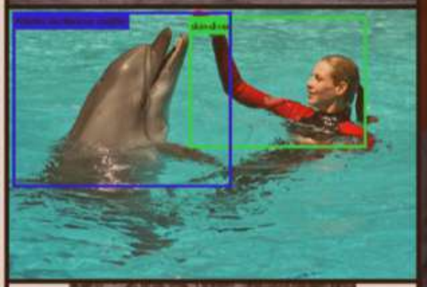
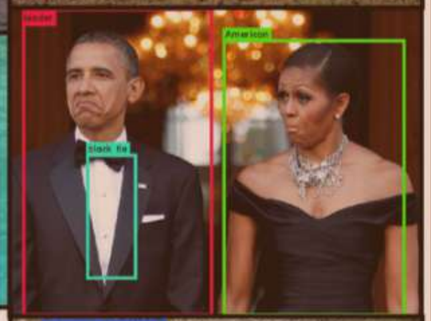
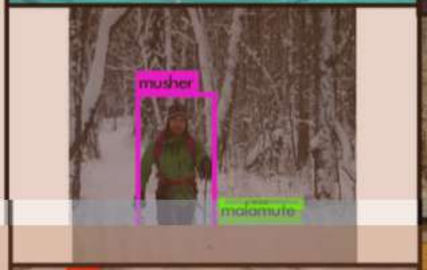
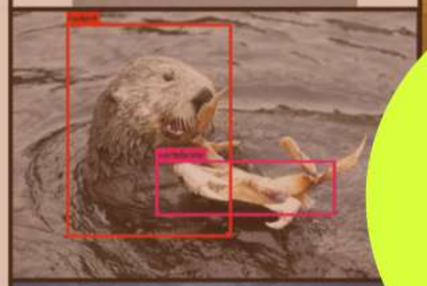
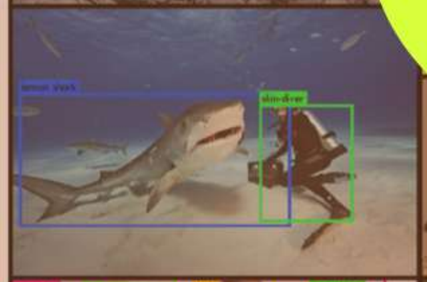
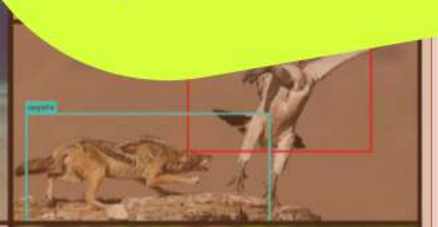
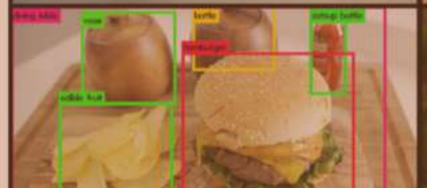
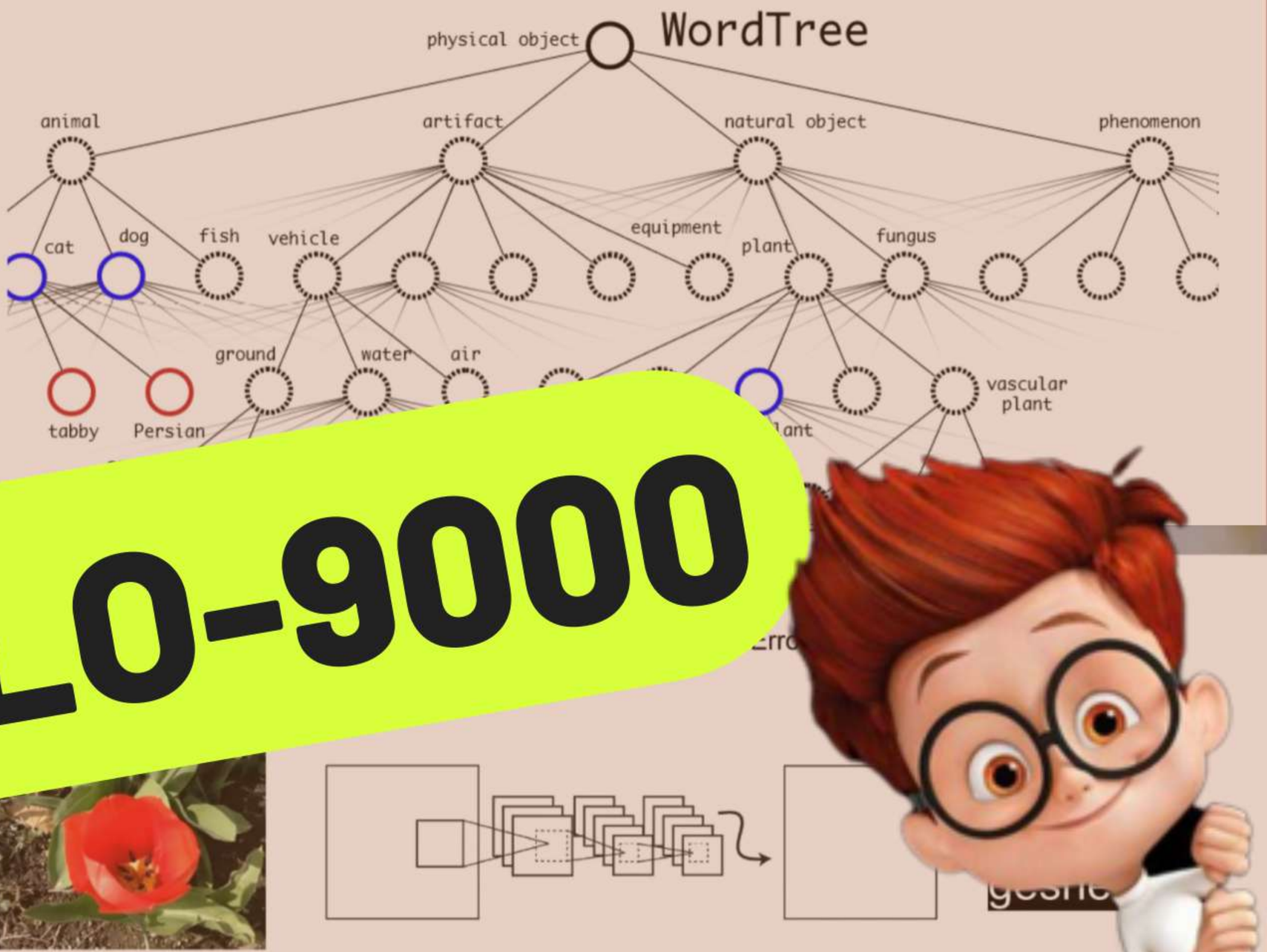
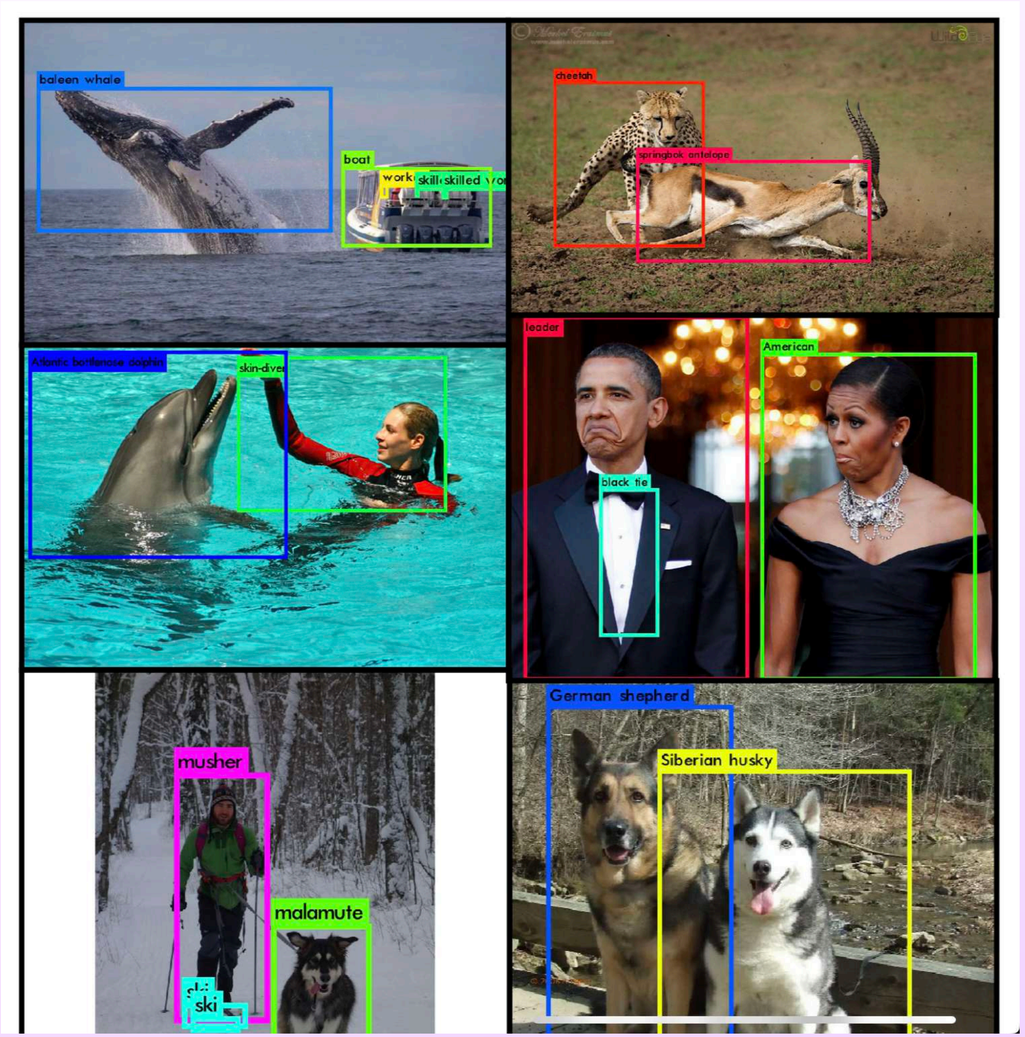




musher
malamute

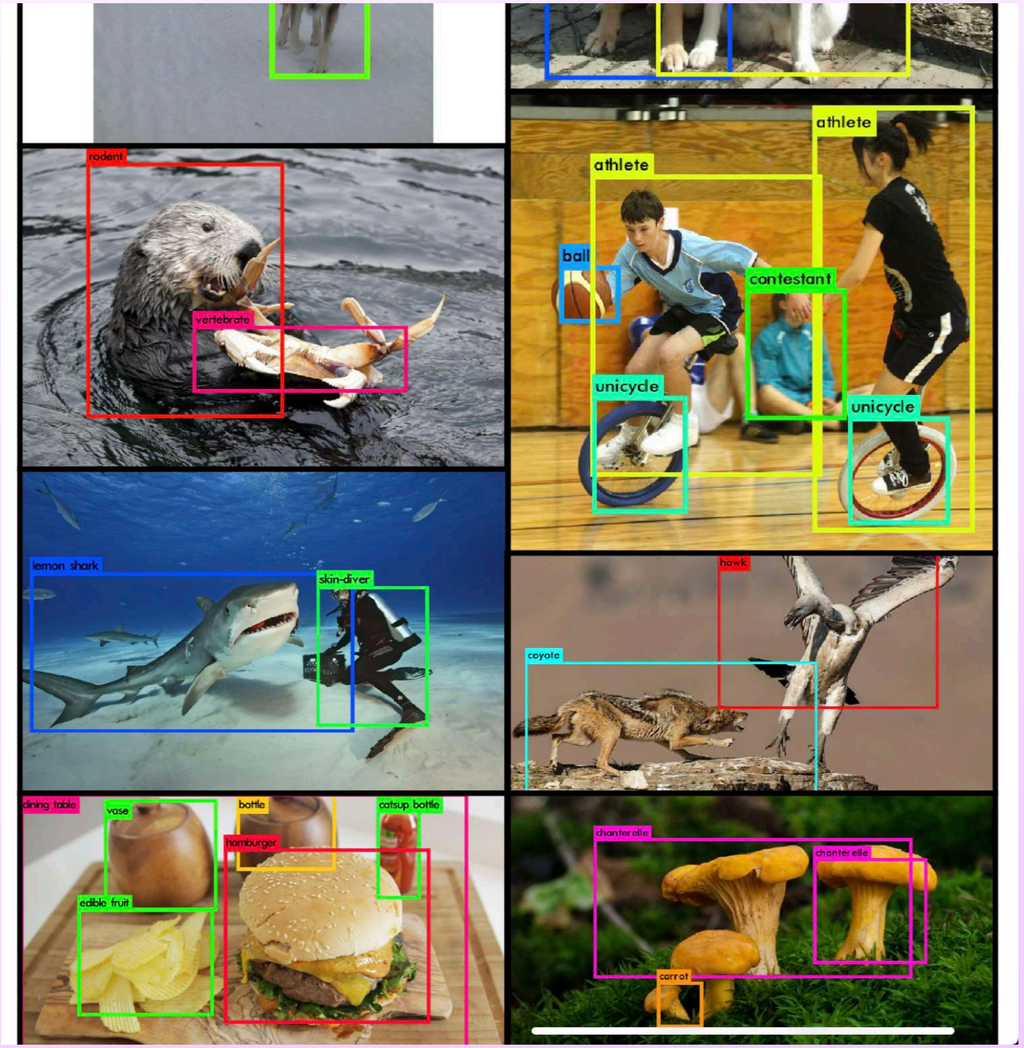
German shepherd
Siberian husky






WordTree
physical object
animal
artifact
natural object
phenomenon
cat
dog
fish
vehicle
equipment
plant
fungus
ground
water
air
vascular plant
tabby
Persian
Classification Error

YOLO-9000

YOLO-9000


baleen whale
boat
cheetah
springbok antelope
Atlantic bottlenose dolphin
skin-diver
leader
black tie
American
musher
malamute
ski
German shepherd
Siberian husky


rodent
vertebrate
athlete
athlete
ball
contestant
unicycle
unicycle
lemon shark
skin-diver
hawk
coyote
dining table
vase
bottle
catsup bottle
hamburger
edible fruit
chanterelle
chanterelle
carrot

Object Detection Datasets

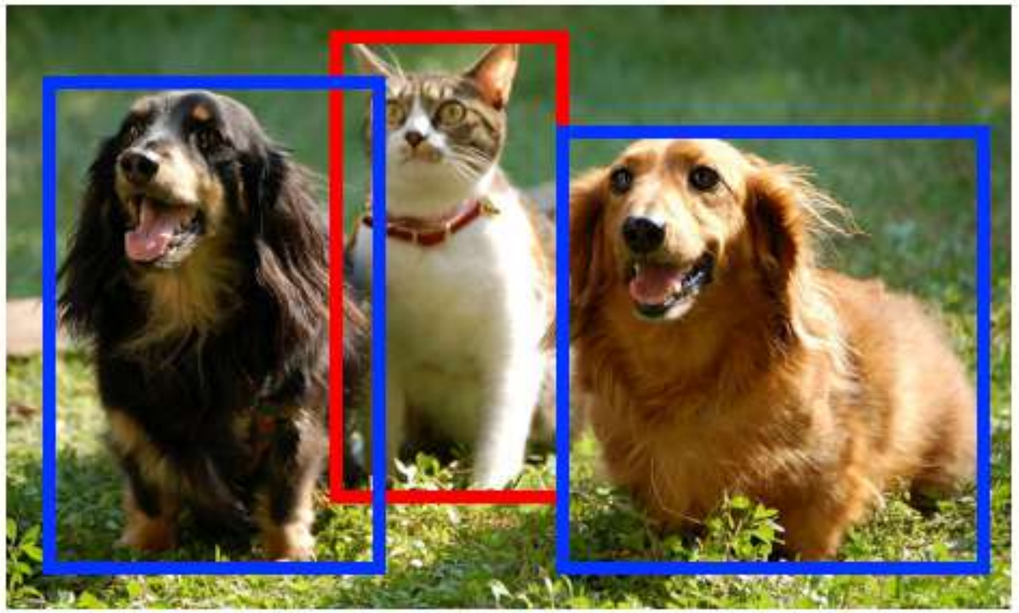
Dataset	COCO 2017	PASCAL VOC (07 + 12)
Number of categories	80	20
Number of training pictures	117,264	16,551
Number of testing pictures	5000	4952
Total sample boxes	902,435	52,090
Total sample boxes / total number of images	7.4	2.4

Annotations

Classification

CAT

Object Detection

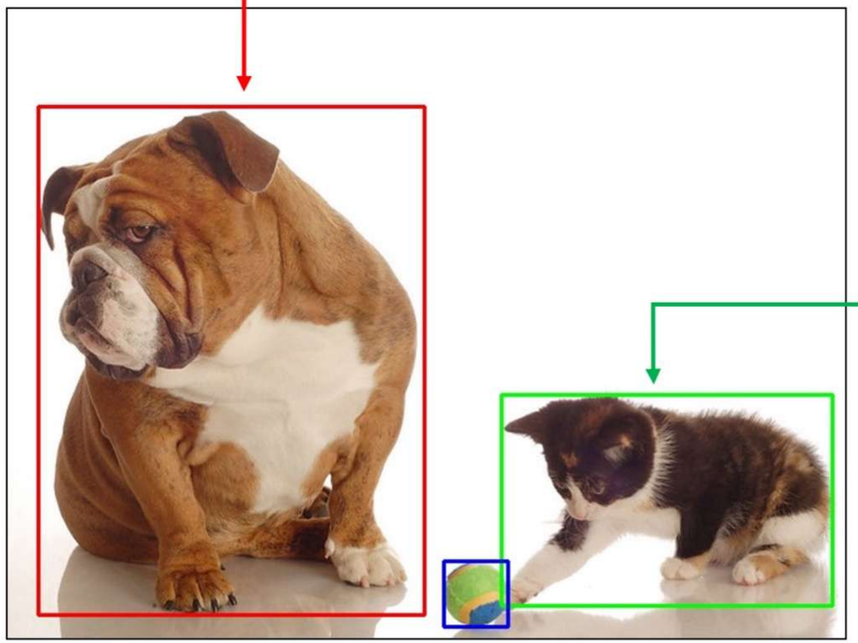


CAT, DOG

Annotations

[23, 74, 295, 388]

[x_min, y_min, width, height]



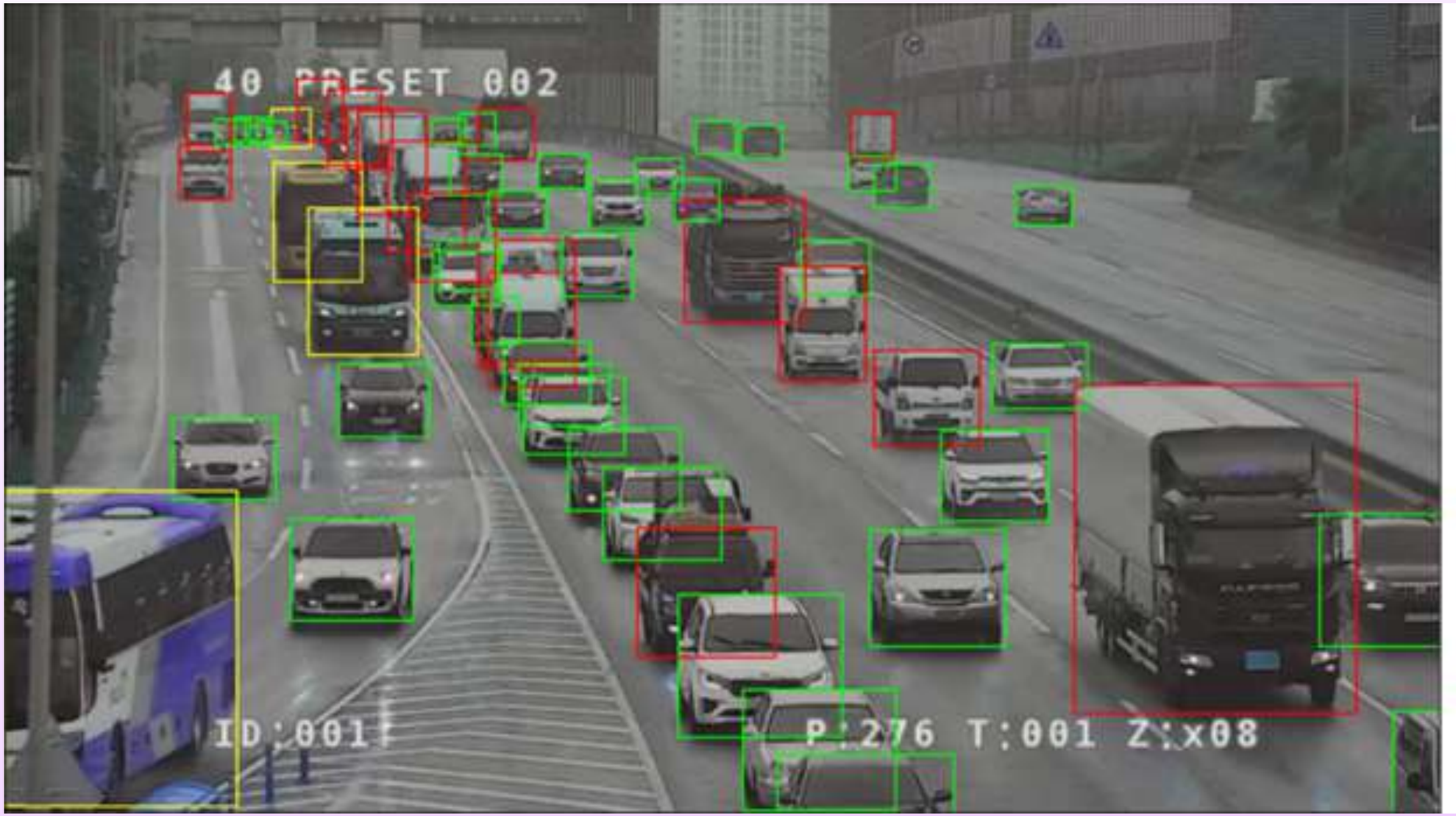
[377, 294, 252, 161]

[x_min, y_min, width, height]

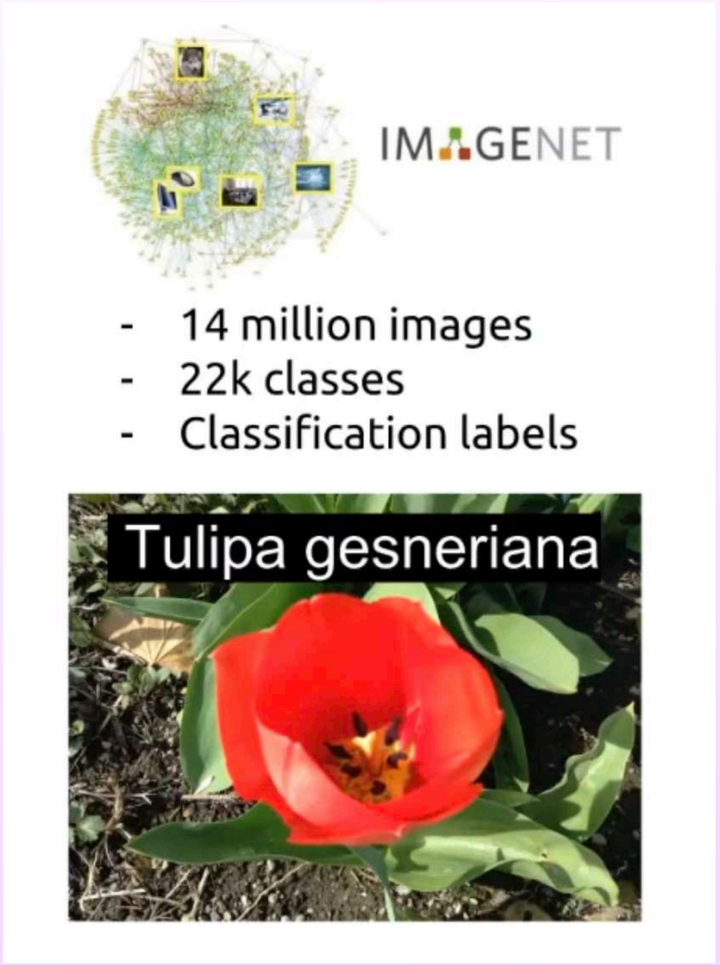
[333, 421, 49, 49]

[x_min, y_min, width, height]

Annotations



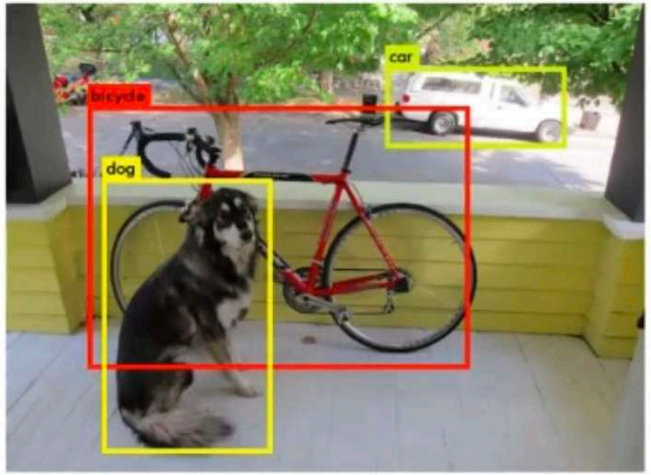
Imagenet Dataset



How to combine these?

COCO
Common Objects in Context

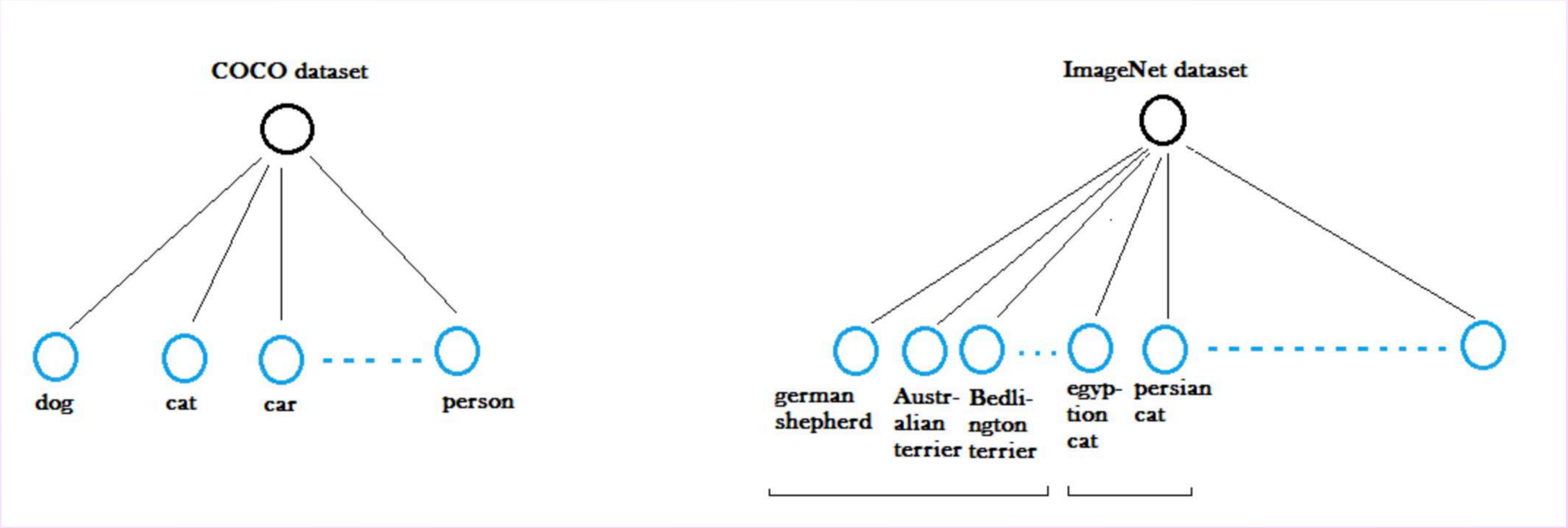
- 100k images
- 80 classes
- Detection labels



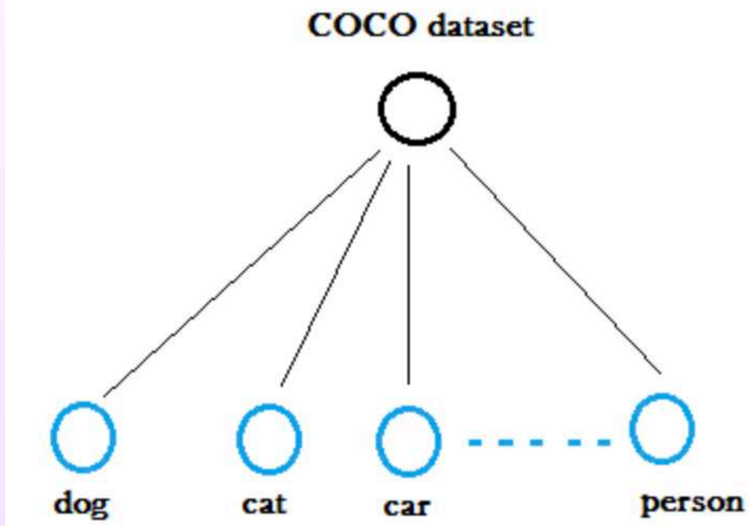
IMAGENET

- 14 million images
- 22k classes
- Classification labels

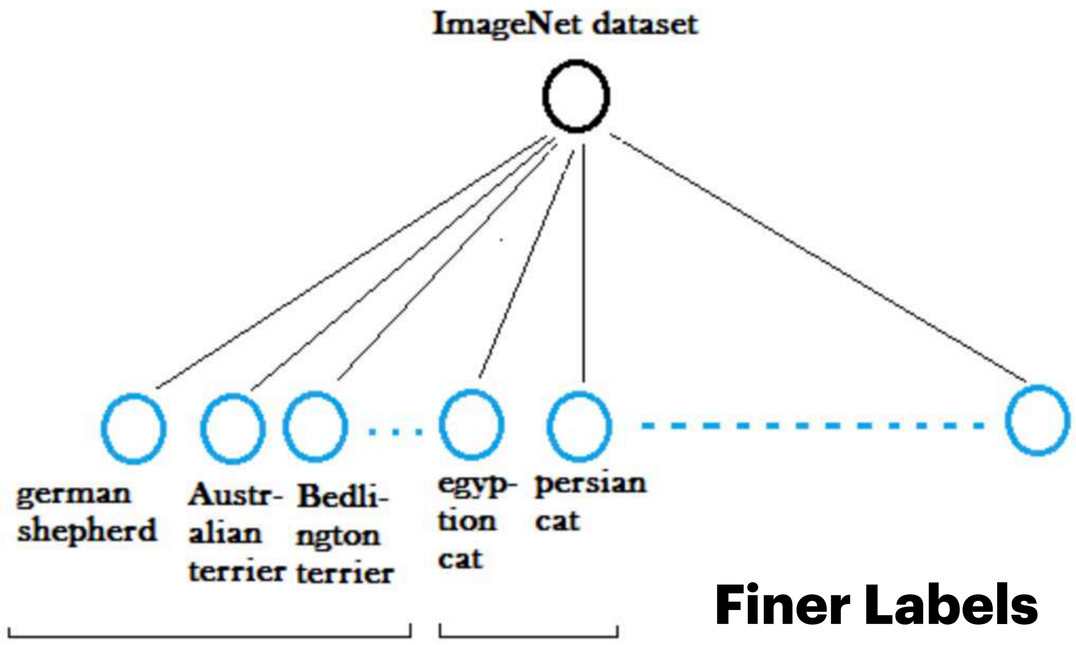
How to combine these?



How to combine these?

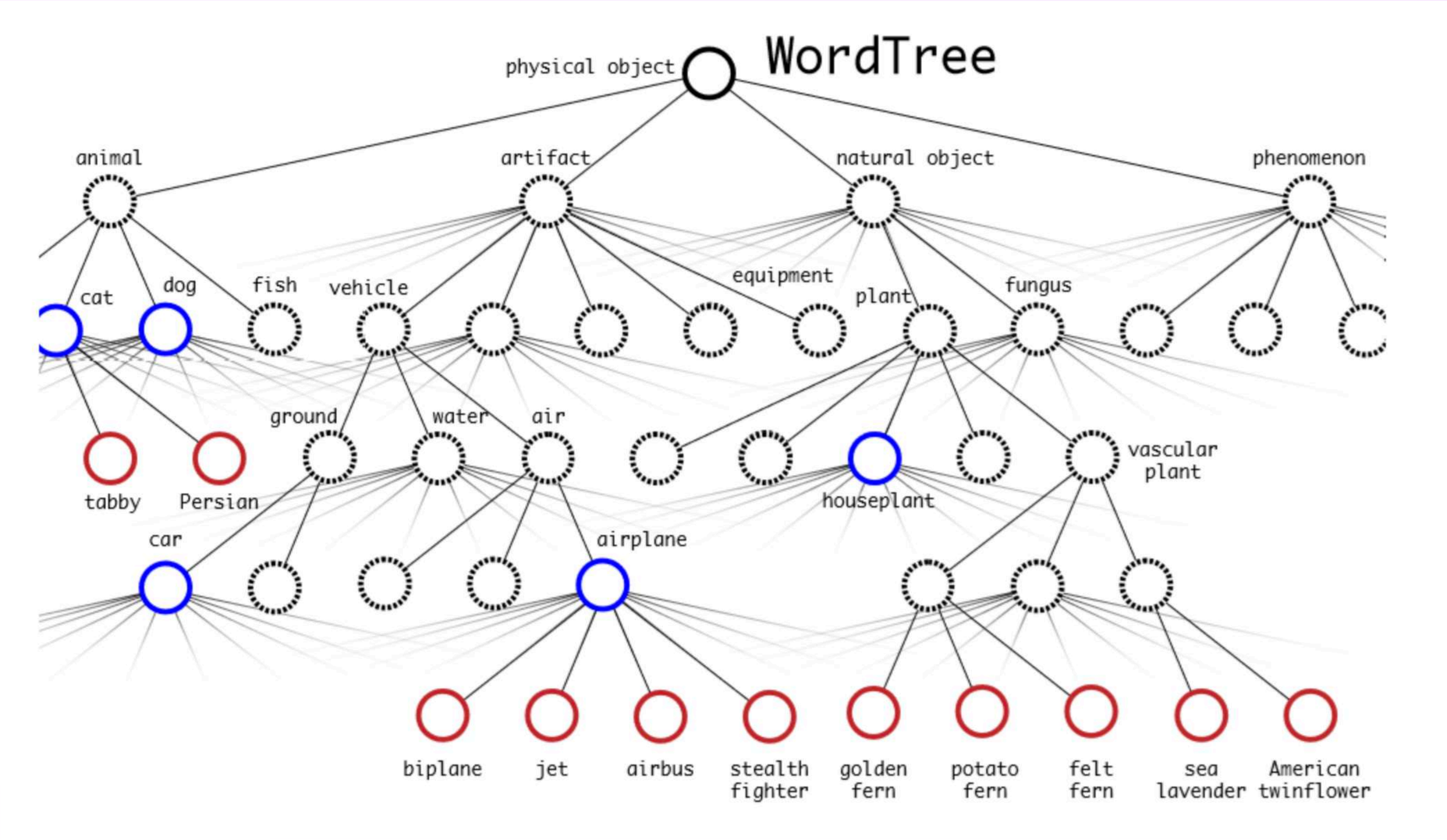


Coarse Labels

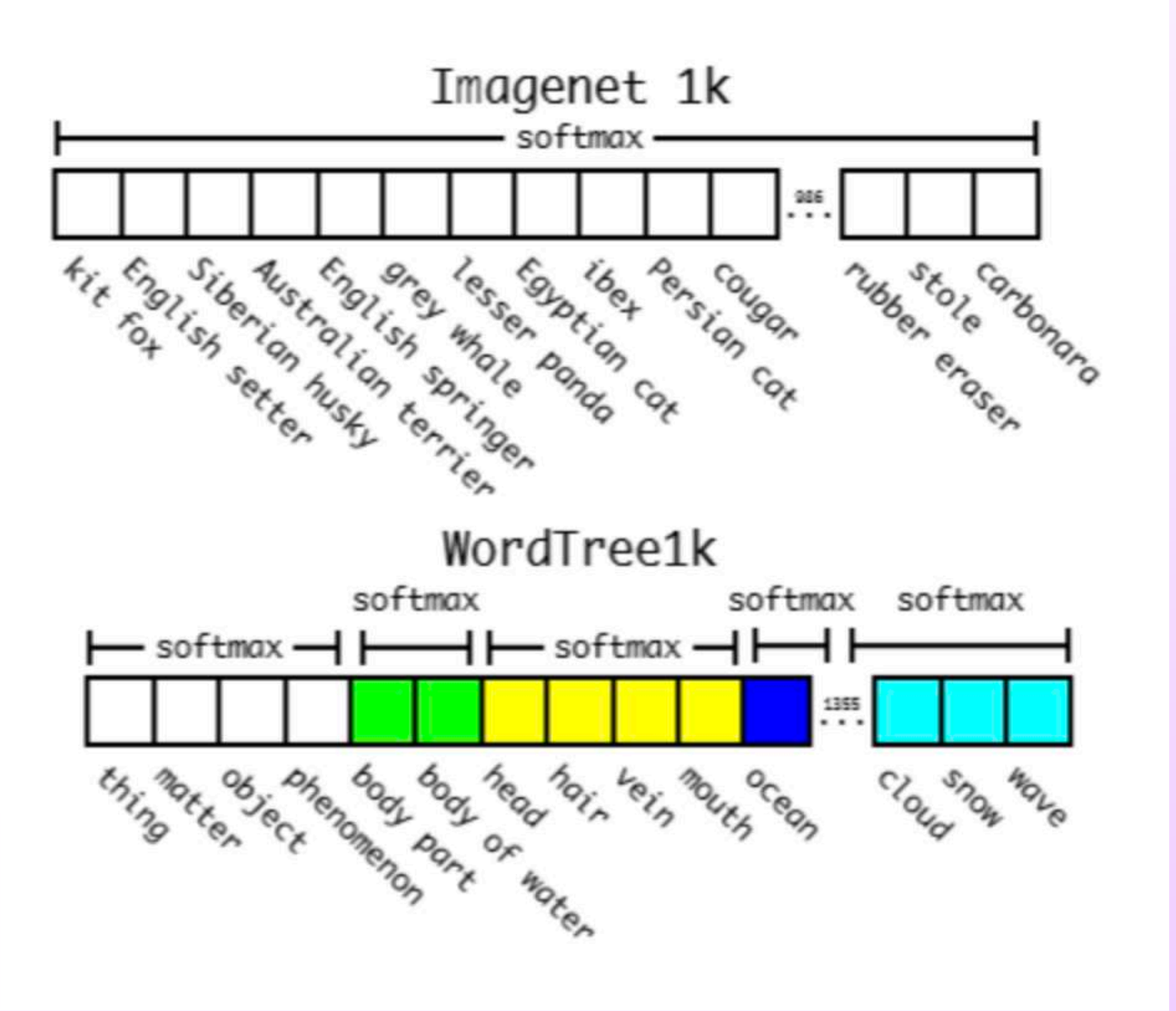


Finer Labels

WordTree 1K


WordTree
physical object
animal
artifact
natural object
phenomenon
cat
dog
fish
vehicle
equipment
plant
fungus
ground
water
air
vascular plant
tabby
Persian
houseplant
car
airplane
biplane
jet
airbus
stealth fighter
golden fern
potato fern
felt fern
sea lavender
American twinflower

Darknet-19 Training

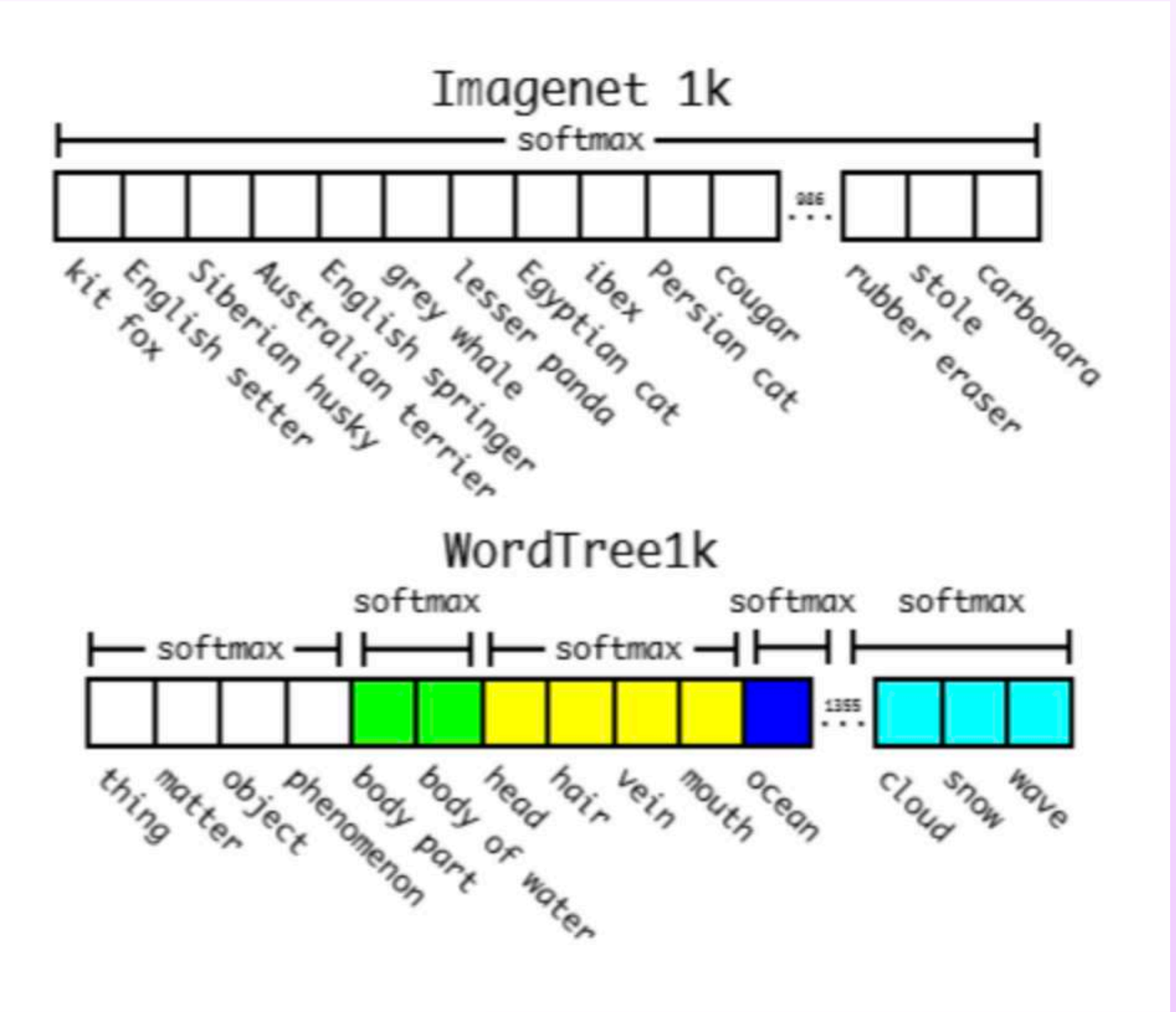


Conditional Probabilities

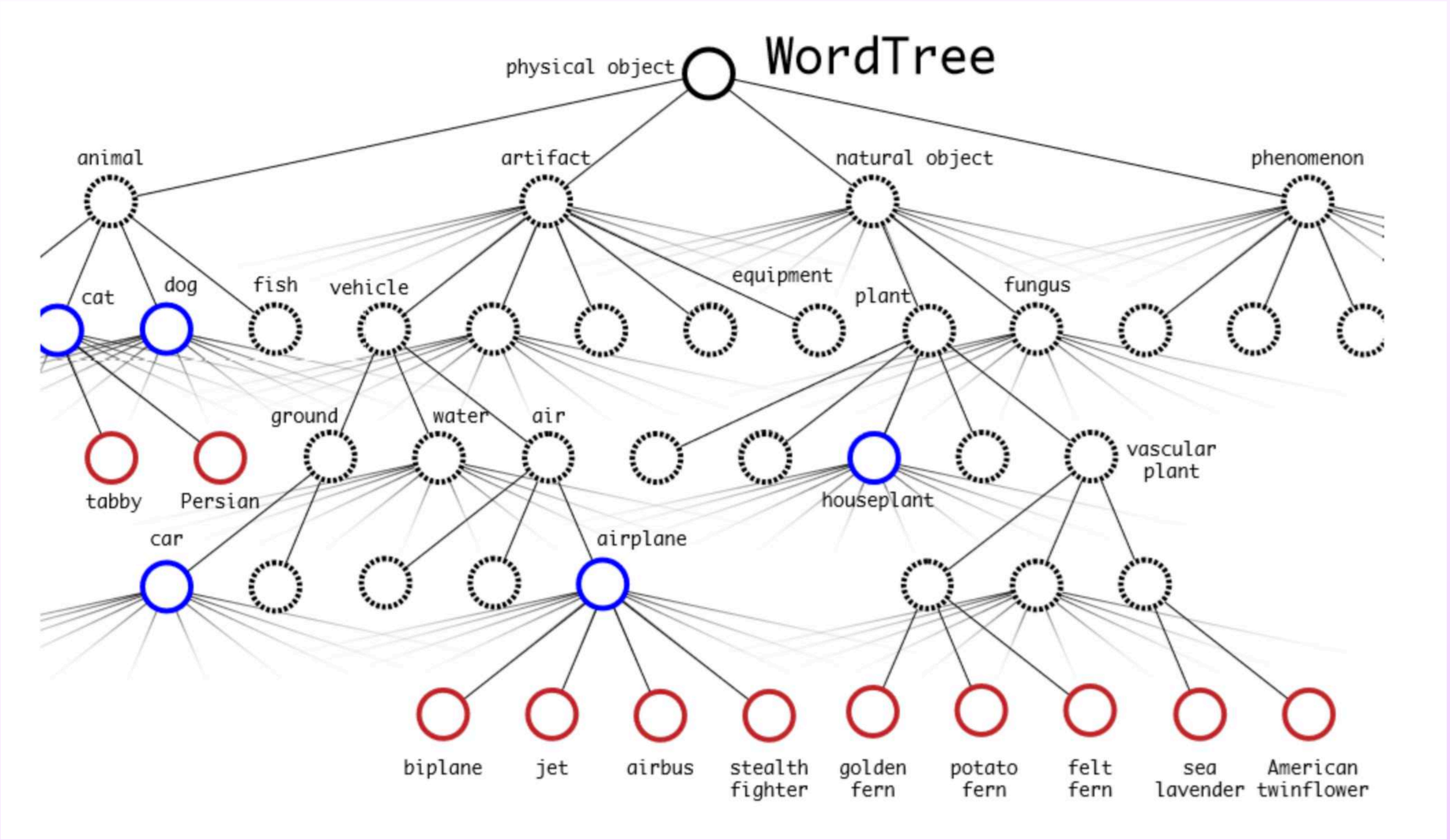
$$Pr(\text{Norfolk terrier}) = Pr(\text{Norfolk terrier}|\text{terrier})$$
$$*Pr(\text{terrier}|\text{hunting dog})$$
$$*\ldots*$$
$$*Pr(\text{mammal}|Pr(\text{animal})$$
$$*Pr(\text{animal}|\text{physical object})$$

Darknet-19 Training

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

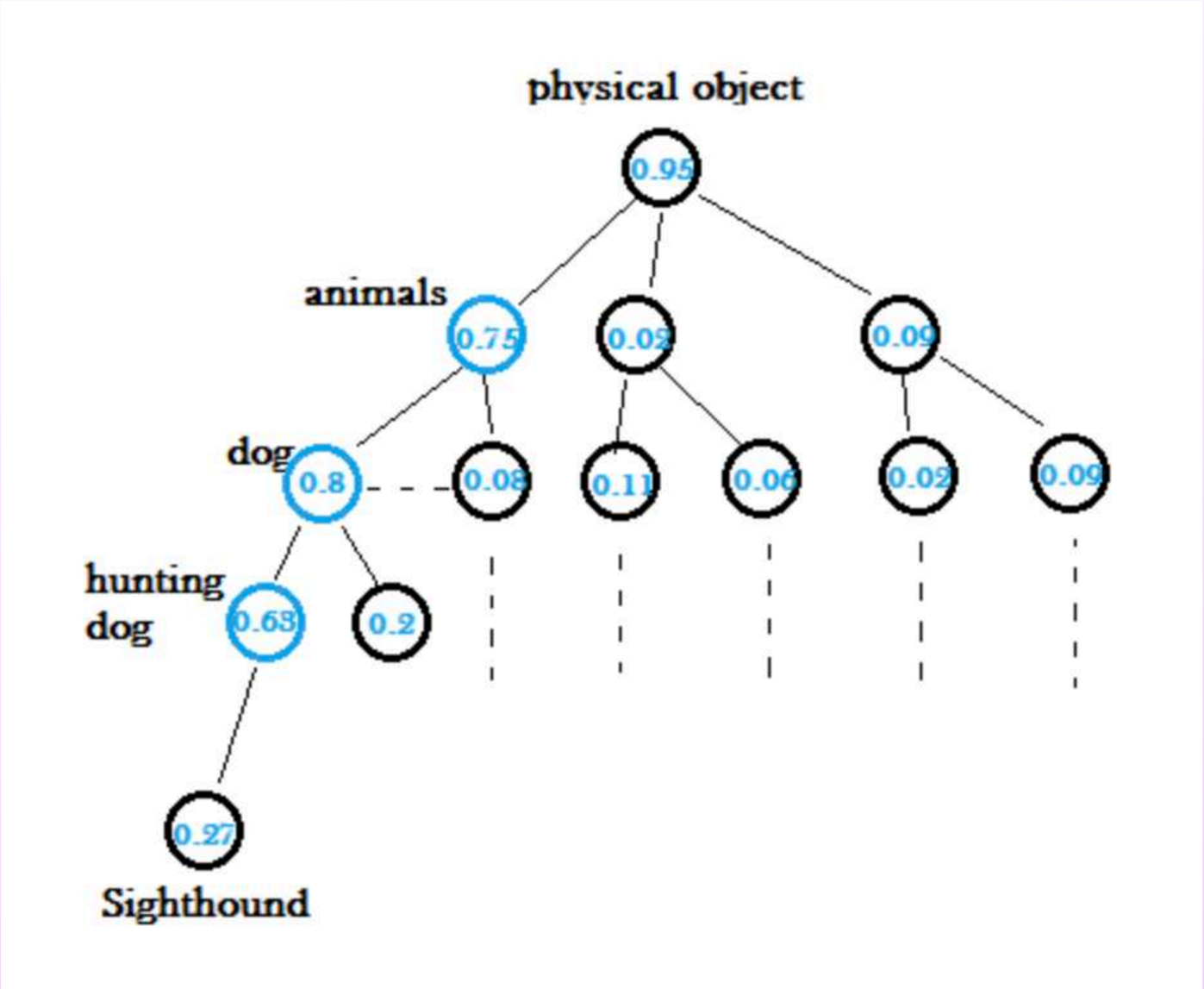


Darknet-19 Training



What is the final class prediction?

Darknet-19 Training

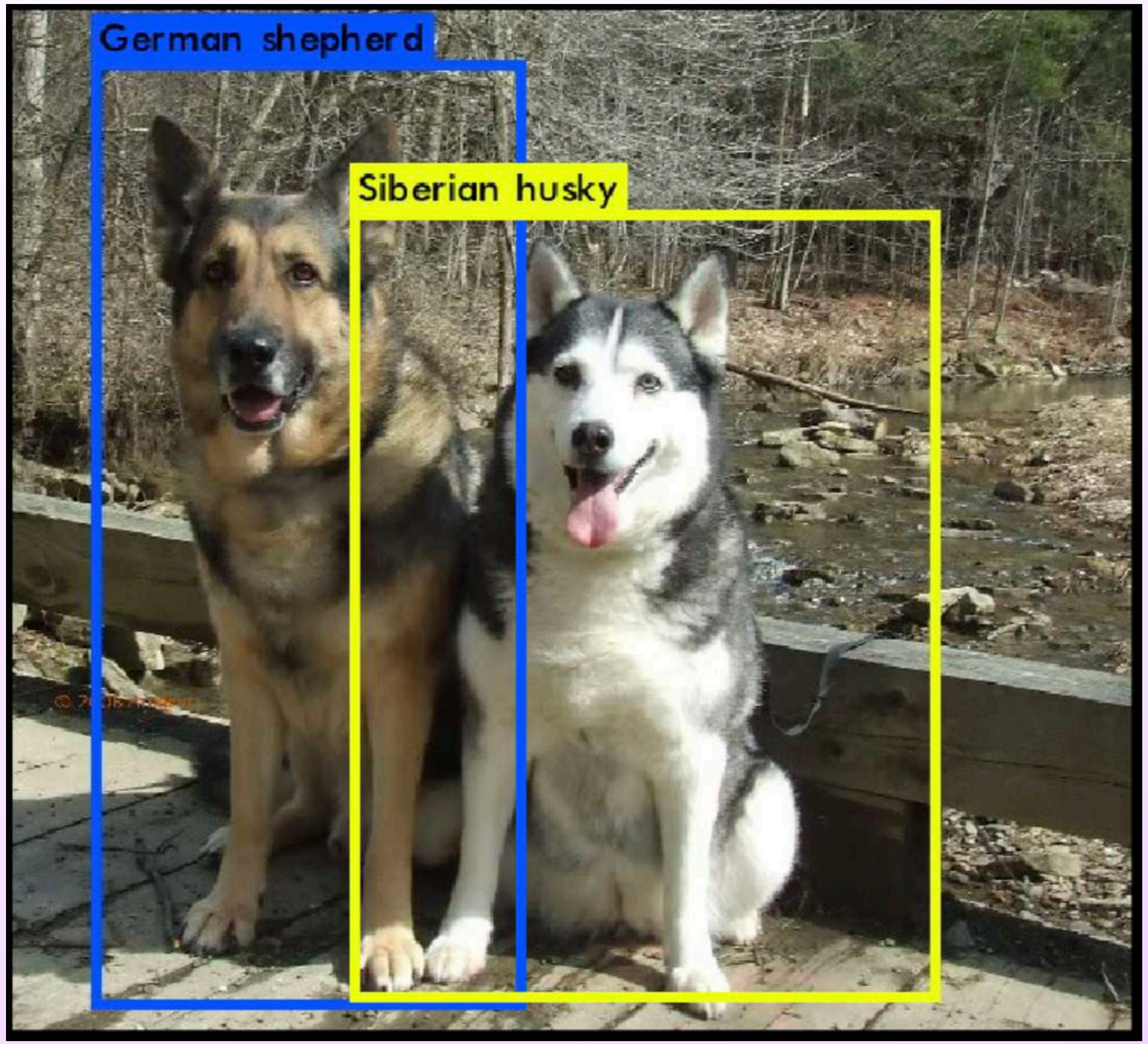


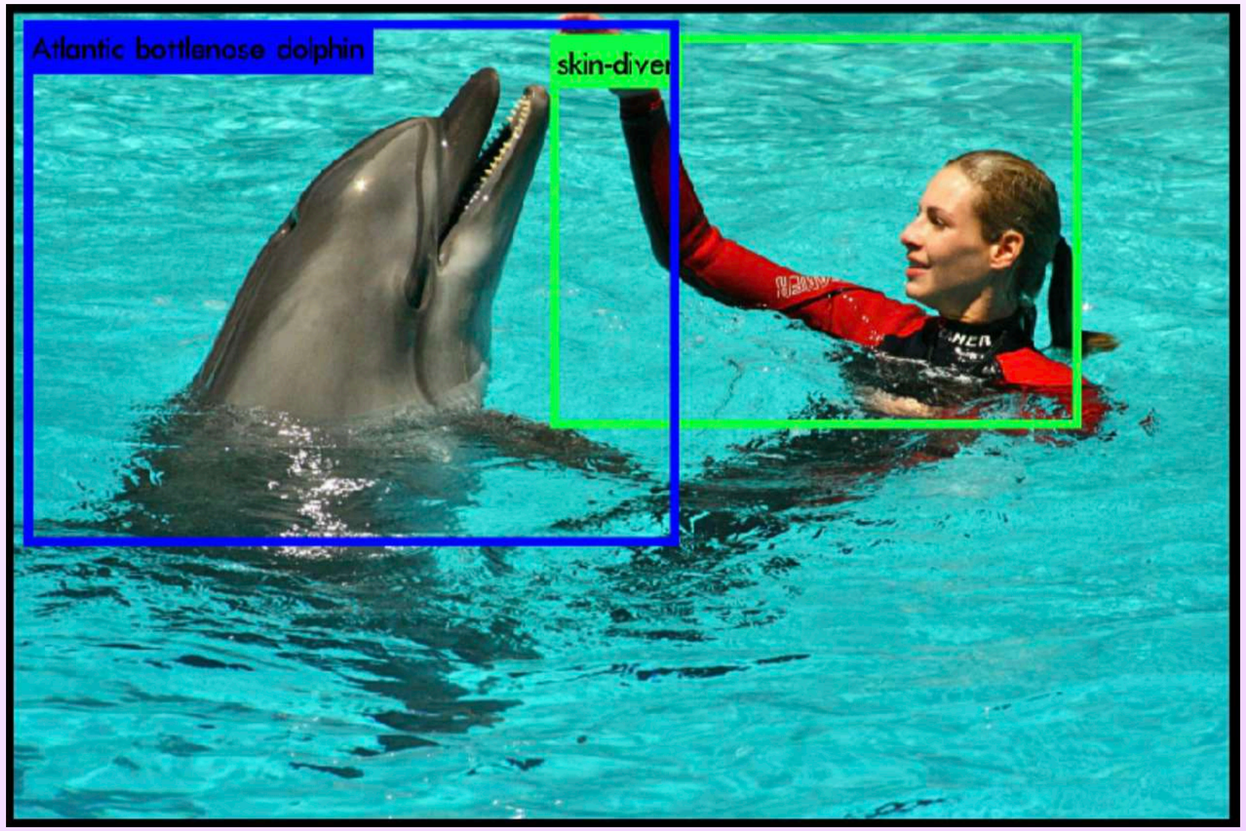
YOLO9000 Training

- 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

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

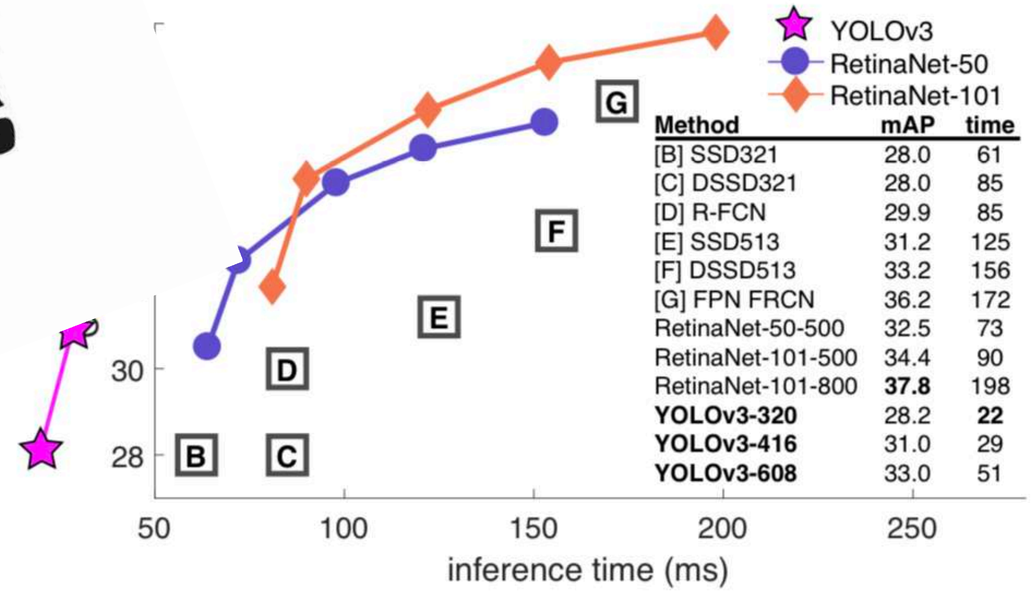
Joseph Redmon Ali Farhadi
Univers' 'shington

Abst

We presen
of lit
w
as
at t
gooa
pared
mance
https: *udie.com/yolo/.*

ok
is quite
a Titan X, com-
RetinaNet, similar perfor-
always, all the code is online at

Thank You!



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

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