

JOSEPH

ALI

REDMON

FARHADI

RETURN IN.....

YOLO9000

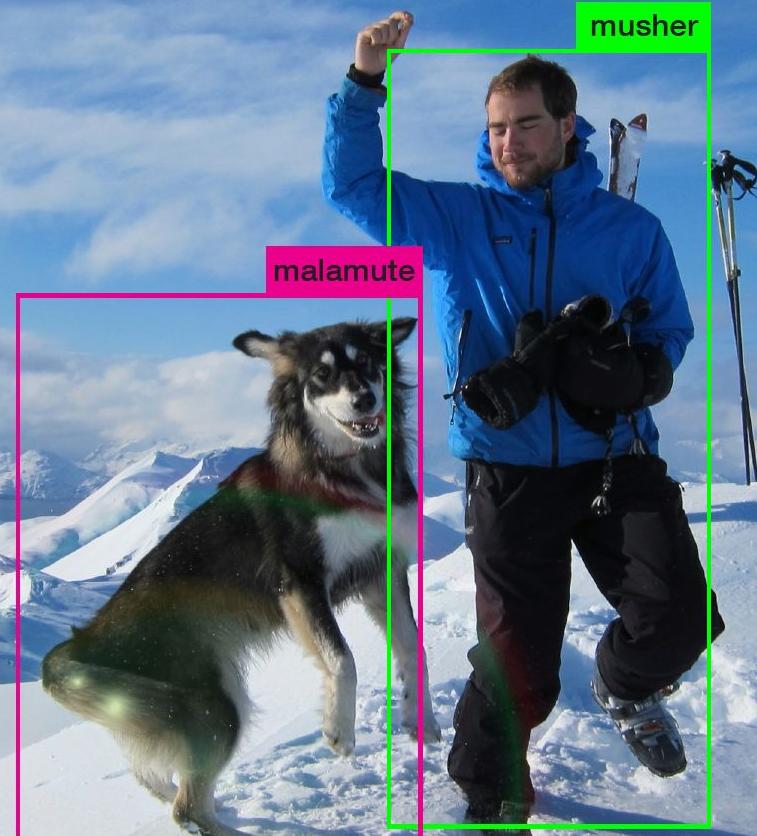
Better, Faster, Stronger

NOW PLAYING IN A DEMO NEAR YOU

A UNIVERSITY OF WASHINGTON PRODUCTION IN ASSOCIATION WITH XNOR.AI AND THE ALLEN INSTITUTE FOR ARTIFICIAL INTELLIGENCE
MODELS BY DARKNET: OPEN SOURCE NEURAL NETWORKS

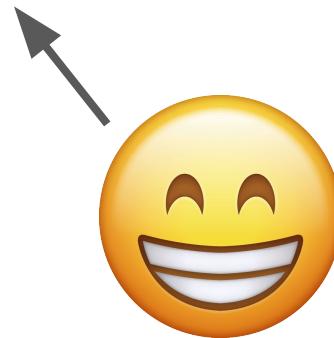
@DARKNETFOREVER #YOLO9000

pjreddie.com/yolo

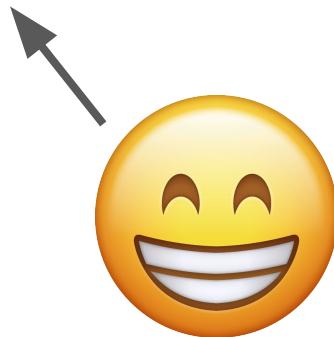
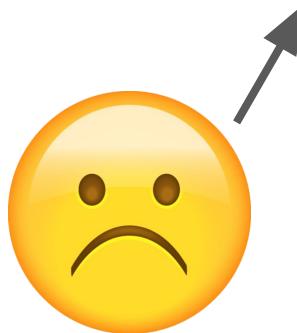


	Pascal 2007 mAP	Speed	
DPM v5	33.7	.07 FPS	14 s/img
R-CNN	66.0	.05 FPS	20 s/img
Fast R-CNN	70.0	.5 FPS	2 s/img
Faster R-CNN	73.2	7 FPS	140 ms/img
YOLO	63.4	45 FPS	22 ms/img

	Pascal 2007 mAP	Speed	
DPM v5	33.7	.07 FPS	14 s/img
R-CNN	66.0	.05 FPS	20 s/img
Fast R-CNN	70.0	.5 FPS	2 s/img
Faster R-CNN	73.2	7 FPS	140 ms/img
YOLO	63.4	45 FPS	22 ms/img



	Pascal 2007 mAP	Speed	
DPM v5	33.7	.07 FPS	14 s/img
R-CNN	66.0	.05 FPS	20 s/img
Fast R-CNN	70.0	.5 FPS	2 s/img
Faster R-CNN	73.2	7 FPS	140 ms/img
YOLO	63.4	45 FPS	22 ms/img



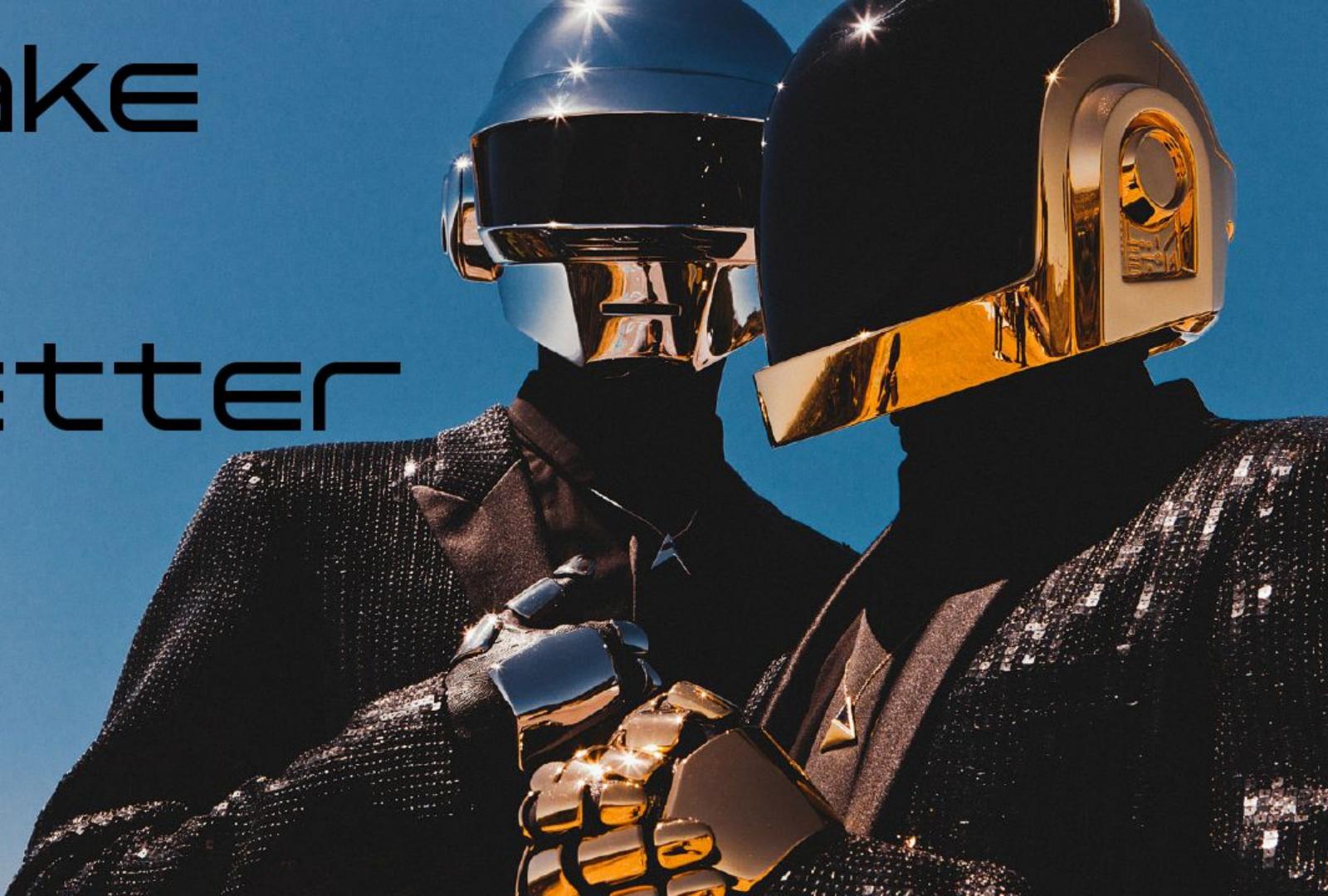


CVPR 2016: Mon 2016-06-27 14:26:43

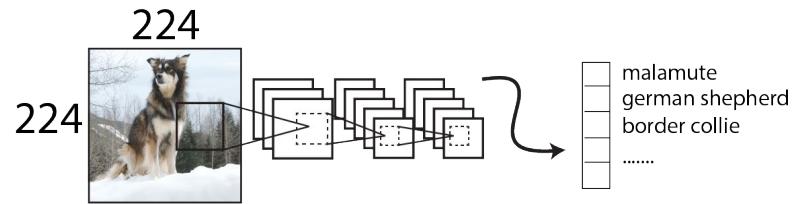
"Work it harder,
Make it better,
Do it faster,
Makes us **stronger** -"



Make
It
Better



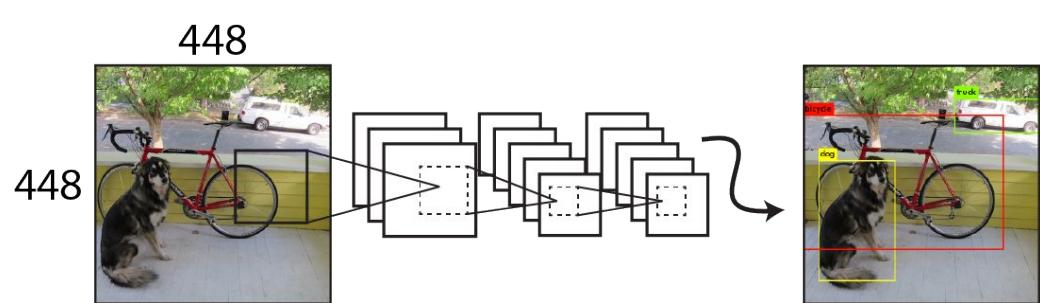
Train on ImageNet



Resize network

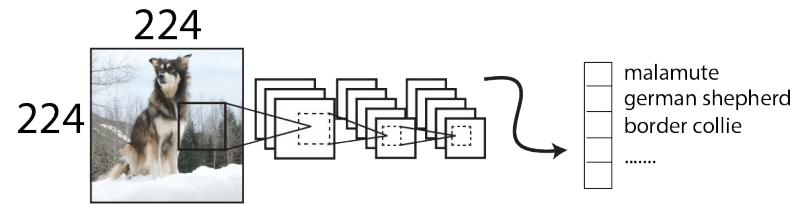


Fine-tune on detection

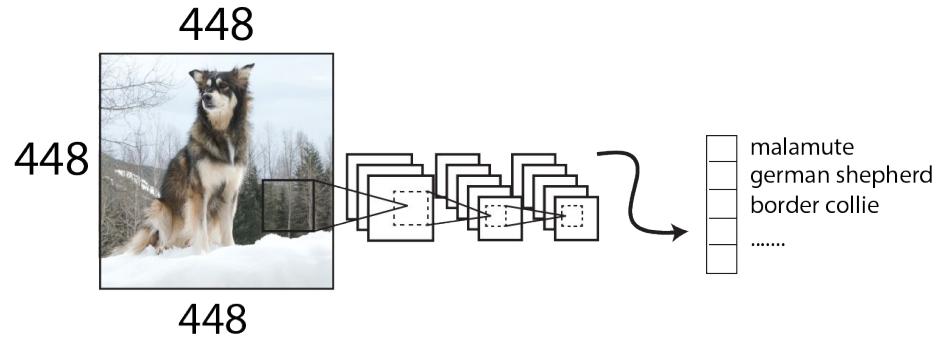


Fine-tune 448x448 Classifier: +3.5% mAP

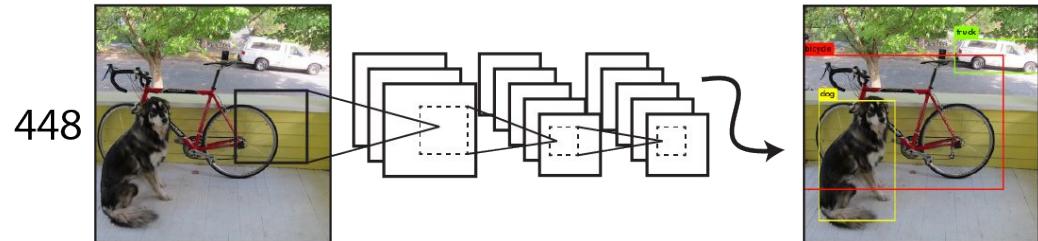
Train on ImageNet



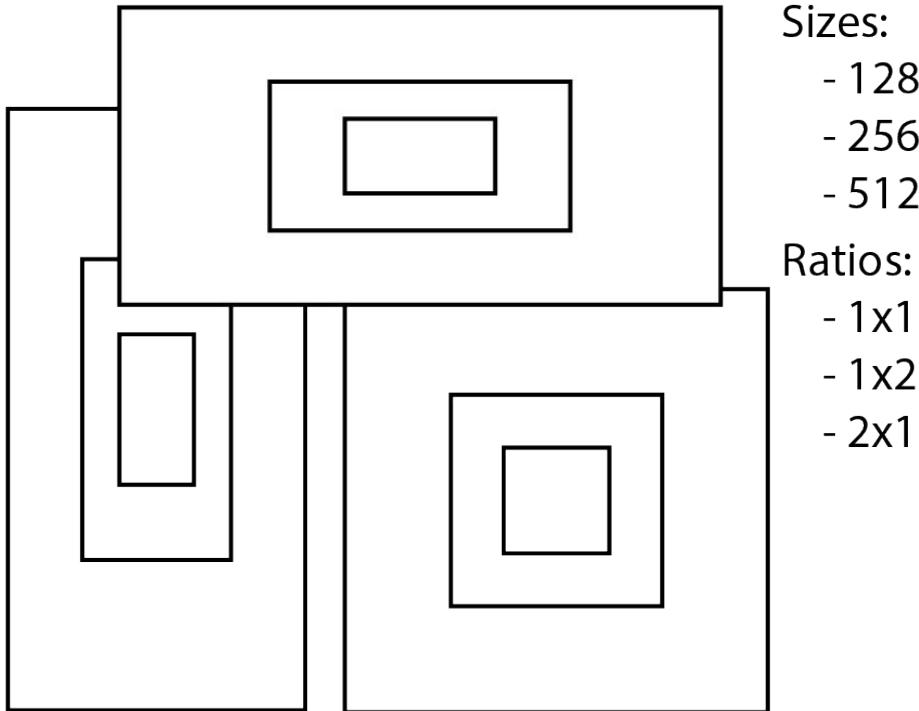
Resize, fine-tune
on ImageNet



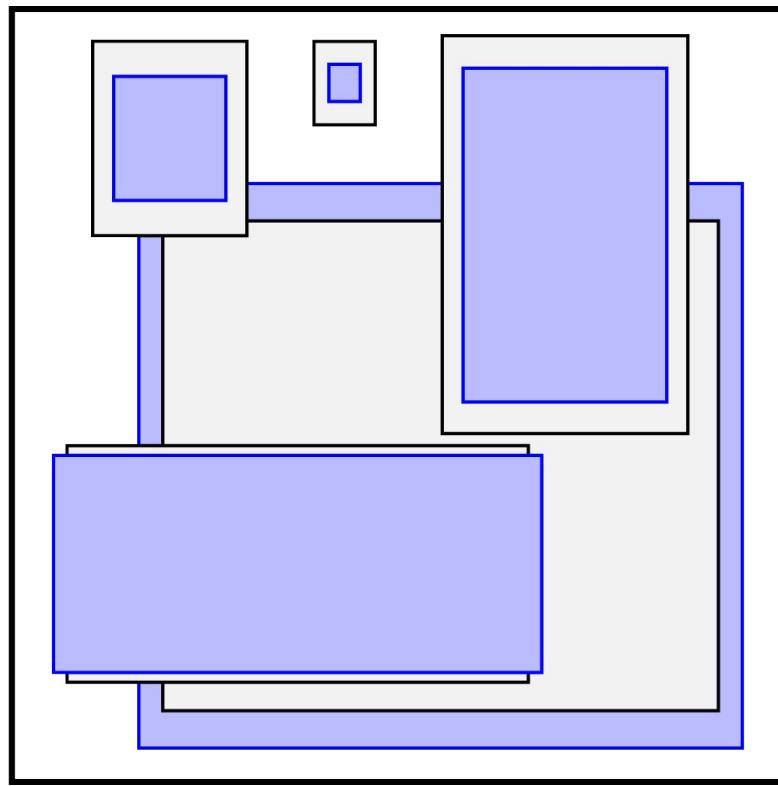
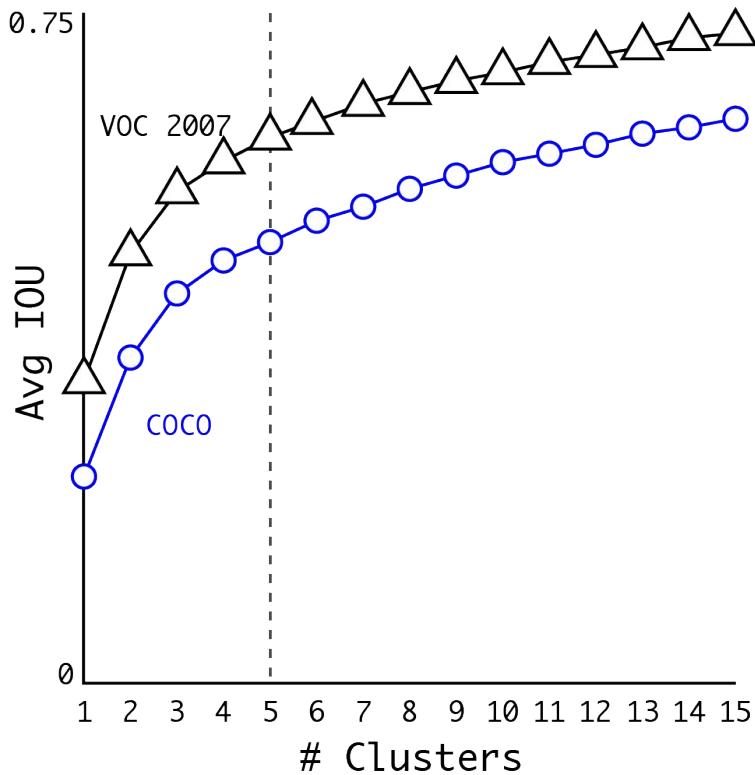
Fine-tune on detection



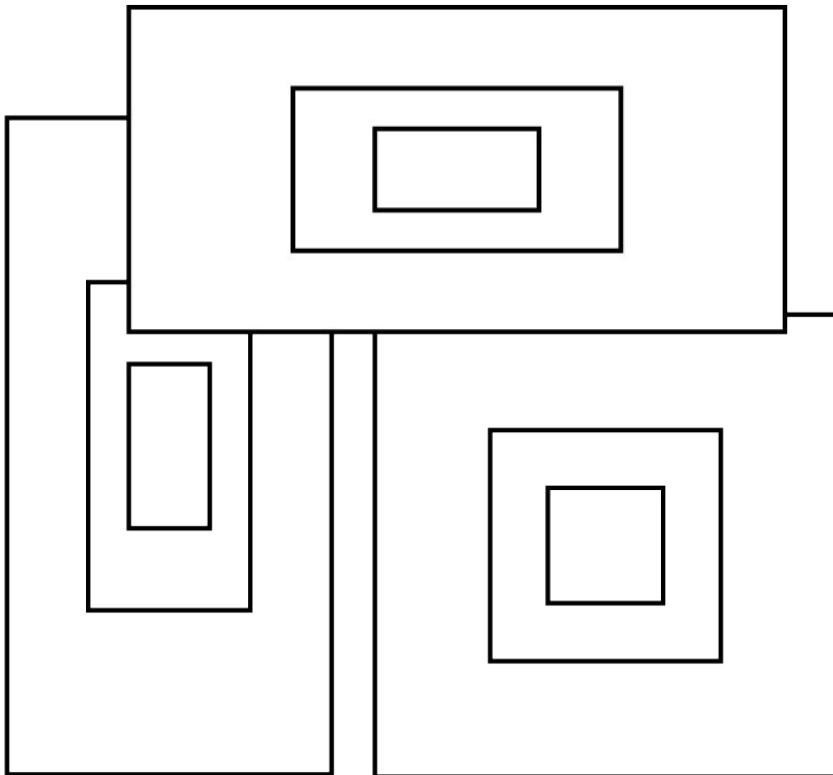
Anchor boxes use static initialization



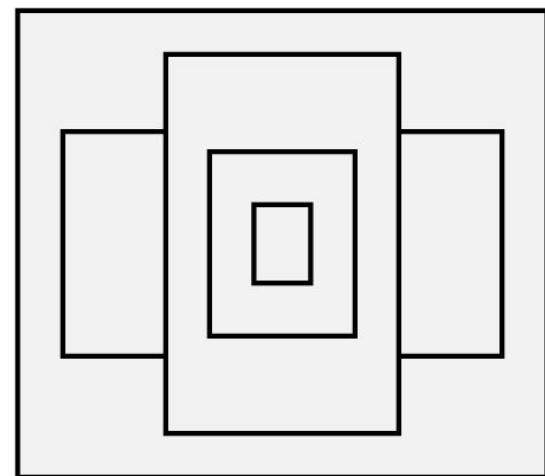
We use k-means to find better initializations



Anchor Boxes



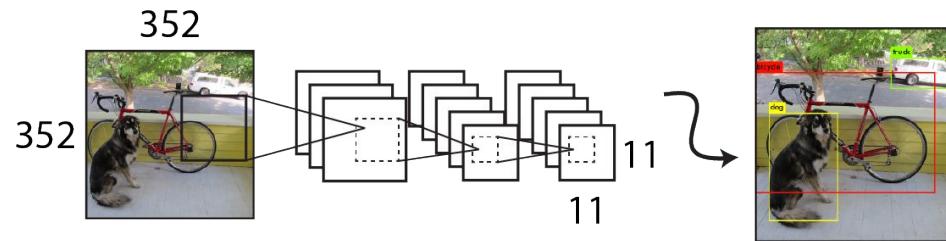
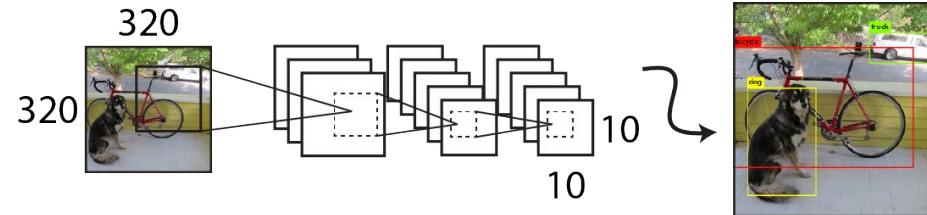
Dimension Clusters



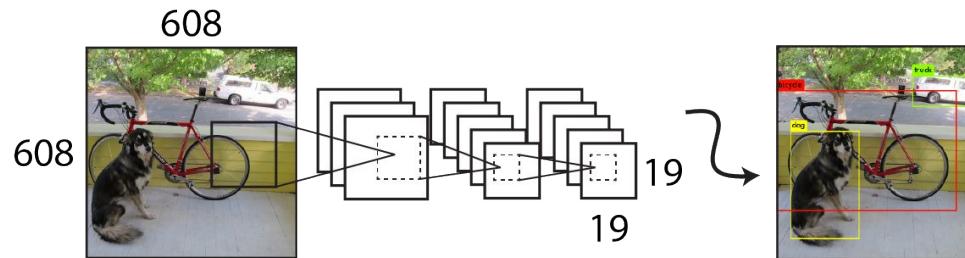
Dimension Clusters: +5% mAP

Box Generation	#	Avg IOU
Cluster SSE	5	58.7
Cluster IOU	5	61.0
Anchor Boxes [15]	9	60.9

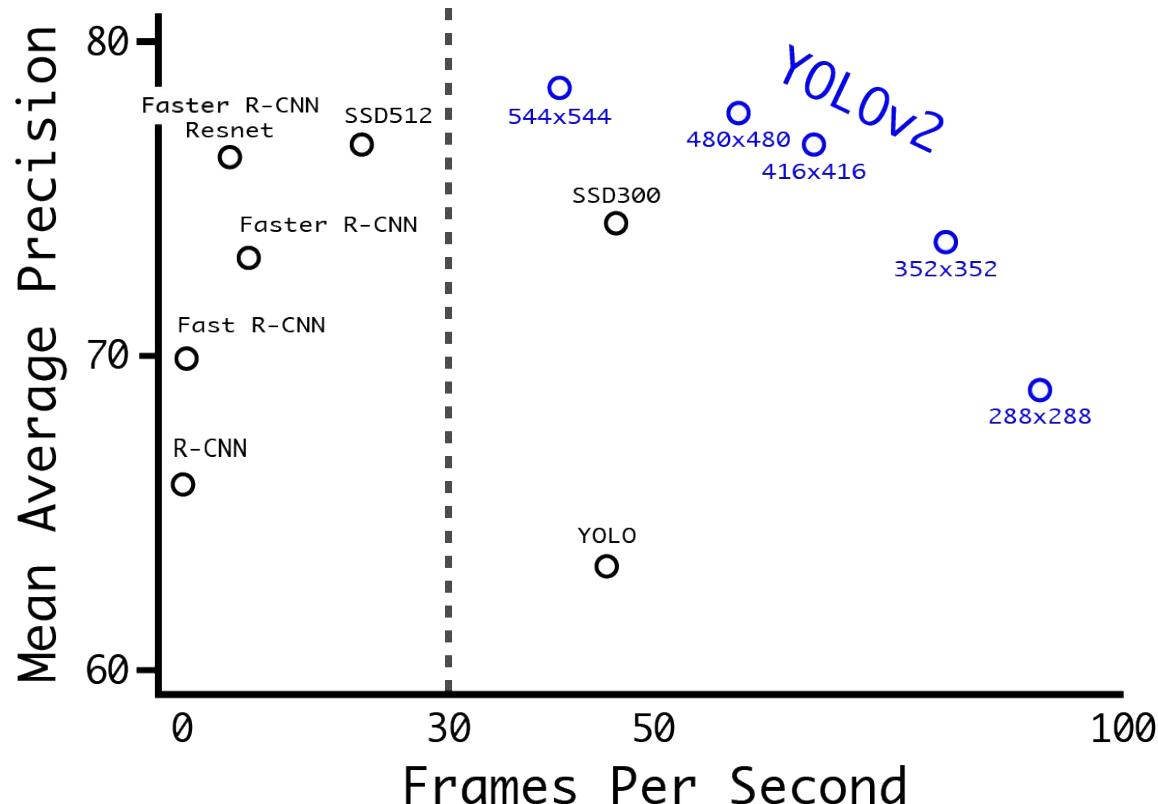
Multi-scale training: +1.5% mAP

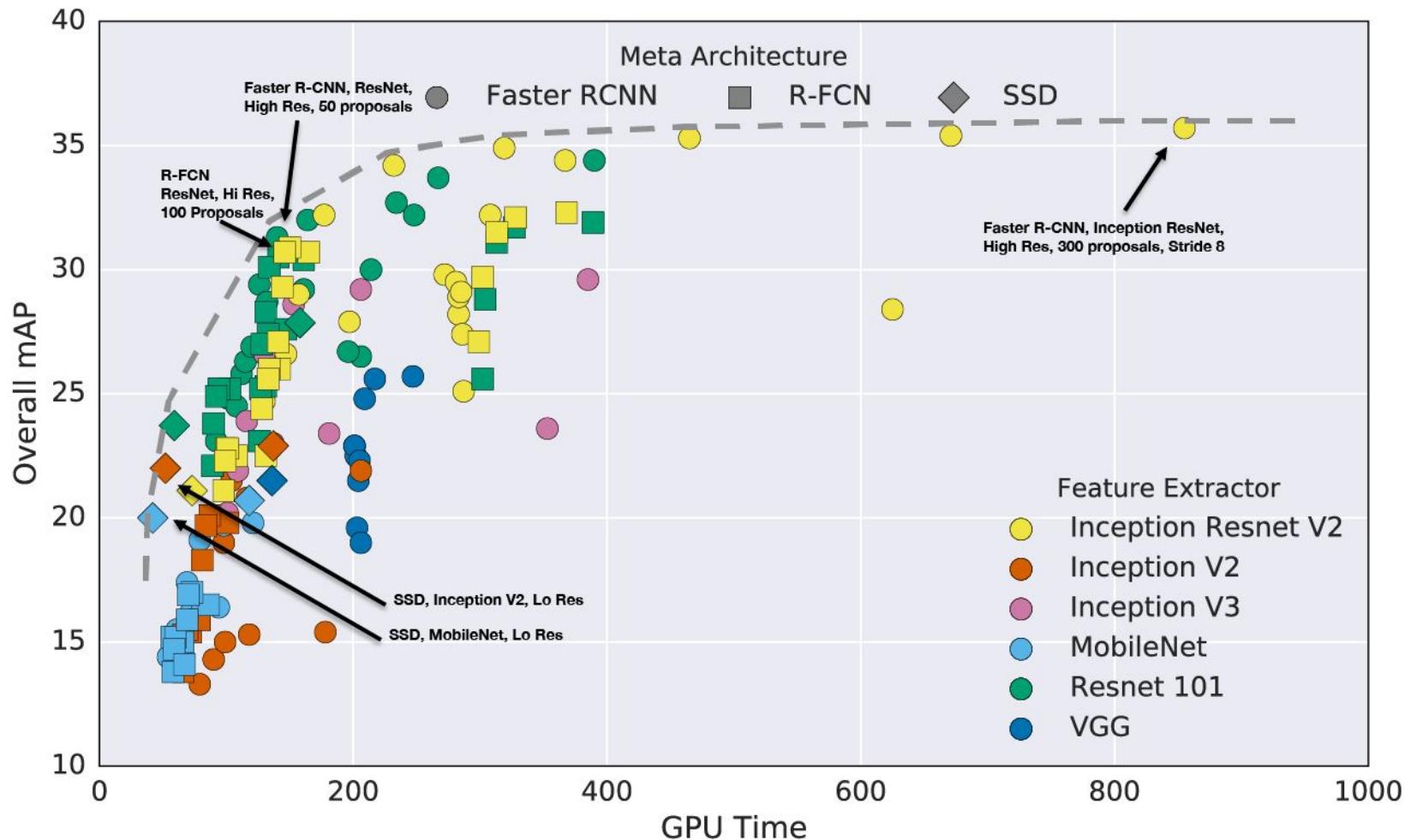


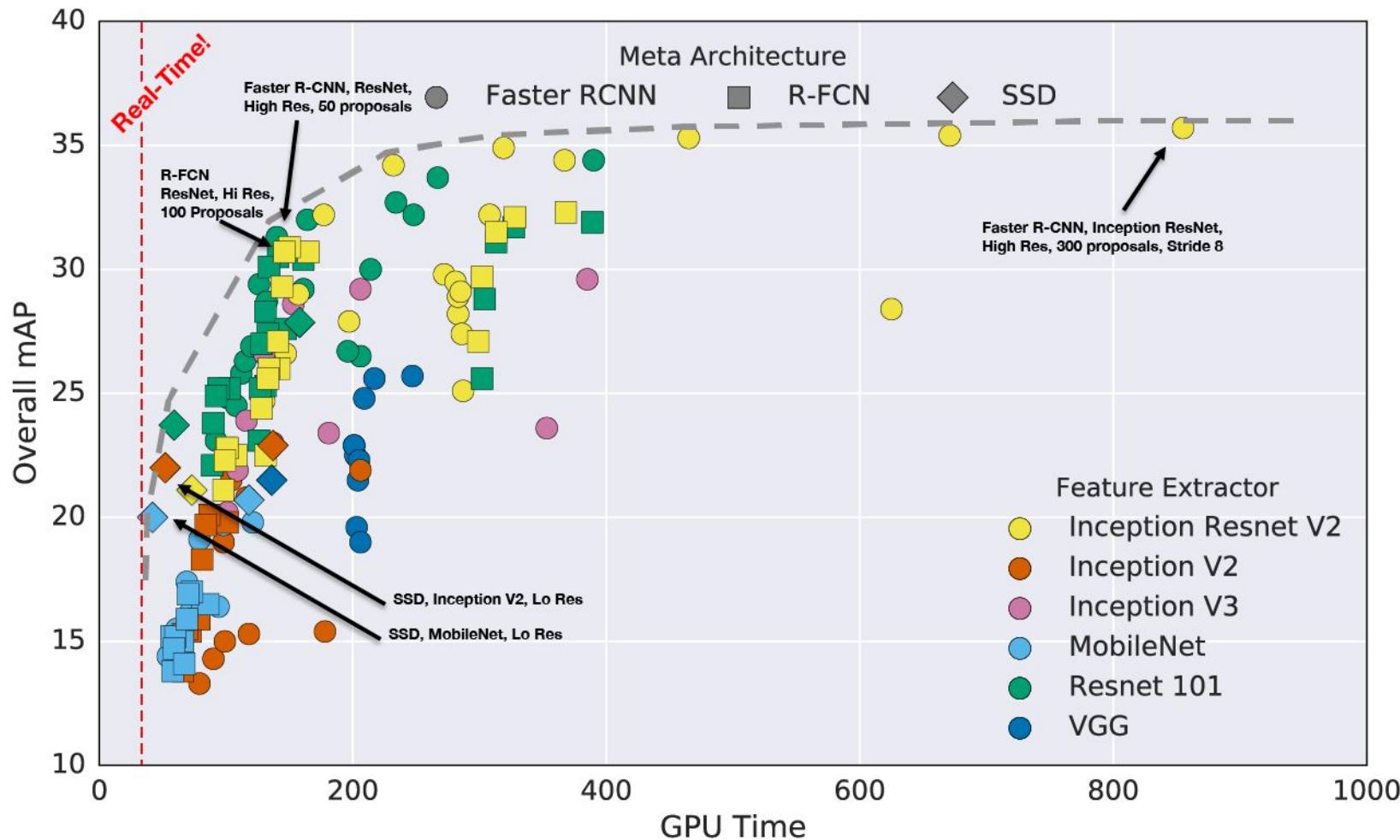
.....

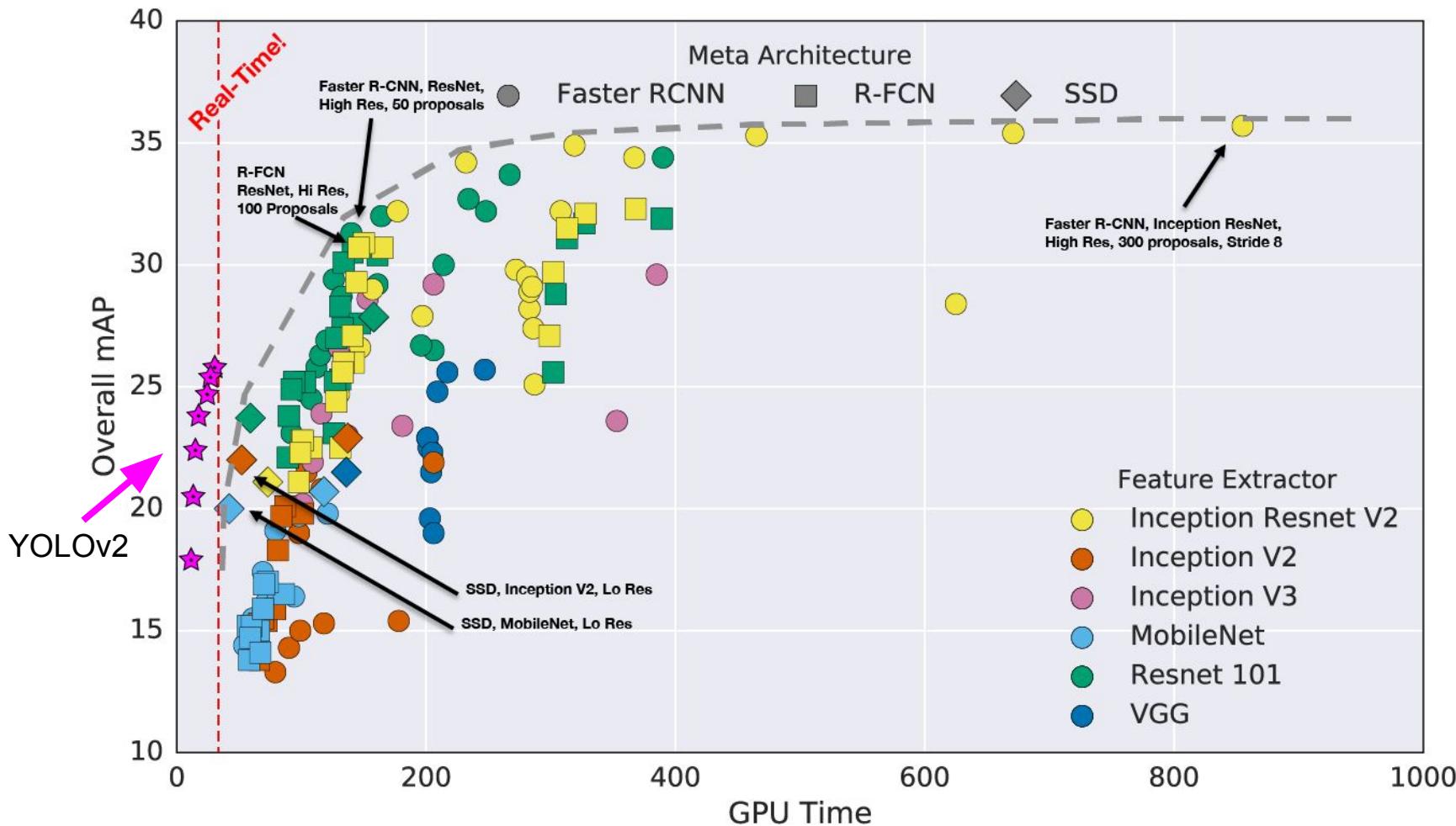


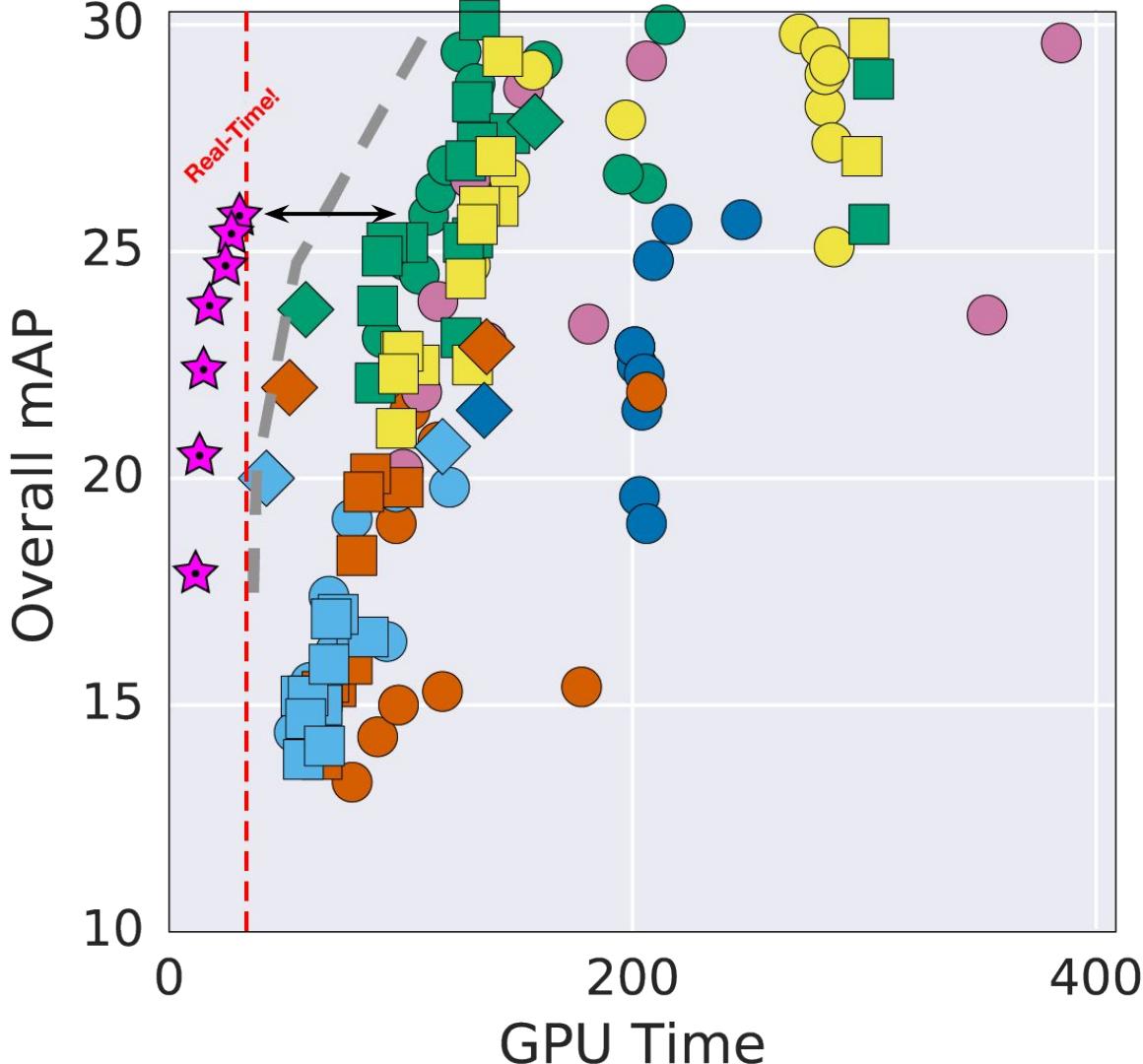
YOLOv2: Fast, Accurate Detection











A promotional image featuring two Daft Punk helmets. The helmet on the left has a rainbow-colored light display on its visor. The helmet on the right has the word "HUMAN" displayed in red LED-style letters on its visor. Both helmets are mounted on black leather jackets. The background is a dark, moody gradient.

Do
It
Faster

Speed is not just parameter counts or FLOPs

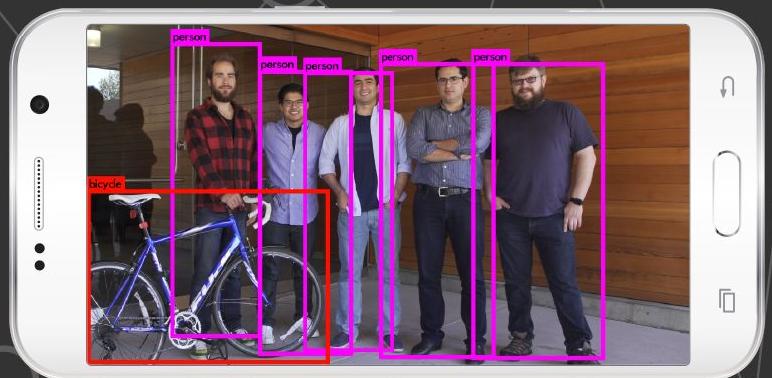
	Top 1	Top 5	FLOPs	GPU Speed
VGG-16	70.5	90.0	30.95 Bn	100 FPS
Extraction (YOLOv1)	72.5	90.8	8.52 Bn	180 FPS
Resnet50	75.3	92.2	7.66 Bn	90 FPS

Darknet19: A good balance of speed and accuracy

	Top 1	Top 5	FLOPs	GPU Speed
VGG-16	70.5	90.0	30.95 Bn	100 FPS
Extraction (YOLOv1)	72.5	90.8	8.52 Bn	180 FPS
Resnet50	75.3	92.2	7.66 Bn	90 FPS
Darknet19	74.0	91.8	5.58 Bn	200 FPS

XNOR.AI

Artificial Intelligence on Any Device



iPhone CPU



\$5 Deep Learning
Machine



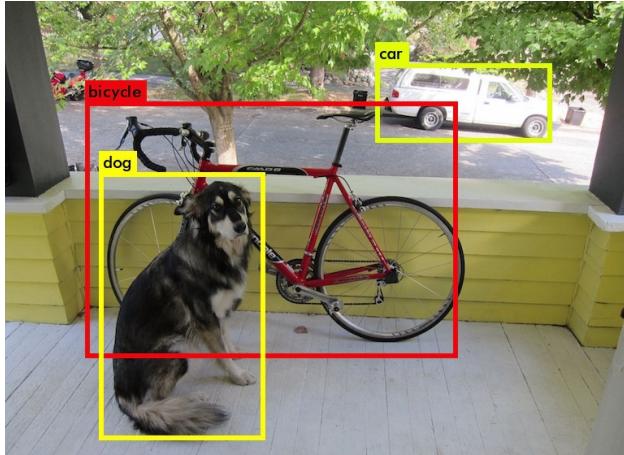
Makes
Us
Stronger







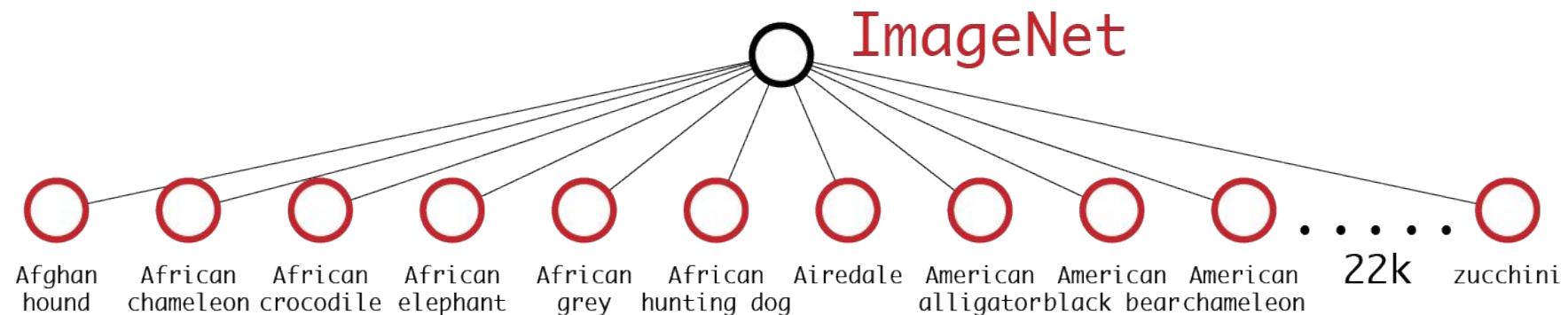
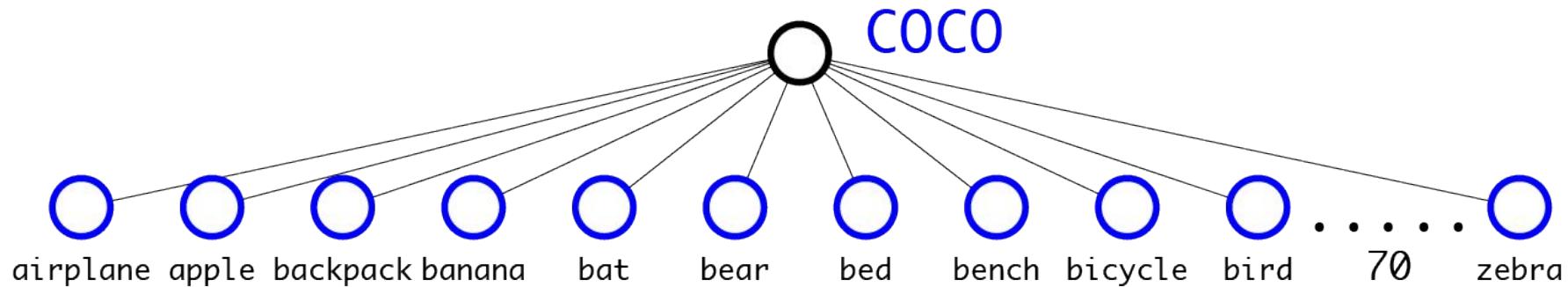
- 100k images
- 80 classes
- Detection labels



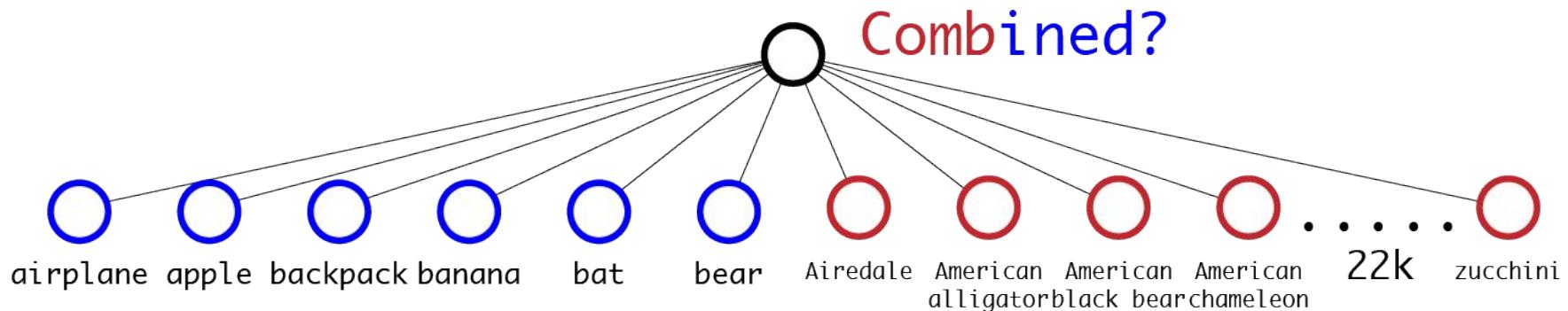
- 14 million images
- 22k classes
- Classification labels



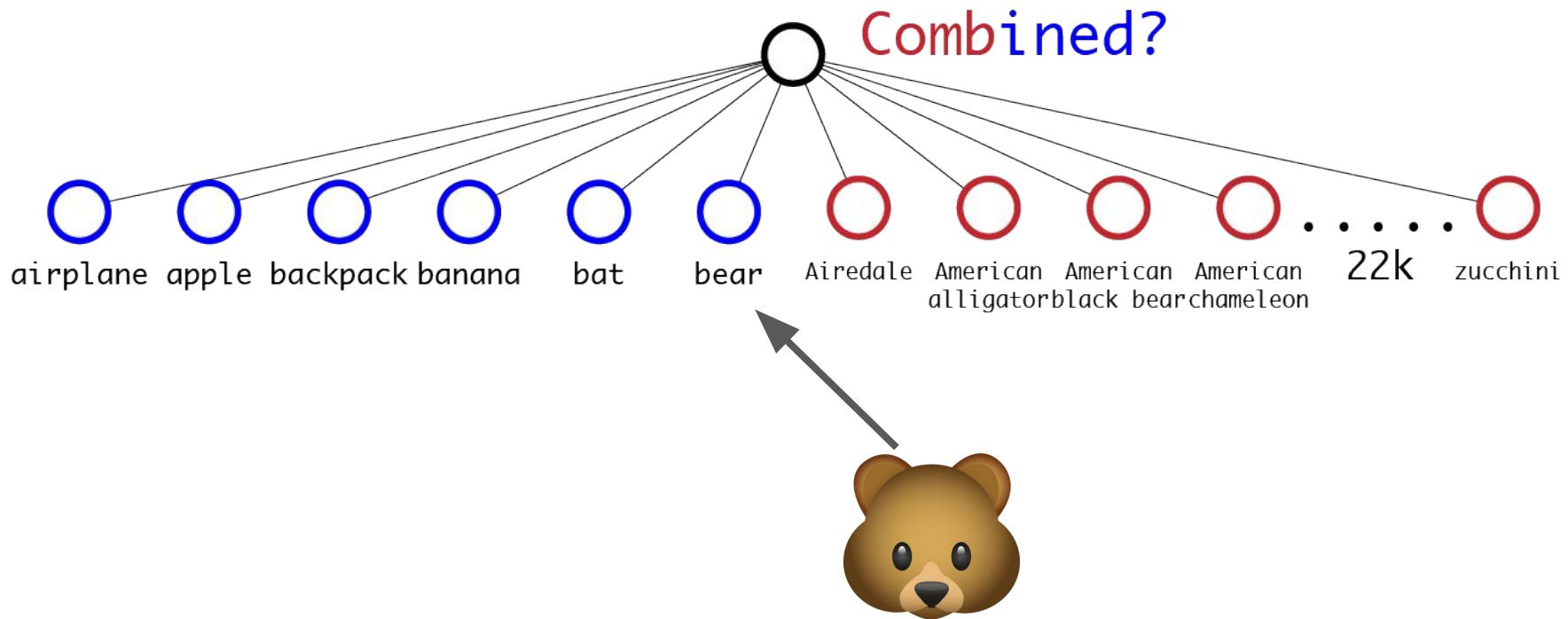
Typically use softmax over all classes



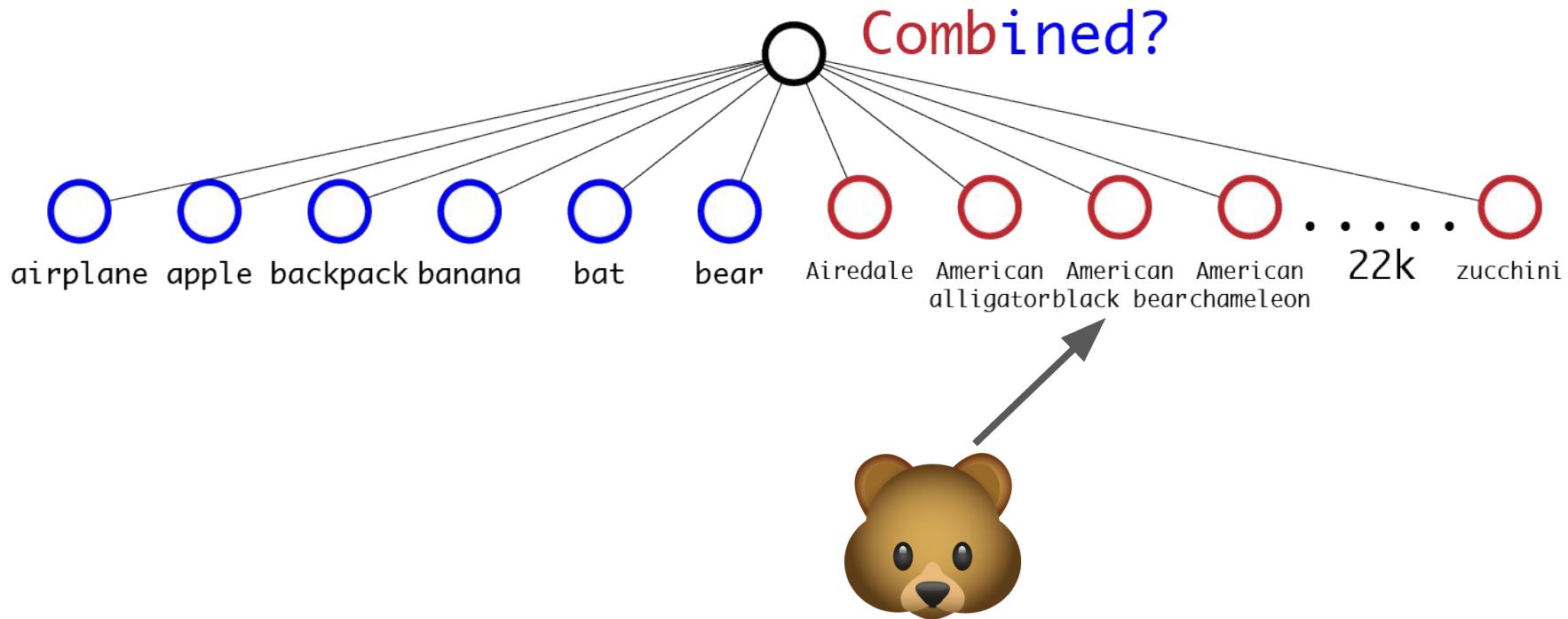
Can't just mash classes together...



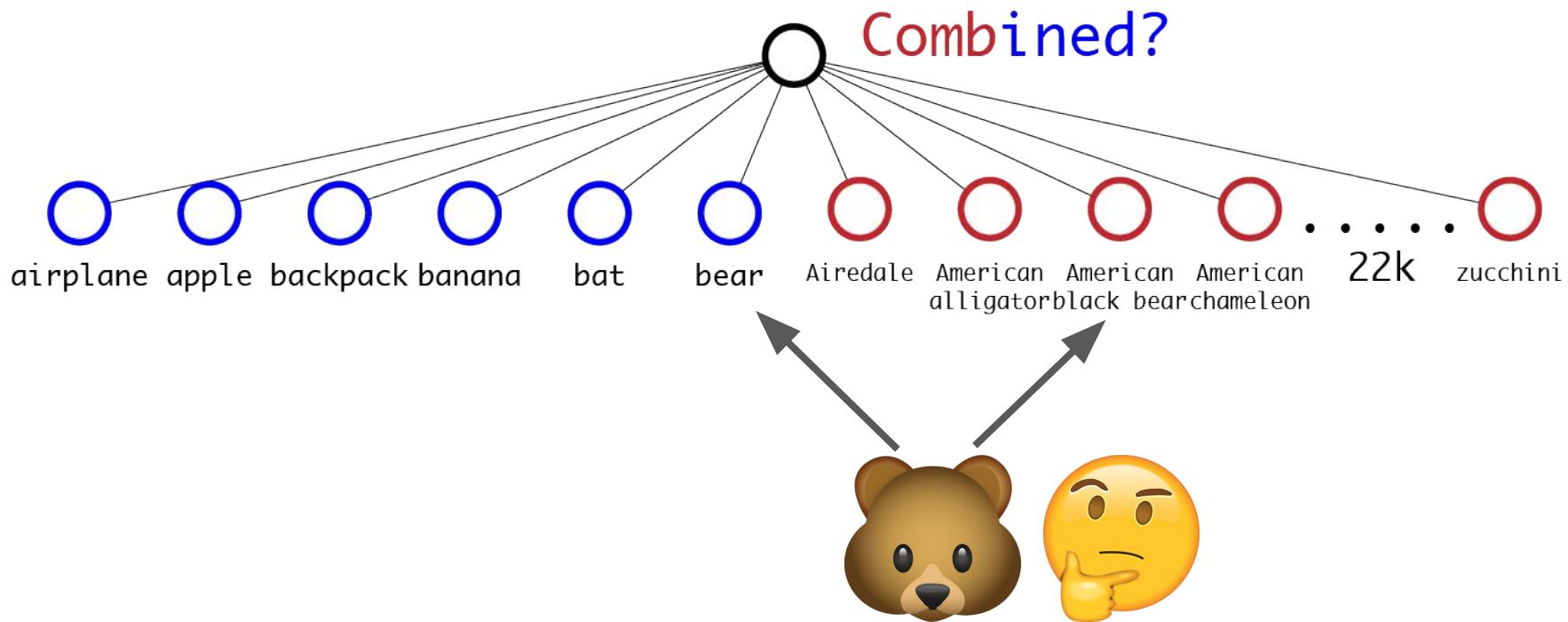
Can't just mash classes together...



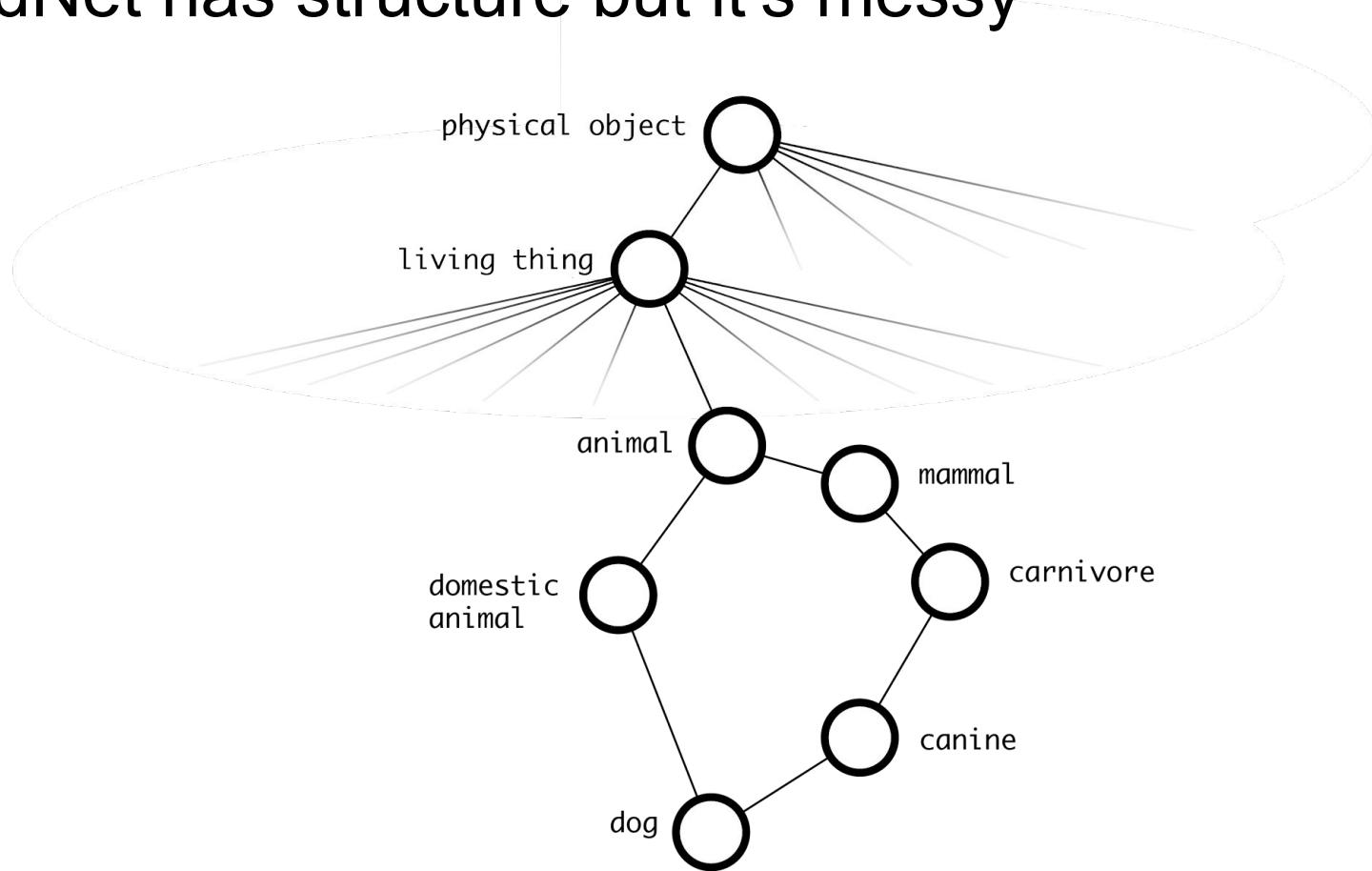
Can't just mash classes together...



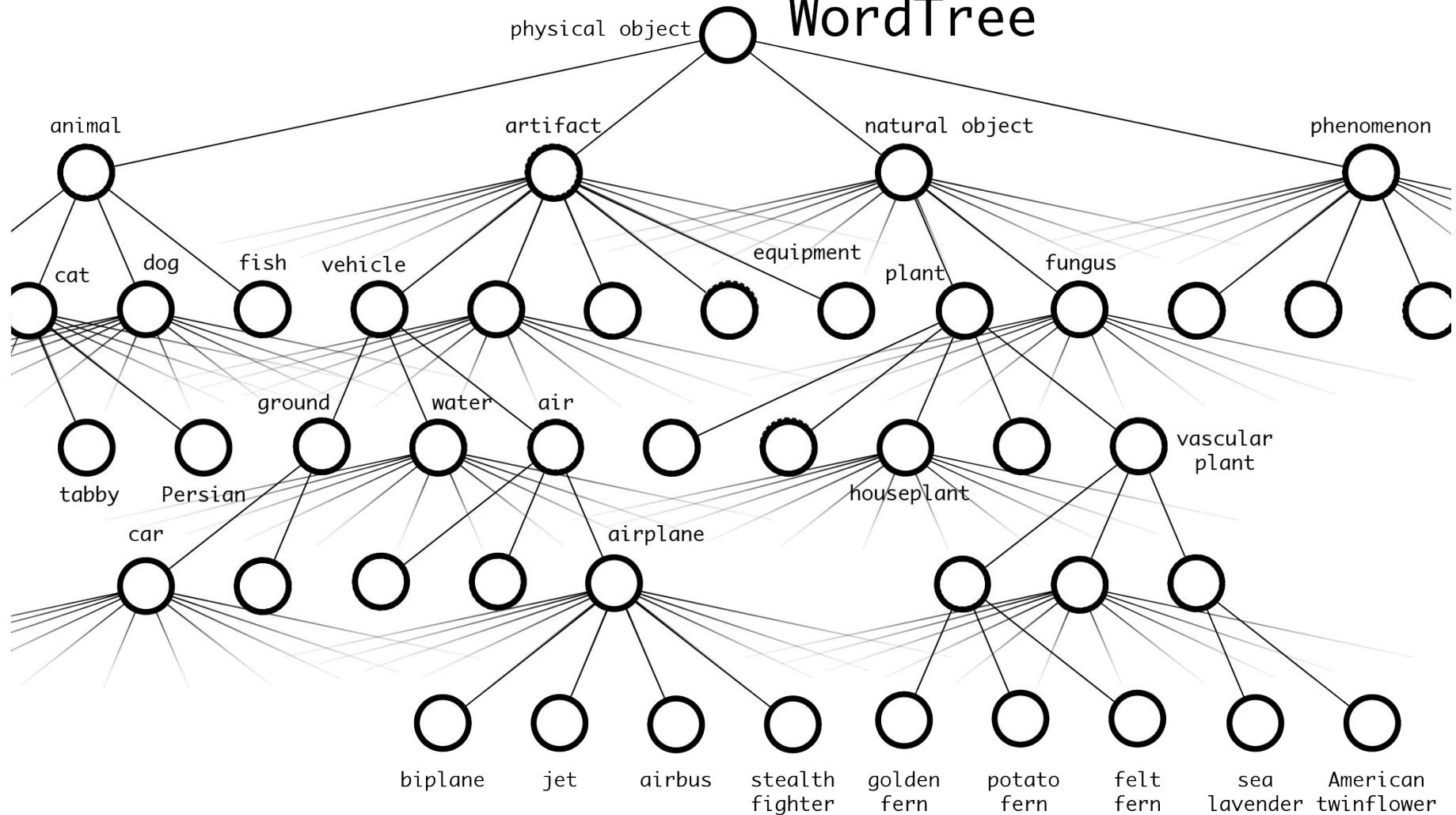
Can't just mash classes together...



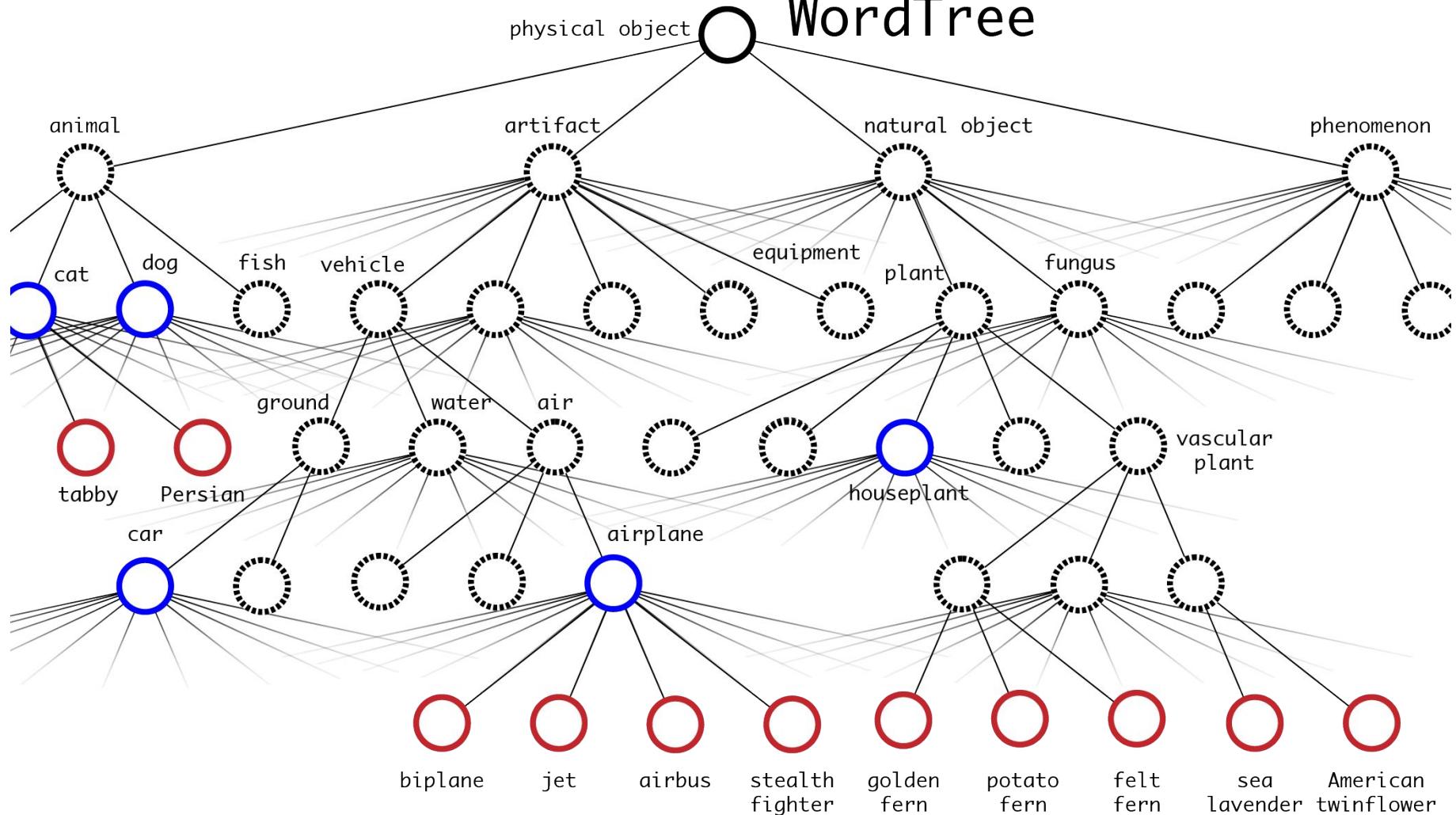
WordNet has structure but it's messy



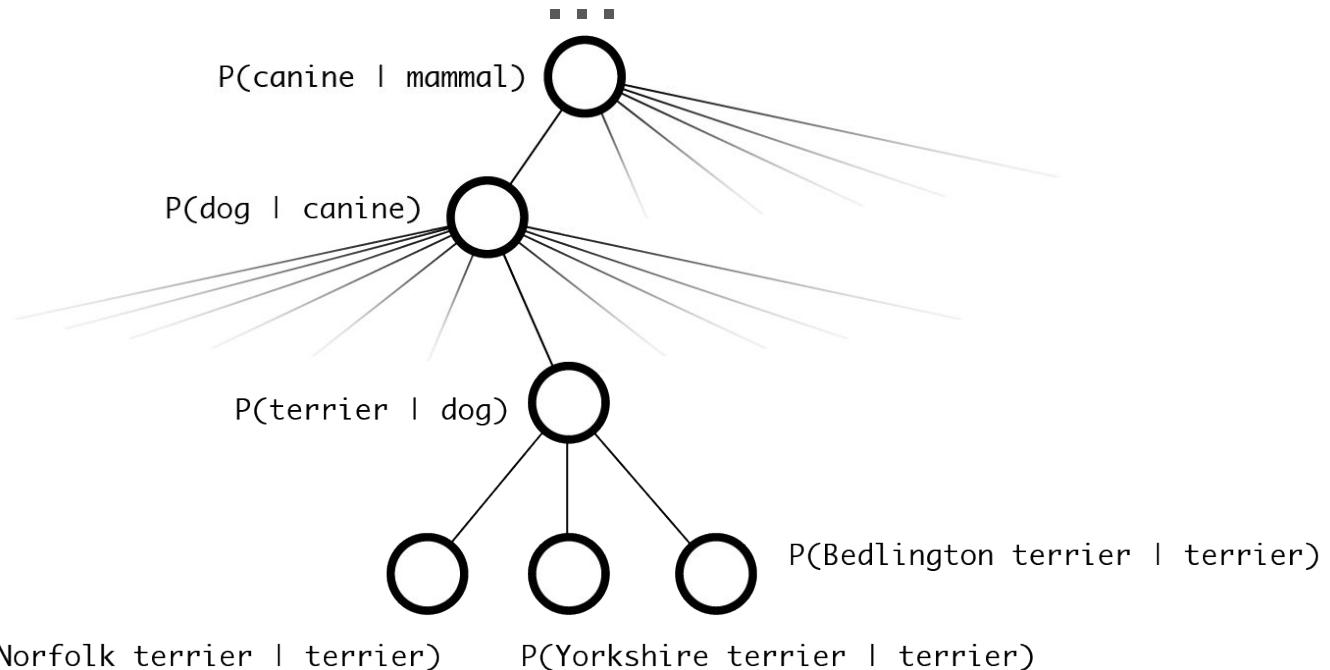
WordTree



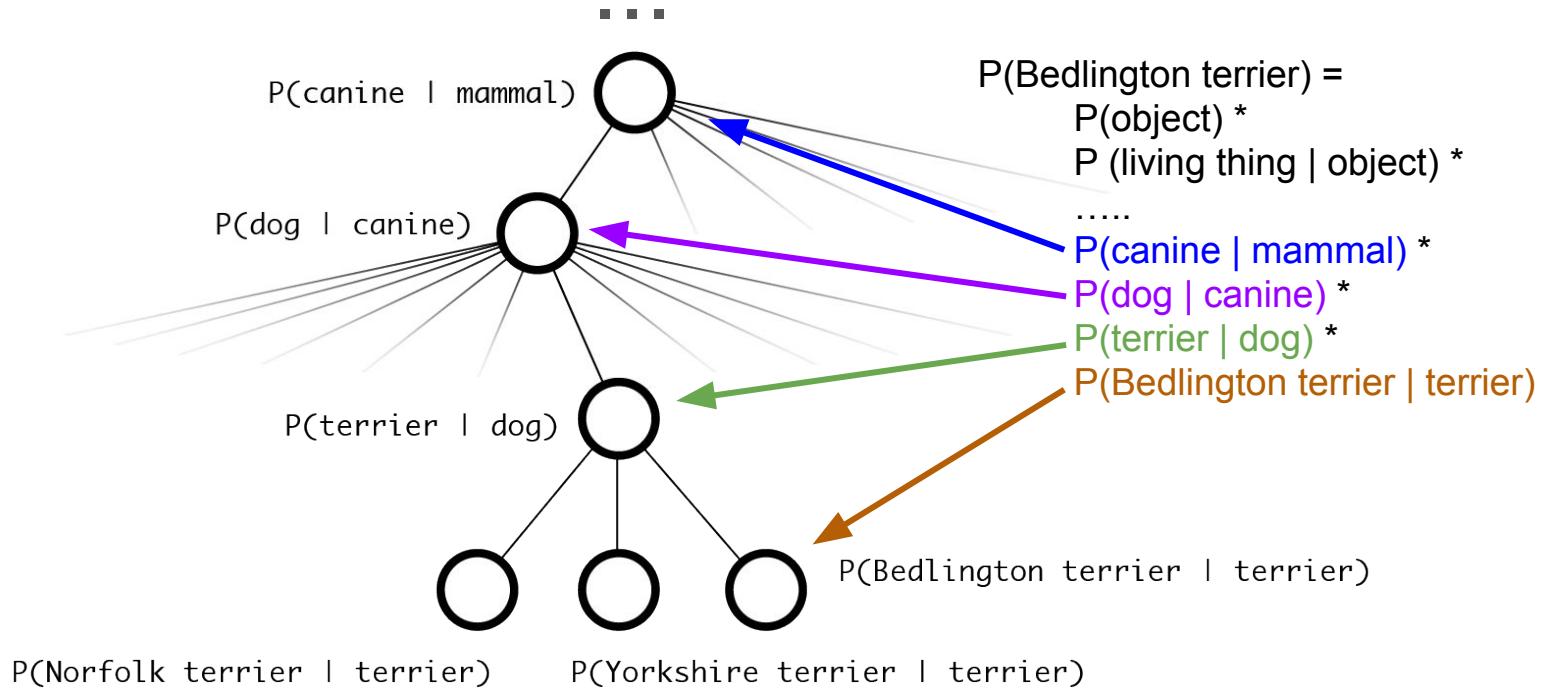
WordTree



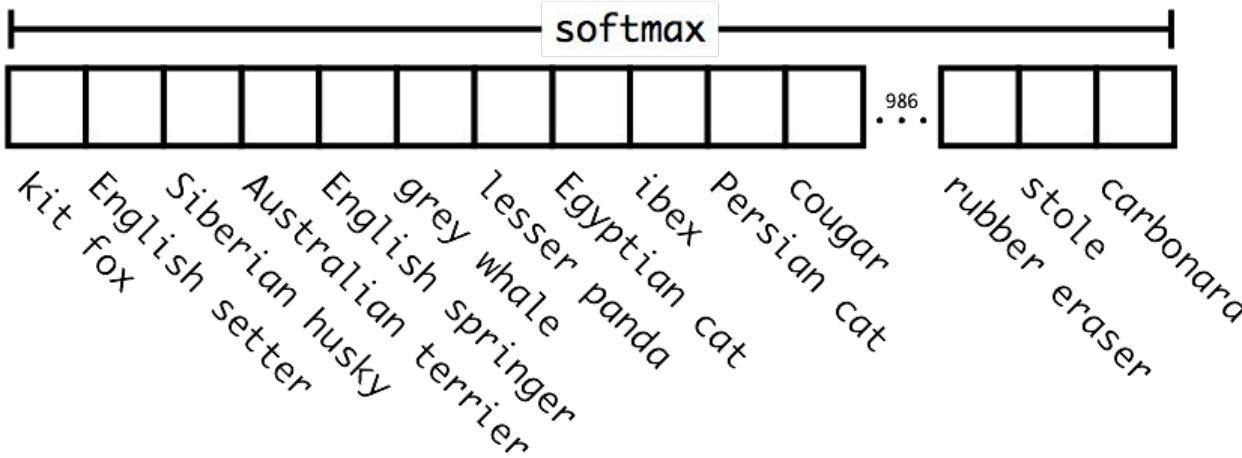
Each node is a conditional probability



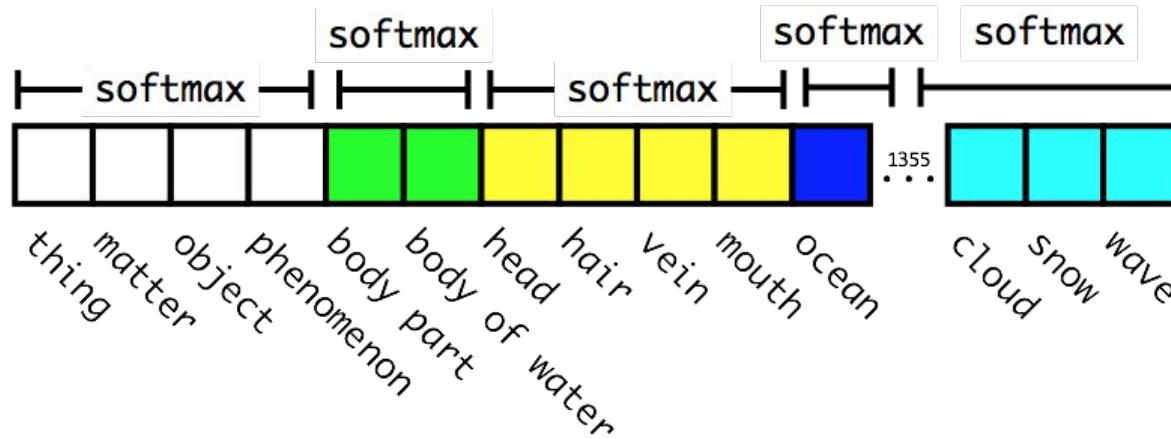
Each node is a conditional probability



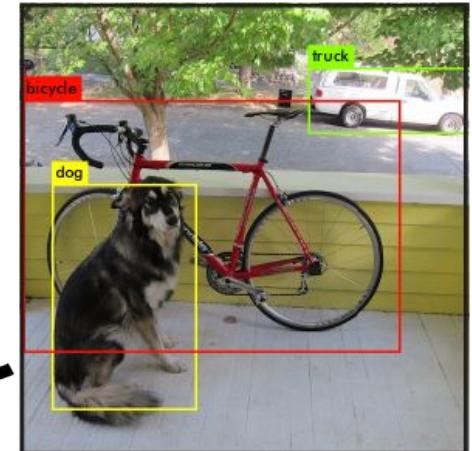
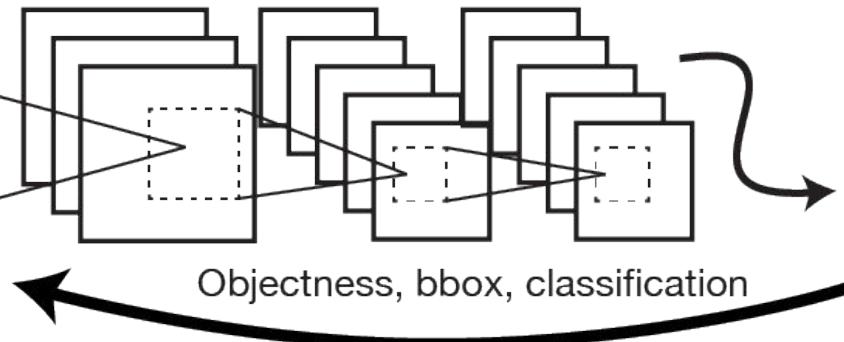
Imagenet 1k



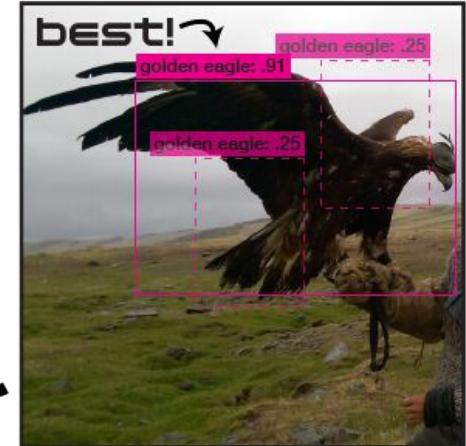
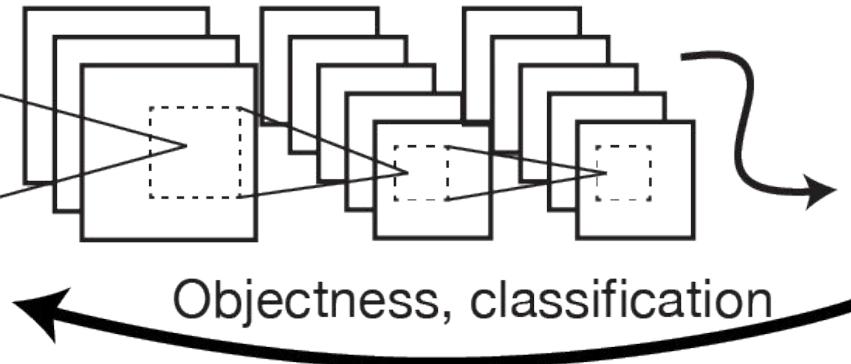
WordTree1k



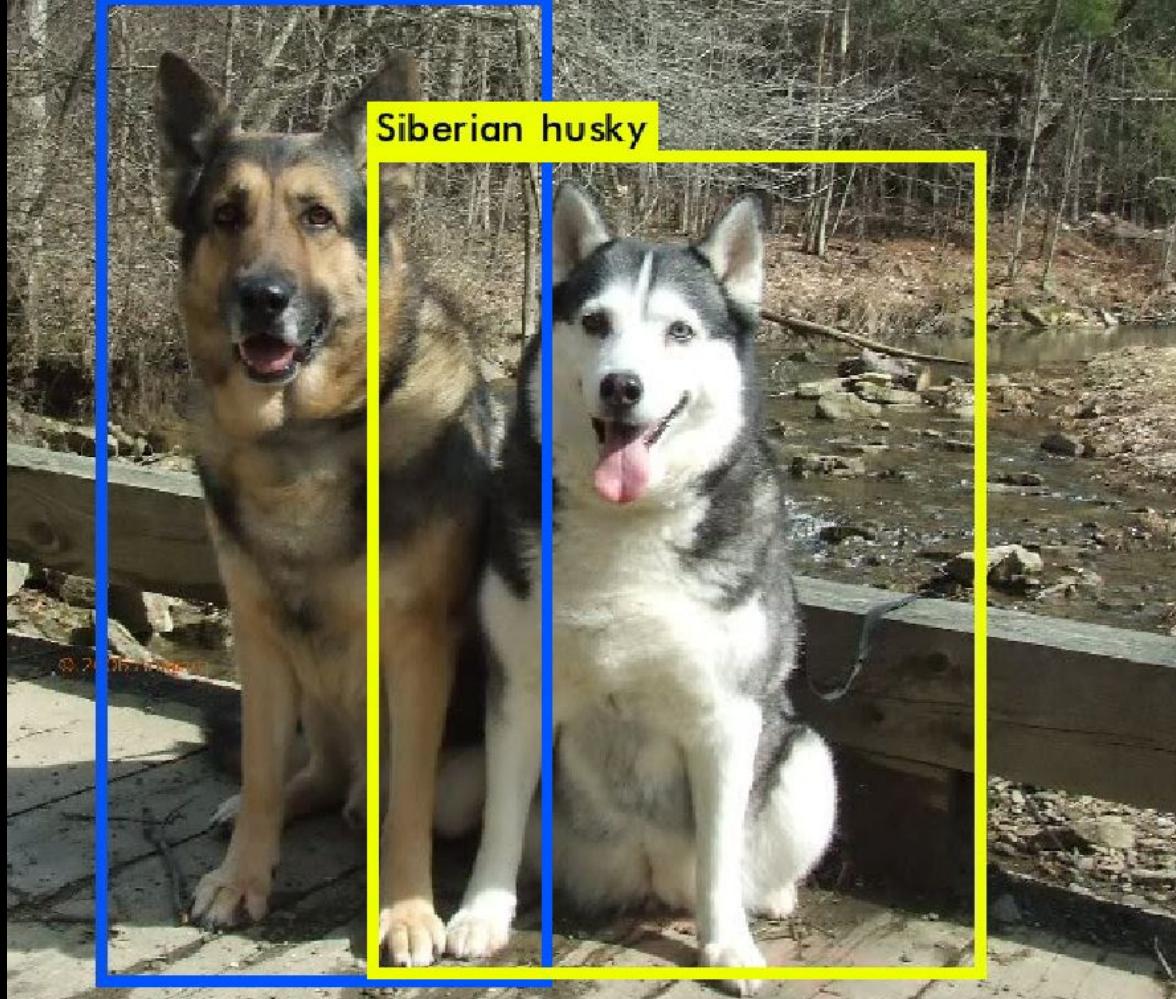
When you see detection data, backpropagate objectness, bbox, and classification error.



When you see classification data, pick best prediction, backpropagate objectness, classification error.

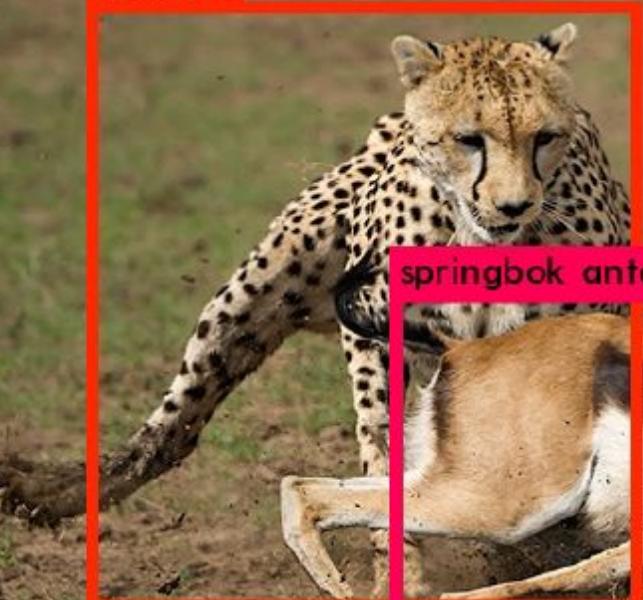


German shepherd



Siberian husky

cheetah



springbok antelope



dining table

vase

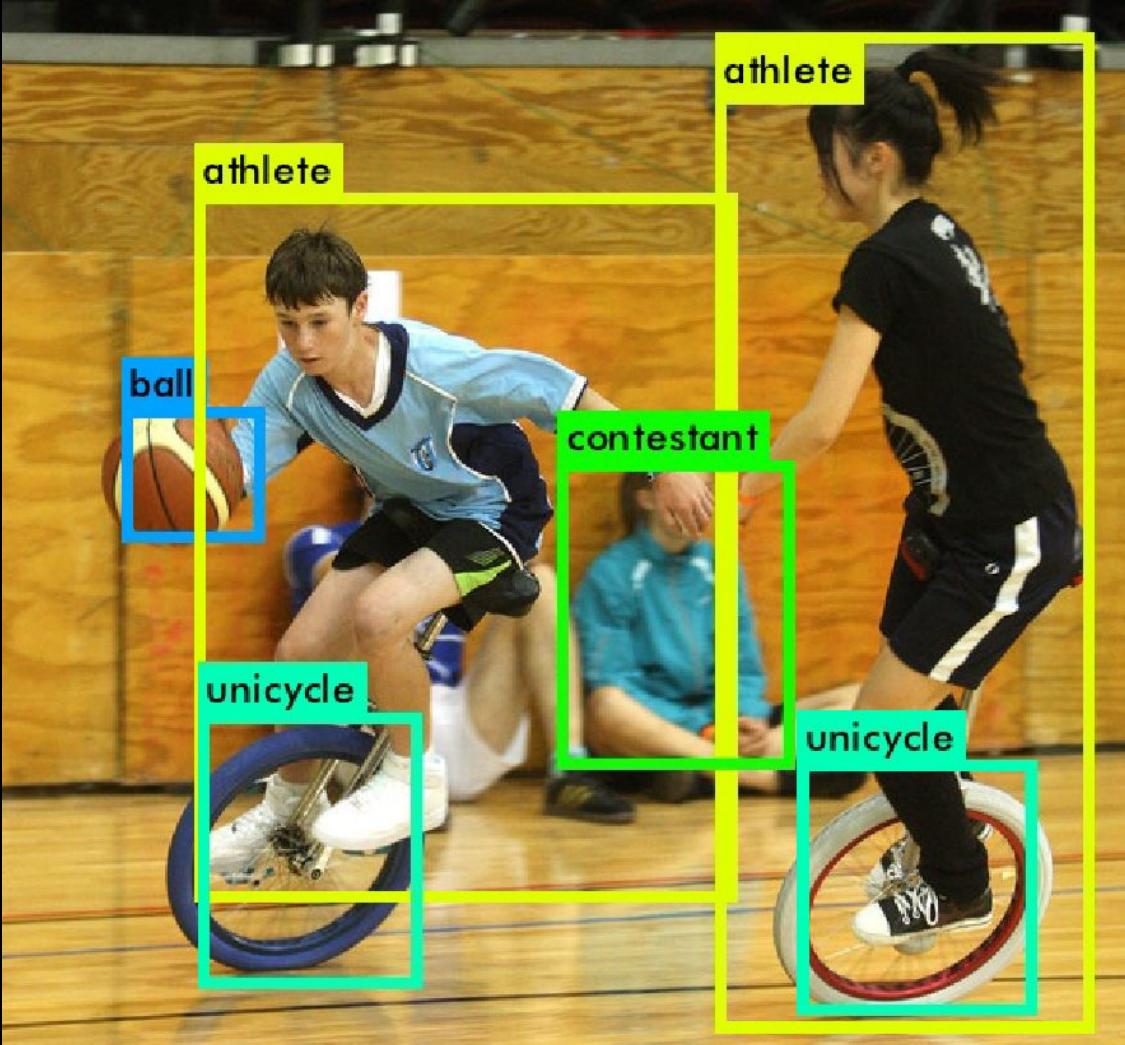
bottle

catsup bottle

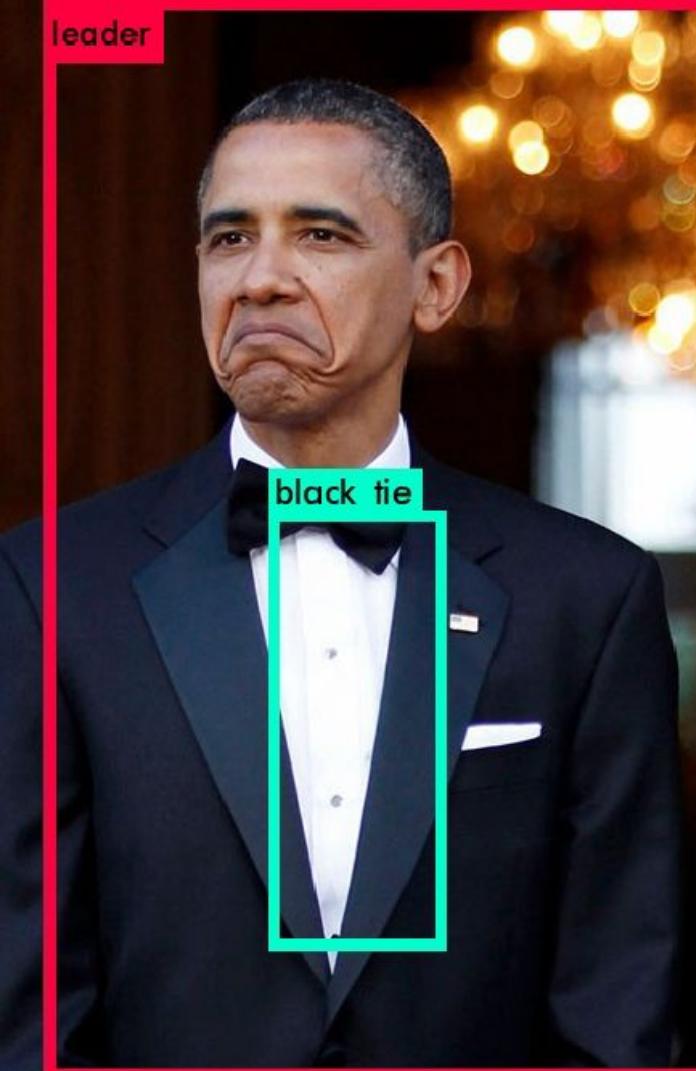
hamburger

edible fruit





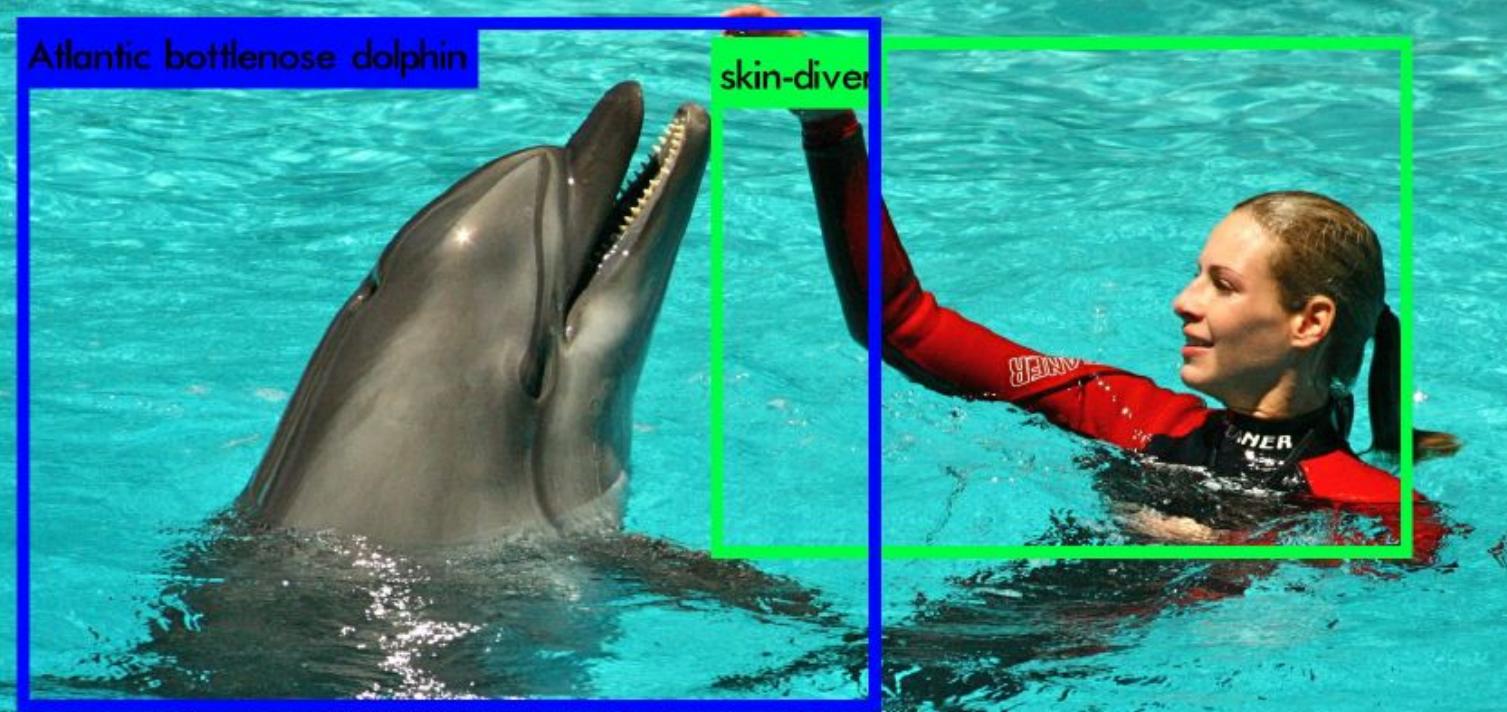
leader



American



Atlantic bottlenose dolphin



skin-diver

baleen whale



boat

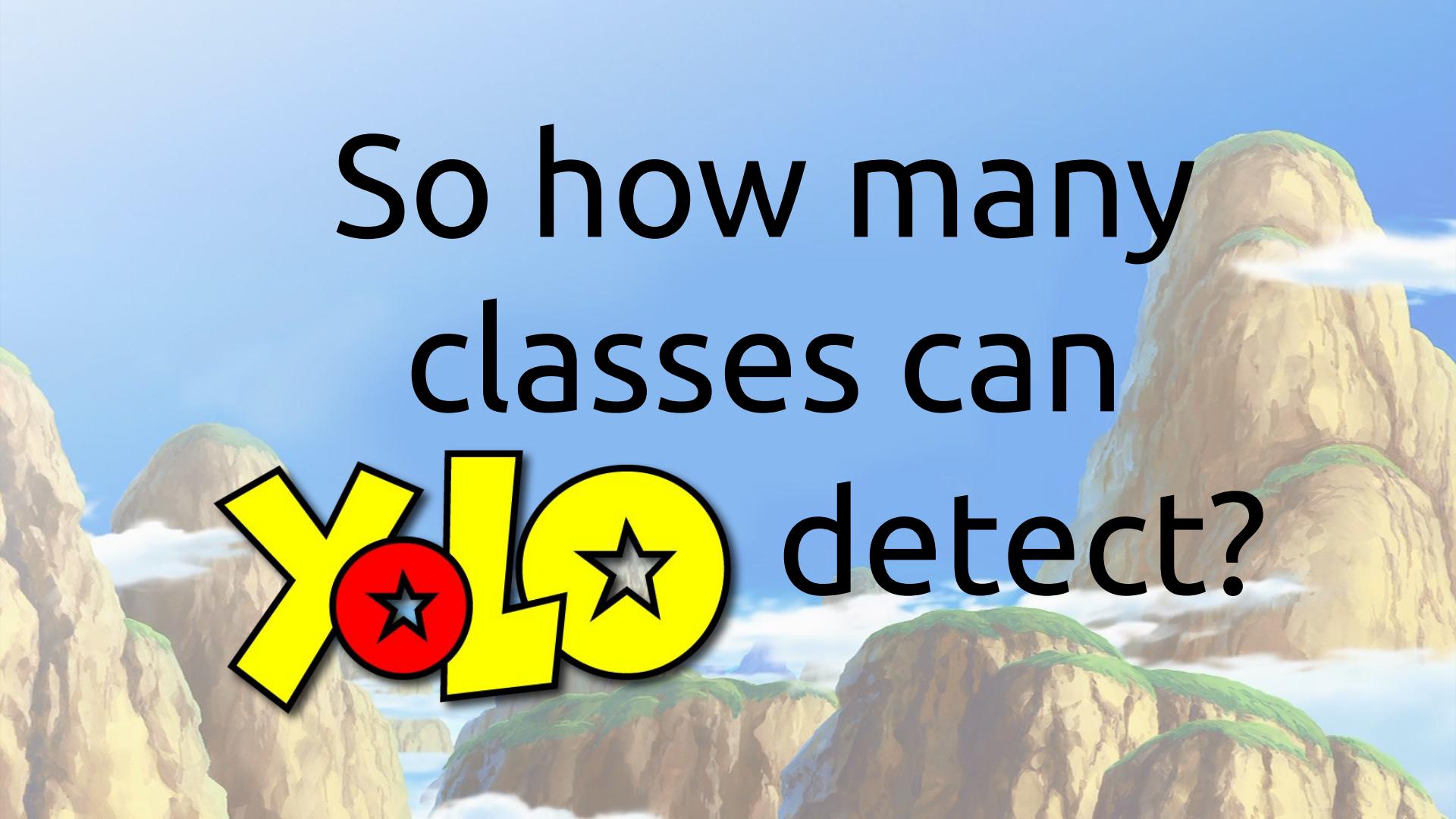


chanterelle

chanterelle

carrot



The background of the slide features a scenic landscape with several large, craggy mountains. The sky is a clear, pale blue with wispy white clouds.

So how many
classes can
YOLO detect?







Our Work Is

NEVER OVER



Code, models, and updates:

<https://pjreddie.com/yolo/>

XNOR.AI