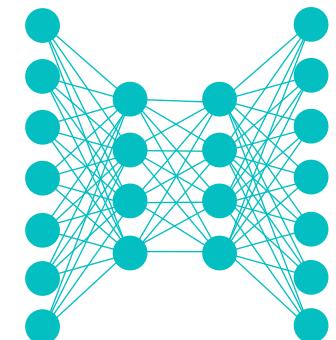


# Lecture Notes for **Neural Networks** **and Machine Learning**



Fully Convolutional Learning II:  
Object Detection

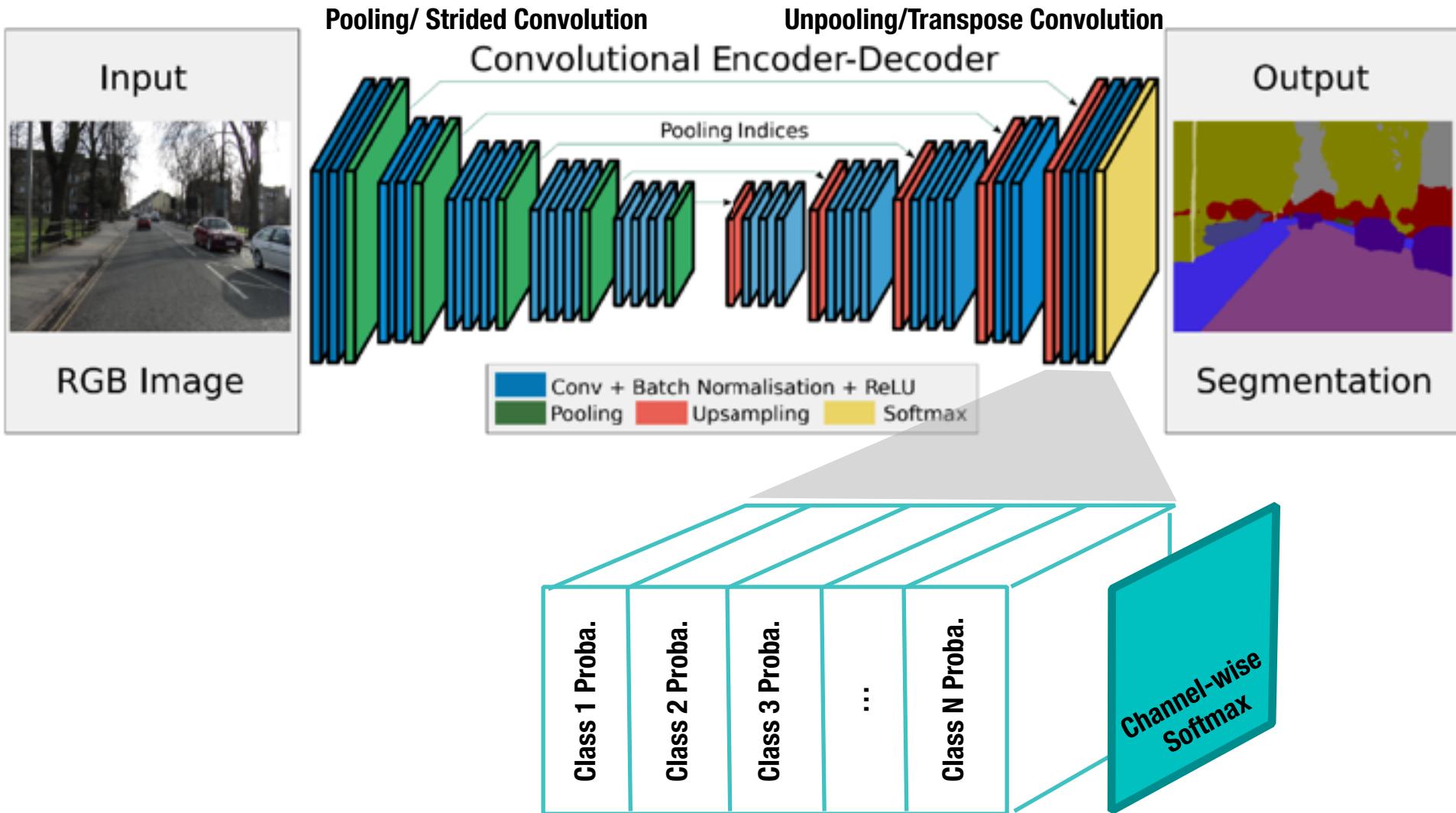


# Logistics and Agenda

- Logistics
  - Next lab due soon
  - Lab grading updates
- Agenda
  - Full Convolutional Architectures
    - ◆ Semantic Segmentation Basics (last time)
    - ◆ Object Detection (this time):
      - RCNN, YOLO
    - ◆ Instance Segmentation (next time, probably):
      - Mask-RCNN, YOLACT

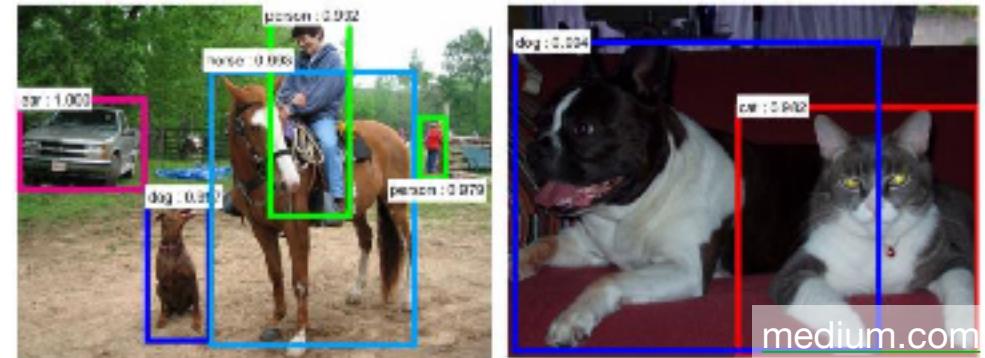


# Last Time



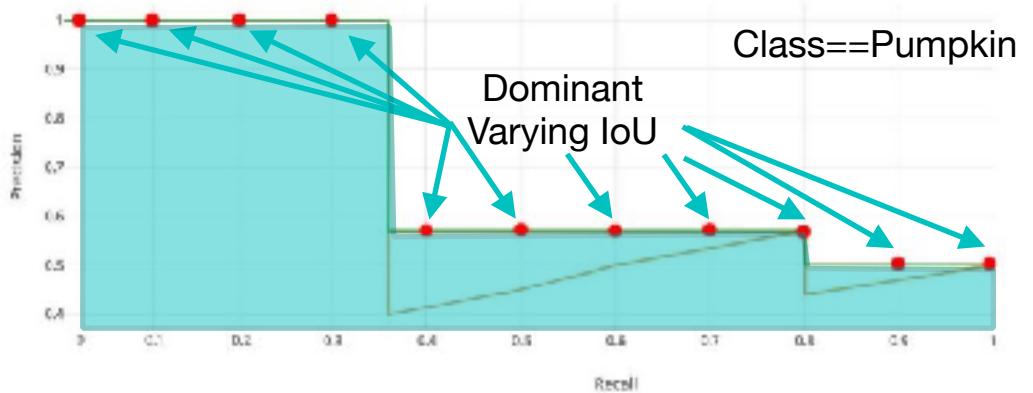
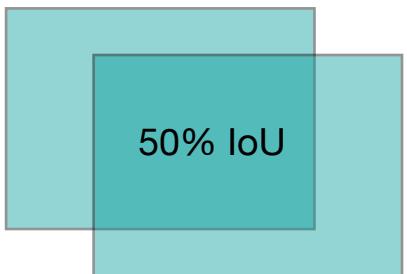
# This time... Object Detection Methods

- Semantic segmentation is not great in terms of mIoU (some classes are at 50%)
- How to adapt these techniques to get bounding boxes, not semantic segmentations?
  - Could this be easier? More stable?
  - More consistent labeling?
  - Suitable for “higher risk” tracking applications?



4

# First: Measuring Performance



$$\text{Recall} = \frac{\text{TP}}{\text{Union}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- mAP( $\text{IoU}=x\%$ )
  - if  $\text{IoU} > X\%$ , check if correct
    - ◆ else not correct
  - Usually ~50%, 75%, 90%
  - Define precision for each class, take average
- mAP(%), sometimes just AP
  - Formulate precision/recall curve for a class at varying levels of IoU (50%-95%)
  - Calculate dominating points
  - Take area under curve
  - Take average area over all classes (macro or micro averaging can be done, usually macro)



# Object Detection with RCNN

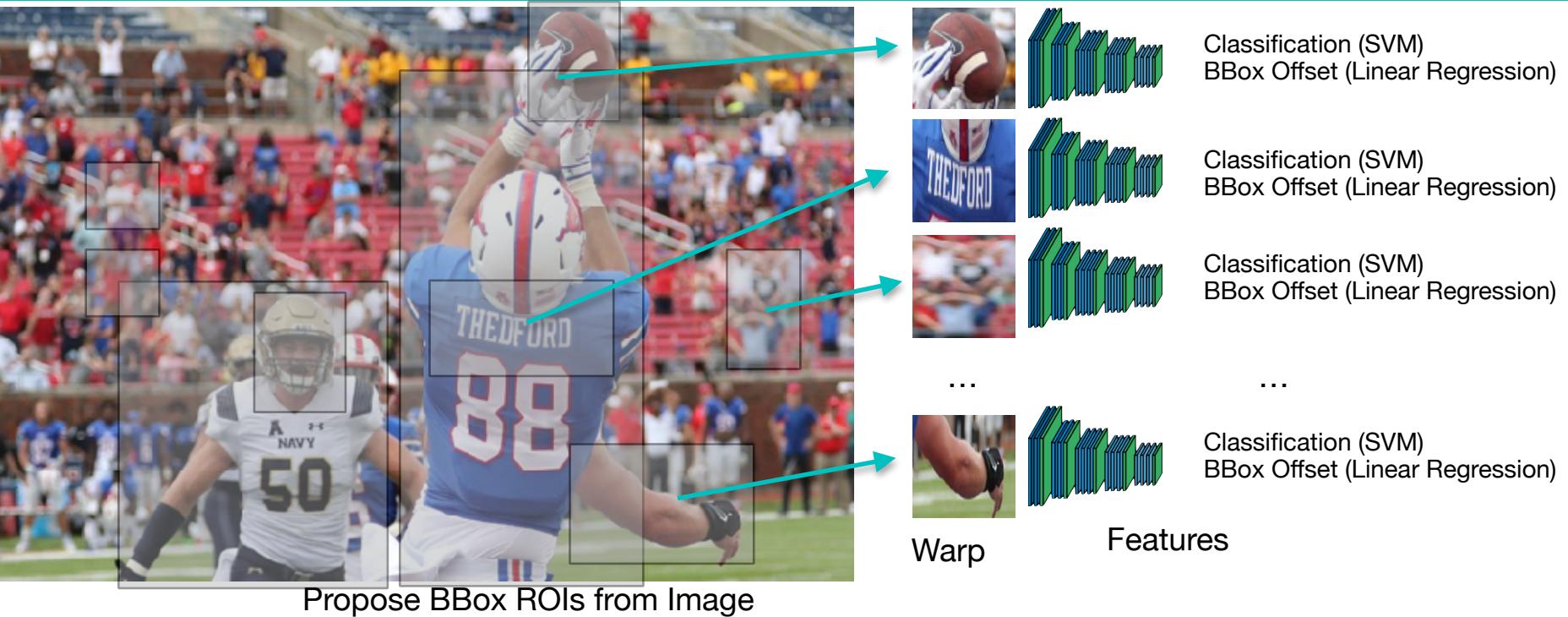


**A history in naming one network five different times  
with five different papers  
each time changing one thing about the architecture**

**Research!**



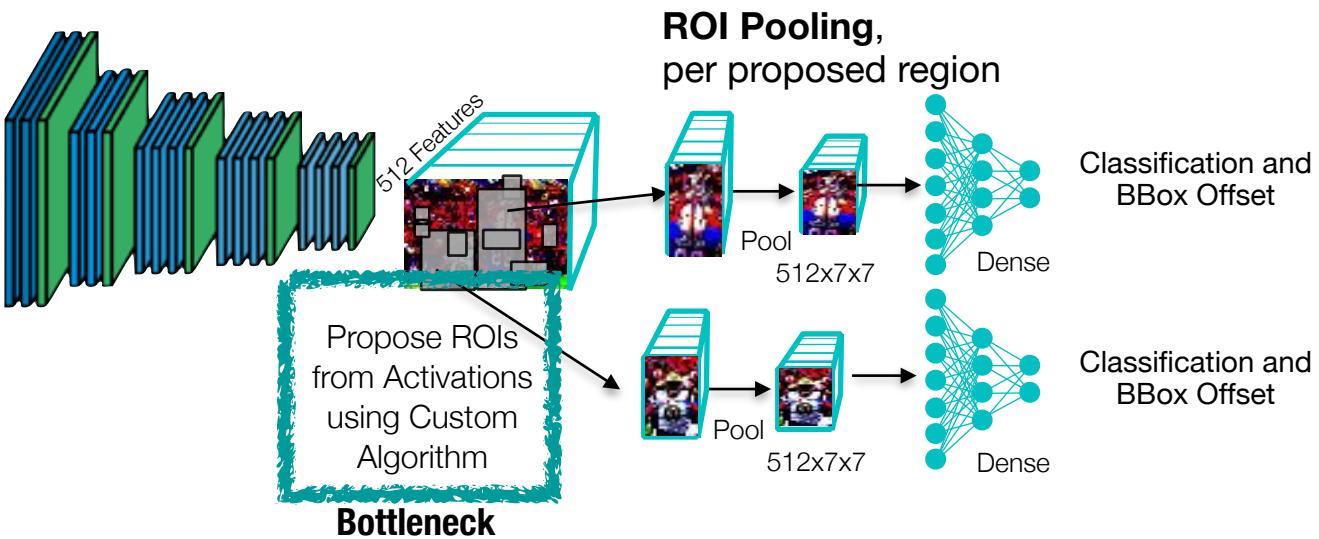
# 2014: R-CNN



- Too Slow to Be Useful
- SVM and BBox Regression Trained Separately
- Fine Tuned Existing ConvNet (for Warped Images)
- ~50 Seconds per Image when Deployed



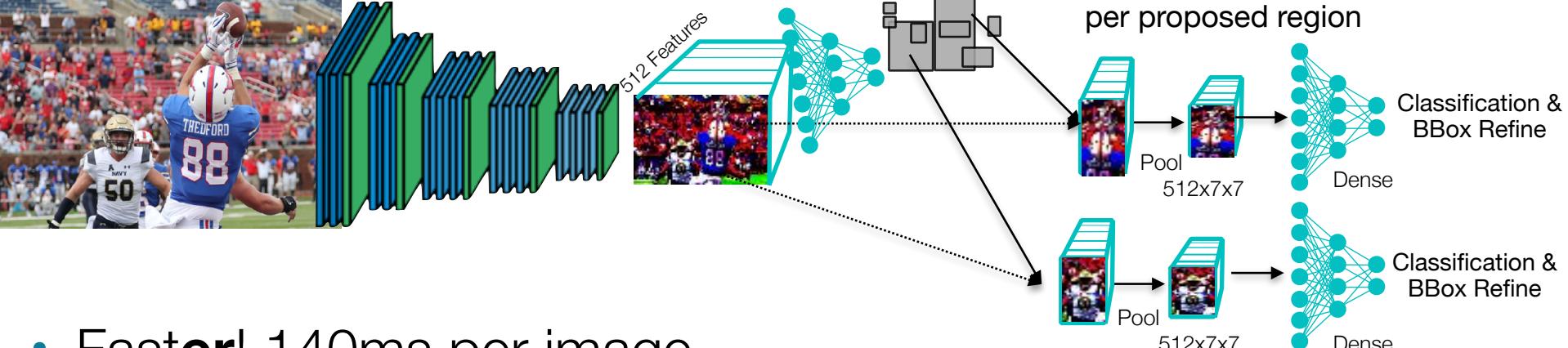
# 2015: Fast R-CNN



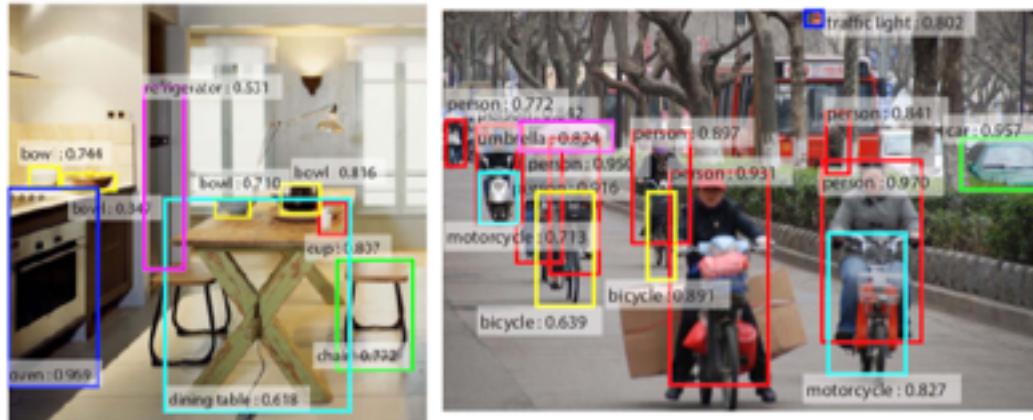
- Fast! 2.3 seconds per image (not ~50)
- But still not real time...



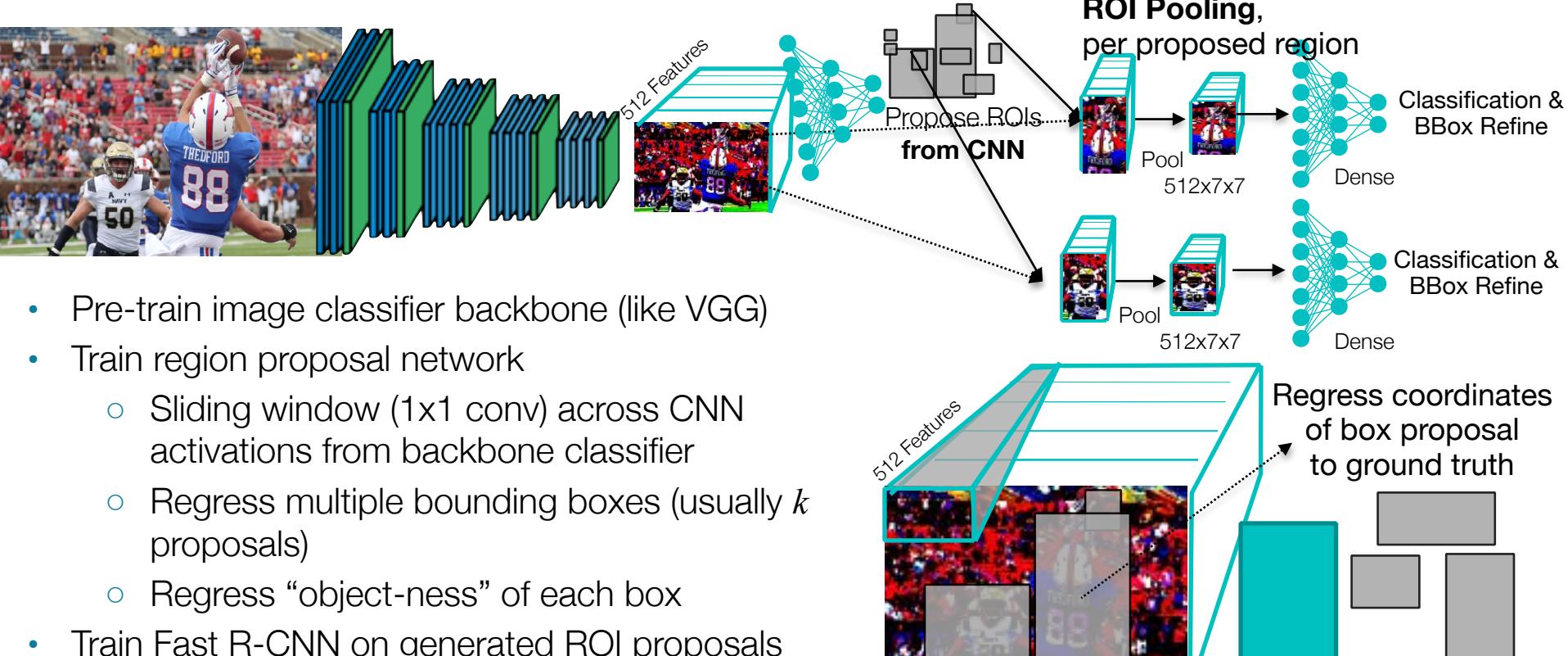
# 2015: Faster R-CNN



- **Faster!** 140ms per image (7 FPS)
- Highly Accurate



# 2015: Faster R-CNN, Training



$$l_{box} = \sum_i \hat{p}_i \left[ (x - \hat{x}_i)^2 + (y - \hat{y}_i)^2 + (\log w - \log \hat{w}_i)^2 + (\log h - \log \hat{h}_i)^2 \right]$$

$$l_{class} = \sum_c CE(c, \hat{c})$$

$$l_{obj} = \sum_i CE(p_i, \hat{p}_i)$$

Ren et al. Faster R-CNN, 2015, November 10



# Object Detection with YOLO



**Joseph Redmon** @pjreddie • Feb 20, 2020



"We shouldn't have to think about the societal impact of our work because it's hard and other people can do it for us" is a really bad argument.



**Roger Grosse** @RogerGrosse

Replying to @kevin\_zakkia and @hardmaru

To be clear, I don't think this is a positive step. Societal impacts of AI is a tough field, and there are researchers and organizations that study it professionally. Most authors do not have expertise in the area and won't do good enough scholarship to say something meaningful.



**Joseph Redmon**  
@pjreddie

I stopped doing CV research because I saw the impact my work was having. I loved the work but the military applications and privacy concerns eventually became impossible to ignore.



**Roger Grosse** @RogerGrosse

Replying to @skoularidou

What's an example of a situation where you think someone should decide not to submit their paper due to Broader Impacts reasons?

10:09 AM · Feb 20, 2020



• RELATED, BUT DIFFERENT APPROACH

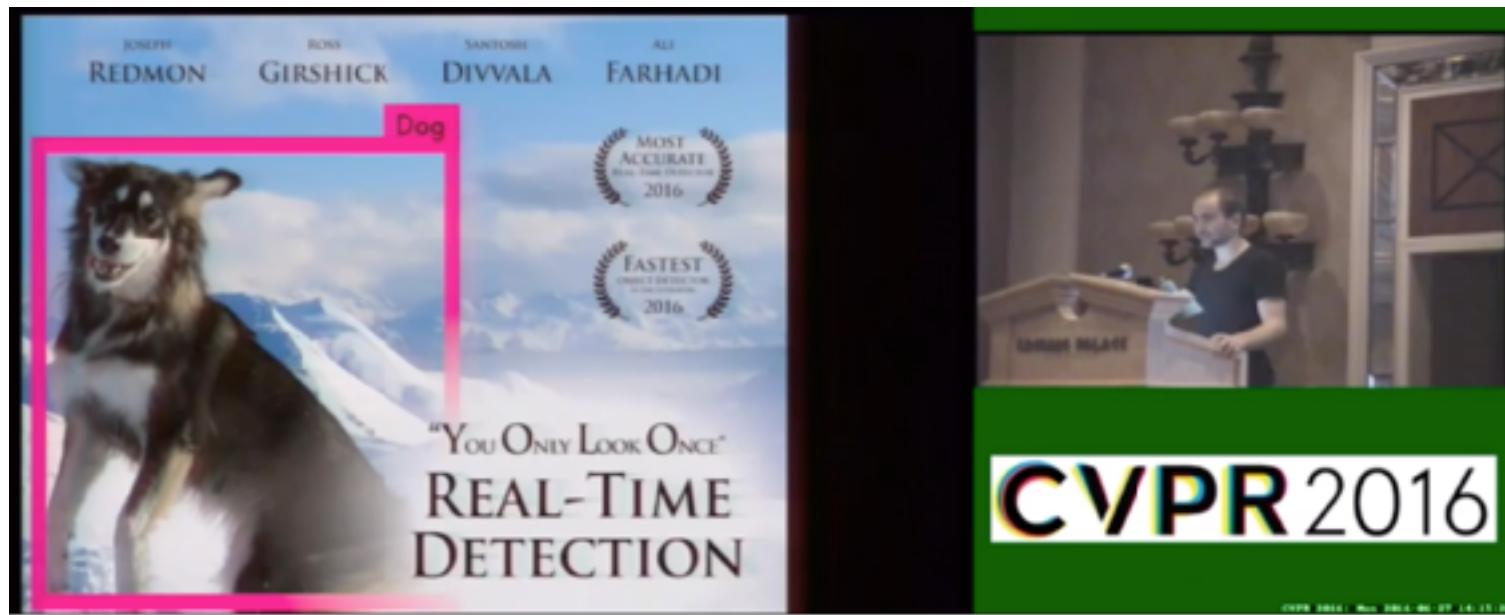


• Researchers started working on a



# Don't want to listen to me explain it?

- Check out Joseph Redmon's Talk at CVPR 2016:
  - <https://www.youtube.com/watch?v=NM6lrxy0bxS>
  - This is how you give a Technical presentation, plus he is from the school where I did graduate school... so he is clearly superior by grace of proximity...



# 2018: Everybody has a Gimmick



## YOLO9000: Better, Faster, Stronger

Joseph Redmon<sup>\*†</sup>, Ali Farhadi<sup>\*†</sup>  
University of Washington<sup>\*</sup>, Allen Institute for AI<sup>†</sup>  
<http://pjreddie.com/yolo9000/>



- YOLO ~40-60 FPS
- Slightly more Accurate than **Faster R-CNN**

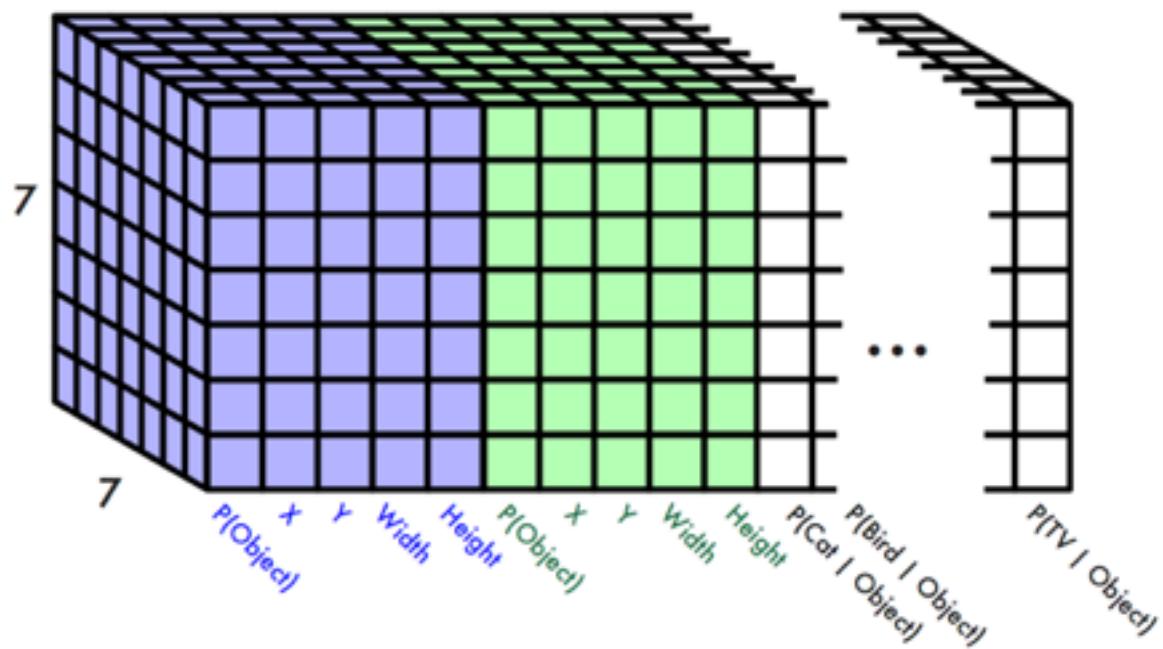


# The YOLO Output Tensor



The Output Tensor

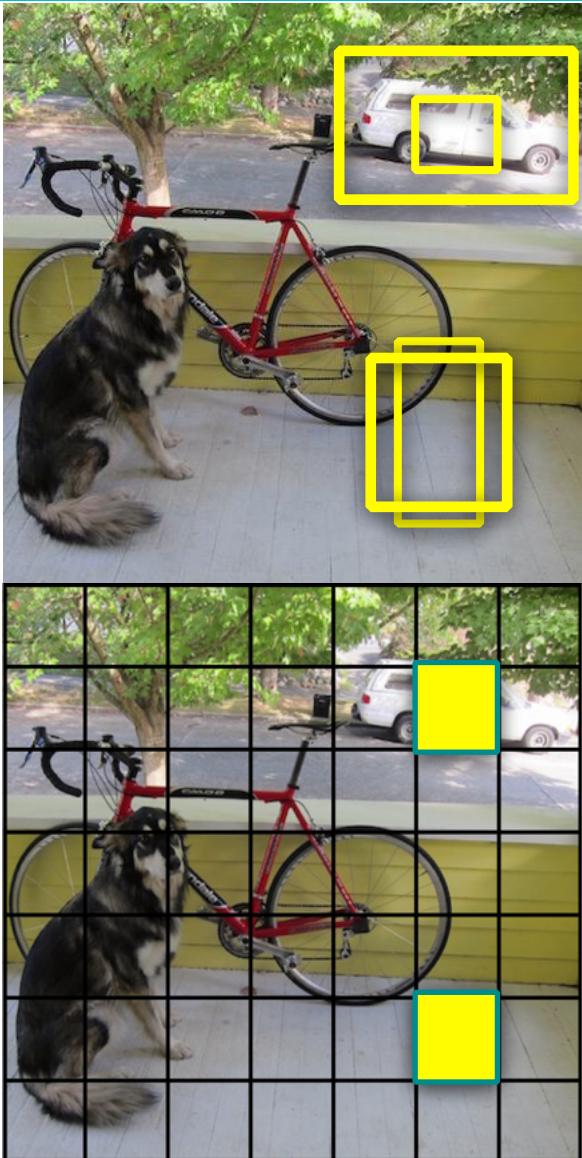
First Bounding Box      Second Bounding Box      Class Probabilities



Redmon and Farhadi, YOLO9000: Better, Faster, Stronger, 2016,  
December 25 — Merry Christmas?

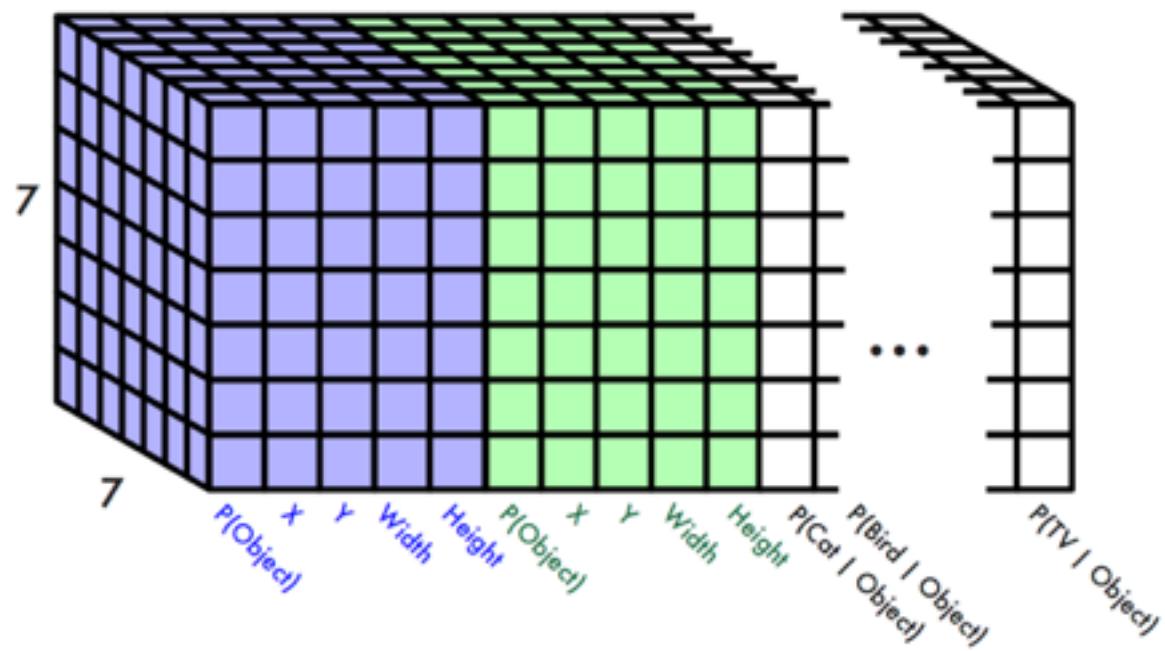


# The YOLO Output Tensor



**The Output Tensor**

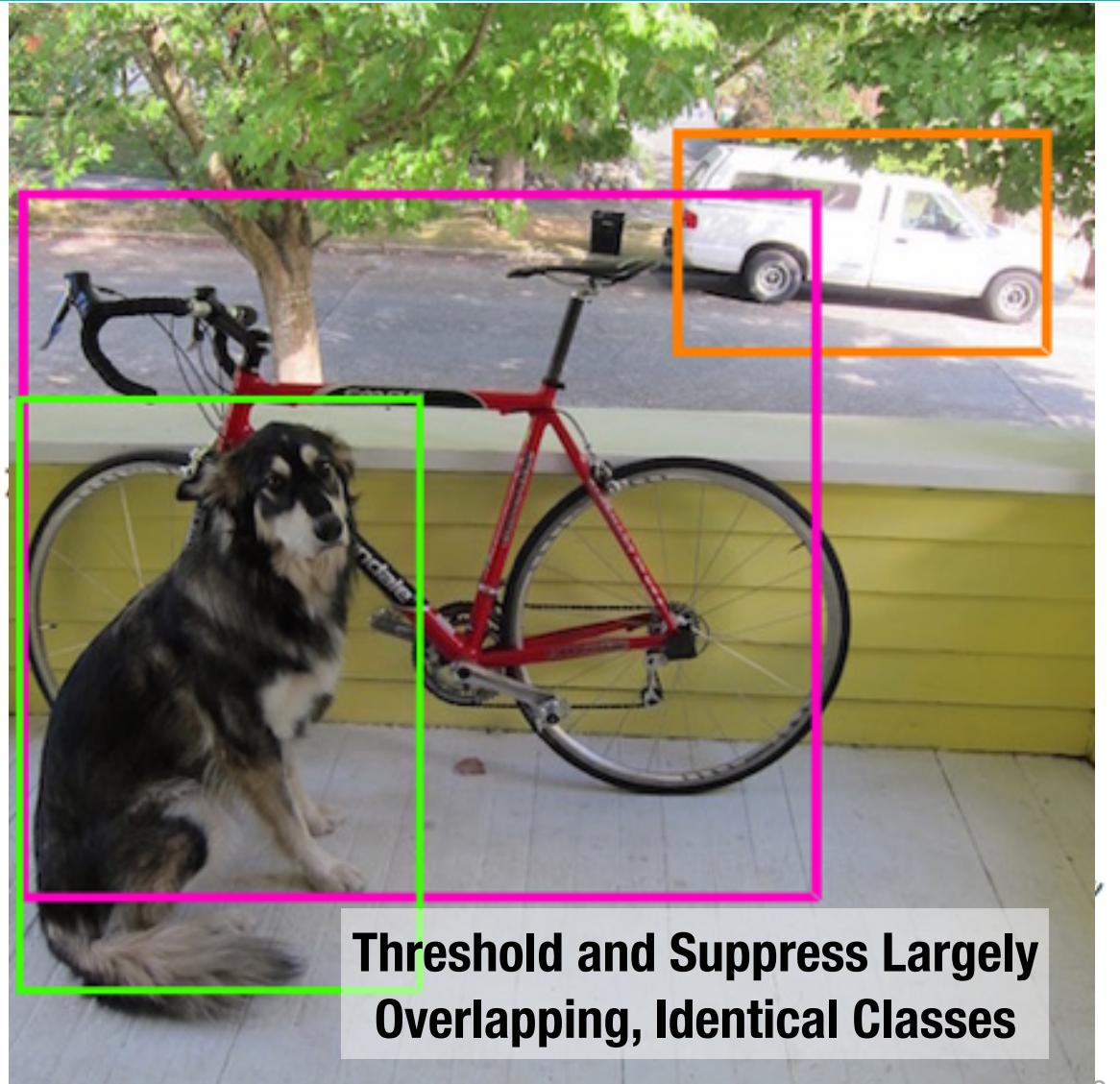
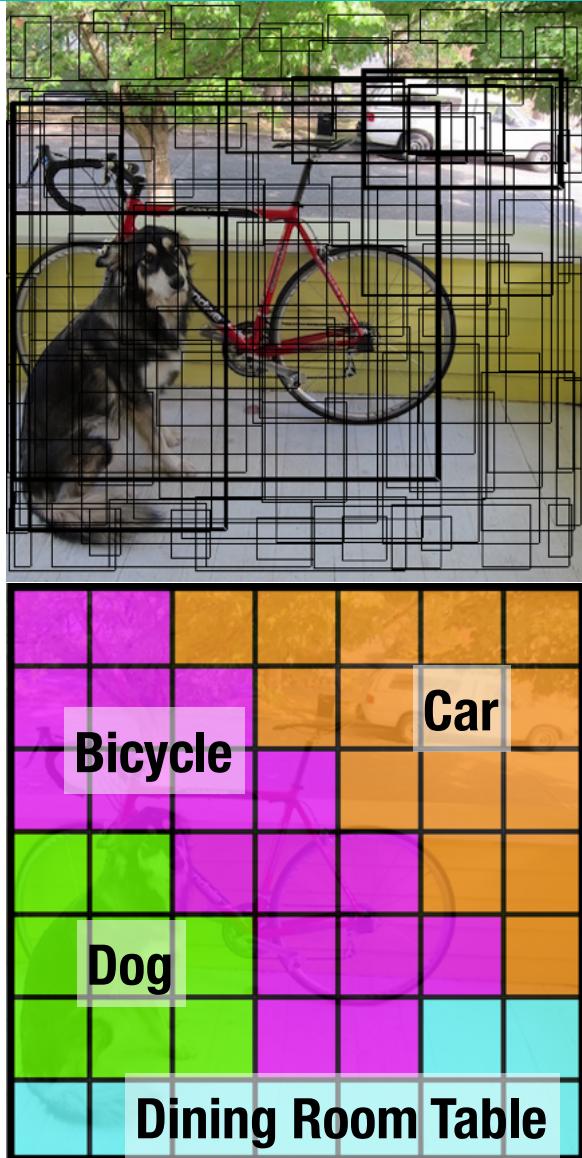
First Bounding Box      Second Bounding Box      Class Probabilities



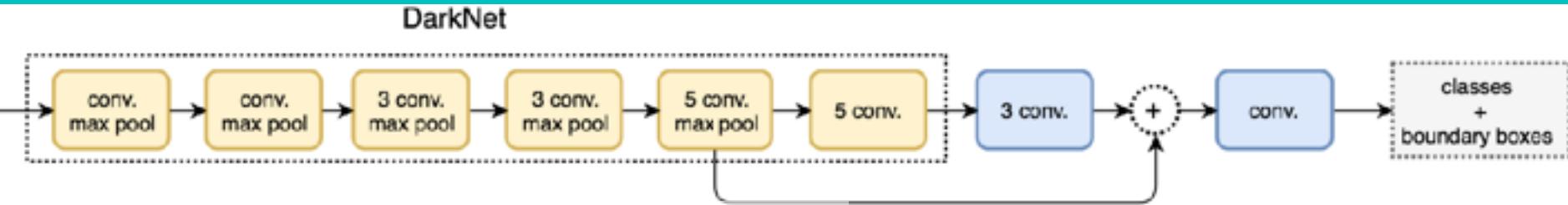
Redmon and Farhadi, YOLO9000: Better, Faster, Stronger, 2016,  
December 25 — Merry Christmas?



# The YOLO Output Tensor



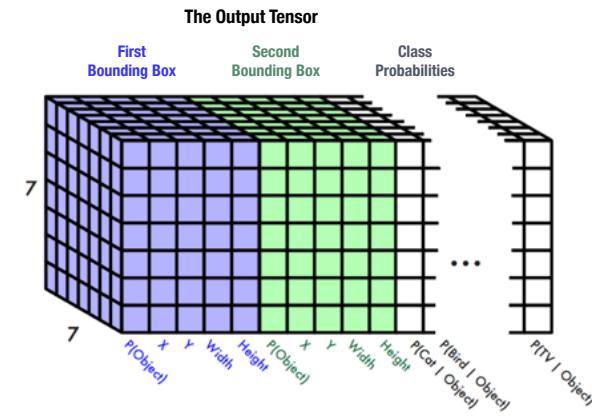
# The YOLO Architecture



**Trained from Traditional Image Dataset.  
Architecture usually: DarkNet on ImageNet**

	pjreddie guys one of my beehives died :-(
	guys one of my beehives died :-(
	SELU activation and yolo openimages
	GUYS I THINK MAYBE IT WAS BROKEN ON OPENCV IDK
	YO DAWG, I HEARD YOU LIKE LICENSES
	generate own license, totally legal :verified:

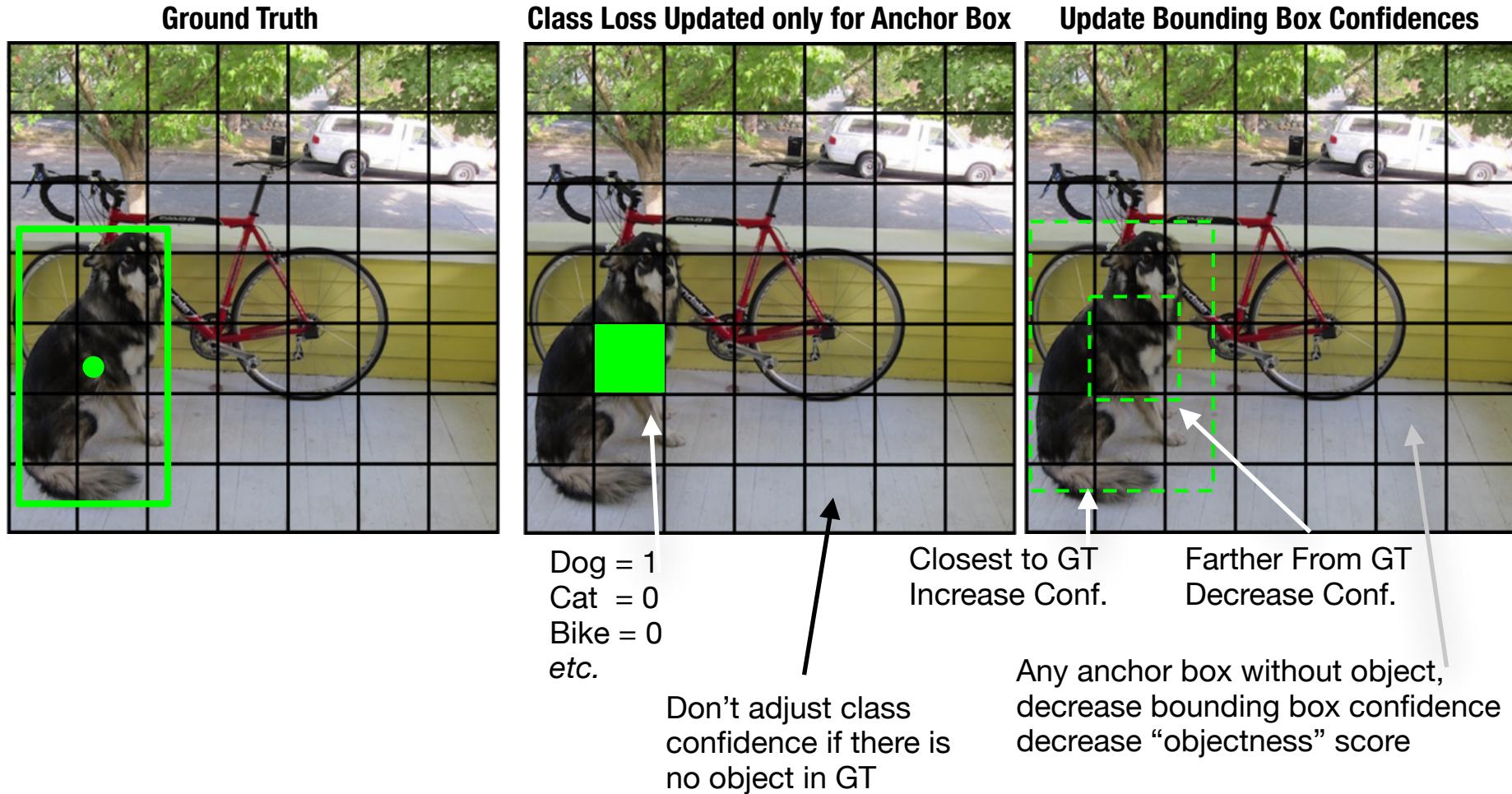
**Last layers:  
Trained on images  
with bounding boxes**



[https://medium.com/@jonathan\\_hui/real-time-object-detection-with-yolo-yolov2-28b1b93e2088](https://medium.com/@jonathan_hui/real-time-object-detection-with-yolo-yolov2-28b1b93e2088)

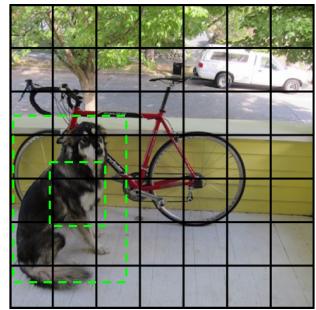


# Training the YOLO Architecture

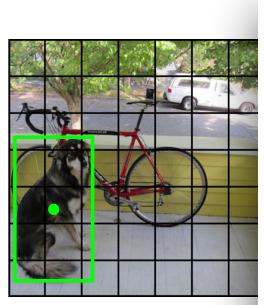


# The YOLO Loss Function

## Update Bounding Box



$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[ (x_i - \hat{x}_{ij})^2 + (y_i - \hat{y}_{ij})^2 \right] \\ + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[ (\sqrt{w_i} - \sqrt{\hat{w}_{ij}})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_{ij}})^2 \right]$$



$S \times S$  cells,  $i^{\text{th}}$  cell

$B$  boxes per cell,  $j^{\text{th}}$  box

$\mathbb{1}^{\text{obj}}$  indicator function, from GT

$\hat{C}$  is confidence per box

$\hat{p}(c)$  softmax output, per class

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (\hat{C}_i - \hat{C}_{ij})^2$$

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} (\hat{C}_i - \hat{C}_{ij})^2$$

## Class Loss



$$+ \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

[https://medium.com/@jonathan\\_hui/real-time-object-detection-with-yolo-yolov2-28b1b93e2088](https://medium.com/@jonathan_hui/real-time-object-detection-with-yolo-yolov2-28b1b93e2088)

Localization Loss

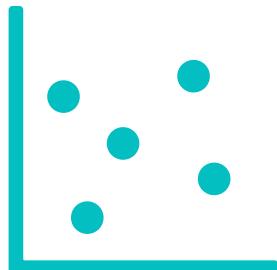
Object Detection Loss

Classification Loss



# Lecture Notes for **Neural Networks** **and Machine Learning**

FCN Learning: Detection



**Next Time:**  
Instance Segmentation  
**Reading:** None

