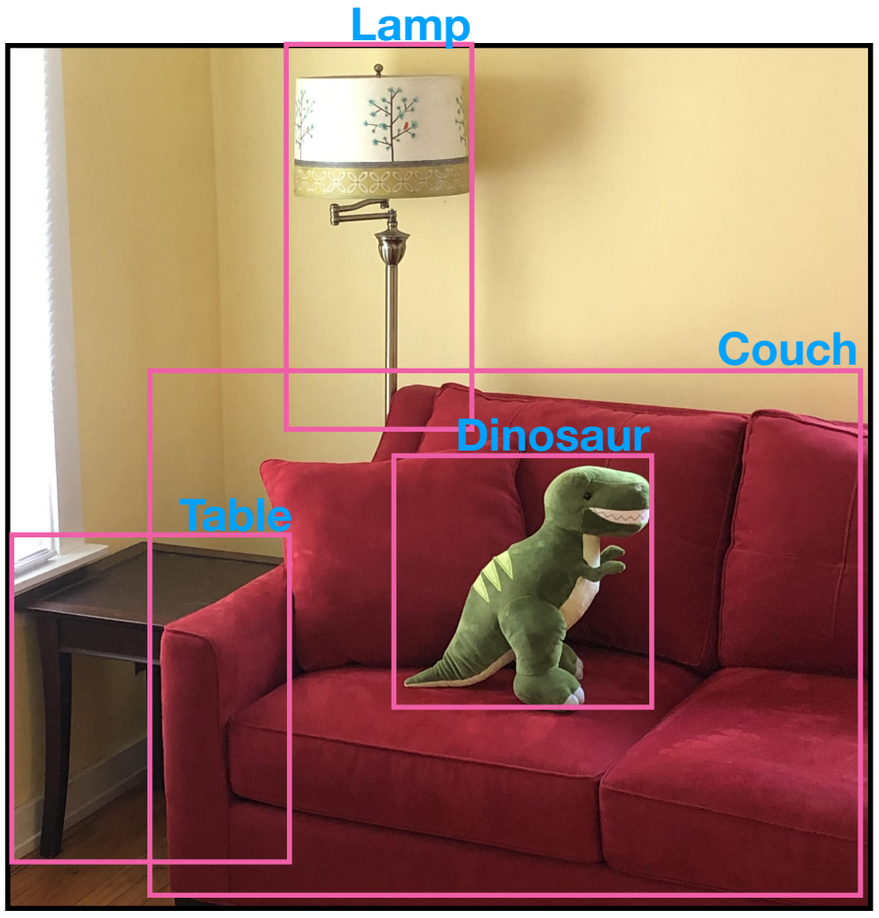
YOLO Trained on Multiple Data Sets



Object Detection Training Data Sets:

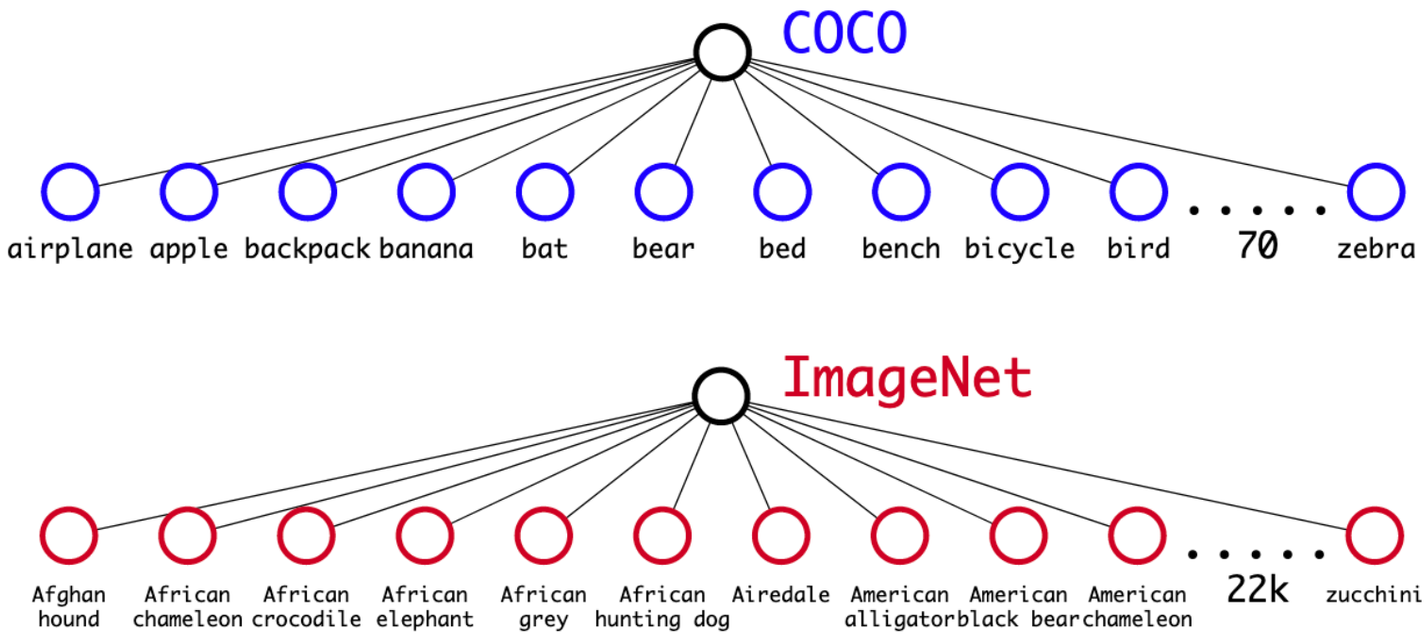
- Need many labelled images:
  - Bounding boxes and object type labels for each object in image
- YOLO is trained on the COCO data set:
  - ~200,000 labeled images
  - 80 object categories: “dog”, “boat”, …

Object Classification Training Data Sets:

- Label object type, not necessarily bounding boxes for objects
- Consider ImageNet:
  - ~14 million labeled images
  - ~22,000 object categories: “Norfolk terrier”, “Yorkshire terrier”, …

YOLO puts these together

YOLO Trained on Multiple Data Sets

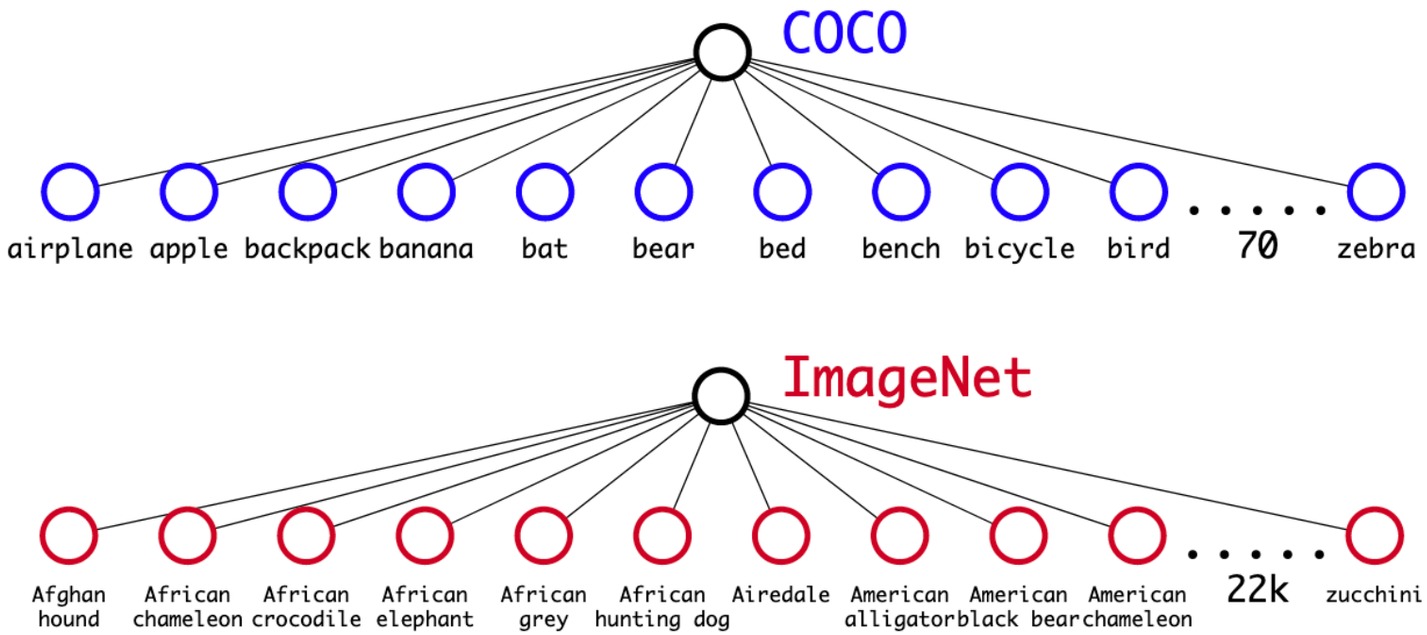


Connecting the data sets with conditional probabilities:

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

Figure from Redmon and Farhadi. “YOLO9000: Better, Faster, Stronger” (2009)

YOLO Trained on Multiple Data Sets



Training on both data sets:

- Batches sample images from both COCO and ImageNet
  - 'we balance the data set by oversampling COCO so that ImageNet is only larger by a factor of 4:1.'

Figure from Redmon and Farhadi. "YOLO9000: Better, Faster, Stronger" (2009)

YOLO Trained on Multiple Data Sets

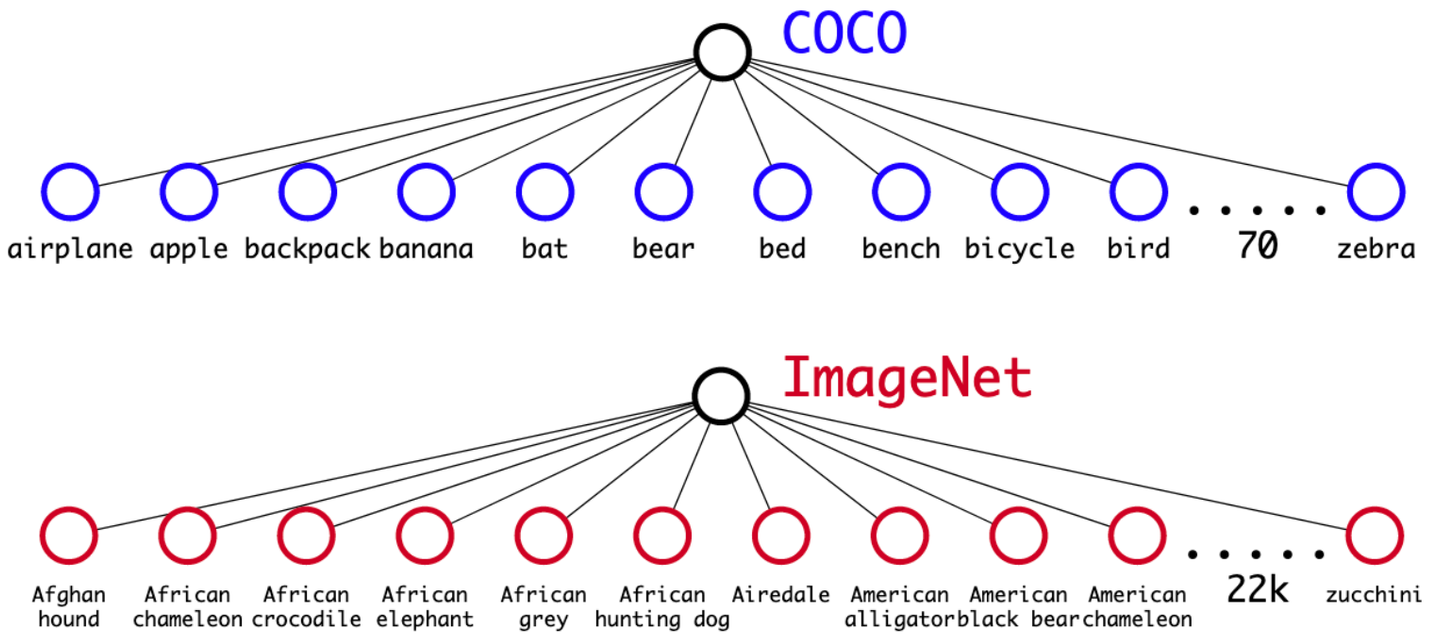
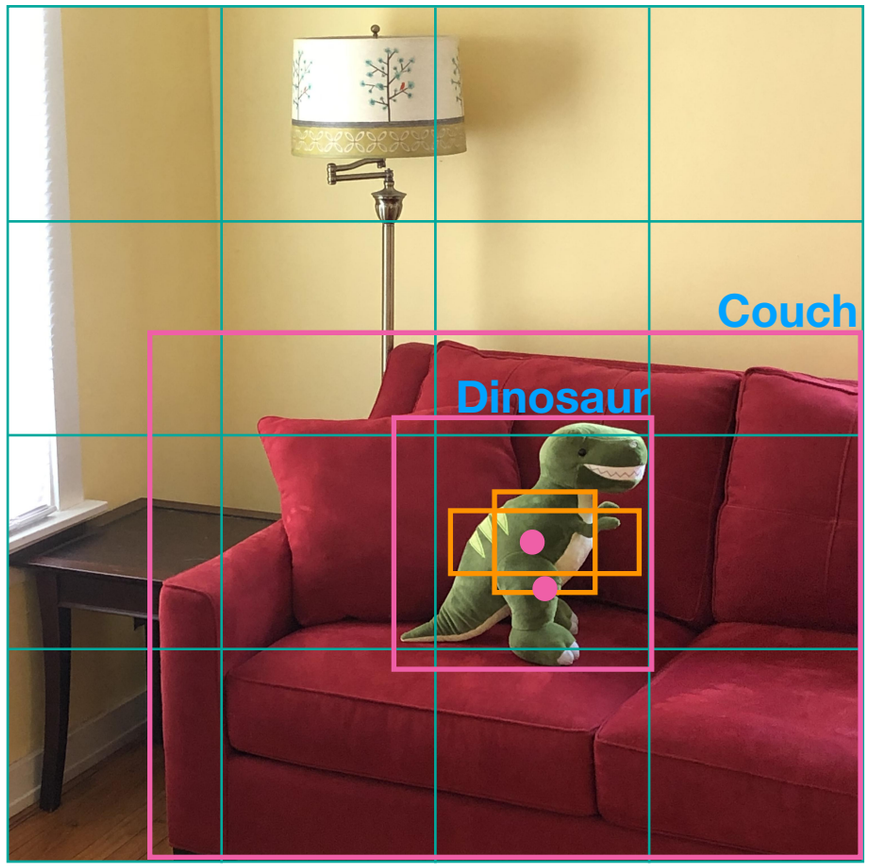


Figure from Redmon and Farhadi. “YOLO9000: Better, Faster, Stronger” (2009)

Training on both data sets:

- Batches sample images from both COCO and ImageNet
  - ‘we balance the data set by oversampling COCO so that ImageNet is only larger by a factor of 4:1.’
- Backpropagation updates only as appropriate to sampled image
  - ‘When our network sees a detection image we backpropagate loss as normal.’
  - ‘For classification loss, we only backpropagate loss at or above the corresponding level of the label. For example, if the label is “dog” we do[n’t] assign any error to predictions further down in the tree, “German Shepherd” versus “Golden Retriever”, because we do not have that information.’

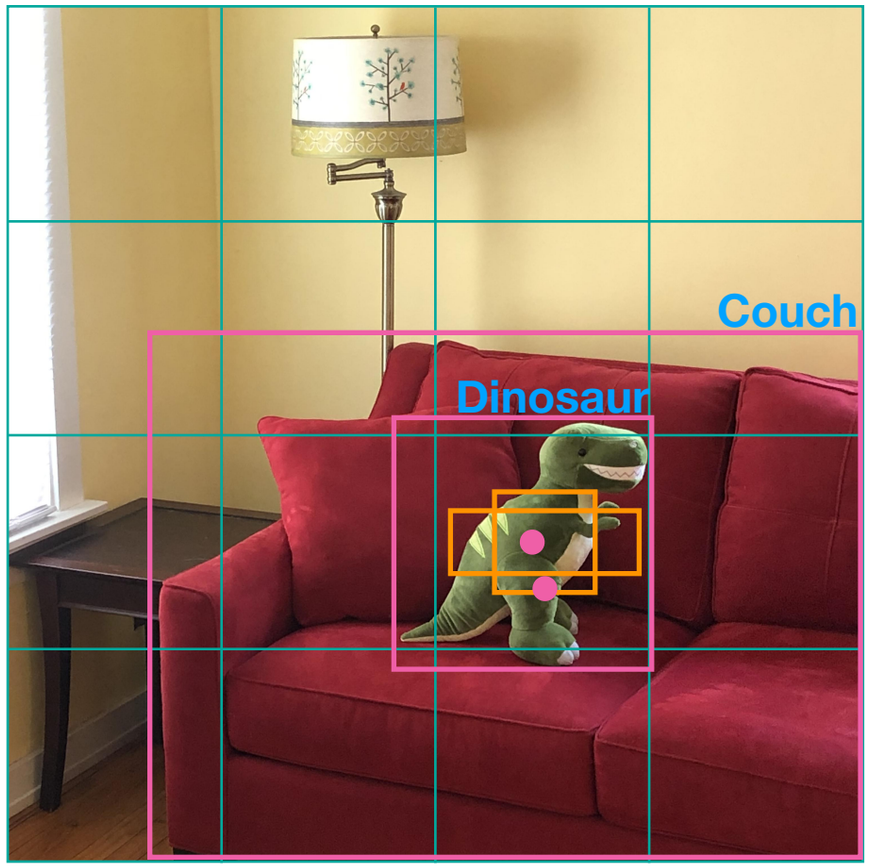
Related Issue: Many Sigmoids, not One Softmax



Multi-class classification:

- In general, image has one class (dinosaur, lamp, table, couch, …)
- Softmax activation: class probabilities sum to 1
- In context of YOLO, each combination of cell and anchor box has one associated class

Related Issue: Many Sigmoids, not One Softmax



Multi-class classification:

- In general, image has one class (dinosaur, lamp, table, couch, …)
- Softmax activation: class probabilities sum to 1
- In context of YOLO, each combination of cell and anchor box has one associated class

Multi-label classification:

- Each image may have multiple associated classes
- Separate sigmoid activation for each class; class probabilities don’t need to sum to 1
- In context of YOLO, each combination of cell and anchor box may have more than one associated class