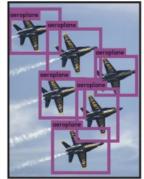


You Only Look Once: Unified, Real-Time Object Detection Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi

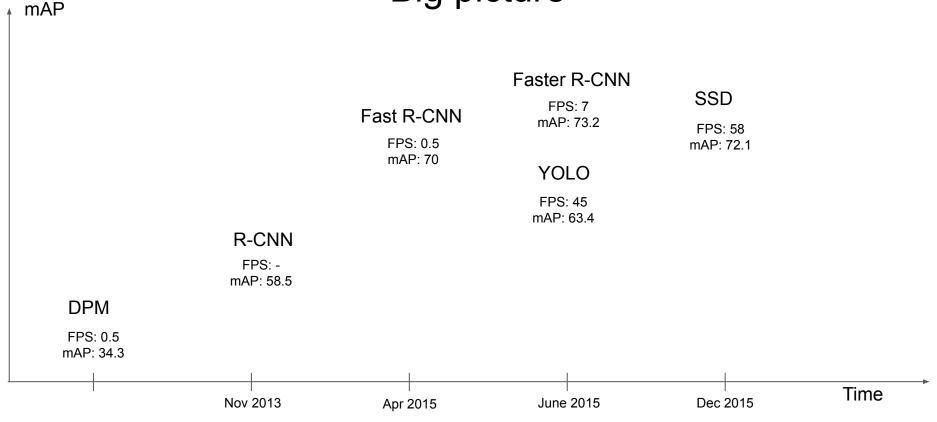




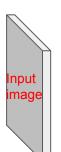




Big picture







448x448x3





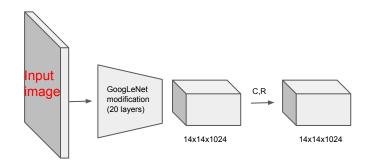
Input GoogLeNet modification (20 layers)

14x14x1024

448x448x3



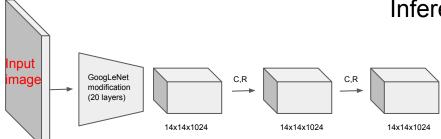




448x448x3







448x448x3



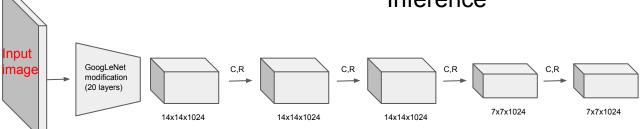


Input image GoogLeNet modification (20 layers) 14x14x1024 C,R C,R Tx7x1024

448x448x3







448x448x3





Inference GoogLeNet modification (20 layers) 14x14x1024 C,R Tx7x1024 Tx7x1024 FC,R Tx7x1024 Tx7x1024 4096x1





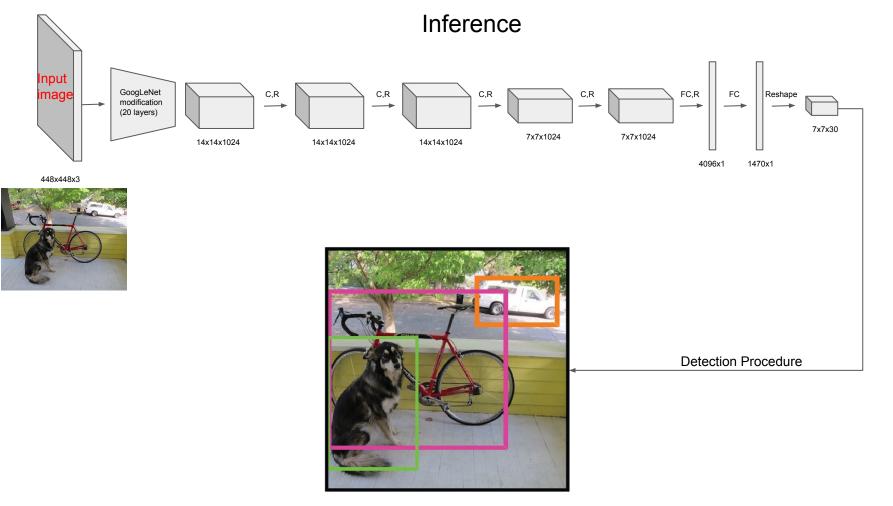
Inference Input GoogLeNet modification image C,R C,R FC,R FC (20 layers) 7x7x1024 7x7x1024 14x14x1024 14x14x1024 14x14x1024 4096x1 1470x1

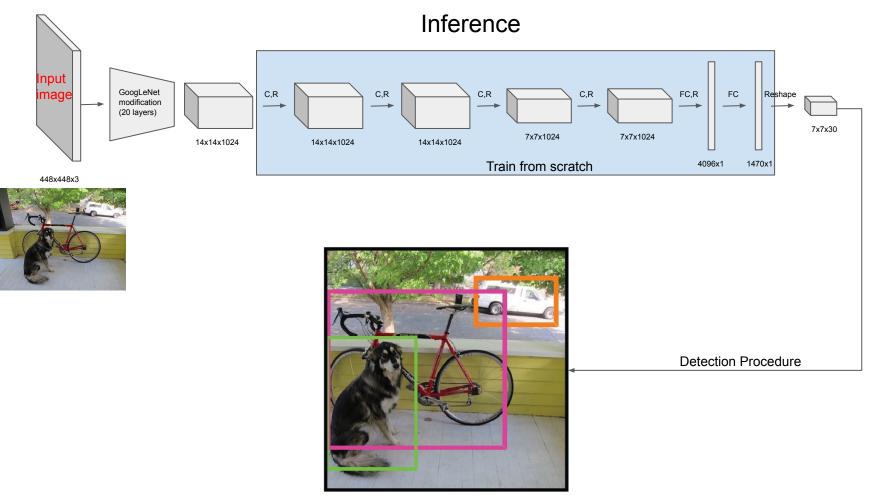




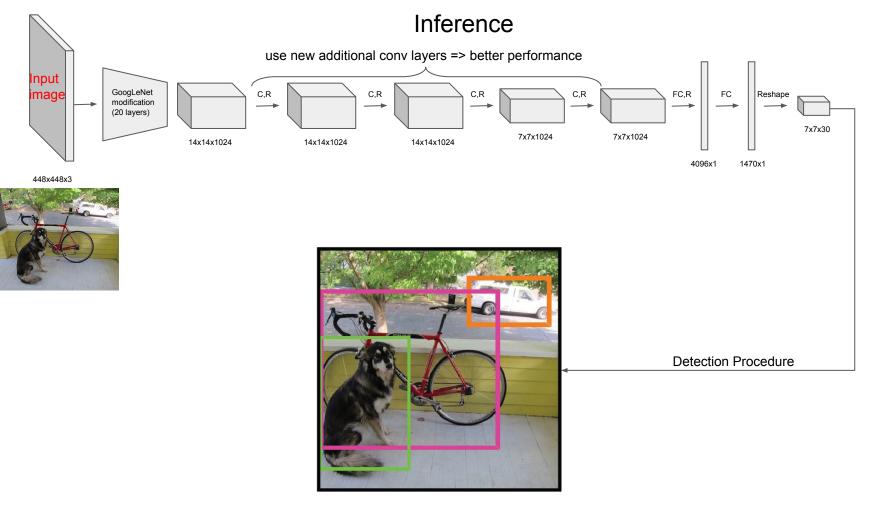
Inference Input GoogLeNet modification image C,R C,R FC,R FC Reshape (20 layers) 7x7x30 7x7x1024 7x7x1024 14x14x1024 14x14x1024 14x14x1024 4096x1 1470x1

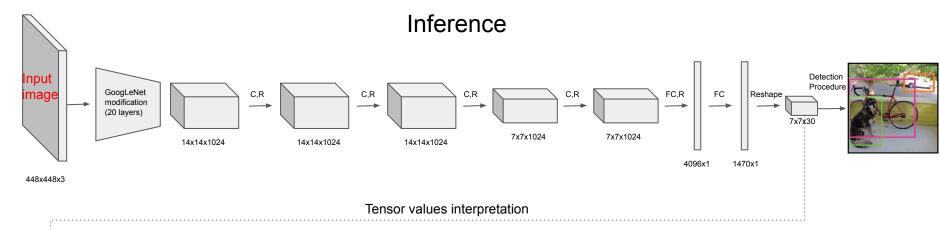


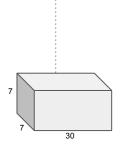


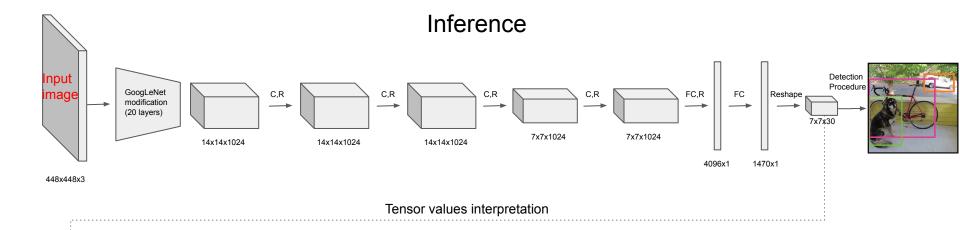




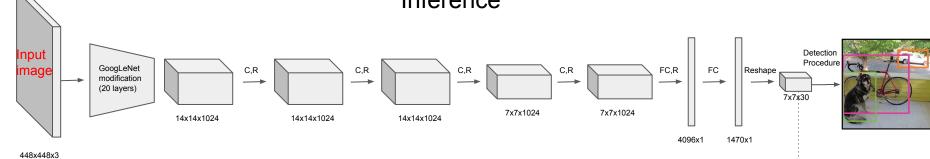




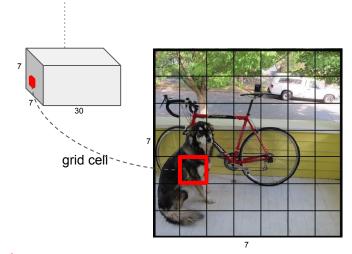




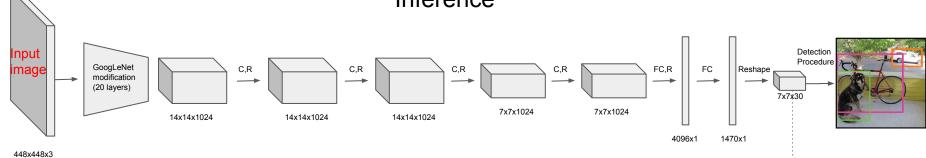




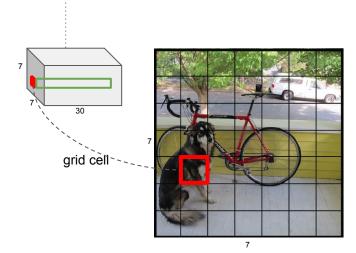
Tensor values interpretation



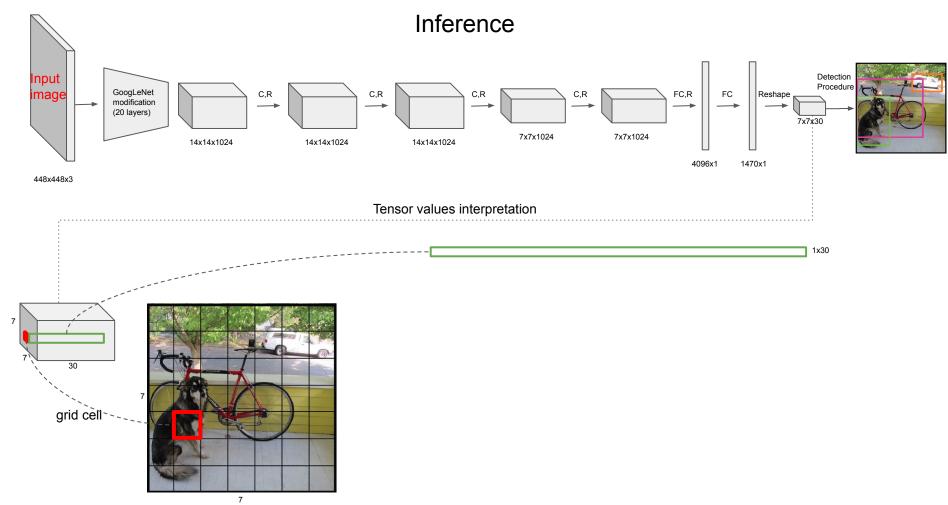




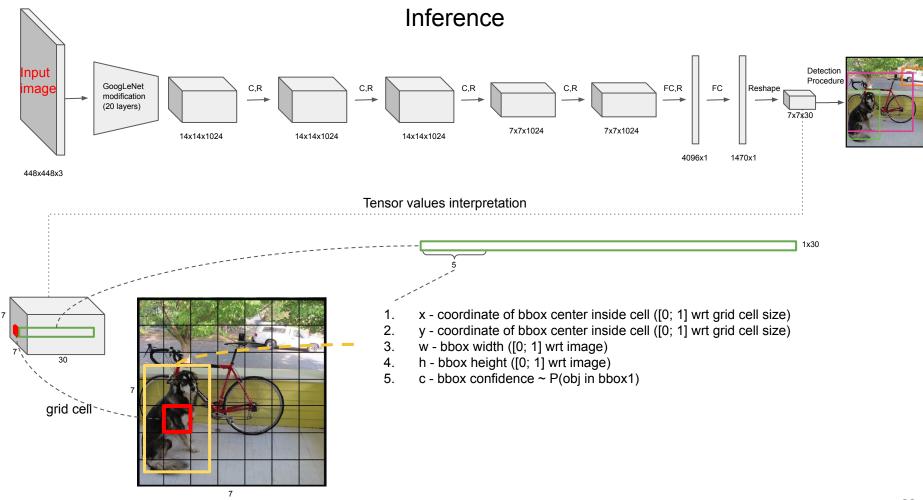
Tensor values interpretation

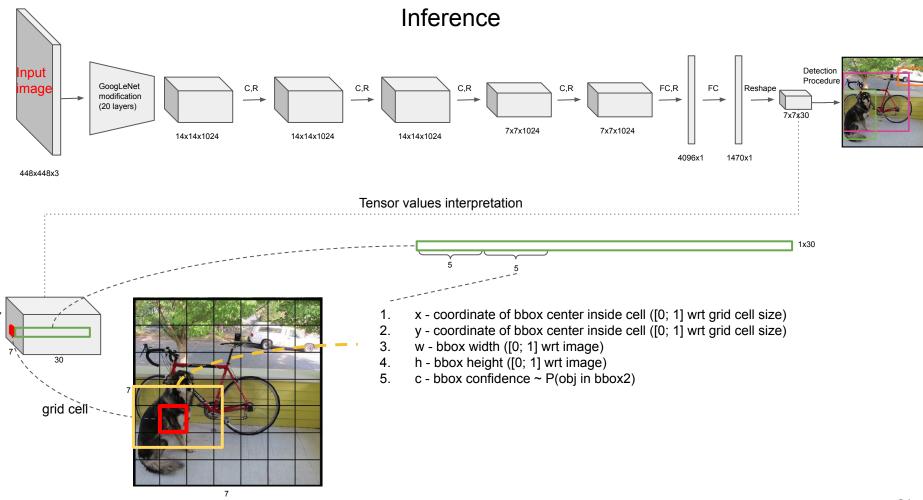


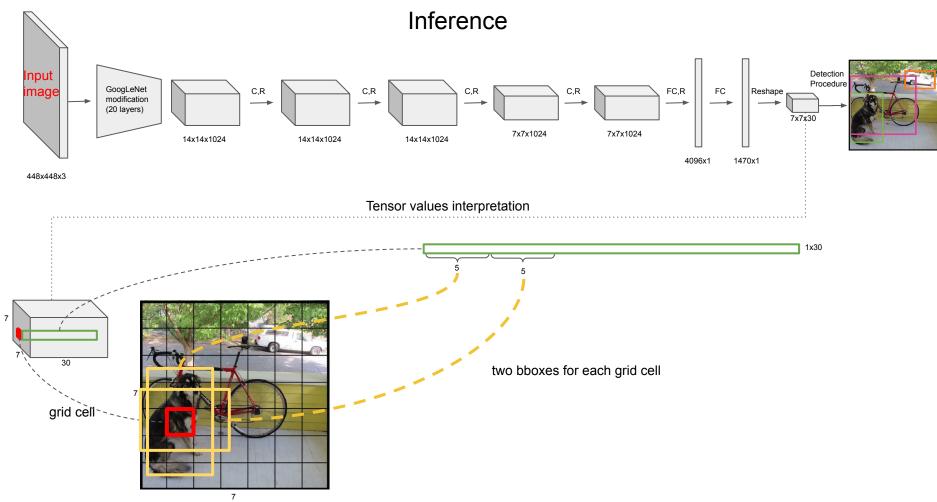




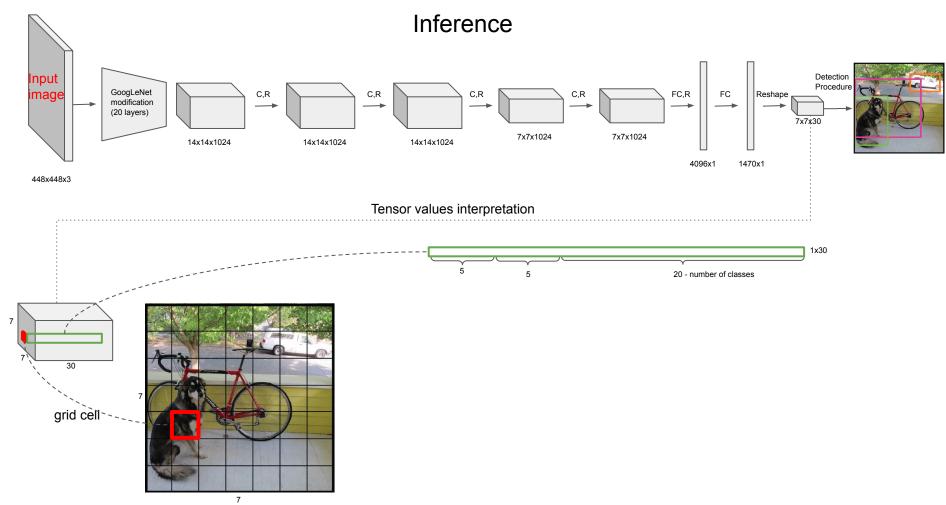




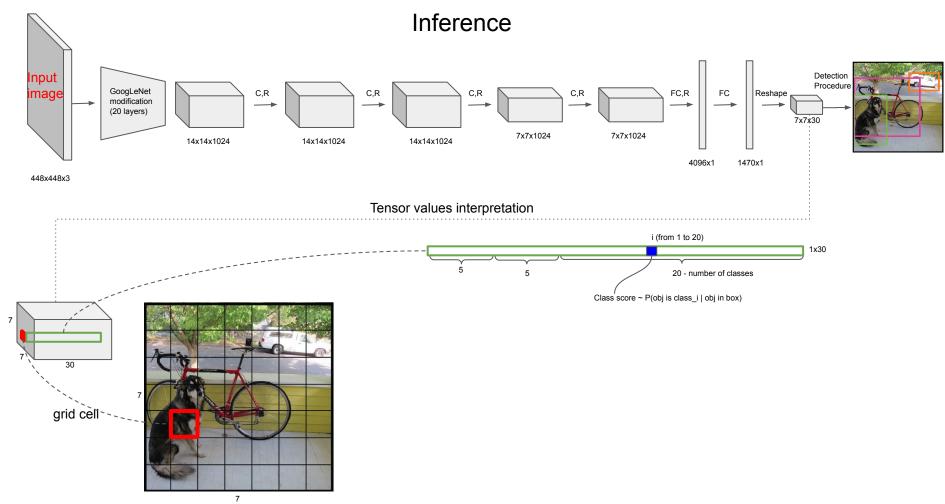


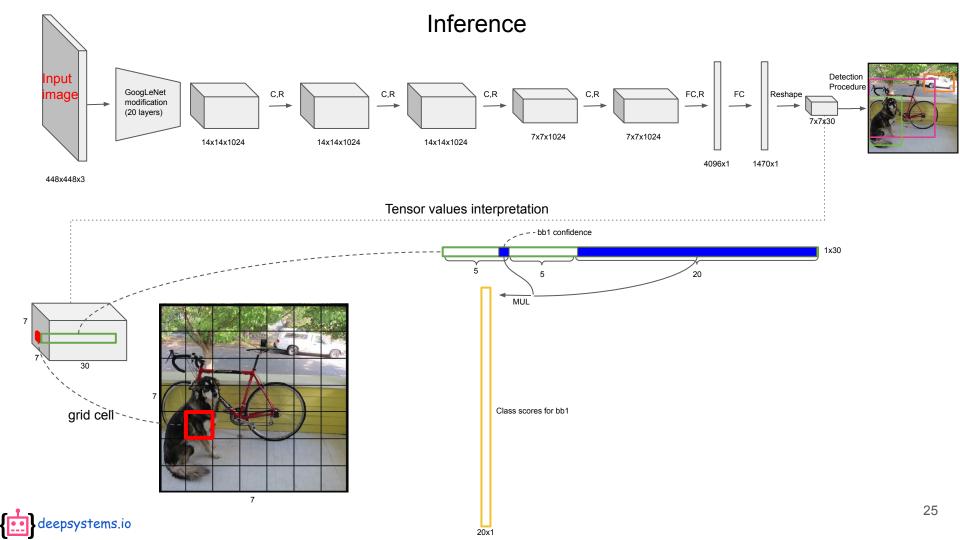


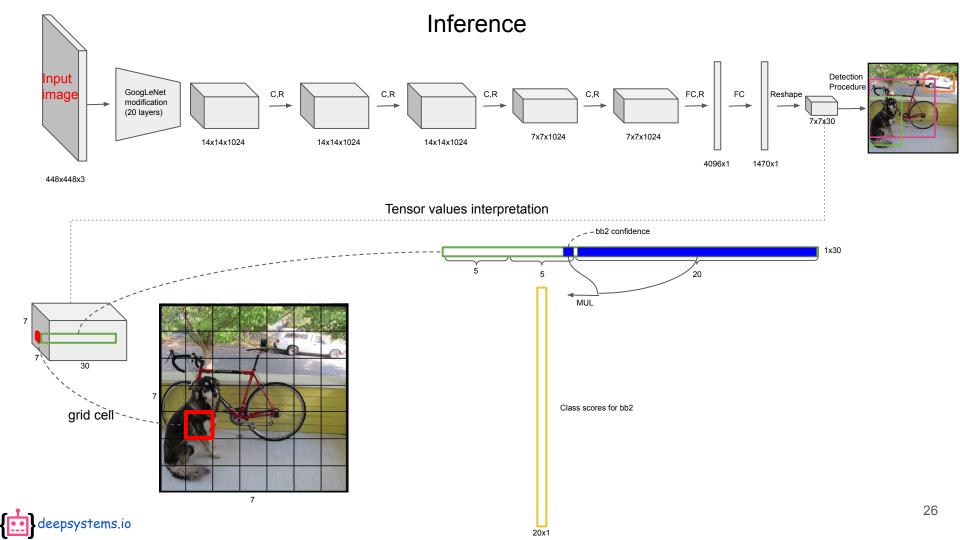


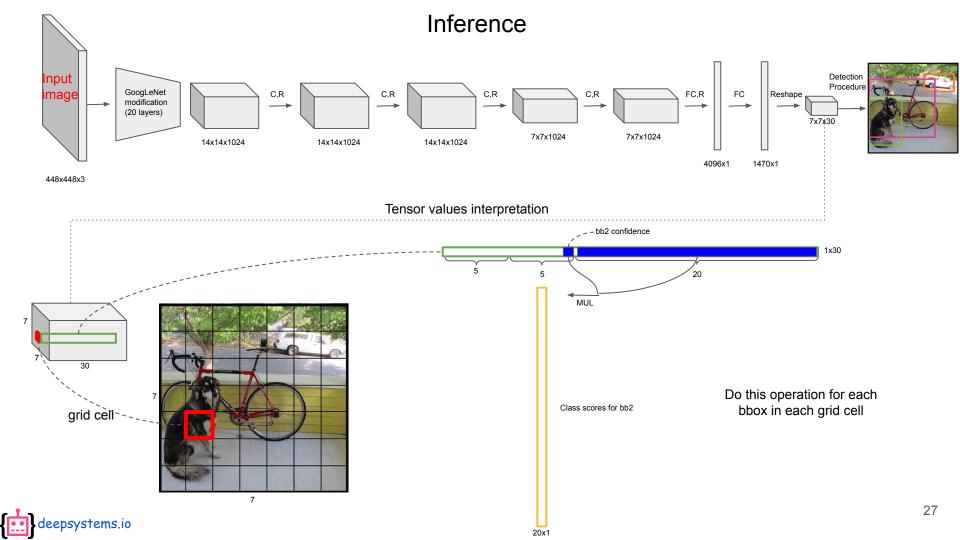


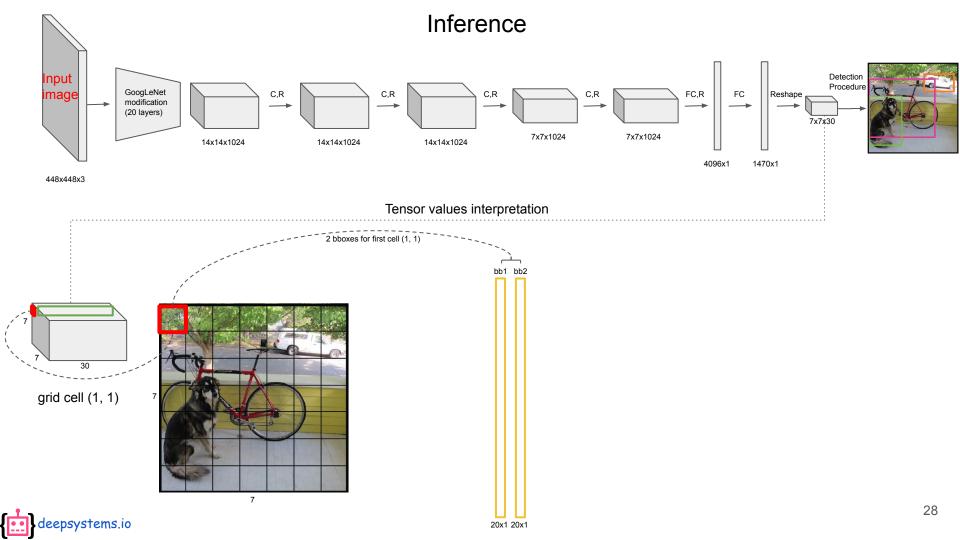


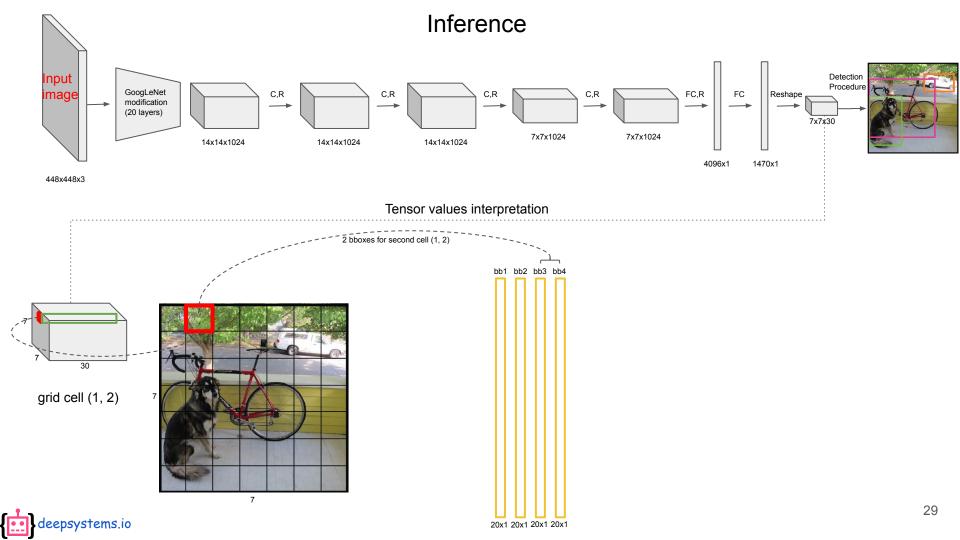


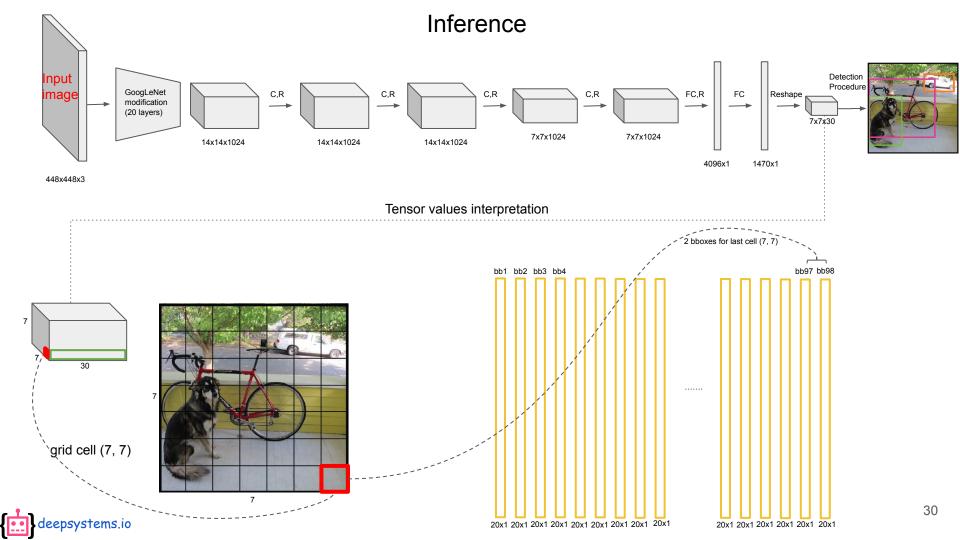


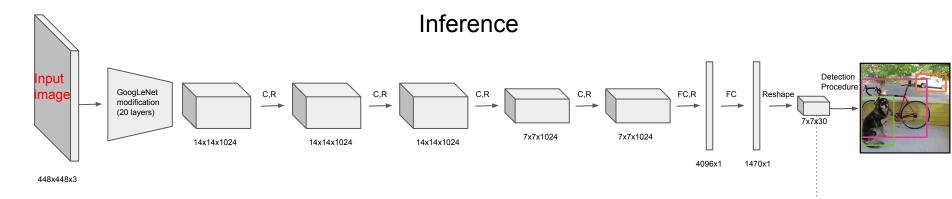




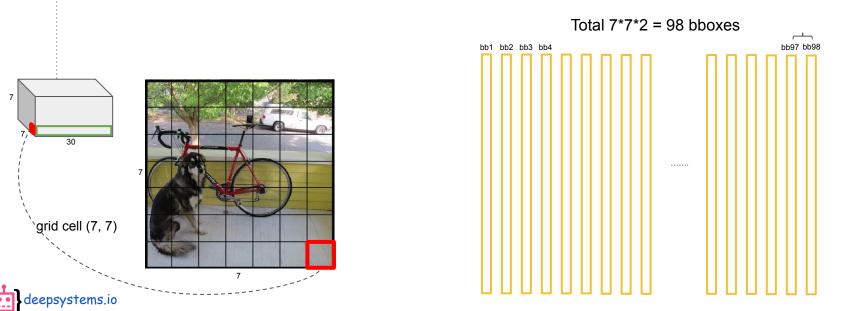




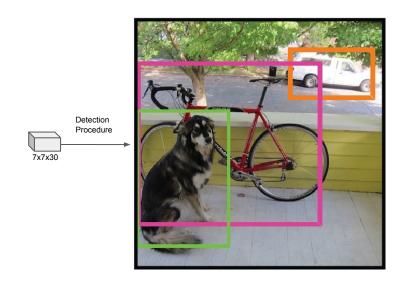




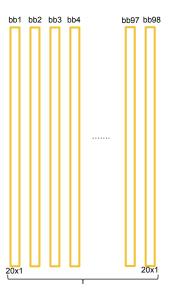




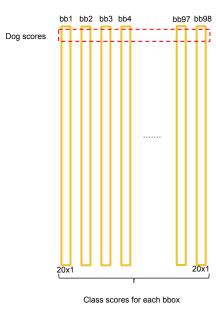
Look at detection procedure



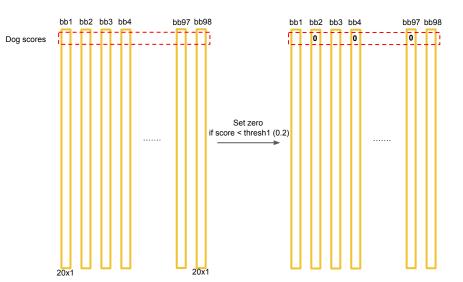


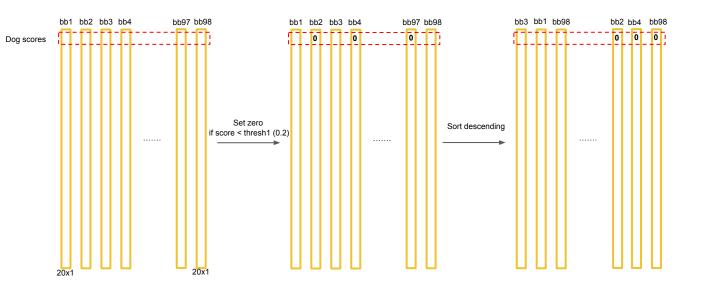


Class scores for each bbox

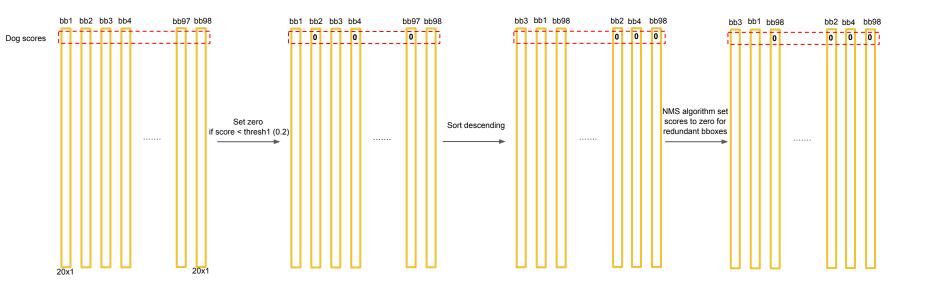




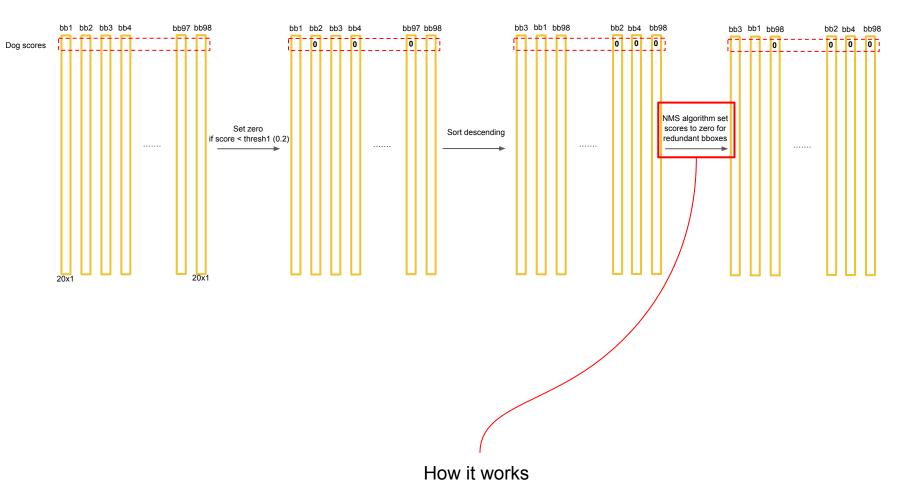








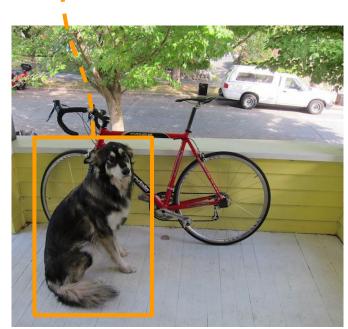




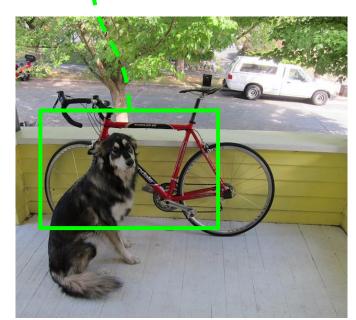




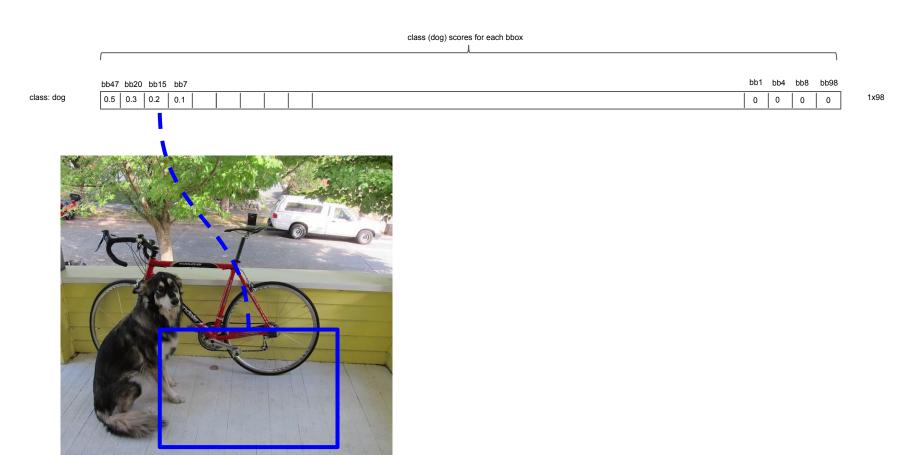




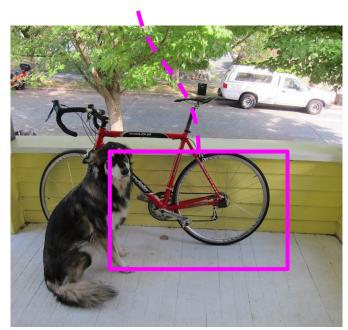






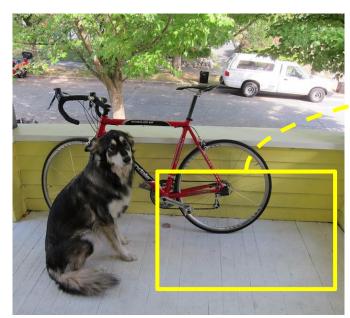






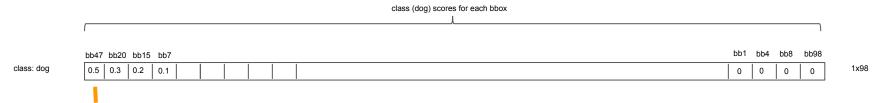


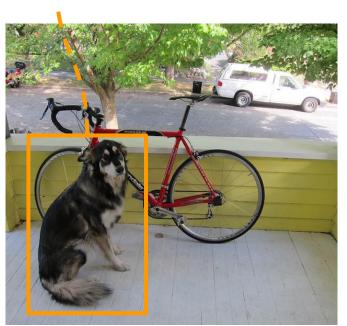
class (dog) scores for each bbox





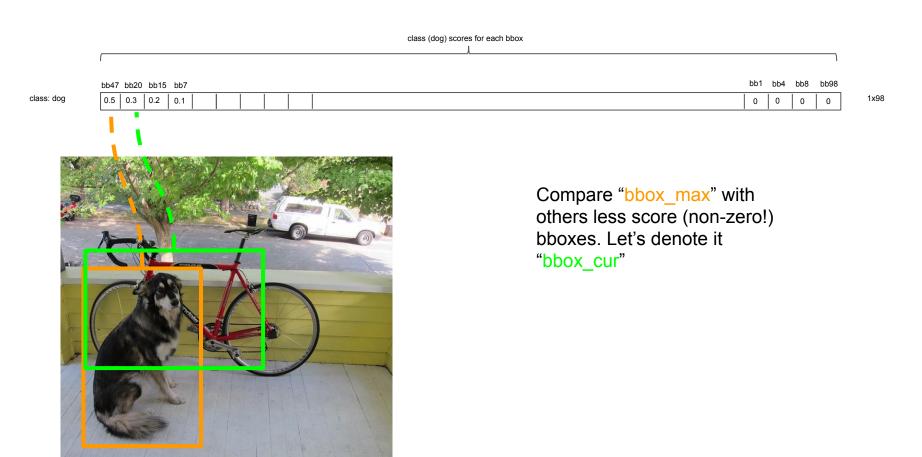
1x98



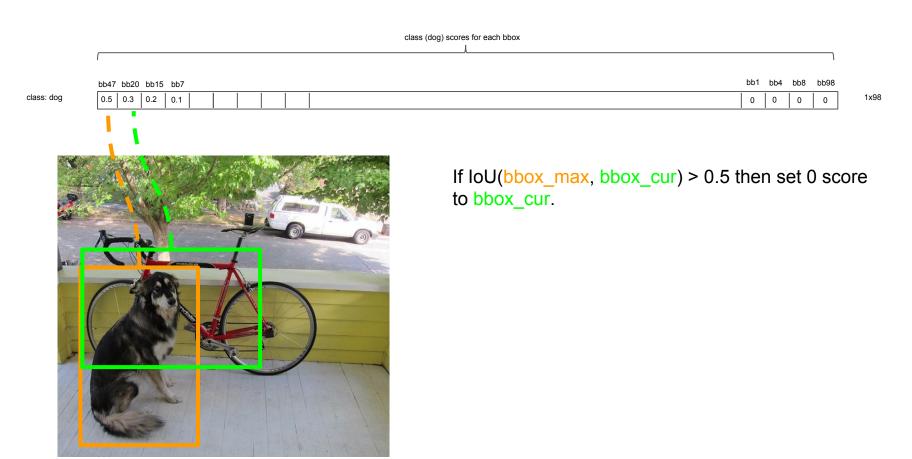


Get bbox with max score. Let's denote it "bbox_max"

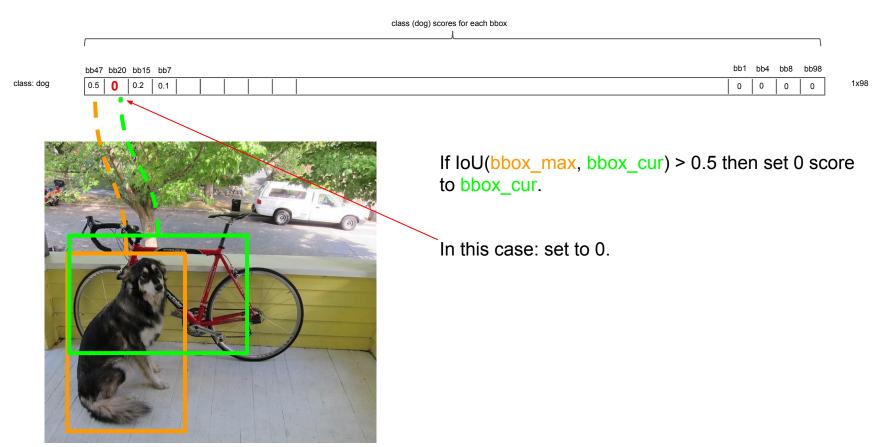




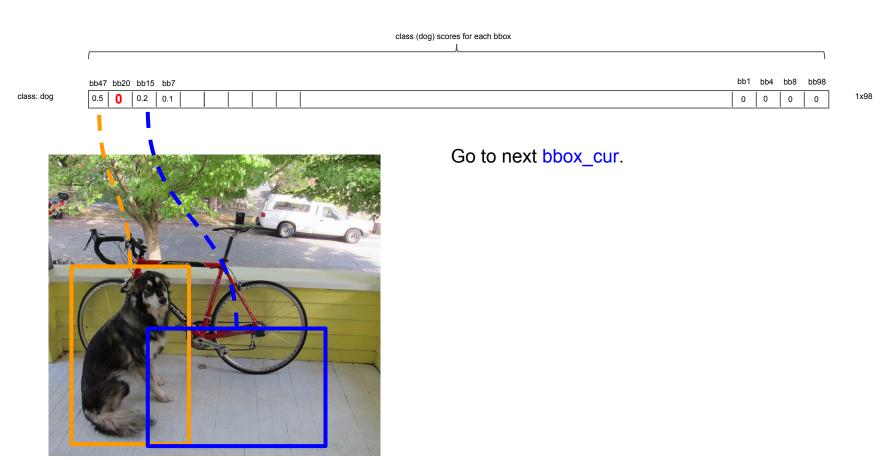




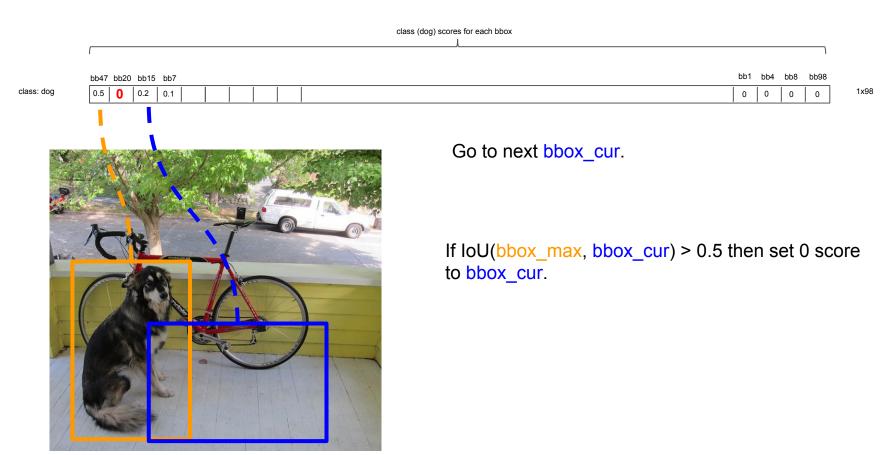




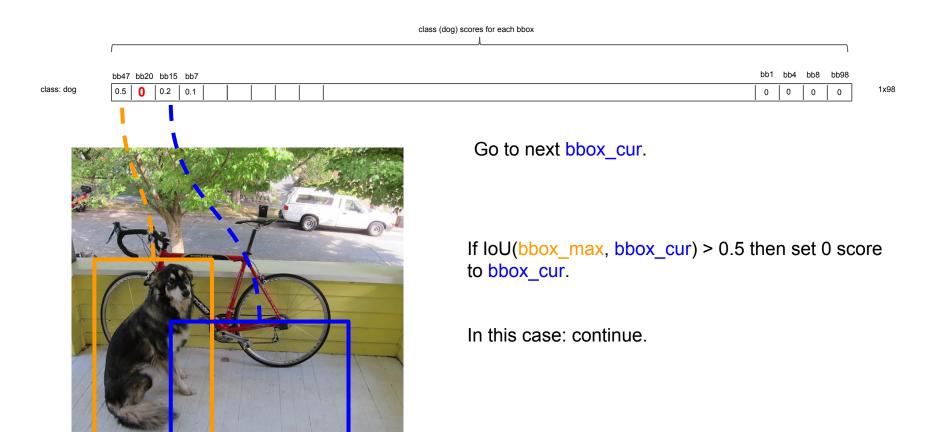




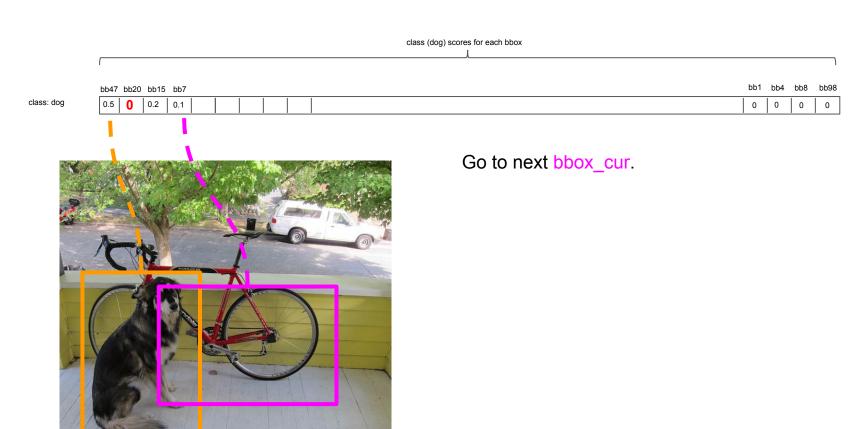






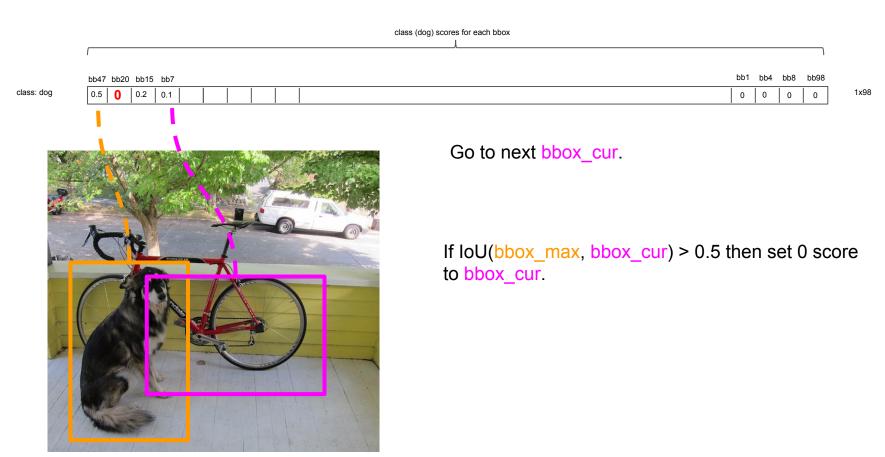








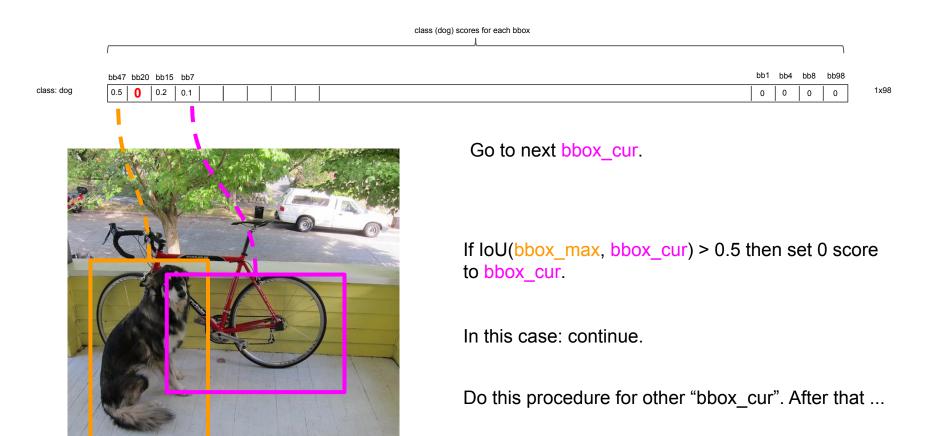
1x98



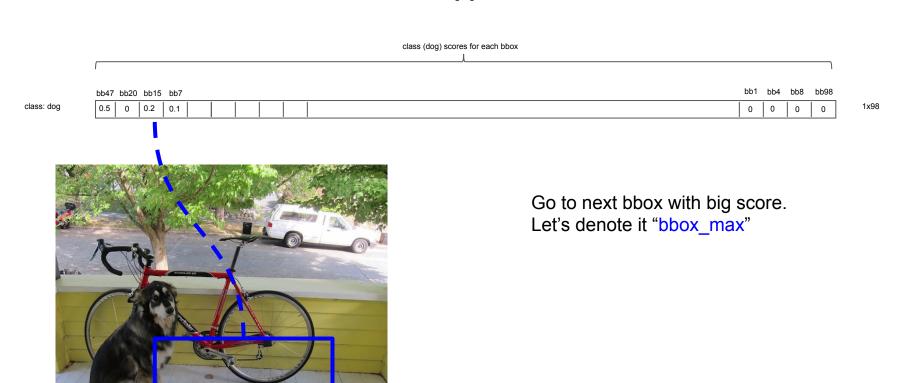








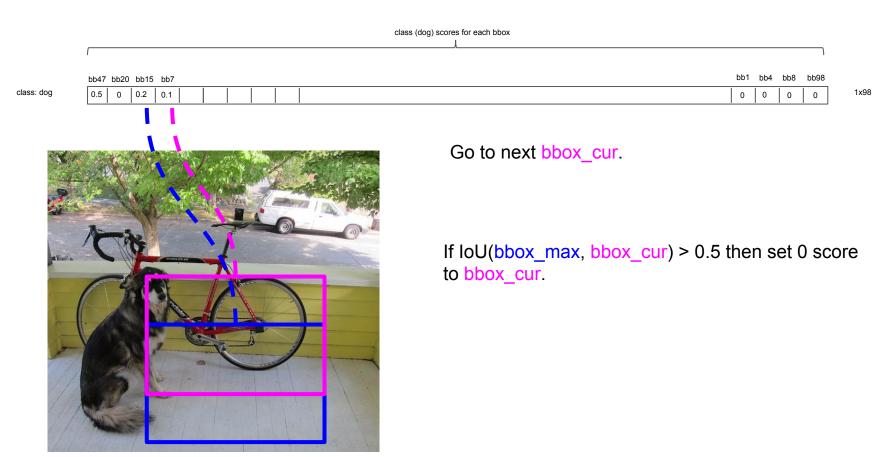




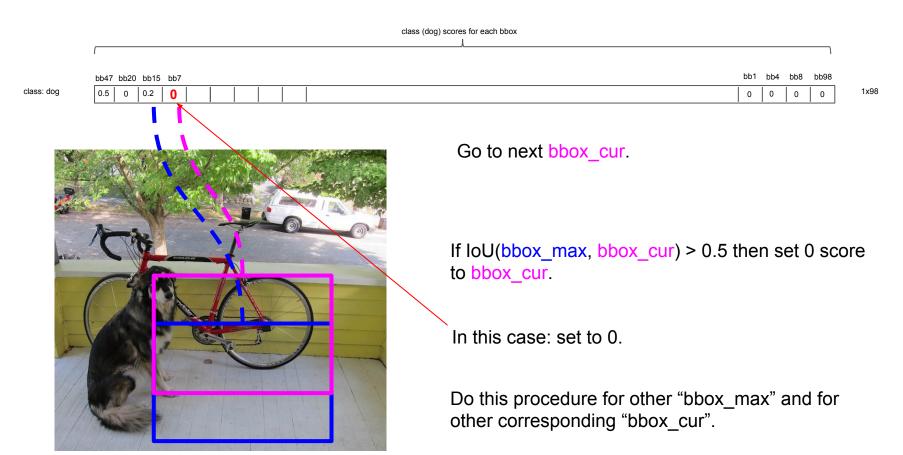








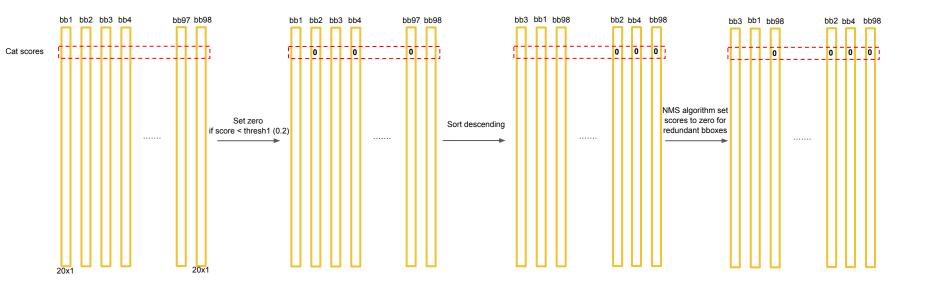




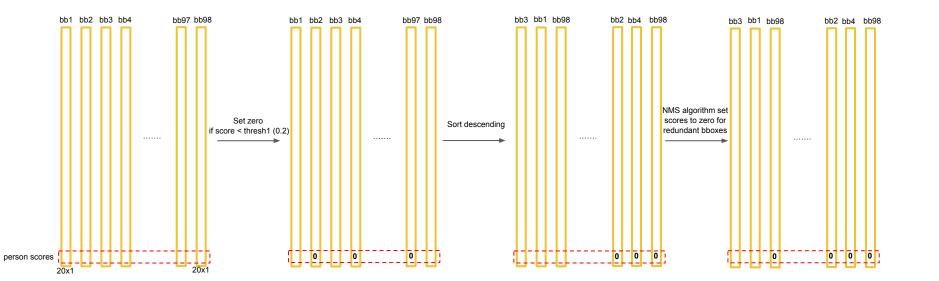




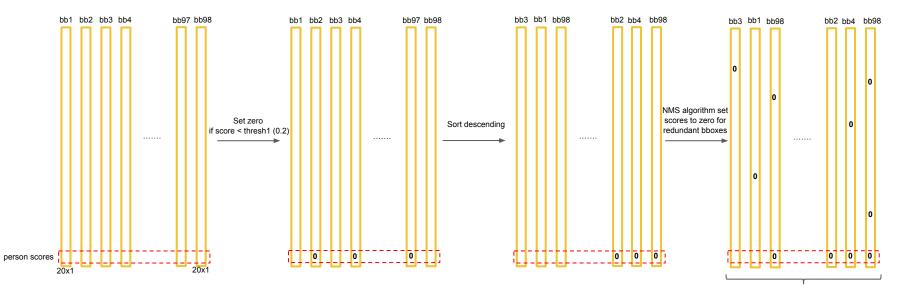






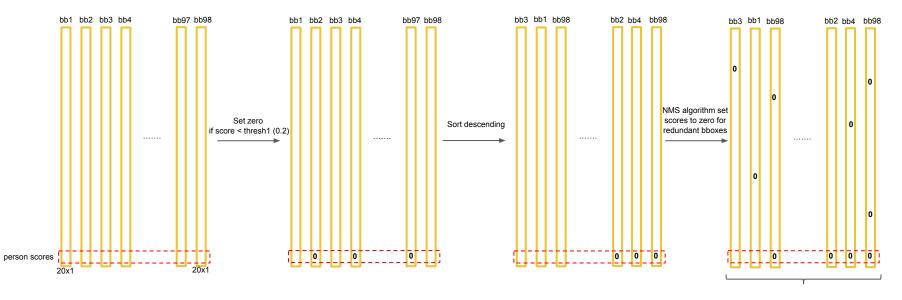






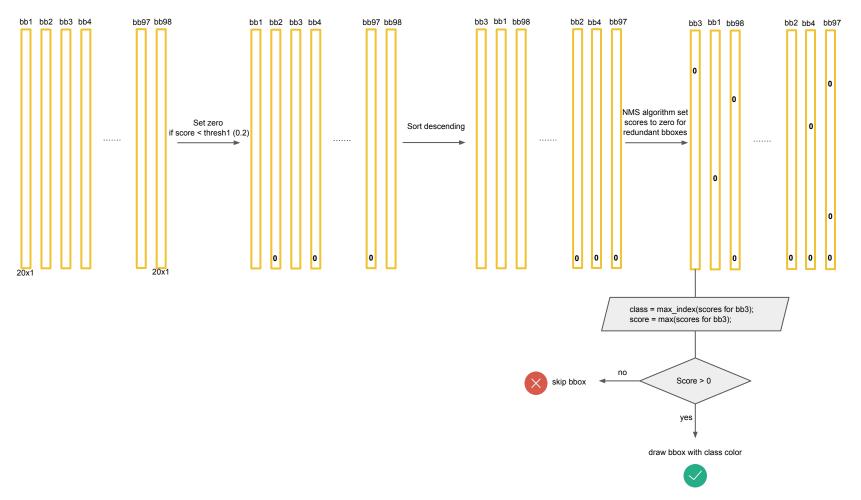
After this procedure - a lot of zeros



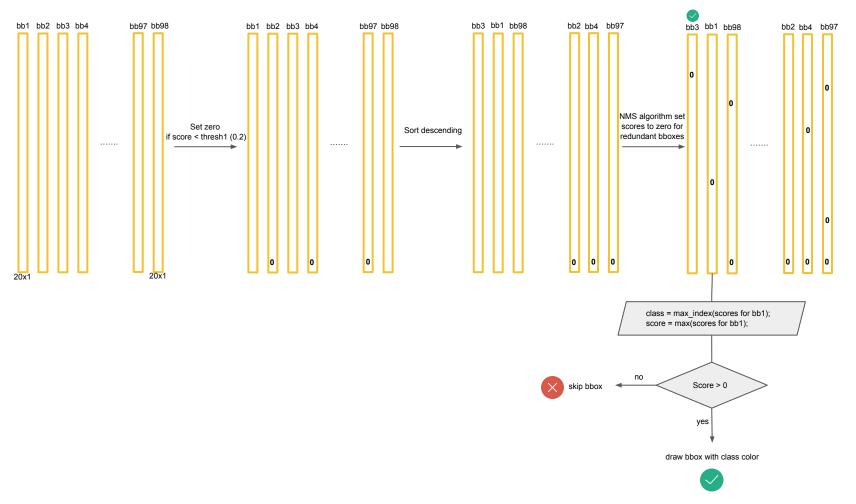


Select bboxes to draw by class score values

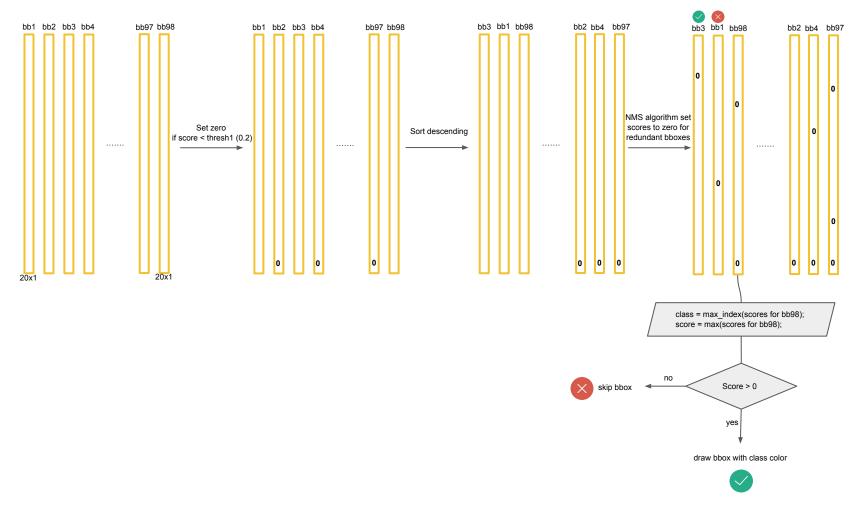




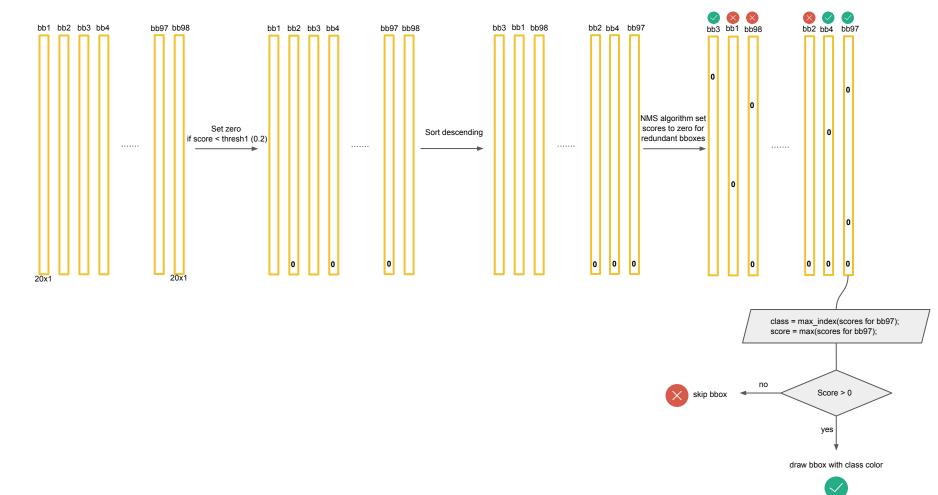




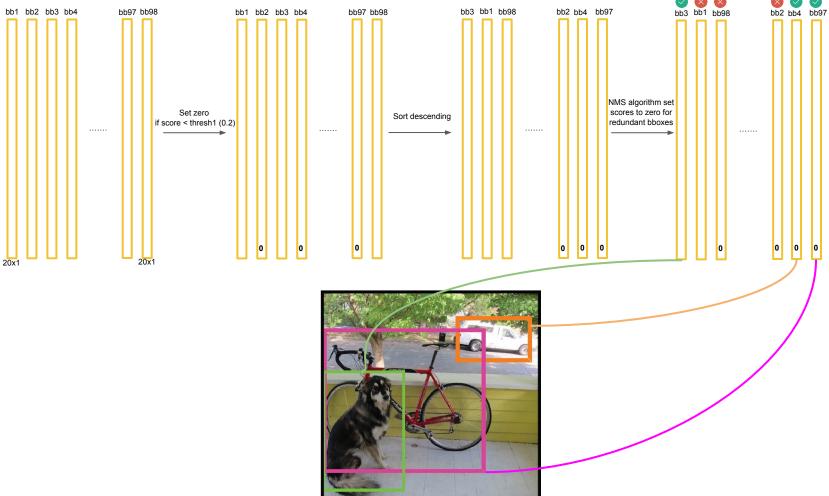












Key Points

- 1. Fast: YOLO 45 fps, YOLO-tiny 155 fps.
- End-to-end training.
- 3. Makes more localization errors but is less likely to predict false positives on background
- 4. Performance is lower than the current state of the art.
- 5. Combined Fast R-CNN + YOLO model is one of the highest performing detection methods.
- 6. Learns very general representations of objects: it outperforms other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains











Links

- Arxiv: https://arxiv.org/abs/1506.02640
- Blog: http://pjreddie.com/publications/yolo/
- Darknet: https://github.com/pjreddie/darknet
- Caffe: https://github.com/xingwangsfu/caffe-yolo
- Tensorflow:
 - Test+train: https://github.com/thtrieu/yolotf
 - Test: https://github.com/qliese581qq/YOLO tensorflow



Thank you!

Our website: <u>deepsystems.io</u>

Our team is looking for business partners to make exciting deep learning solutions.

