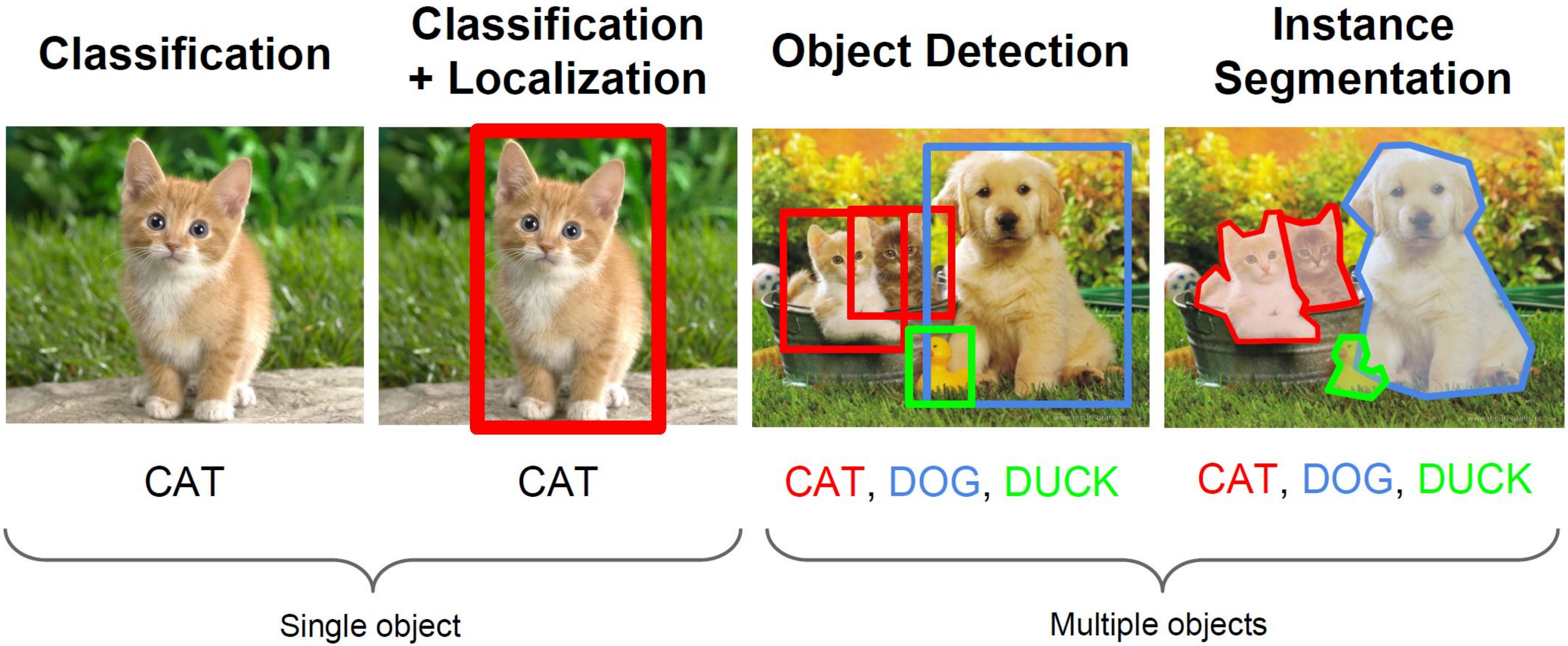
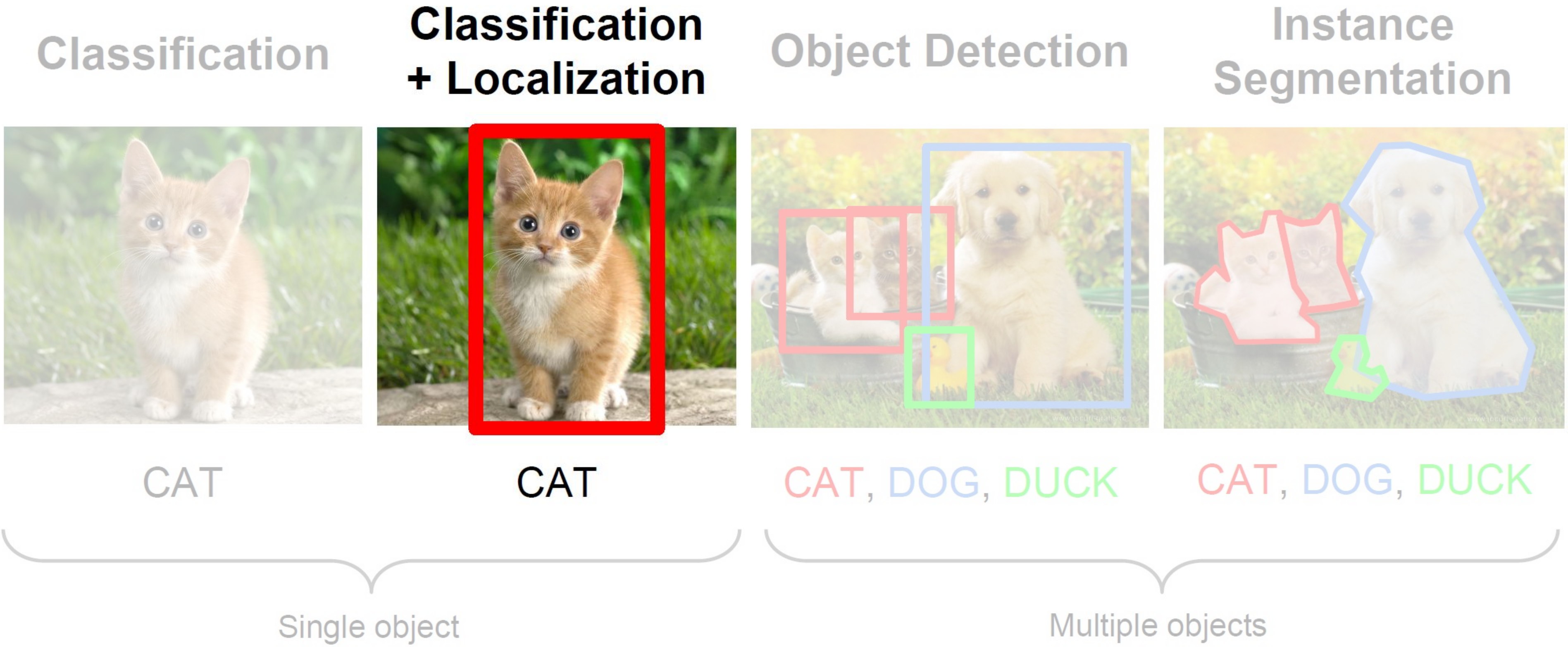
Introduction to Object Detection

Computer Vision Tasks



Source: CS231n Object detection http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Computer Vision Tasks



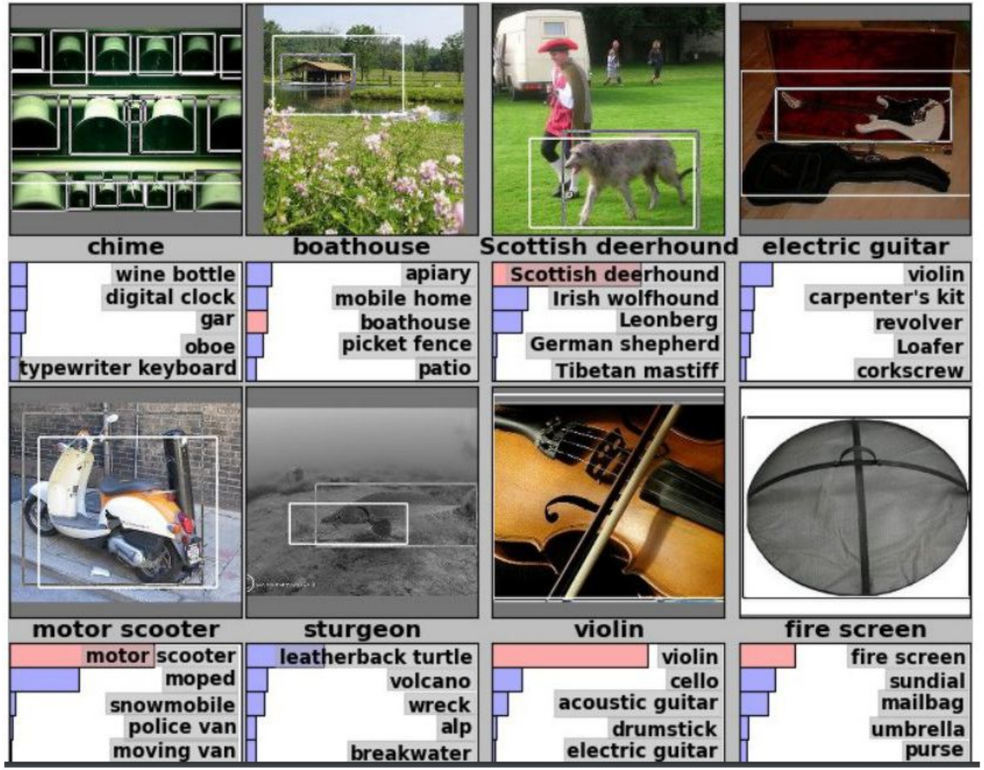
Source: CS231n Object detection http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Classification + Localization

- Classification:
  - Input: Image
  - Output: Class label
  - Loss: Cross entropy (Softmaxlog)
  - Evaluation metric: Accuracy
- Localization:
  - Input: Image
  - Output: Box in the image (x, y, w, h)
  - Loss: L2 Loss (Euclidean distance)
  - Evaluation metric: Intersection over Union
- Classification + Localization:
  - Input: Image
  - Output: Class label + box in the image
  - Loss: Sum of both losses

Classification + Localization: ImageNet Challenge

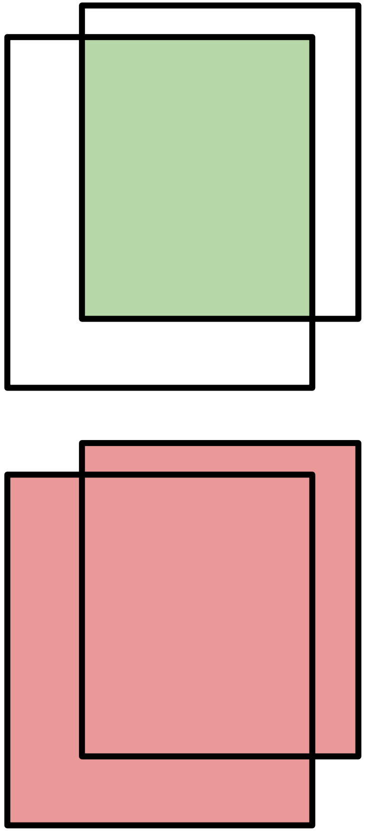
- Dataset
  - 1000 Classes.
  - Each image has 1 class with at least one bounding box.
  - ~800 Training images per class.
- Evaluation
  - Algorithm produces 5 (class + bounding box) guesses.
  - Example is correct if at least one of guess has correct class AND bounding box at least 50% intersection over union.



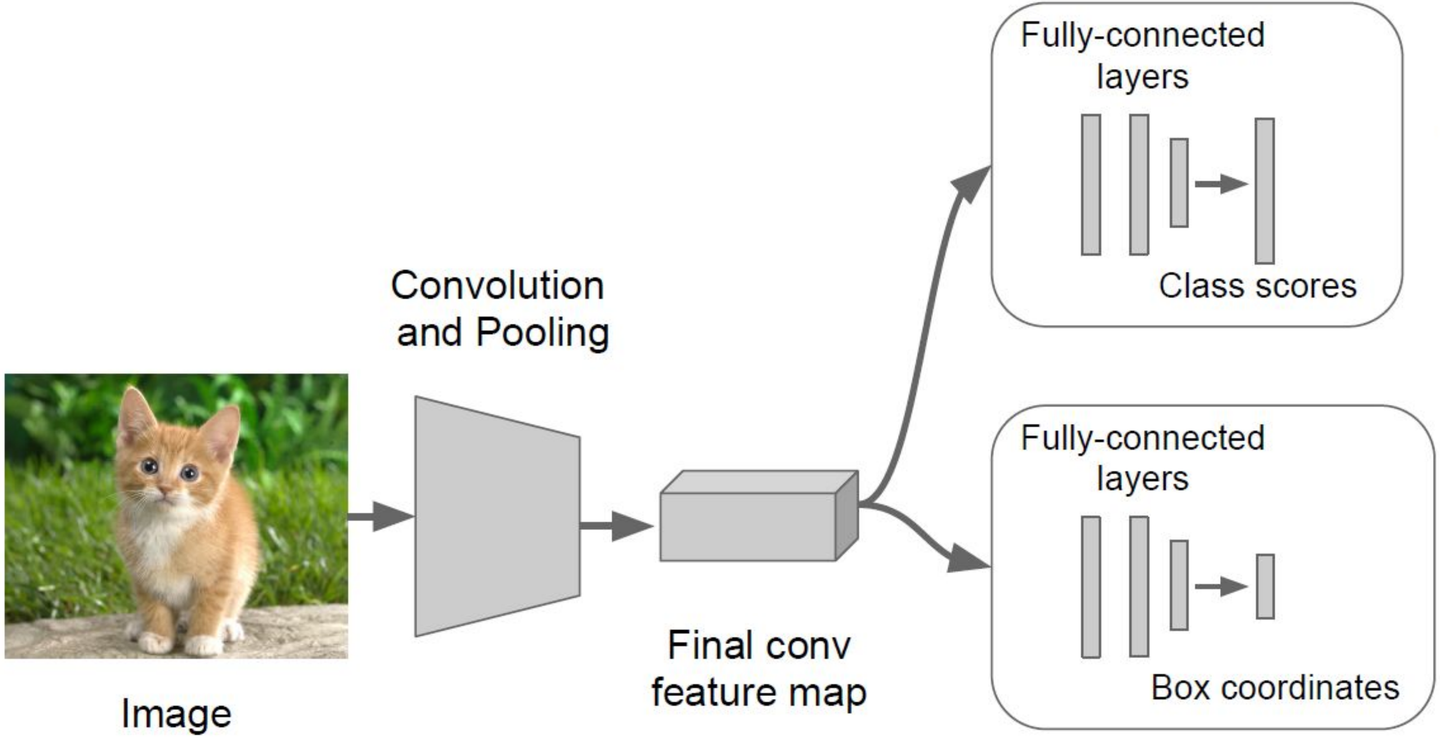
Intersection Over Union (IoU)

- Important measurement for object localization.
- Used in both training and evaluation.

$$\mathrm{IoU}(A,B) = \frac{\mathrm{Intersection}(A,B)}{\mathrm{Union}(A,B)}$$



Classification + Localization: Model



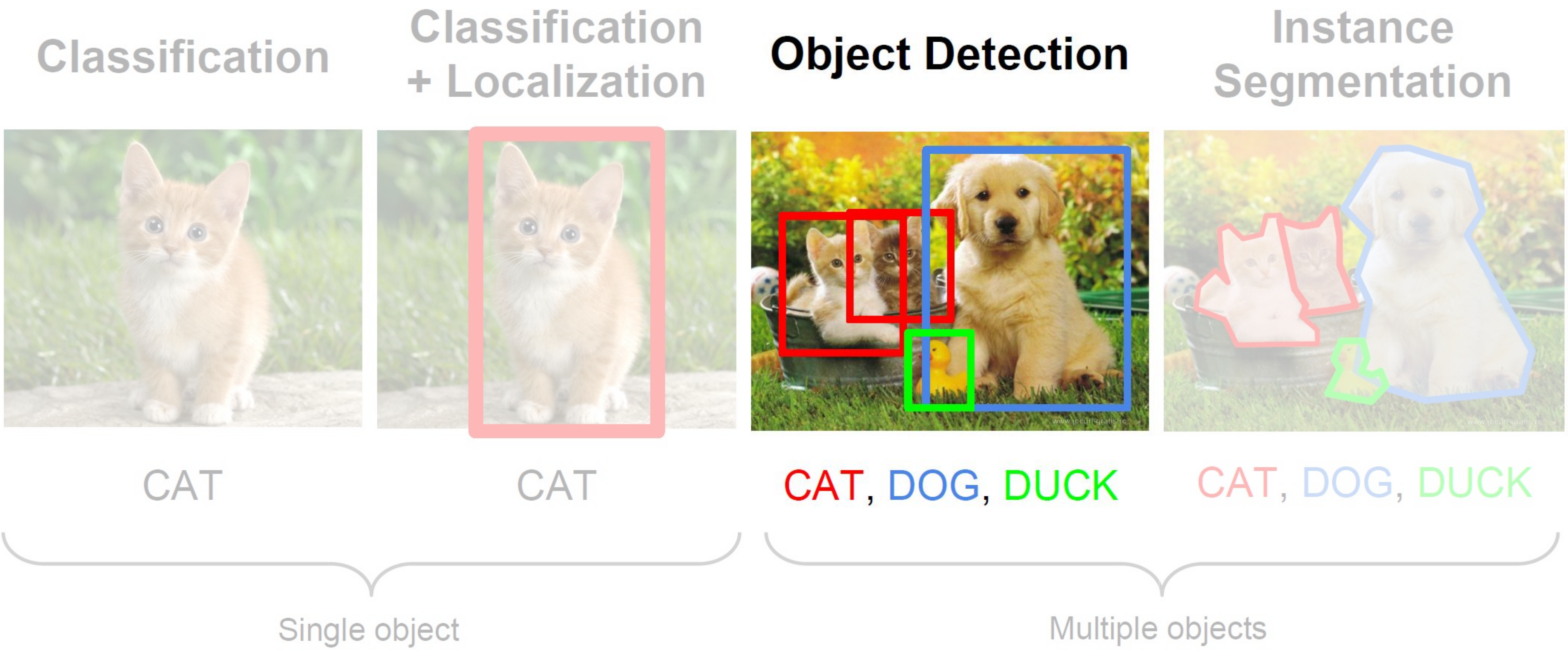
Classification Head:

- C Scores for C classes

Localization Head:

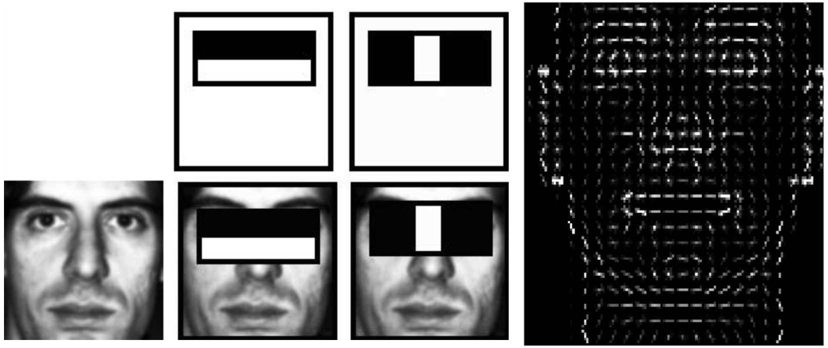
- Class agnostic: (x,y,w,h)
- Class specific: (x,y,w,h) X C

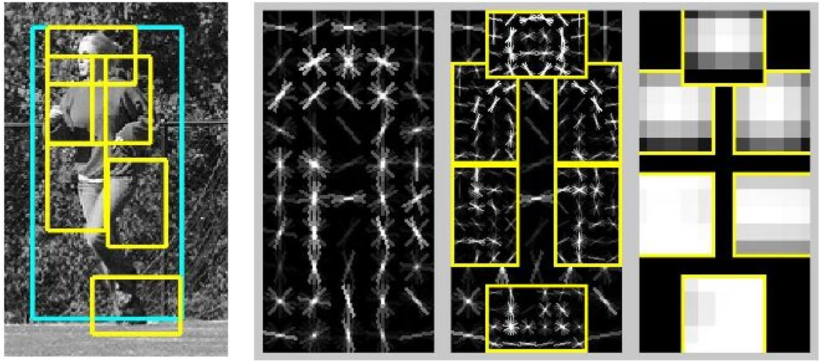
Computer Vision Tasks



Object Detection 2001-2007

- Rapid Object Detection using a Boosted Cascade of Simple Features (2001)
  - Viola & Jones
- Histograms of Oriented Gradients for Human Detection (2005)
  - Dalal & Triggs
- Object Detection with Discriminatively Trained Part Based Models (2010)
  - Felzenszwalb, Girshick, Ramanan
- Fast Feature Pyramids for Object Detection (2014)
  - Dollar





Object Detection 2007-2012

Object Detection Today

Source: Ross Girshick's CVPR 2017 Tutorial http://deeplearning.csail.mit.edu/instance_ross.pptx

Object Detection: Datasets

2007 — Pascal VOC

- 20 Classes
- 11K Training images
- 27K Training objects

Was de-facto standard, currently used as quick benchmark to evaluate new detection algorithms.

2013 — ImageNet ILSVRC

- 200 Classes
- 476K Training images
- 534K Training objects

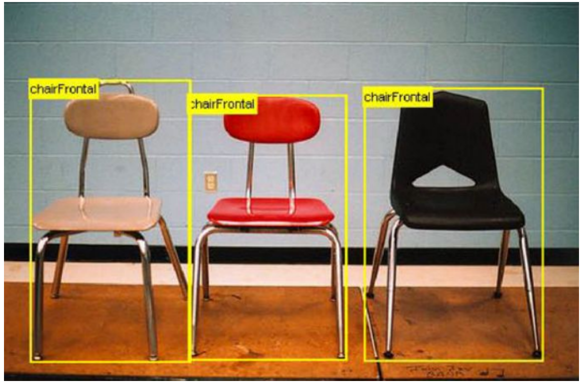
Essentially scaled up version of PASCAL VOC, similar object statistics.

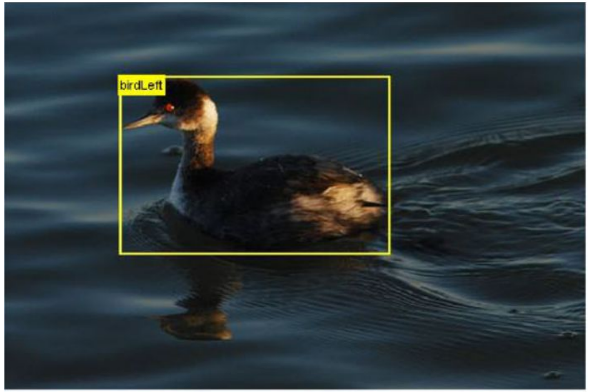
2015 — MS COCO

- 80 Classes
- 200K Training images
- 1.5M Training objects

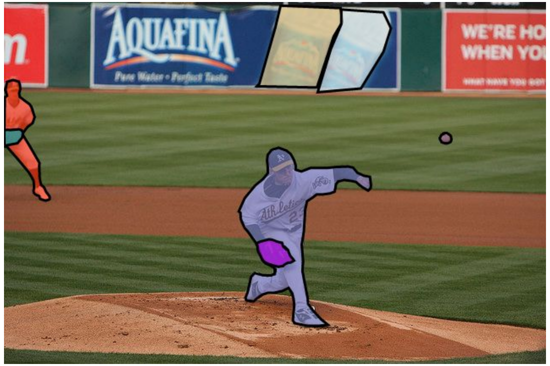
More categories and more object instances in every image. Only 10% of images contain a single object category, 60% in Pascal. More small objects than large objects.

Pascal Examples



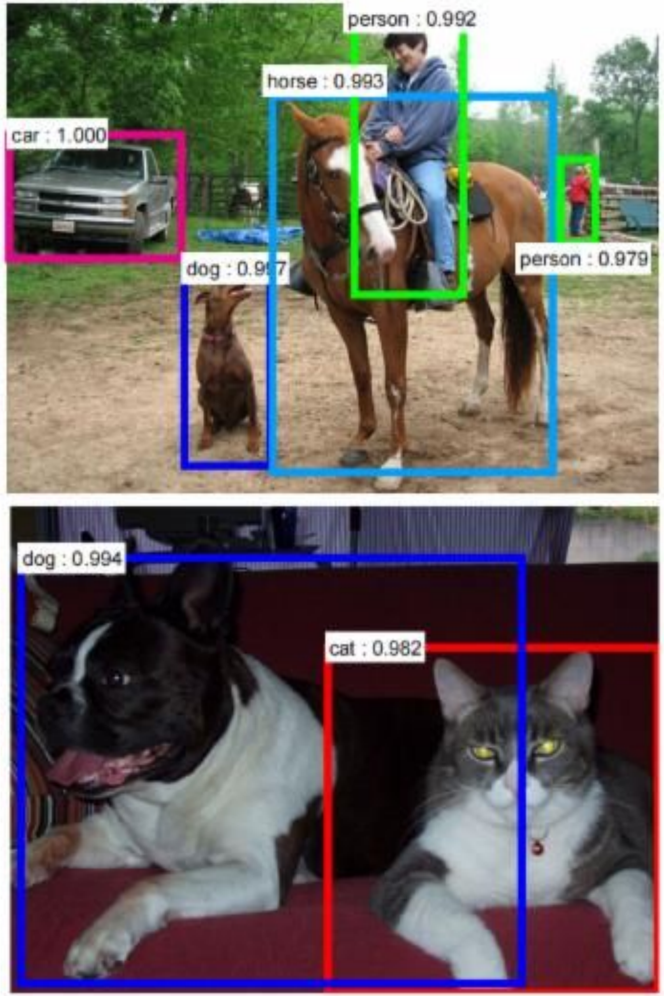


COCO Examples



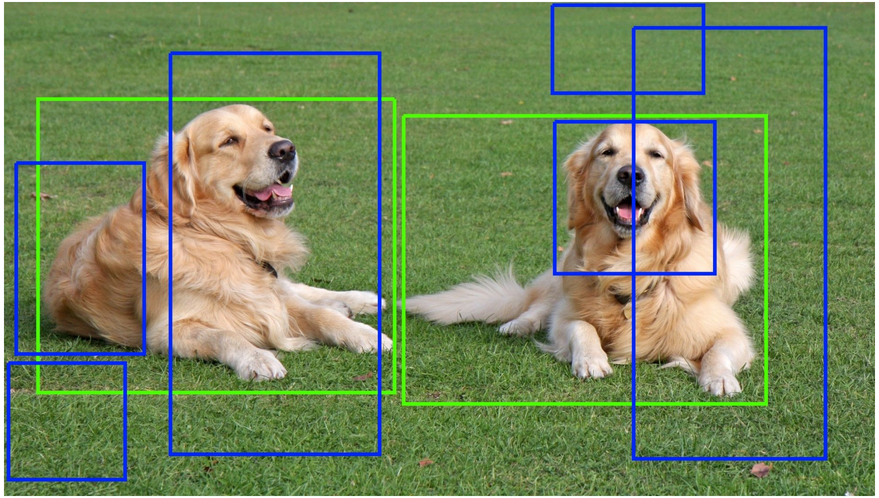
Object Detection

- Input: Image
- Output: For each object class c and each image i, an algorithm returns predicted detections: $\{(b_{ij}, s_{ij})\}_{j=1}^{M}$ ocations b_{ij}th confidence scores . s_{ij}



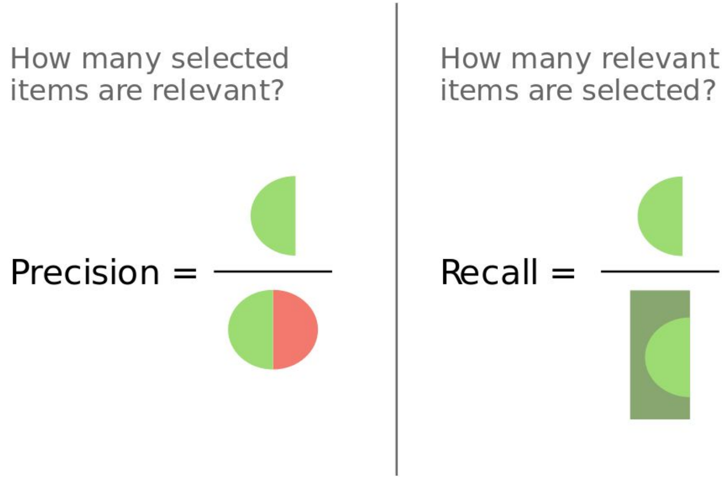
Object Detection: Evaluation

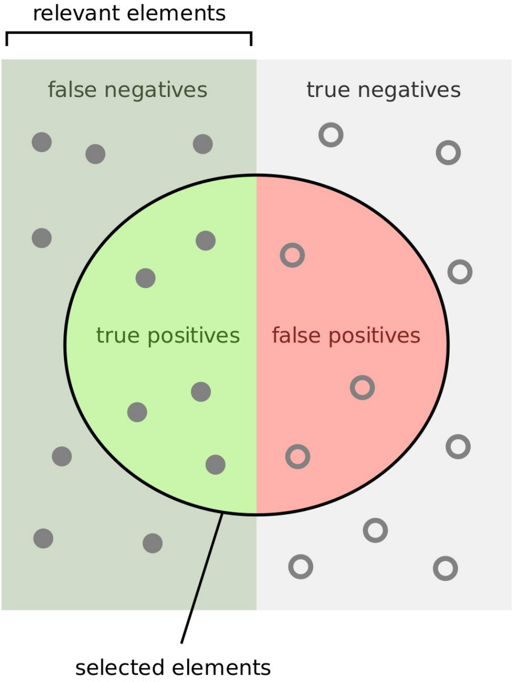
- True positive: correct class prediction **AND** IoU > 50%.
- False positive: wrong class or IoU < 50%.
- False negative: missed (not detected) object
- Only one detection can be matched to an object.
- $z_{ij} = \begin{cases} 1 & b_{ij} \text{ is a True Positive} \\ 0 & b_{ij} \text{ is a False Positive} \end{cases}$



Object Detection: Evaluation

- Mean Average Precision (mAP) across all classes, based on Average Precision (AP) per class, based on **Precision** and **Recall**.



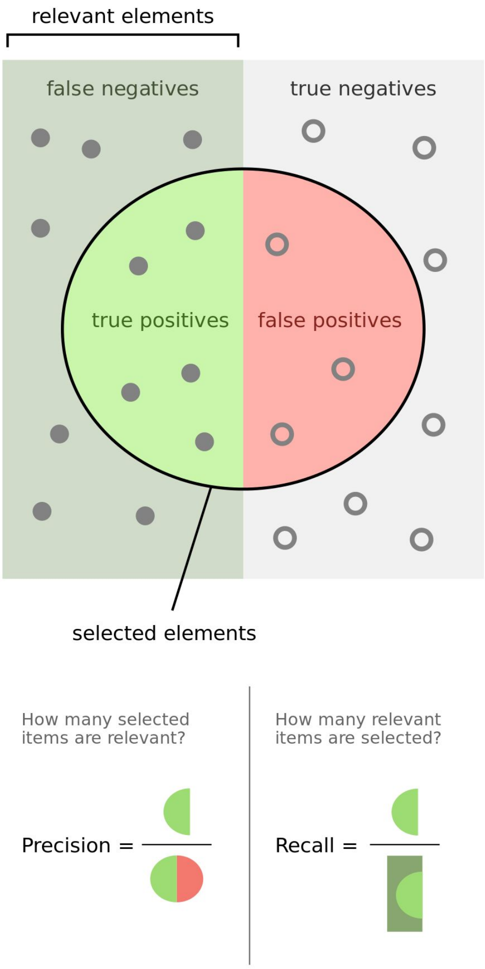


Precision And Recall For a Threshold

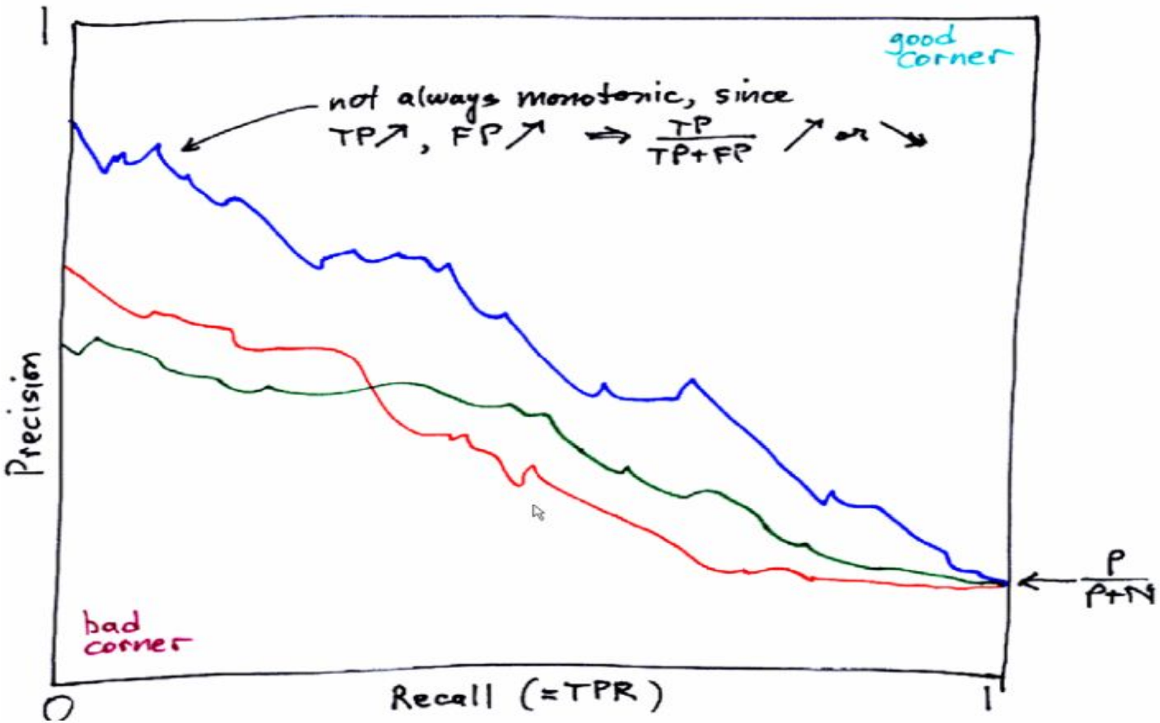
$$Recall(t) = \frac{\sum_{ij} 1[s_{ij} \geq t] z_{ij}}{N}$$

$$Precision(t) = \frac{\sum_{ij} 1[s_{ij} \geq t] z_{ij}}{\sum_{ij} 1[s_{ij} \geq t]}$$

$$z_{ij} = \begin{cases} 1 & b_{ij} \text{ is a True Positive} \\ 0 & b_{ij} \text{ is a False Positive} \end{cases}$$



Precision-Recall Curve



Source: Drawing by Prof. William H. Press, the University of Texas at Austin

Average Precision (AP)

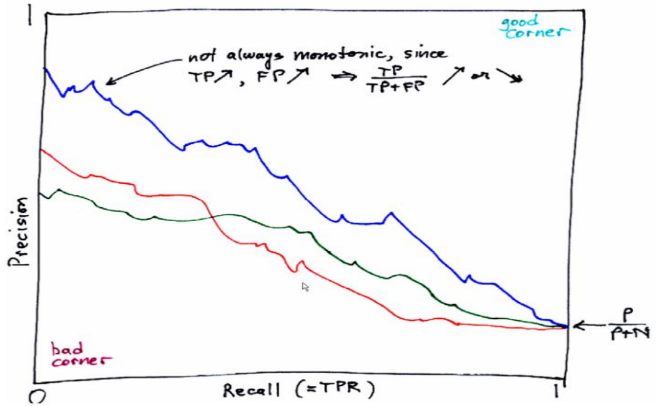
- [In the vision community] AP is the estimated area under the PR curve

$$AveragePrecision = \frac{\sum_{r \in Recall([0,1])} Precision(t_r)}{|Recall([0,1])|}$$

For example:

$$Recall([0,1]) = \{0, 0.1, 0.2, \ldots, 0.9, 1.0\}$$

$$|Recall([0,1])| = 11$$



Mean Average Precision (mAP)

- The winner of each object class is the team with the highest average precision
- The winner of the challenge is the team with the highest mean Average Precision (mAP) across all classes.

$$meanAveragePrecision = \frac{\sum_{c \in C} AveragePrecision_c}{|C|}$$

Object Detection: Evaluation

- Mean Average Precision (mAP) across all classes, based on Average Precision (AP) per class, based on **Precision** and **Recall**.

Table 4: Object detection results on PASCAL VOC 2007 *test* set. The performance is measured by mean of Average Precision (mAP, in %).

Method	**mAP**	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbk	prsn	plant	sheep	sofa	train	tv
DenseNet-161 (k=48)	79.9	80.4	85.9	81.2	72.8	68.0	87.1	88.0	88.8	64.0	83.3	75.4	87.5	87.6	81.3	84.2	54.6	83.2	80.2	87.4	77.2
ResNet-101 [16]	76.4	79.8	80.7	76.2	68.3	55.9	85.1	85.3	**89.8**	56.7	**87.8**	69.4	88.3	**88.9**	80.9	78.4	41.7	78.6	79.8	85.3	72.0
ResNeXt-101 (32 × 4d)	80.1	80.2	86.5	79.4	72.5	67.3	86.9	88.6	88.9	64.9	85.0	76.2	87.3	87.8	81.8	84.1	55.5	84.0	79.7	87.9	77.0
DPN-92 (32 × 3d)	**82.5**	**84.4**	**88.5**	**84.6**	**76.5**	**70.7**	**87.9**	**88.8**	89.4	**69.7**	87.0	**76.7**	**89.5**	88.7	**86.0**	**86.1**	**58.4**	**85.0**	**80.4**	**88.2**	**83.1**

Object Detection: Evaluation

- Today new metrics are emerging
  - Averaging precision over all IoU thresholds: 0.5:0.05:0.95
  - Averaging precision for different object sizes (small, medium, big)
  - Averaging recall as a metric to measure object proposal quality.

Average Precision (AP):	
AP	% AP at IoU=.50:.05:.95 **(determines challenge winner)**
$AP^{IoU=.50}$	% AP at IoU=.50 (PASCAL VOC metric)
$AP^{IoU=.75}$	% AP at IoU=.75 (strict metric)
AP Across Scales:	
AP^{small}	% AP for small objects: area < 32^2
AP^{medium}	% AP for medium objects: 32^2 < area < 96^2
AP^{large}	% AP for large objects: area > 96^2
Average Recall (AR):	
$AR^{max=1}$	% AR given 1 detection per image
$AR^{max=10}$	% AR given 10 detections per image
$AR^{max=100}$	% AR given 100 detections per image
AR Across Scales:	
AR^{small}	% AR for small objects: area < 32^2
AR^{medium}	% AR for medium objects: 32^2 < area < 96^2
AR^{large}	% AR for large objects: area > 96^2

Brodmann17

Looking for brilliant researchers

cv@brodmann17.com