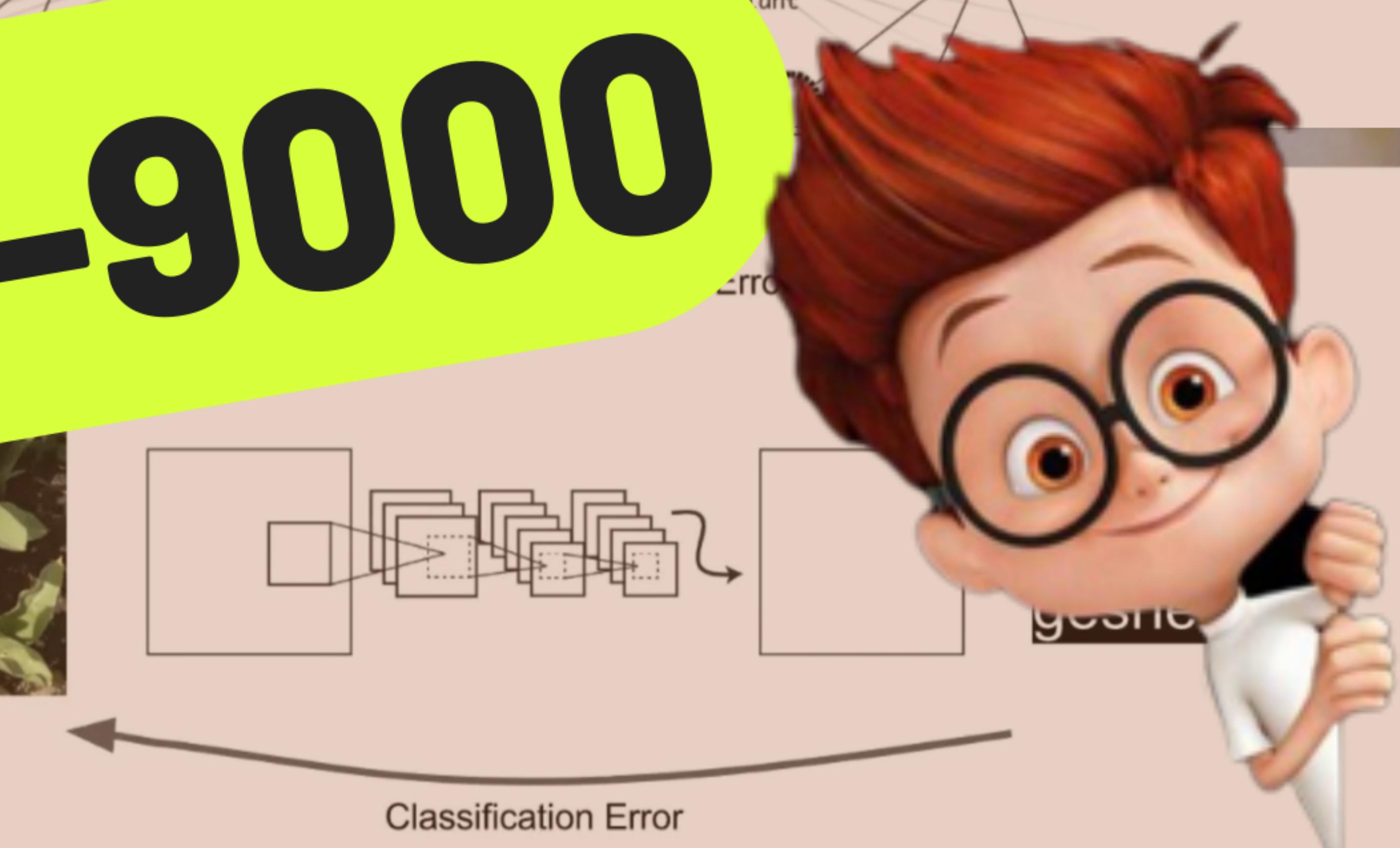
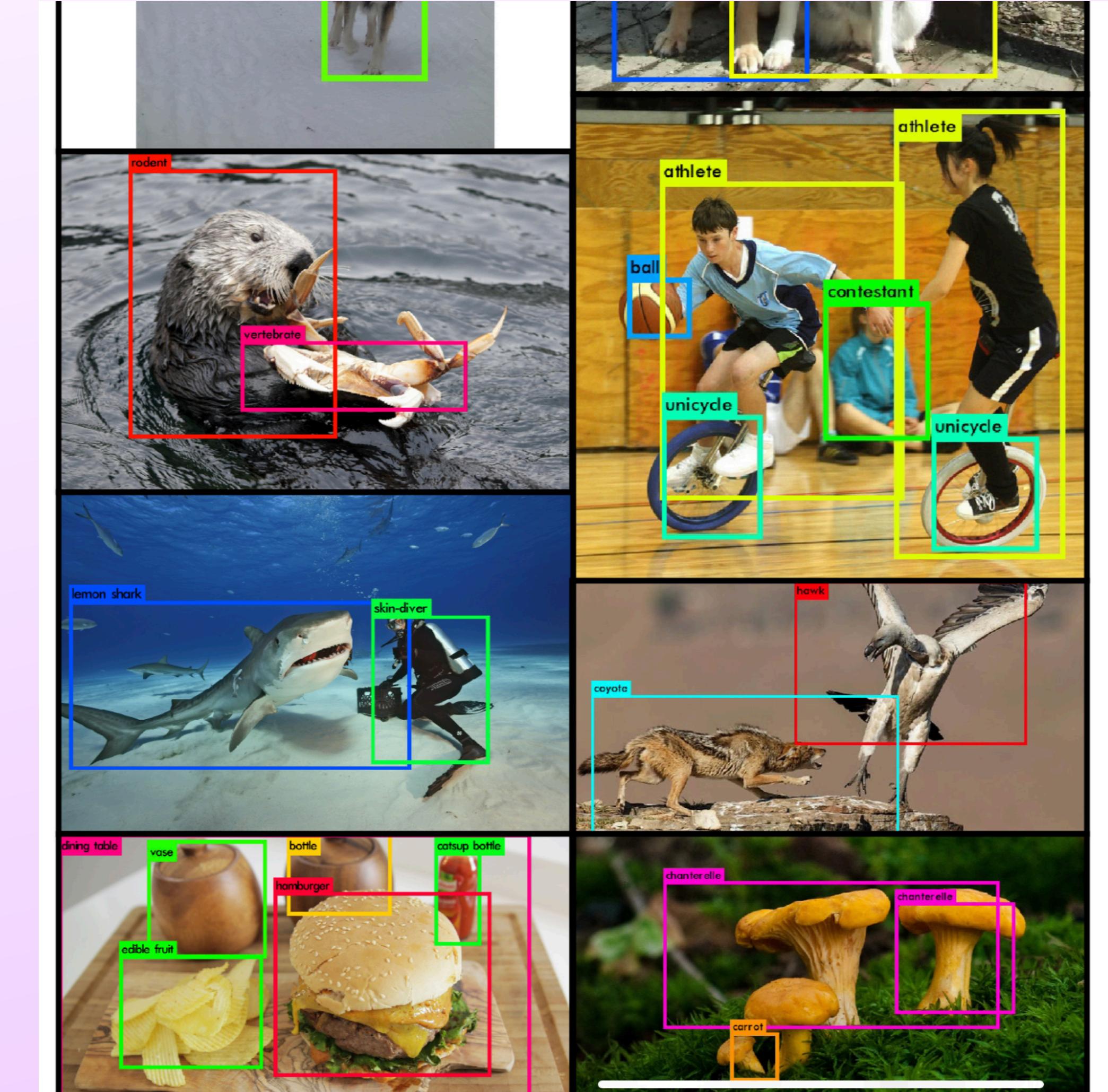
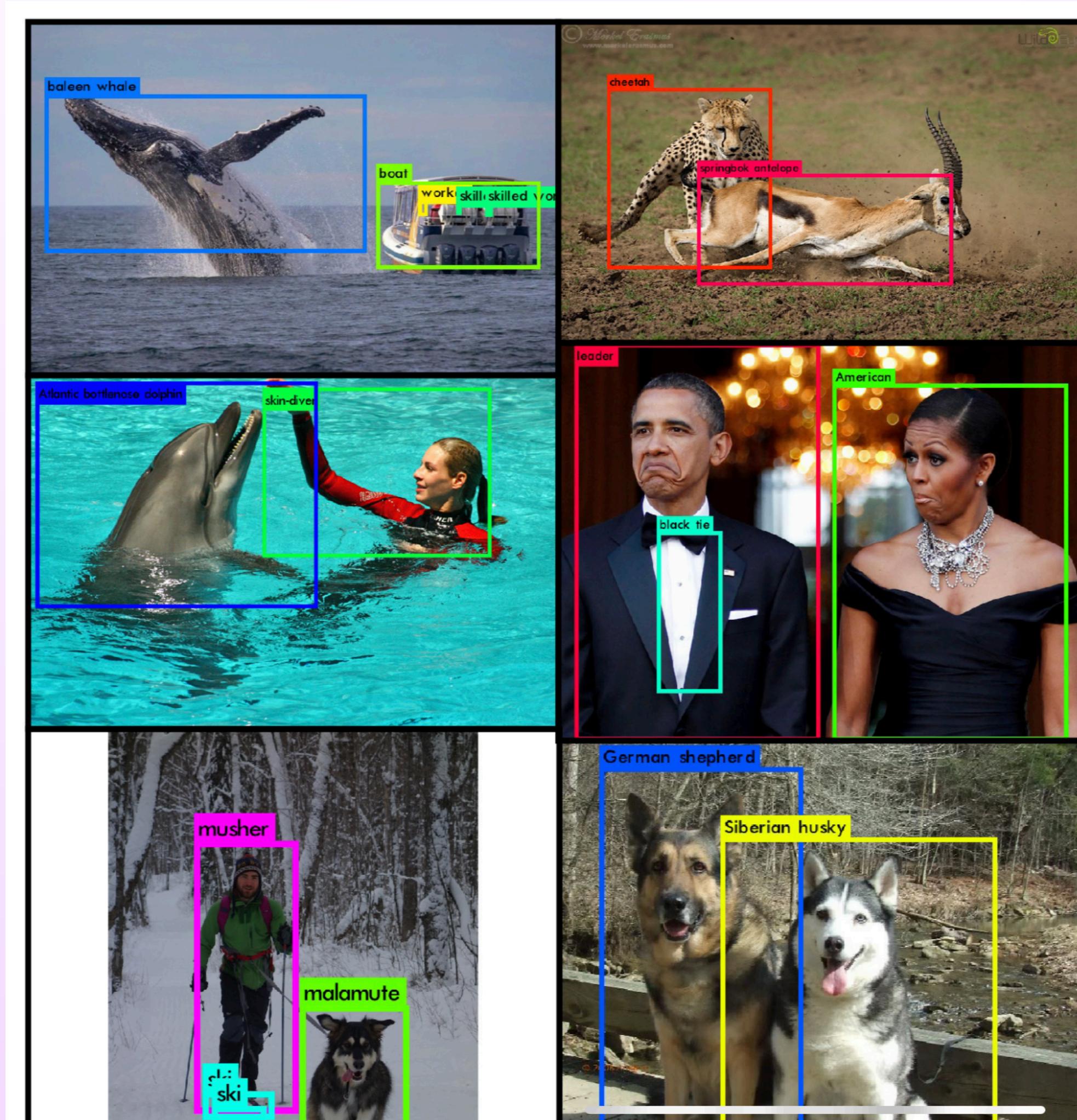


YOLO-9000



YOLO-9000



Object Detection Datasets

Dataset	COCO 2017	PASCAL VOC (07 + 12)
Number of categories	80	20
Number of training pictures	117,264	16,551
Number of testing pictures	5000	4952
Total sample boxes	902,435	52,090
Total sample boxes / total number of images	7.4	2.4

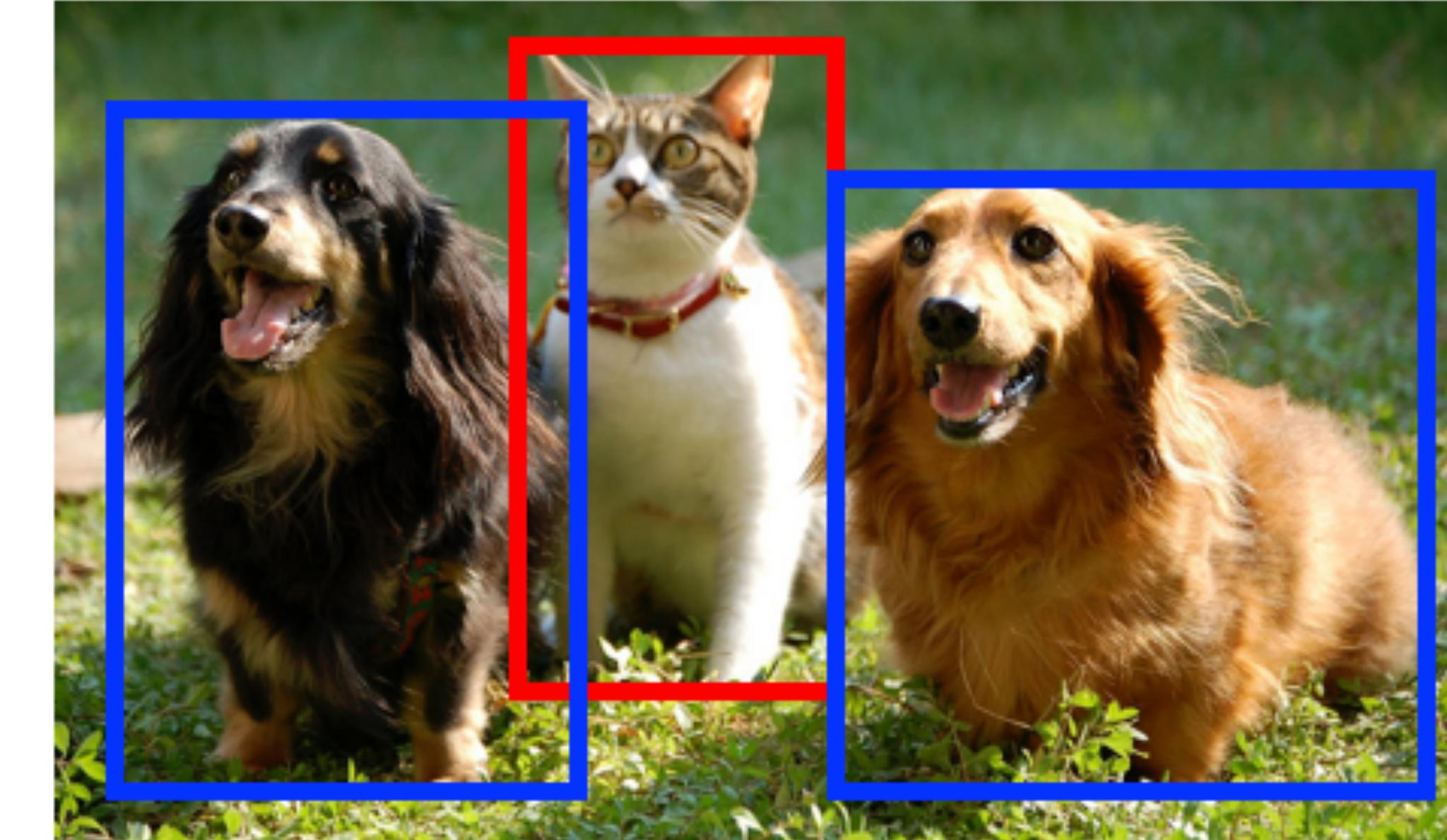
Annotations

Classification



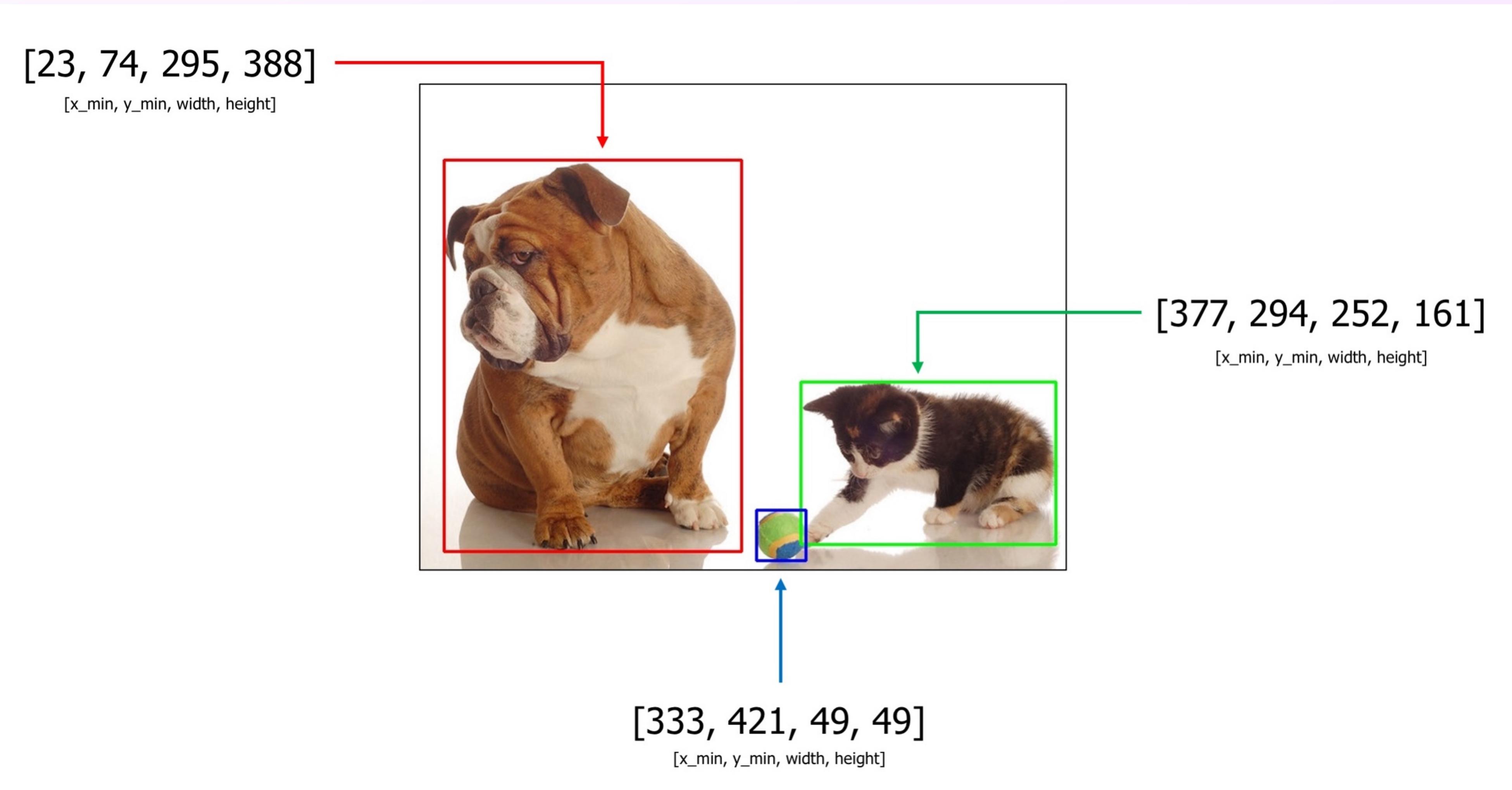
CAT

Object Detection

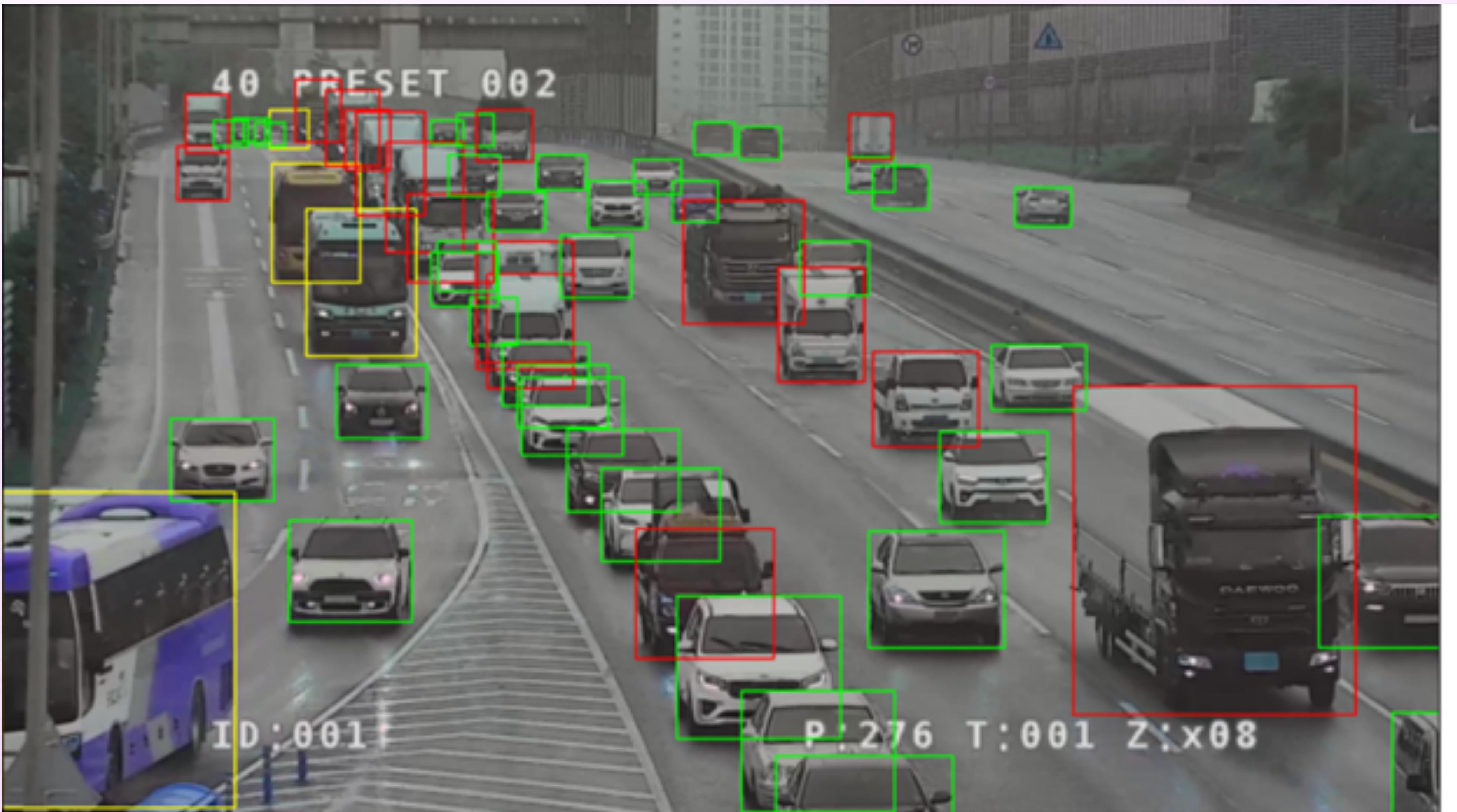


CAT, DOG

Annotations



Annotations



Imagenet Dataset



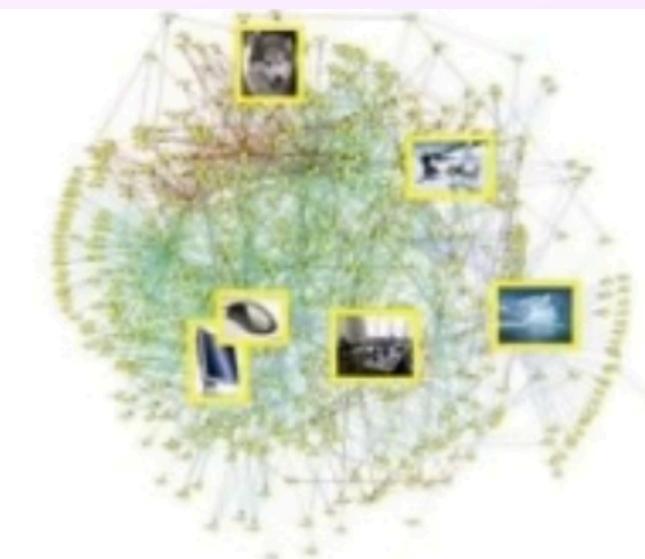
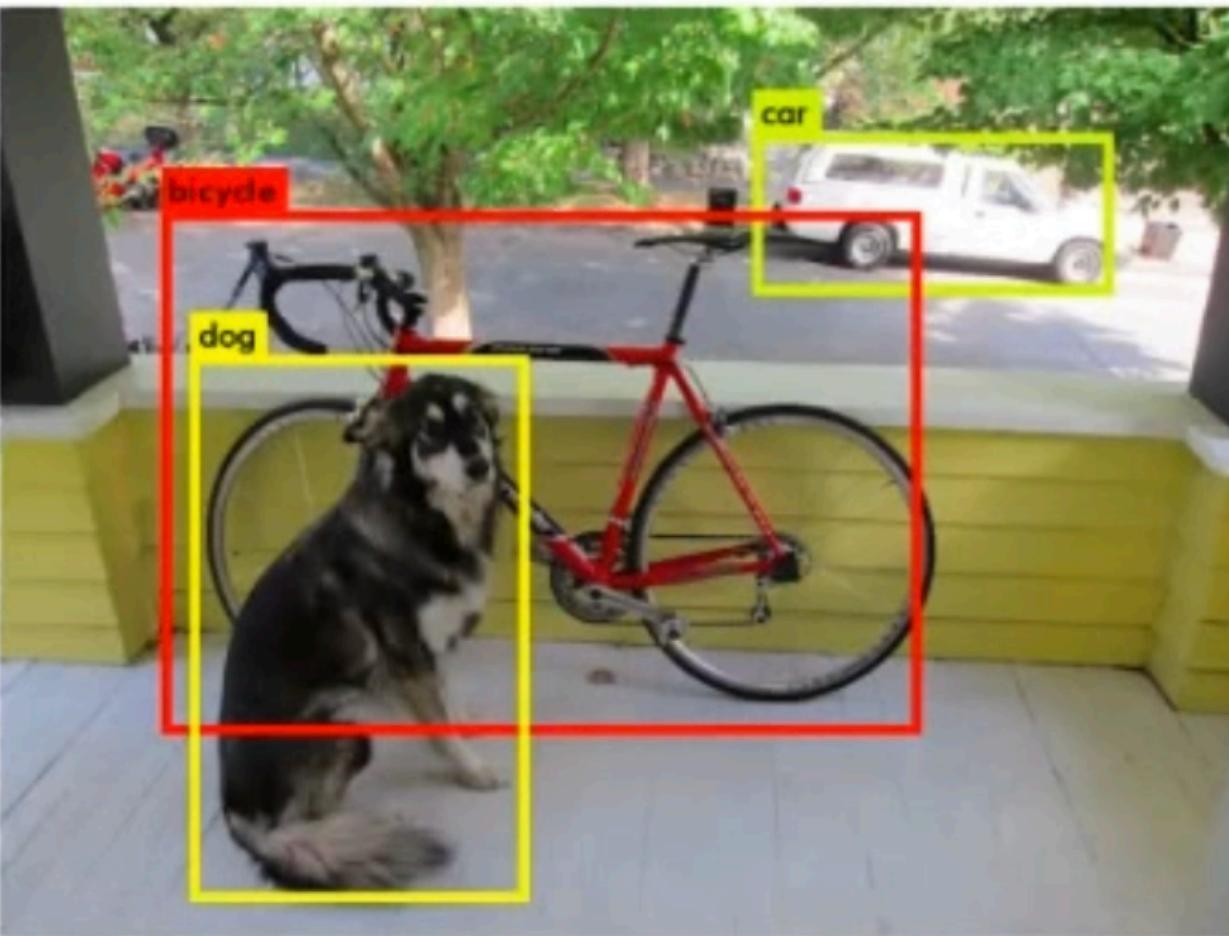
How to combine these?



COCO

Common Objects in Context

- 100k images
- 80 classes
- Detection labels

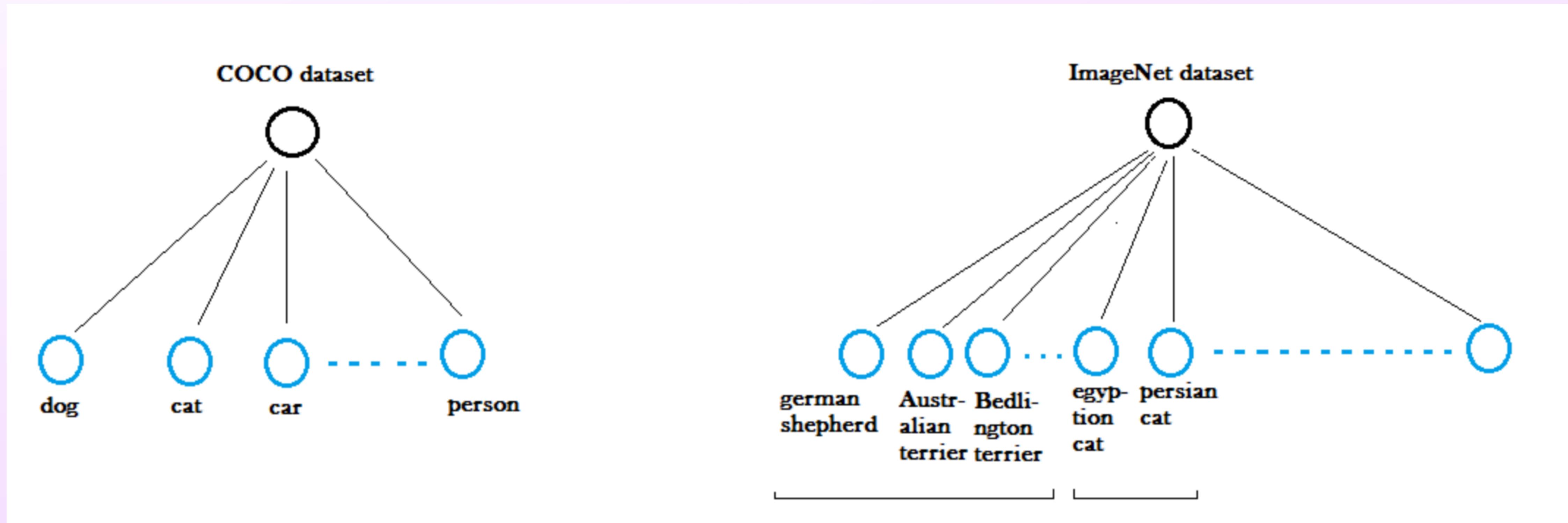


IMAGENET

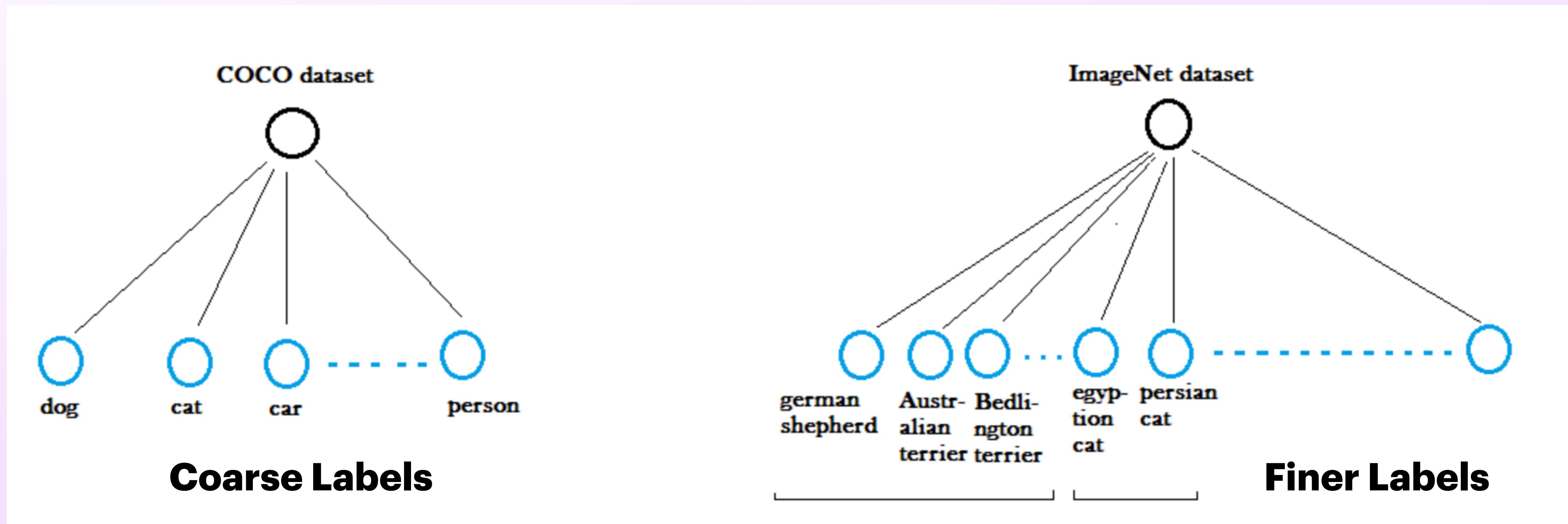
- 14 million images
- 22k classes
- Classification labels



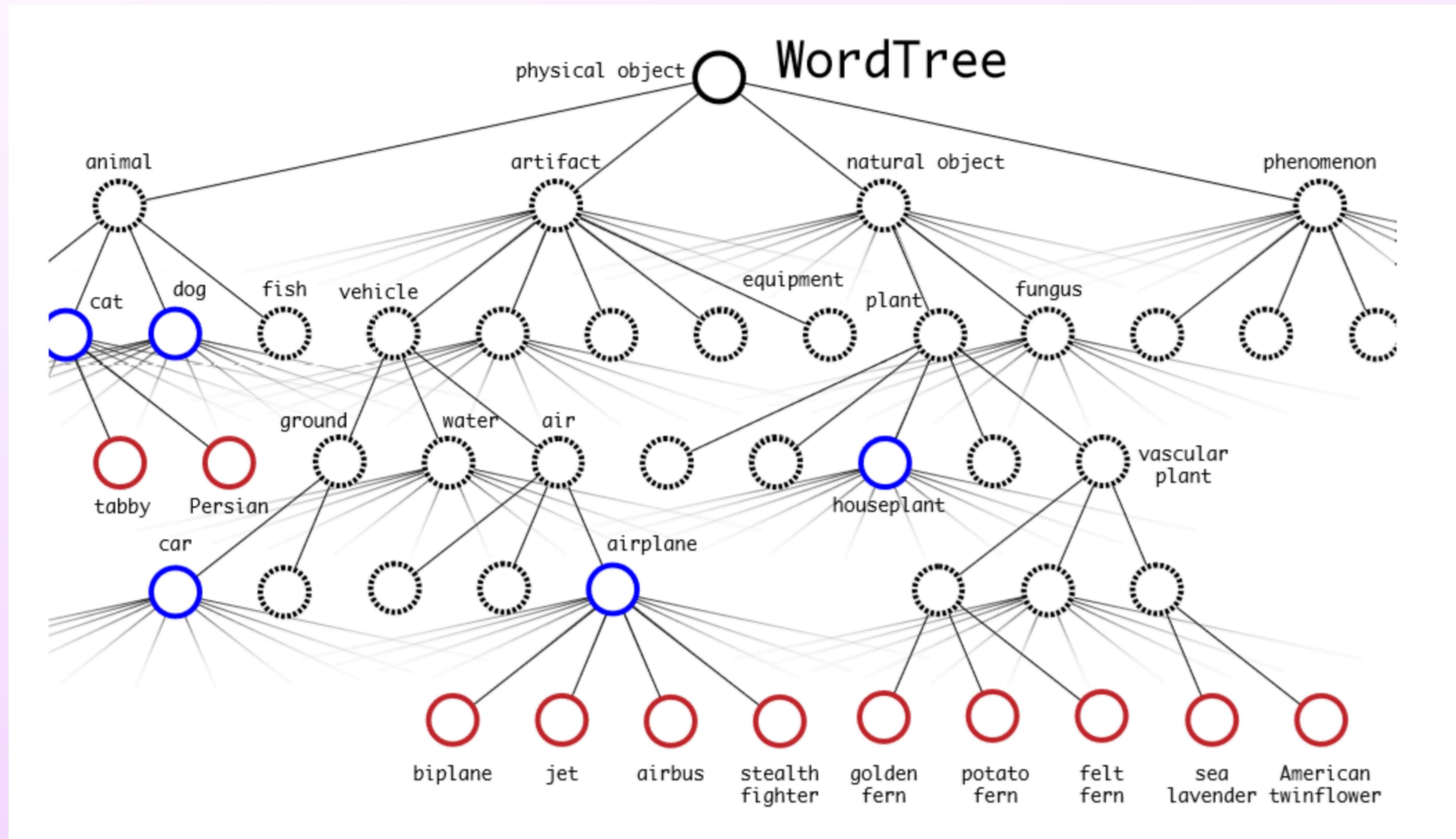
How to combine these?



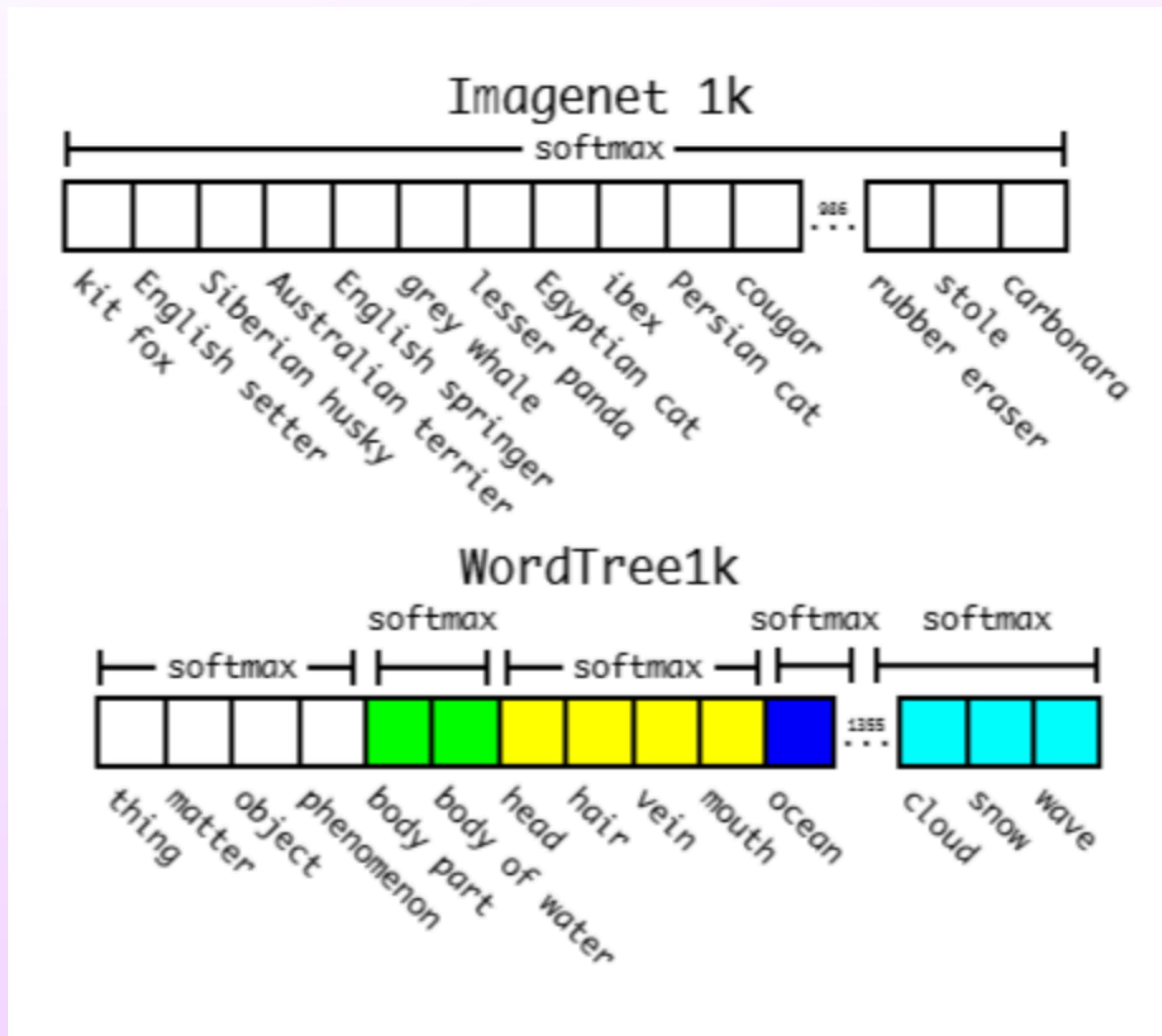
How to combine these?



WordTree 1K



Darknet-19 Training



Conditional Probabilities

$Pr(\text{Norfolk terrier}) = Pr(\text{Norfolk terrier}|\text{terrier})$

* $Pr(\text{terrier}|\text{hunting dog})$

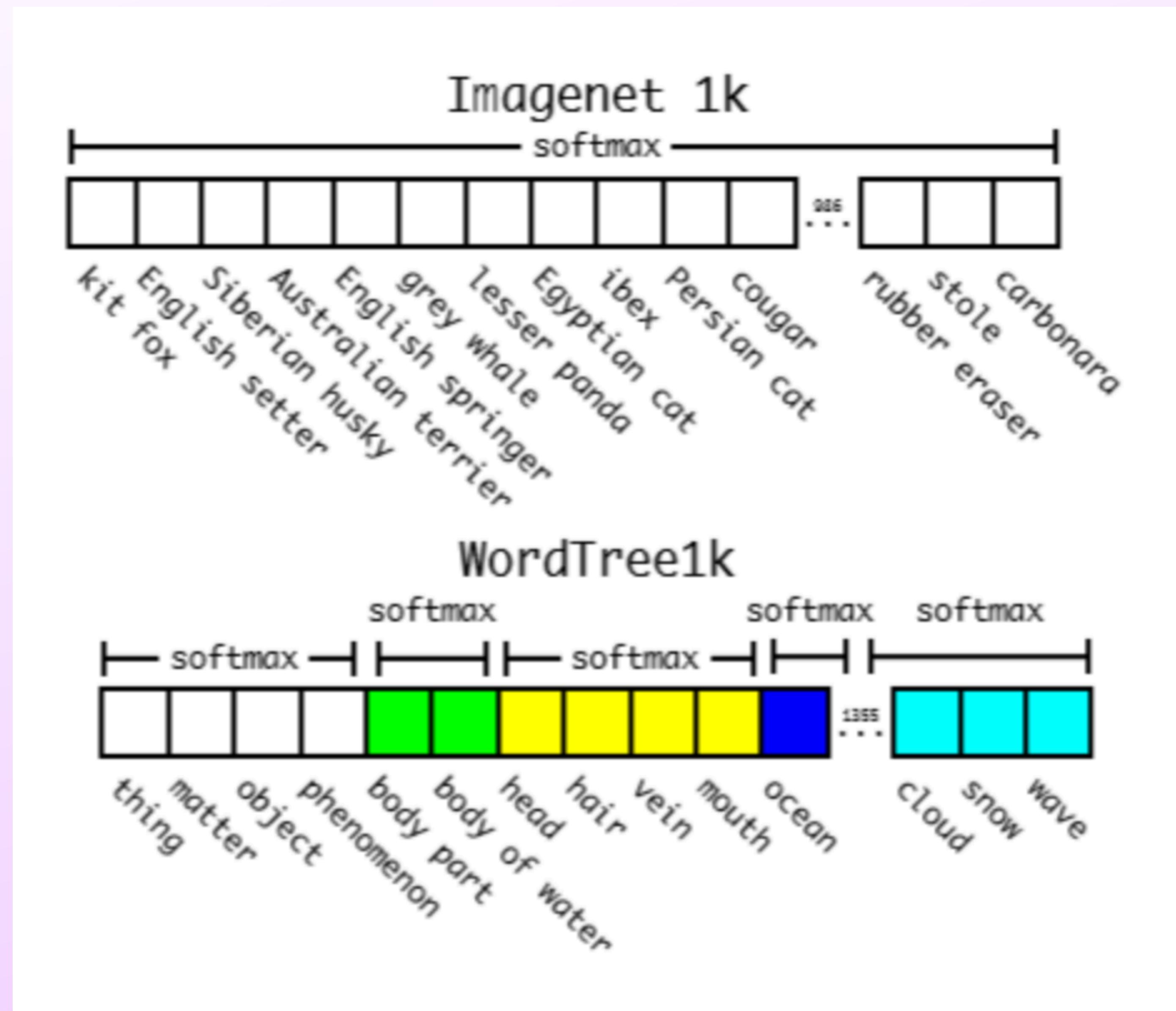
* . . . *

* $Pr(\text{mammal}|Pr(\text{animal})$)

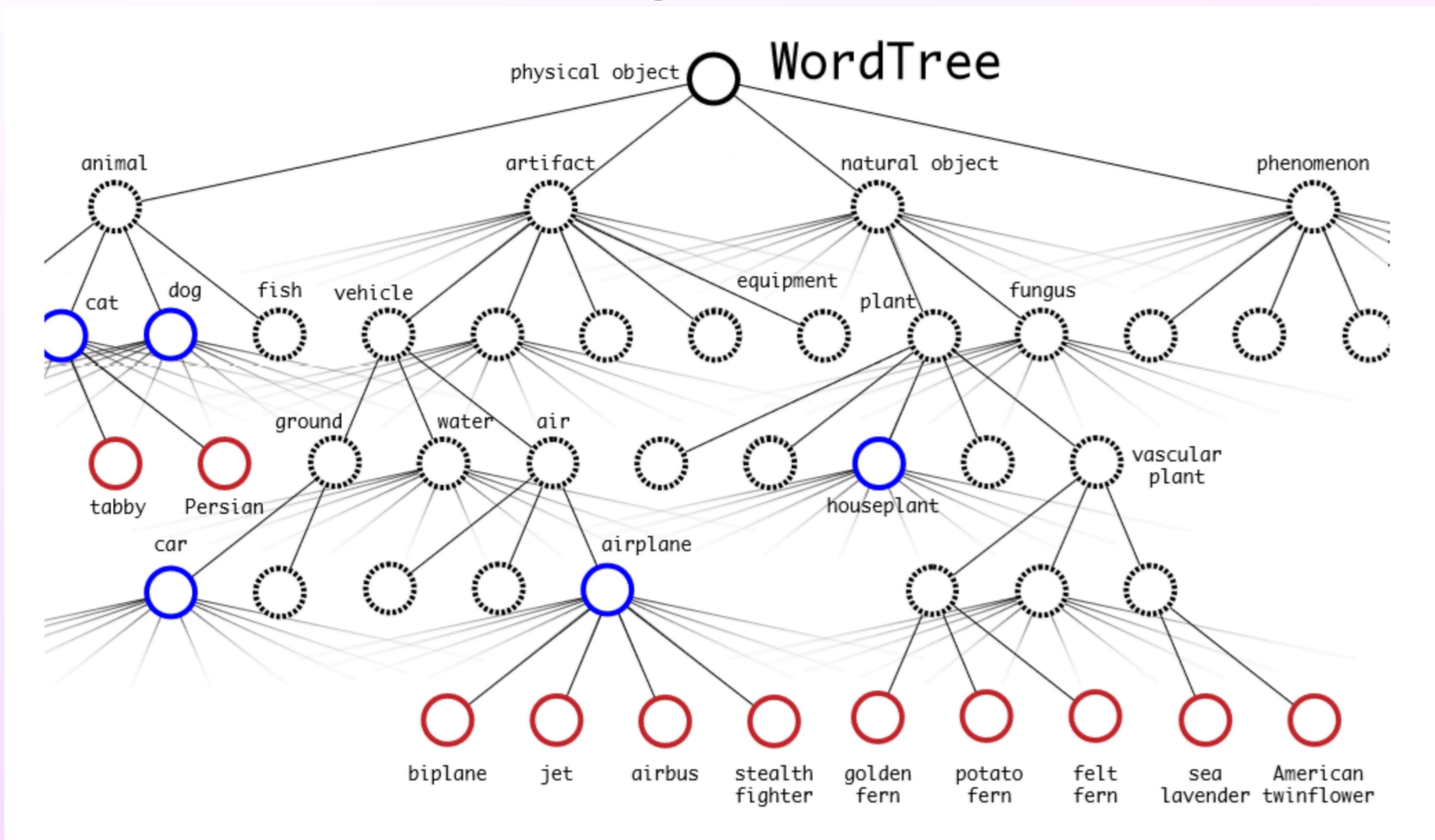
* $Pr(\text{animal}|\text{physical object})$

Darknet-19 Training

- 1369 classes
- Conditional Class Probabilities
- Top-1 ACC: 71.9%
- Top-5 ACC: 90.4%

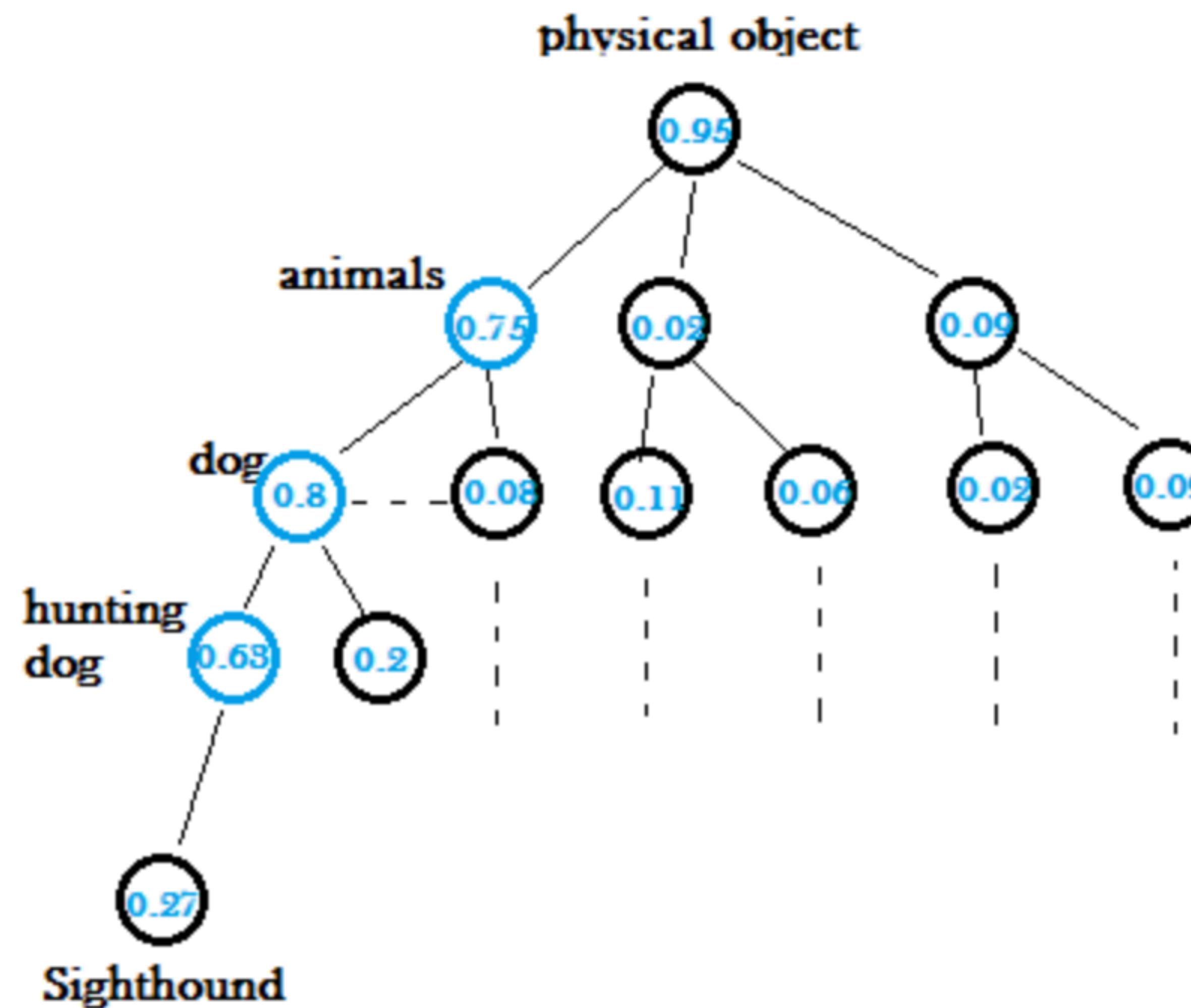


Darknet-19 Training



What is the final class prediction?

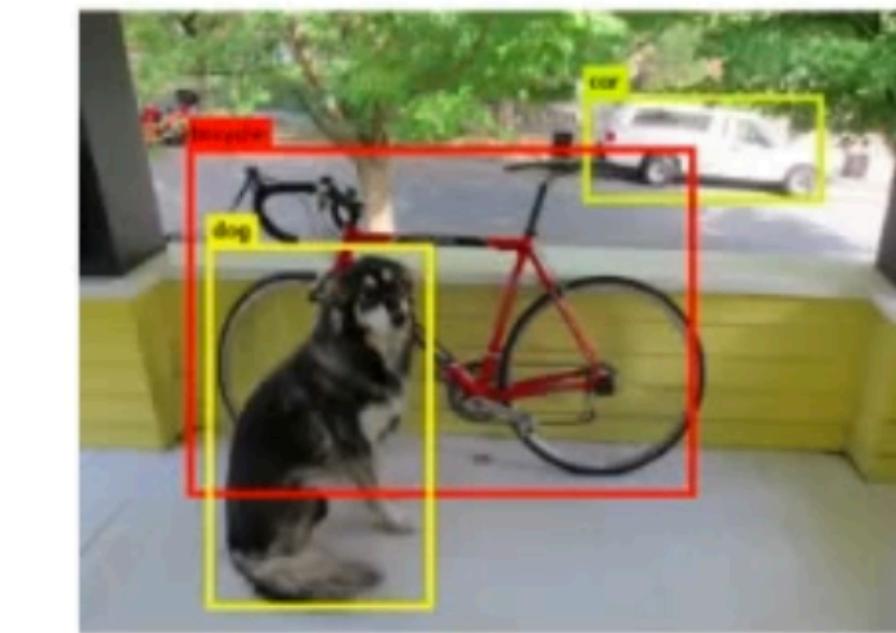
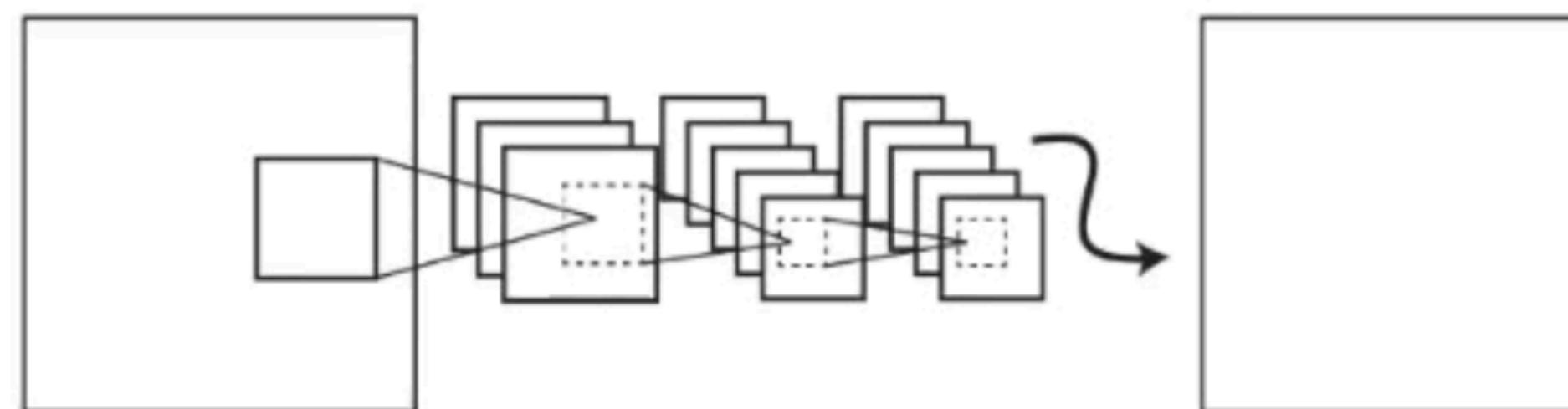
Darknet-19 Training



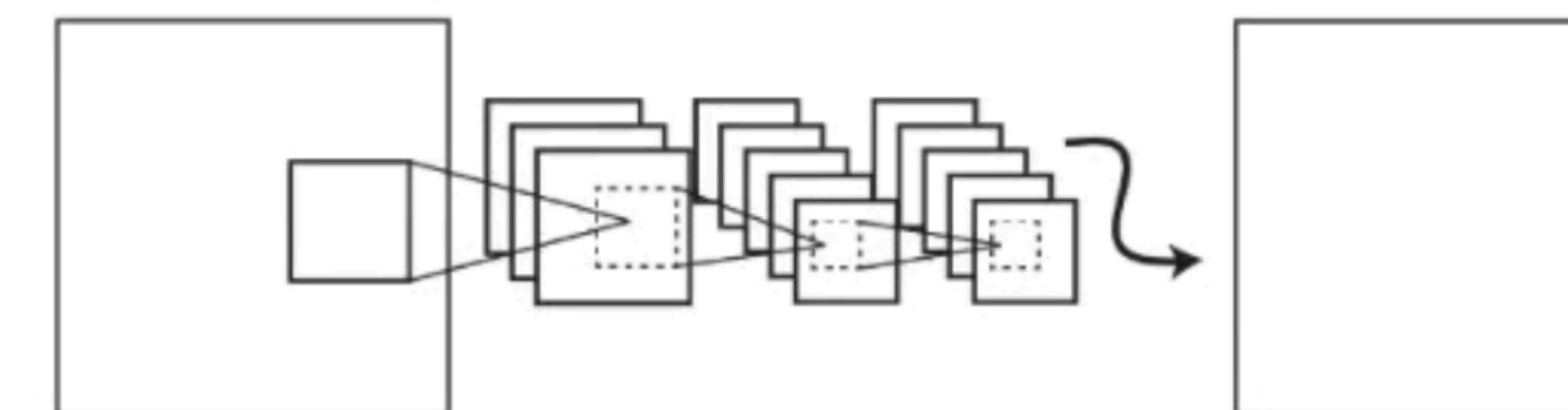
YOLO9000 Training

- Consider top-9000 labels from Imagenet dataset (22k labels)
- Club with COCO labels as hierarchies
- Total Dataset - Wordtree with 9418 labels
- Only few categories have bounding boxes
- Anchor boxes - 3 per grid cell

YOLO9000 Training



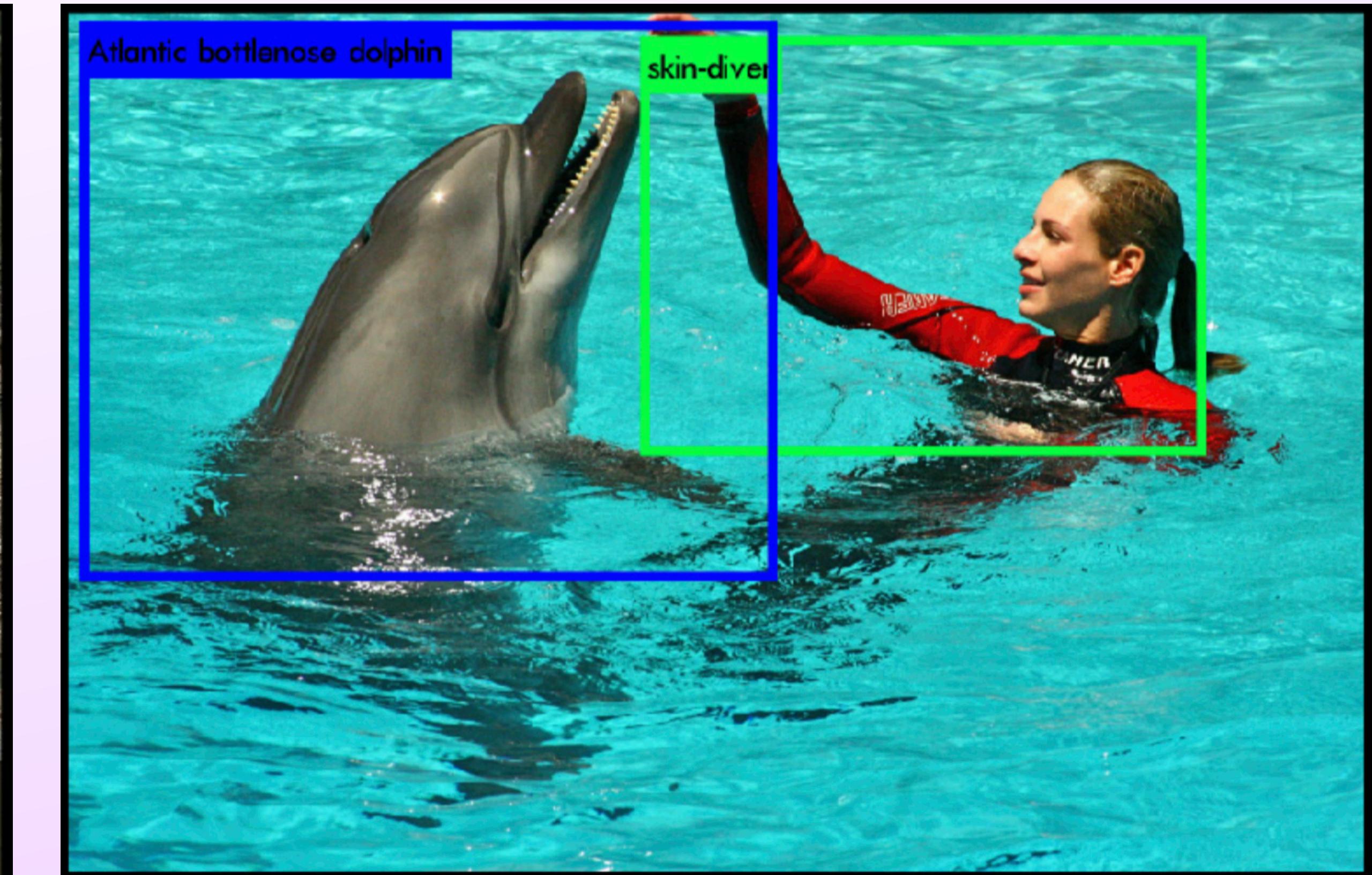
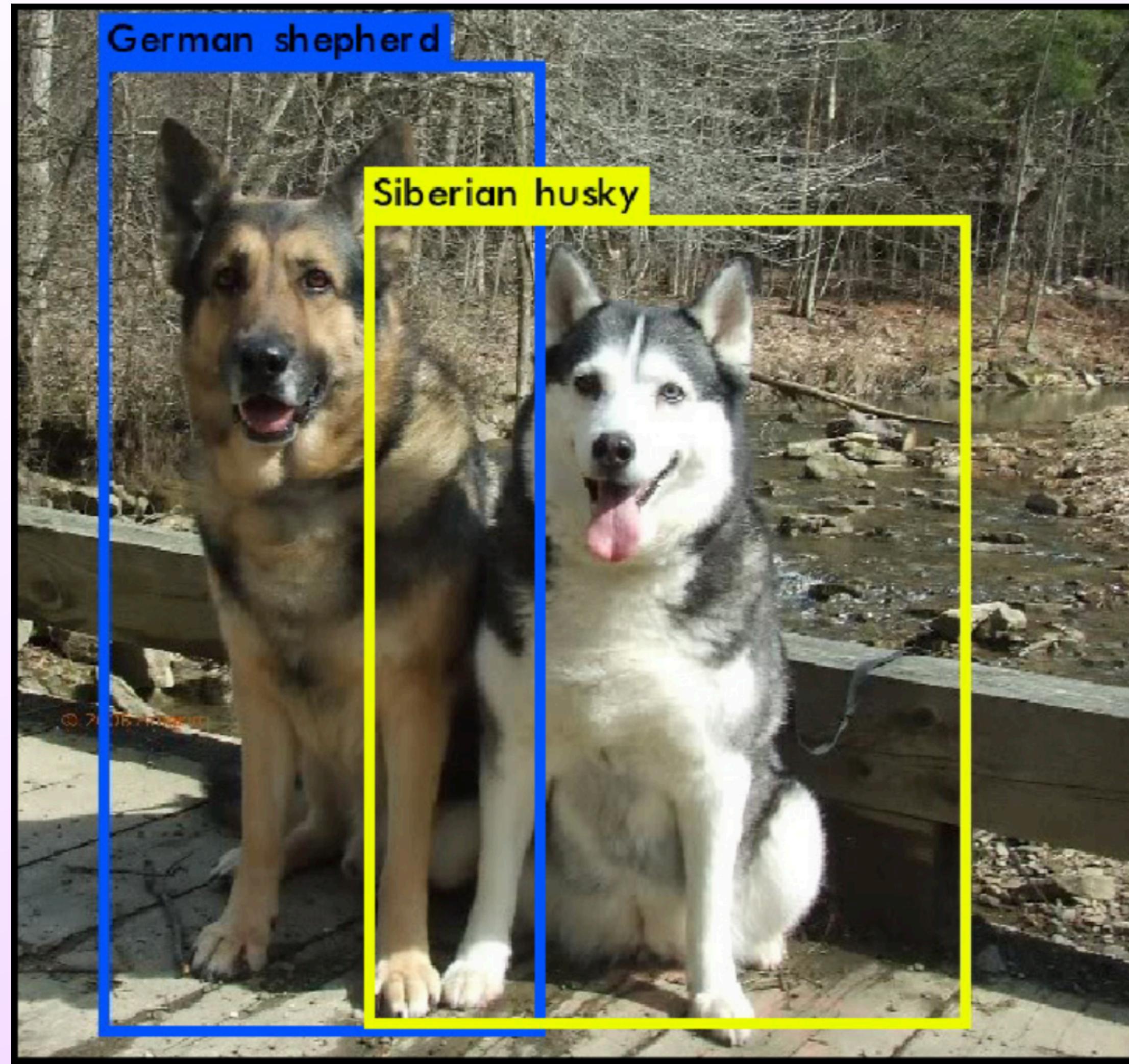
Detection + Classification Error



Tulipa
gesneriana

Classification Error

YOLO9000 Results



YOLO9000 Results



diaper	0.0
horizontal bar	0.0
rubber eraser	0.0
sunglasses	0.0
swimming trunks	0.0
...	
red panda	50.7
fox	52.1
koala bear	54.3
tiger	61.0
armadillo	61.7

YOLO9000 Results

- 19.7 mAP over 9000 classes
- 16 mAP - new objects that are not in COCO
- Best performing classes - animals
- Worst Performing classes - apparel's likes sunglasses

YOLOv3: An Incremental Improvement

Joseph Redmon

Universit

Ali Farhadi

Washington

We present YOLOv3, an incremental improvement to the YOLO family of neural networks for object detection. We show that YOLOv3 is quite fast (~30ms) and accurate (~33.0 mAP) compared to RetinaNet, similar performance to R-FCN, and always faster than SSD and DSSD. All the code is online at <https://pjreddie.com/yolo/>.

Abstract

Thank You!

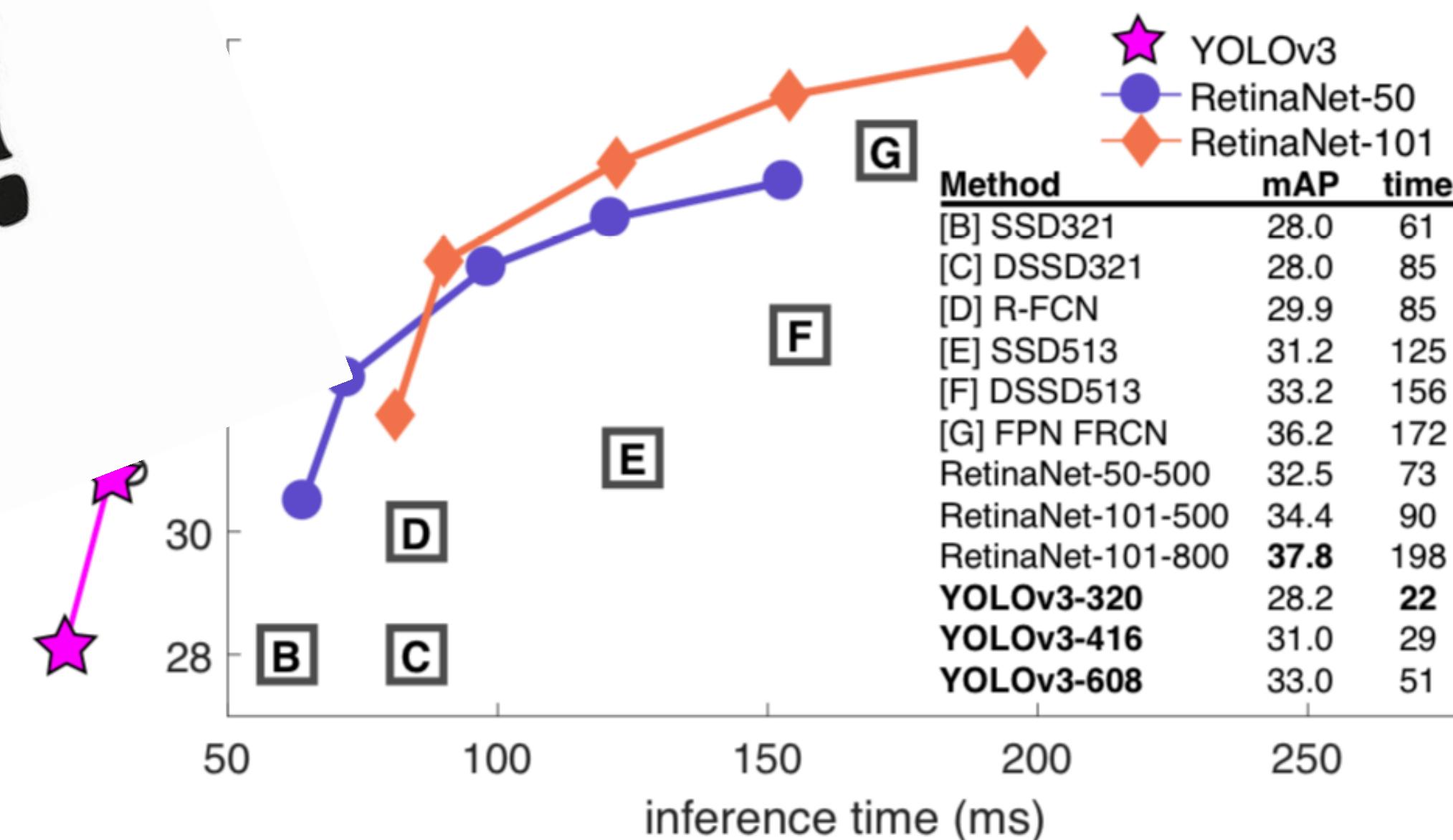


Figure 1. We adapt this figure from the Focal Loss paper [9]. YOLOv3 runs significantly faster than other detection methods with comparable performance. Times from either an M40 or Titan X, they are basically the same GPU.

1. Introduction