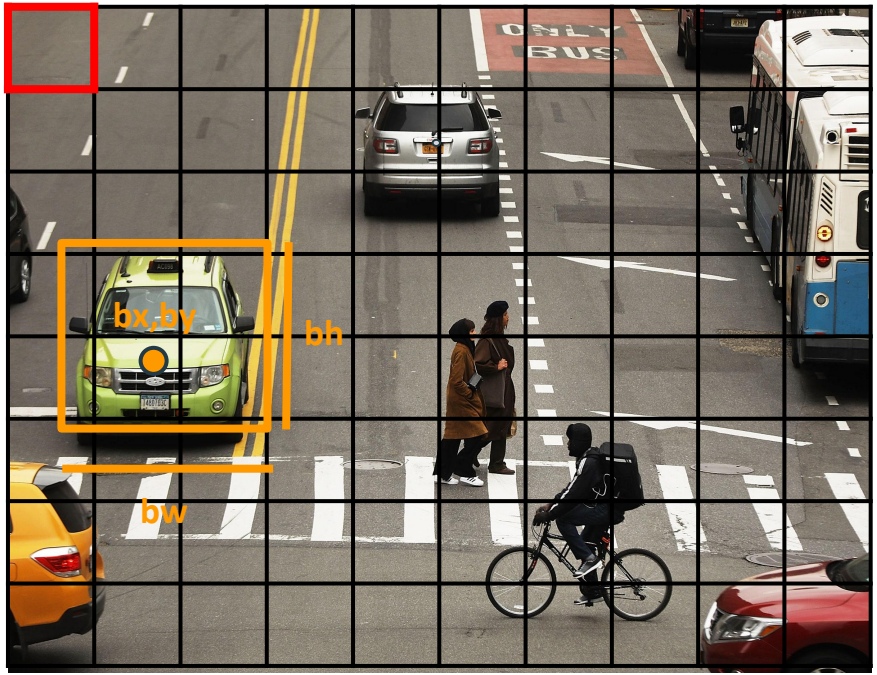
You Only Look Once(YOLO)





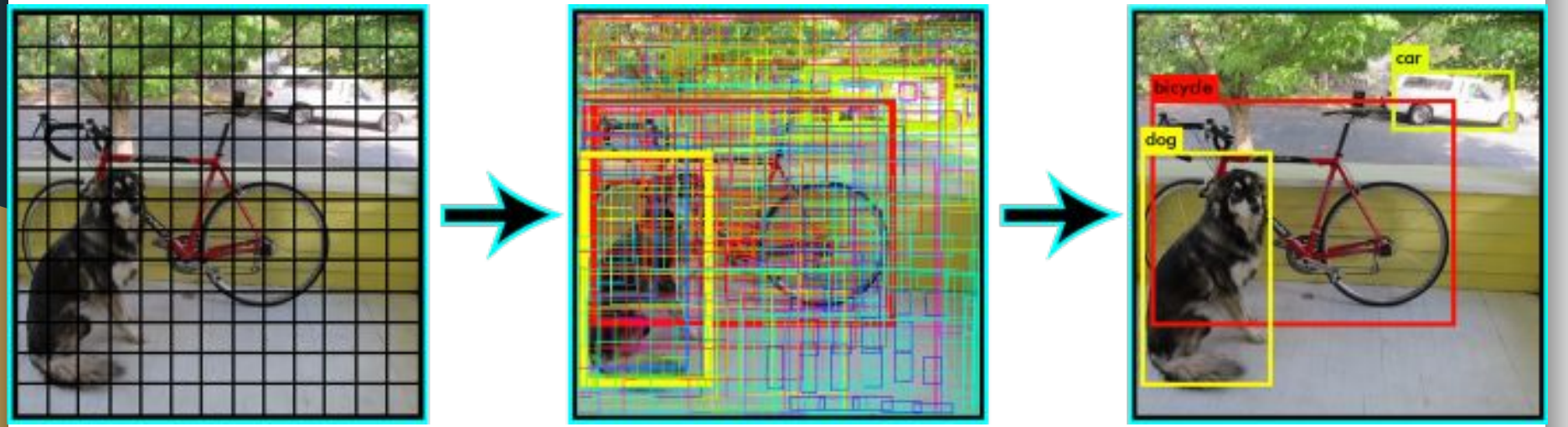
Output label of a Grid Cells

(bx, by):Mid Point, (bh,bw):Height and Width, Assuming two Classes c1:car, c2: bicycle



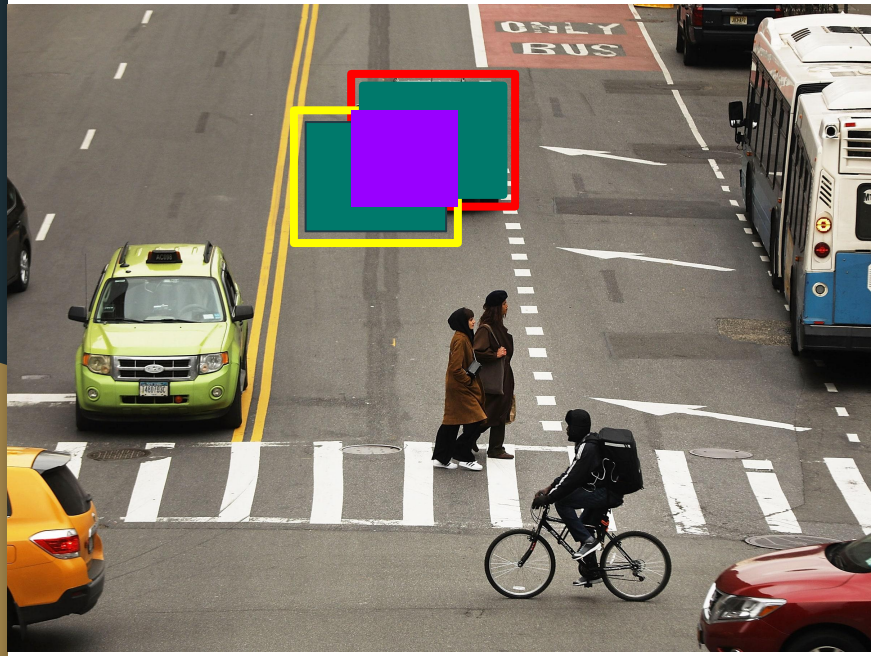
Output Format	Output Cell(0,0)	Output Cell(4,1)
pc	0	1
bx	?	bx
by	?	by
bh	?	bh
bw	?	bw
c1	?	1
c2	?	0

Model Output



Source

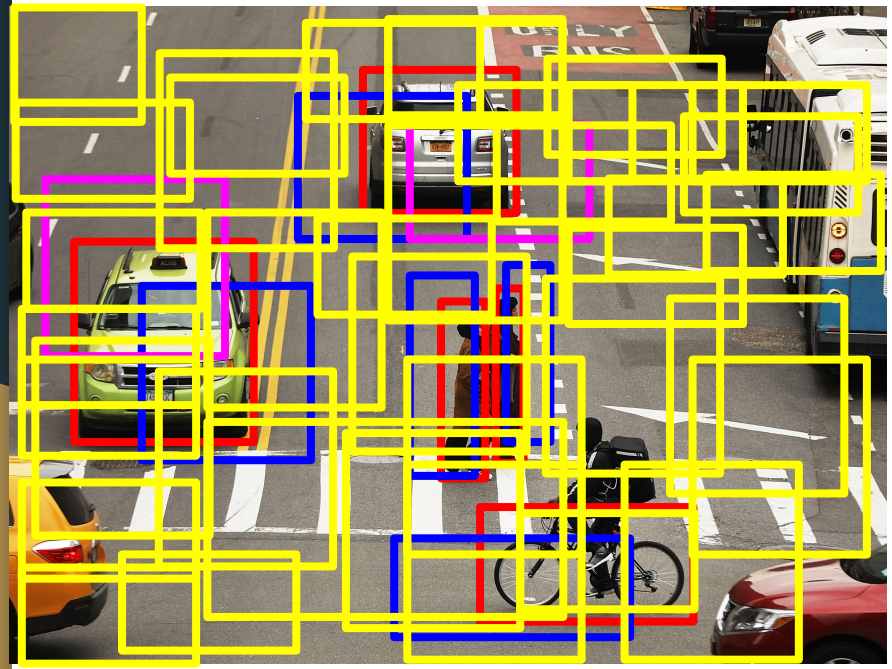
Intersection Over Union(IOU)



IOU = (size of intersection)/(size of union)

IOU = (area of **purple region**)/(area of **green region**)

Non-Max Suppression Algorithm

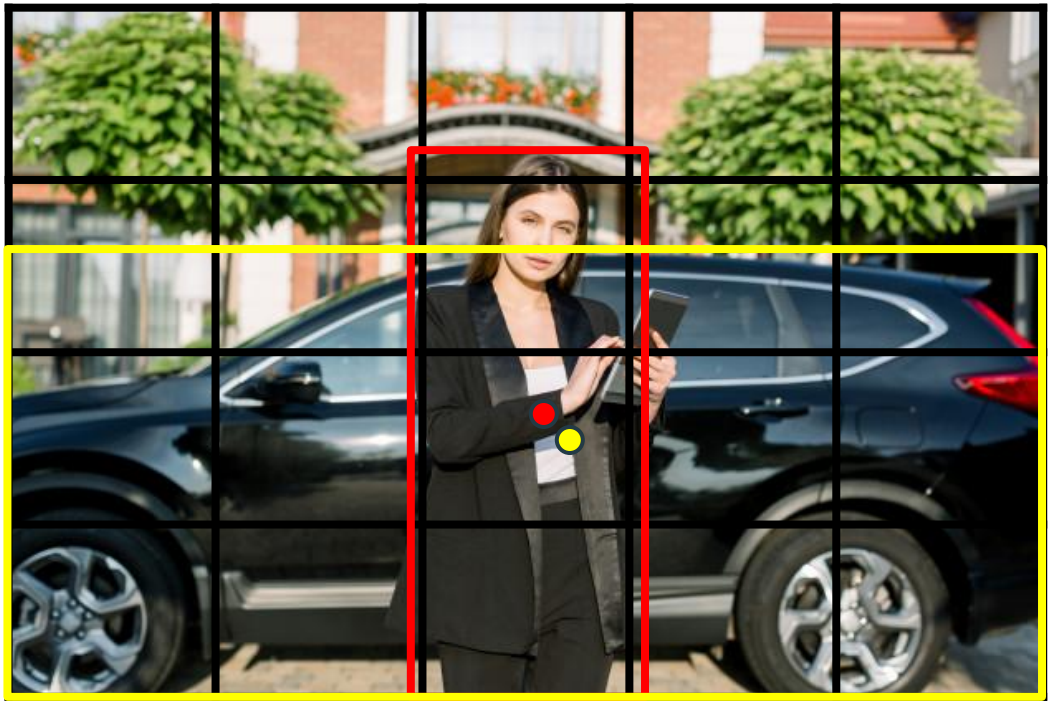


Discard all the bounding boxes having pc < 0.6

While there are any remaining boxes:

- Pick the box having largest pc output that as the prediction.
- Discard any remaining box with IOU >= 0.5 with the box output in the previous step

Anchor Boxes



Assuming two classes c1:human, c2:car

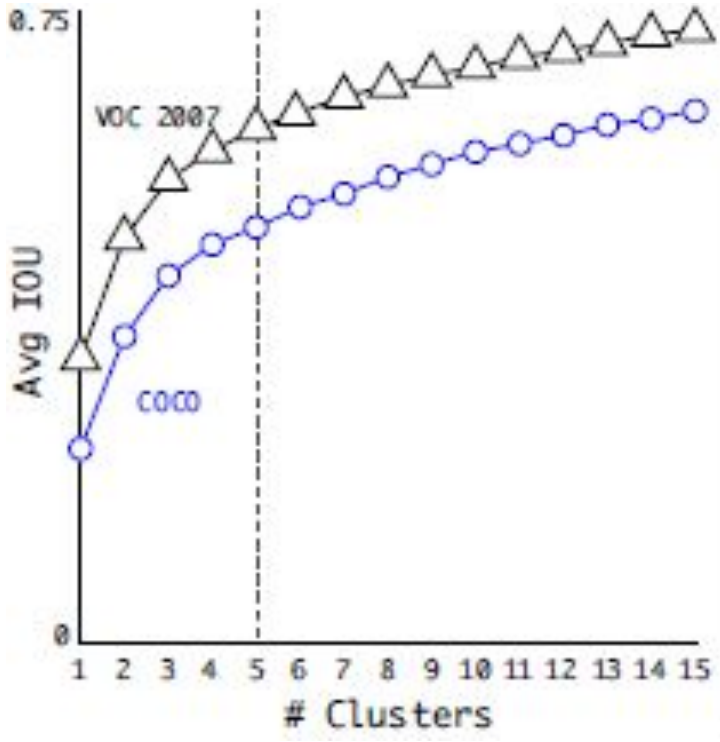
What if Midpoint of two object lie in one grid cell?

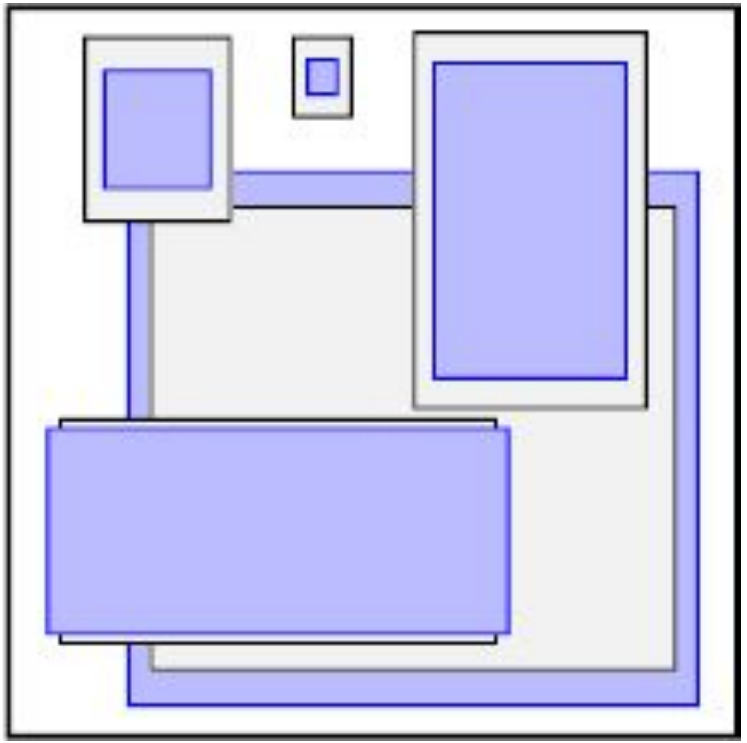
We will use concept of anchor box due to this one grid cell can detect multiple object

Output cell(2,2)

1	1
bx	bx
by	by
bh	bh
bw	bw
1	0
0	1

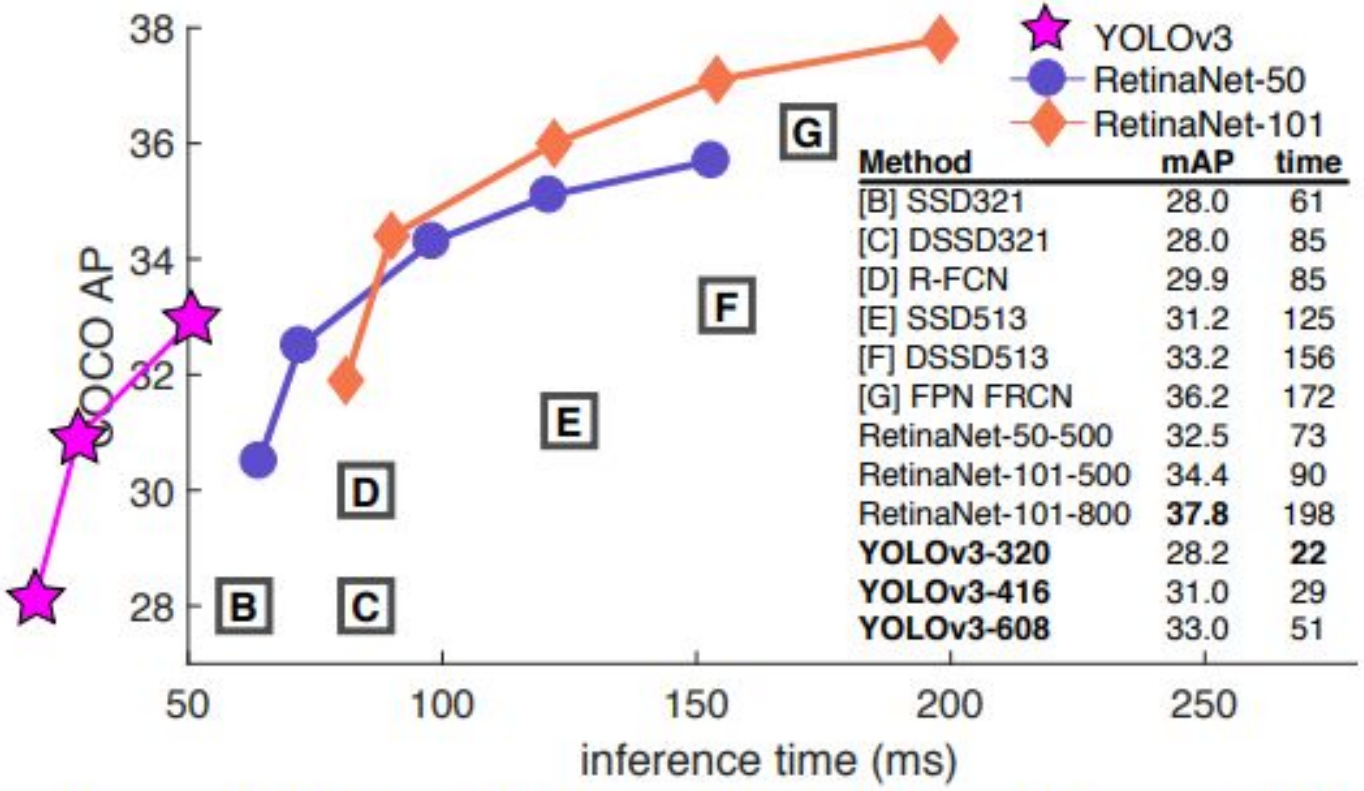
How YOLO Decide Number of Anchor Box





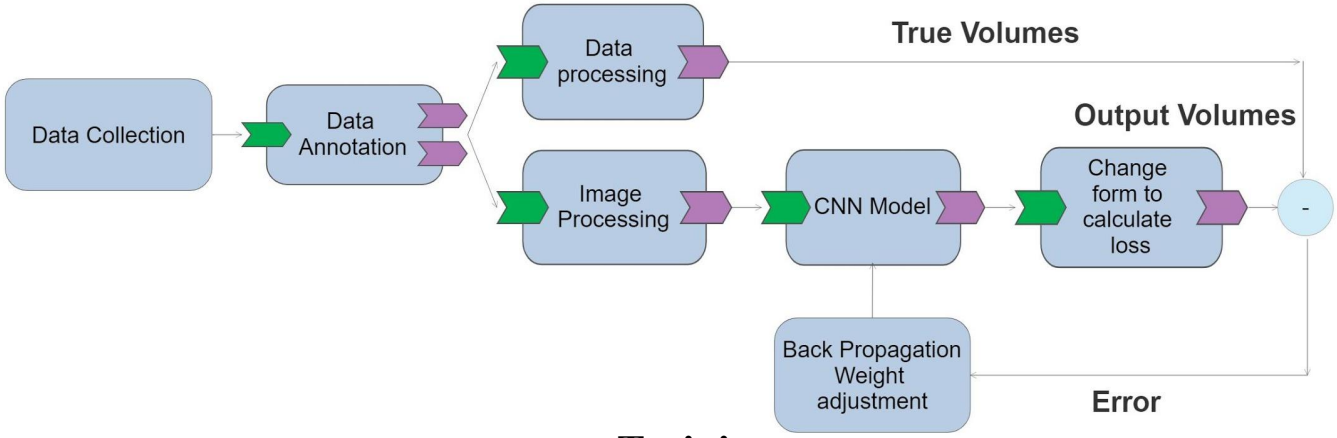
[J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 6517-6525, doi: 10.1109/CVPR.2017.690].

Accuracy and Speed:

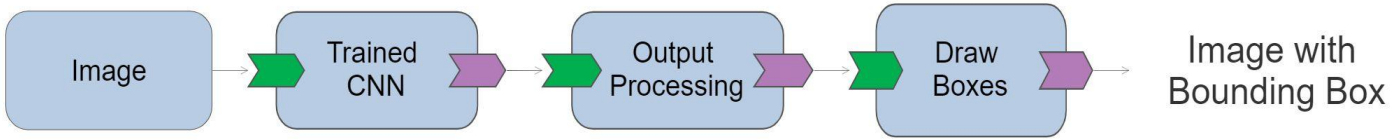


Method	mAP	time
[B] SSD321	28.0	61
[C] DSSD321	28.0	85
[D] R-FCN	29.9	85
[E] SSD513	31.2	125
[F] DSSD513	33.2	156
[G] FPN FRCN	36.2	172
RetinaNet-50-500	32.5	73
RetinaNet-101-500	34.4	90
RetinaNet-101-800	**37.8**	198
YOLOv3-320	28.2	**22**
YOLOv3-416	31.0	29
YOLOv3-608	33.0	51

System Design



Training



Testing


Object Detection
starring
YOLOv3

References:

- Sermanet, Pierre & Eigen, David & Zhang, Xiang & Mathieu, Michael & Fergus, Rob & Lecun, Yann. (2013). OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. International Conference on Learning Representations (ICLR) (Banff).

- R. 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, Columbus, OH, 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81.

- R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

- S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

- J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

References:

- J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 6517-6525, doi: 10.1109/CVPR.2017.690.

- Redmon, Joseph & Farhadi, Ali. (2018). YOLOv3: An Incremental Improvement.

- Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li and S. Hu, "Traffic-Sign Detection and Classification in the Wild," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2110-2118, doi: 10.1109/CVPR.2016.232.

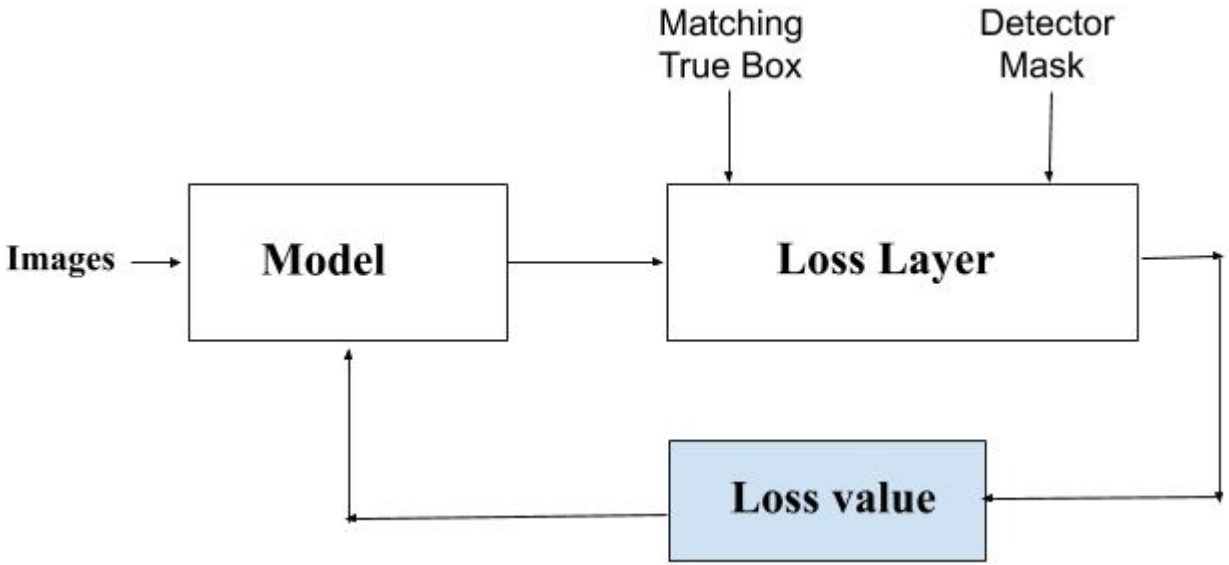
- Rongqiang Qian, Bailing Zhang, Yong Yue, Zhao Wang and F. Coenen, "Robust chinese traffic sign detection and recognition with deep convolutional neural network," 2015 11th International Conference on Natural Computation (ICNC), Zhangjiajie, 2015, pp. 791-796, doi: 10.1109/ICNC.2015.7378092.

ANY QUESTIONS ?

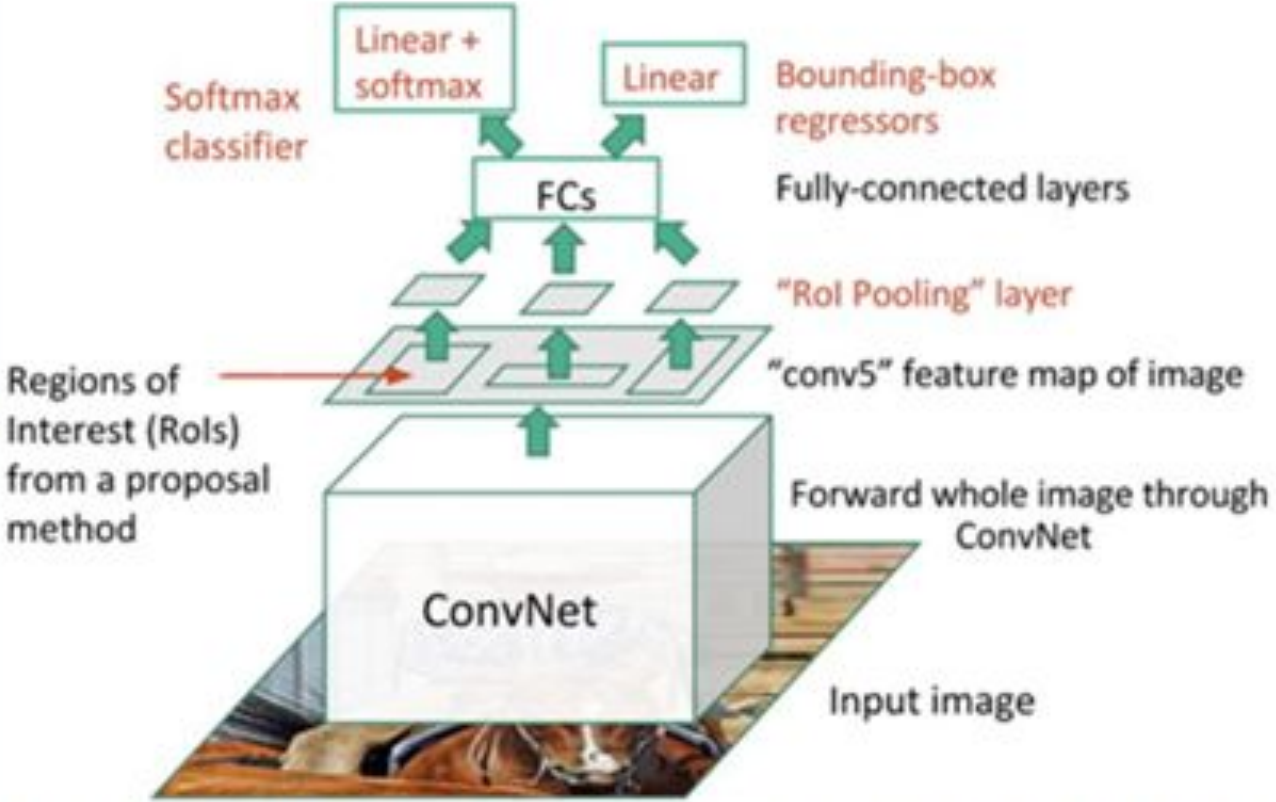
THANK YOU

BACKUP SLIDES

Loss Layer


Matching
True Box
Detector
Mask
Images
Model
Loss Layer
Loss value

Fast R-CNN



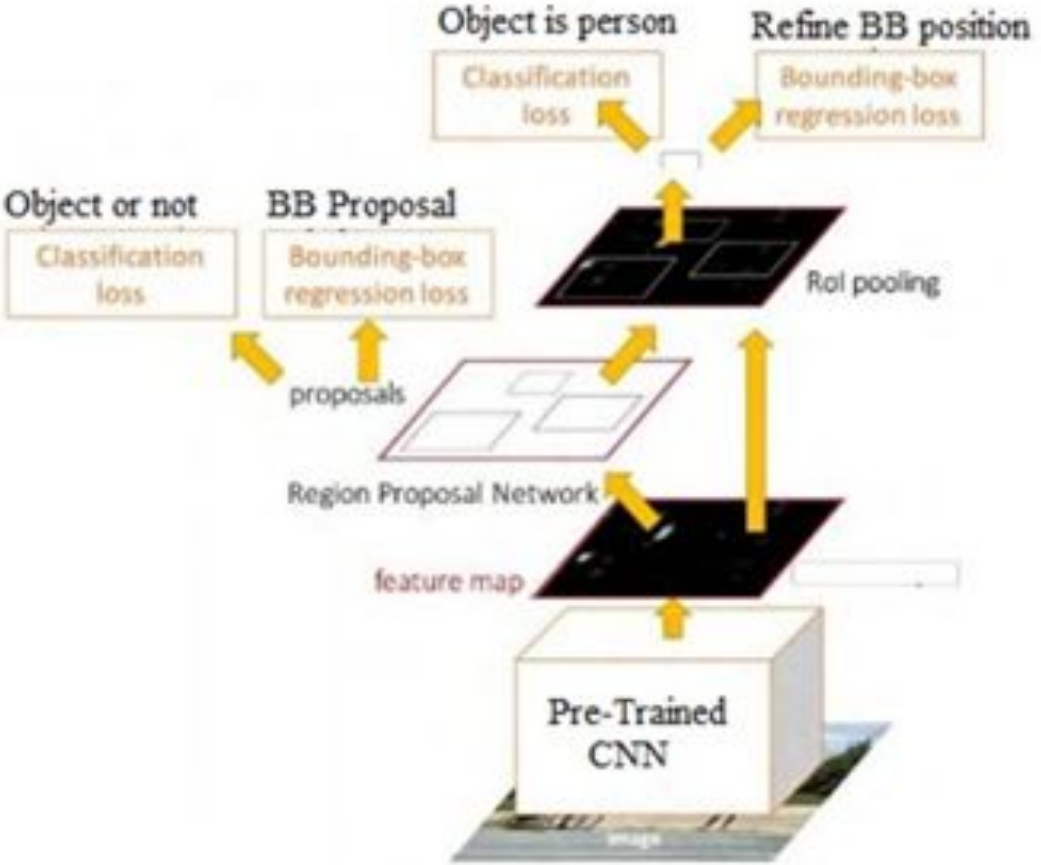
Drawback:

Slow selective search algorithm is used for ROI

[R. Girshick, "Fast R-CNN," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169]

Faster R-CNN

- The main insight of Faster R-CNN was to replace the slow selective search algorithm.
- It introduce the Region Proposal Network



[Ren et. al 2016 Faster R-CNN: Towards real-time object detection with region proposal network]

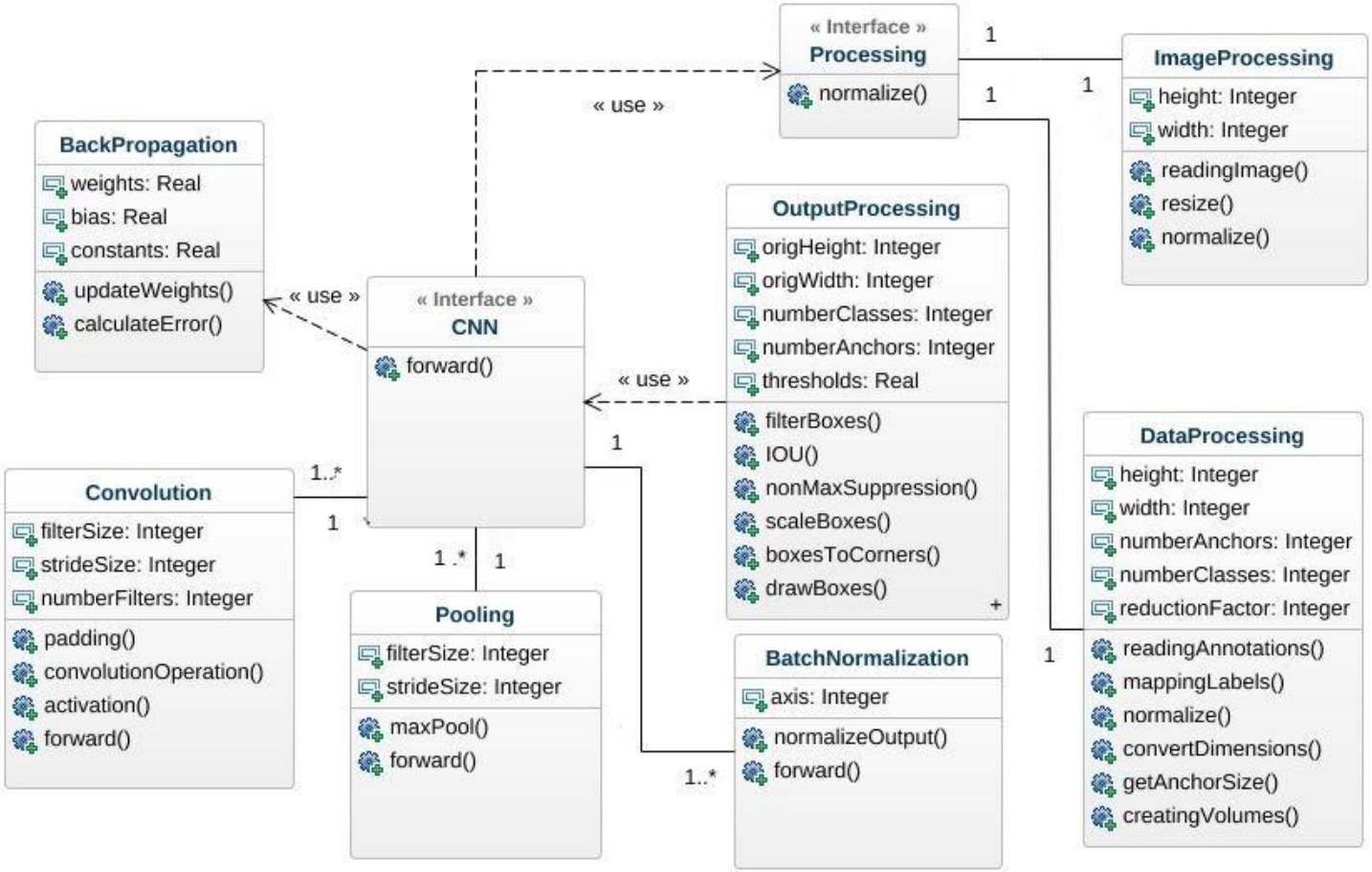
Preprocessing Formula's

1. x_mid = (x_max - x_min) / 2
2. y_mid = (y_max - y_min) / 2
3. height = y_max - y_min
4. width = x_max - x_min

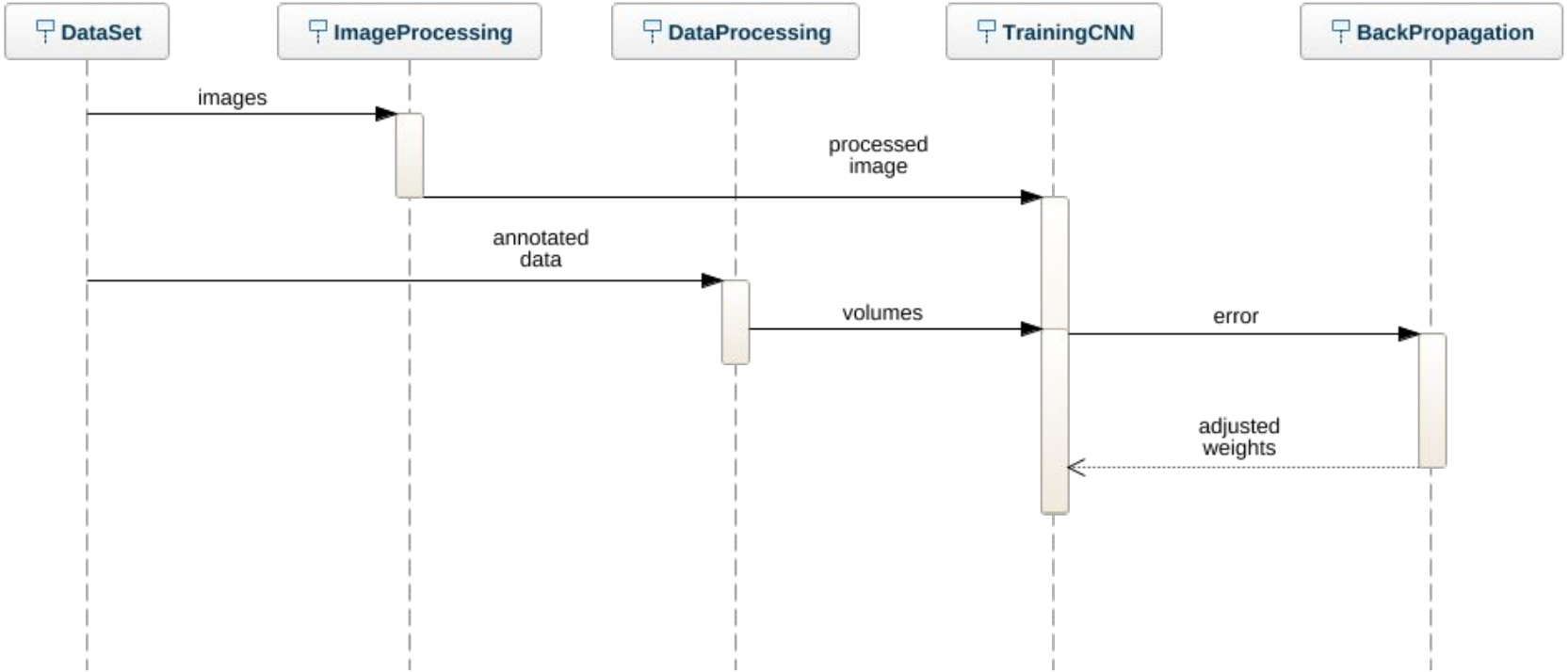
Loss Formula's

1. Classification_loss = λ_{class} * (detector_mask) * (original_class - predicted_class)^2
2. Localization_loss = λ_{coord} * (detector_mask) * (matching true boxes - prediction boxes)
3. Object_loss = λ_{object}* (detector_mask) * (1 - confidence)^2
4. No_object_loss = λ_{no_object}* (1 - object_detections) * (1 - detector_mask) * (confidence)^2
5. Total loss = Classification Loss + Coordinate Loss + Confidence Loss

Class Diagram



Sequence Diagram(Training)



Sequence Diagram(Testing)