

# 目标检测

# Object Detection

目标检测 (Object Detection) = What, and Where

**定位** Localization

**Where?**

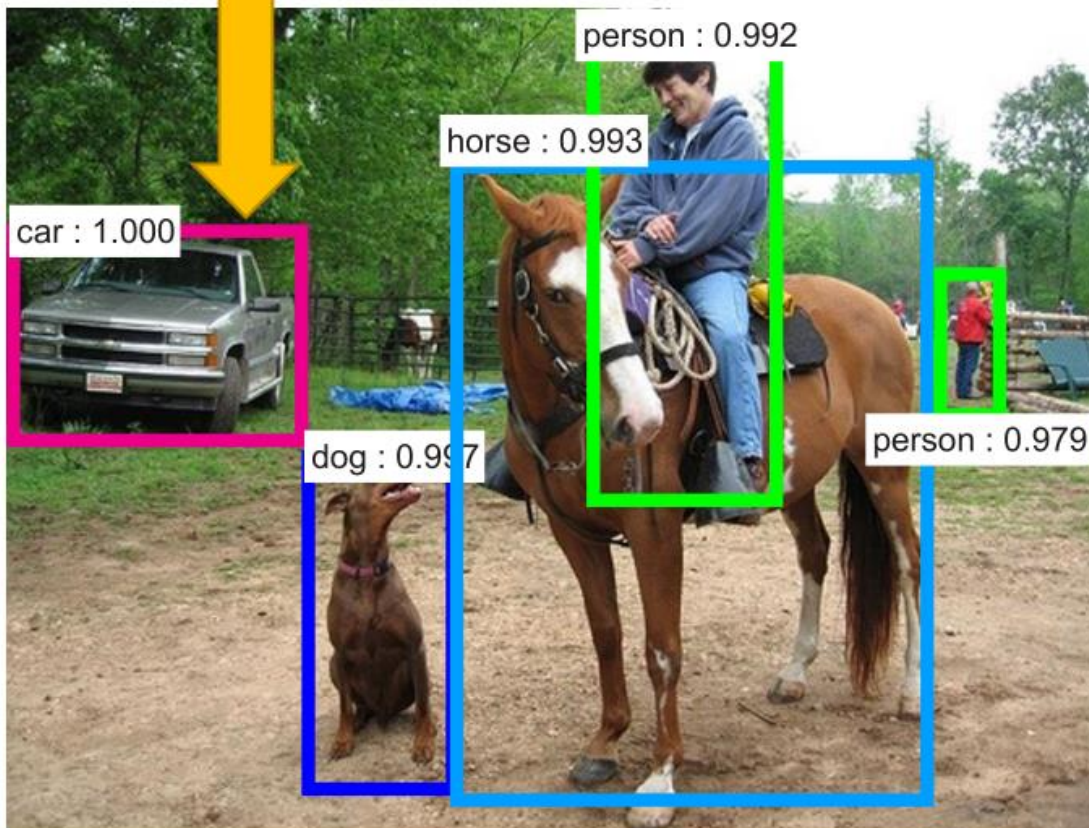
位置 (最小外接矩形, Bounding box)

**识别**

Recognition

**What?**

类别标签 (Category label)

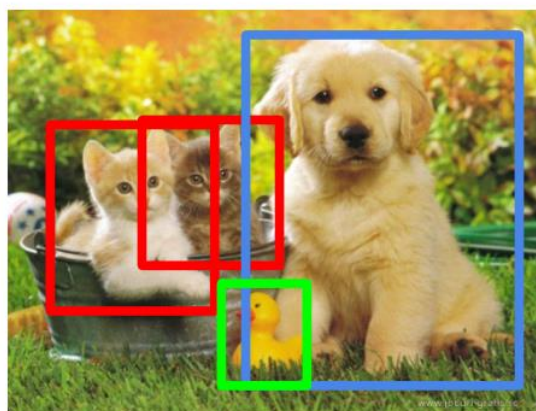


**Classification**

CAT

**Classification  
+ Localization**

CAT

**Object Detection**

CAT, DOG, DUCK

**Instance  
Segmentation**

CAT, DOG, DUCK

Single object

Multiple objects

**定位和检测:**

- **定位**是找到检测图像中带有给定标签的**单个目标**
- **检测**是找到图像中带有给定标签的**所有目标**

# 目标检测数据集

- PASCAL VOC
- ImageNet
- COCO



# PASCAL VOC challenge

PASCAL VOC挑战在2005年至2012年间展开。

该数据集中有20个分类。该数据集包含11530张用于训练和验证的图像，其中感兴趣区域有27450个标定。

以下是数据集中的20个分类：

- 人：人
- 动物：鸟、猫、牛、狗、马、羊
- 车辆：飞机、自行车、船、巴士、汽车、摩托车、火车
- 室内：瓶、椅子、餐桌、盆栽植物、沙发、电视/监视器

每个图像平均有2.4个目标。

20 classes



链接: <http://host.robots.ox.ac.uk/pascal/VOC/voc2012/>

# ImageNet数据集

ILSVRC 2010-2017

- ImageNet拥有用于分类、定位和检测任务评估的数据。
- 与分类数据类似，定位任务有1000个类别。准确率是根据Top 5检测结果计算出来的。
- 对200个目标的检测问题有470000个图像，平均每个图像有1.1个目标。

# COCO

- MS COCO的全称是Microsoft Common Objects in Context，起源于是微软于2014年出资标注的Microsoft COCO数据集，与ImageNet 竞赛一样，被视为是计算机视觉领域最受关注和最权威的比赛之一。
- 在ImageNet竞赛停办后，COCO竞赛就成为是当前目标识别、检测等领域的一个最权威、最重要的标杆，也是目前该领域在国际上唯一能汇集Google、微软、Facebook以及国内外众多顶尖院校和优秀创新企业共同参与的大赛。

<http://cocodataset.org/>

# COCO目标检测挑战

- COCO (Common Objects in Context) 数据集包含20万个图像
- 80个类别中有超过50万个目标标注。它是最广泛公开的目标检测数据库
- 平均每个图像的目标数为7.2





# 性能指标

- Precision, Recall, F1 score
- IoU (Intersection over Union)
- P-R curve (Precision-Recall curve)
- AP (Average Precision)
- mAP (mean Average Precision)
- FPS (Frames Per Second)

# 混淆矩阵 (confusion matrix)

		Prediction	
		Positive	Negative
Actual	True	TP	FN
	False	FP	TN

$$precision = \frac{TP}{TP + FP}$$


$$recall = \frac{TP}{TP + FN}$$

第一位T/F: 表示预测的对错  
第二位P/N: 表示预测的结果

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

- 准确率Precision(查准率)是评估预测的准不准 (看预测列)
- 召回率Recall (查全率) 是评估找的全不全(看实际行)

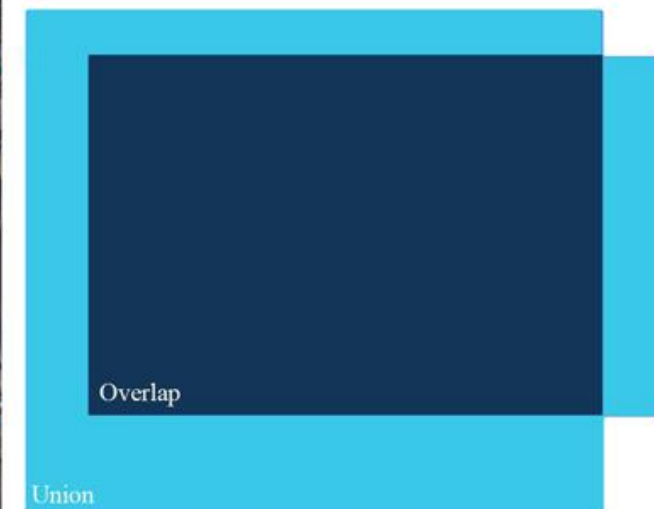
# IoU (Intersection over Union )

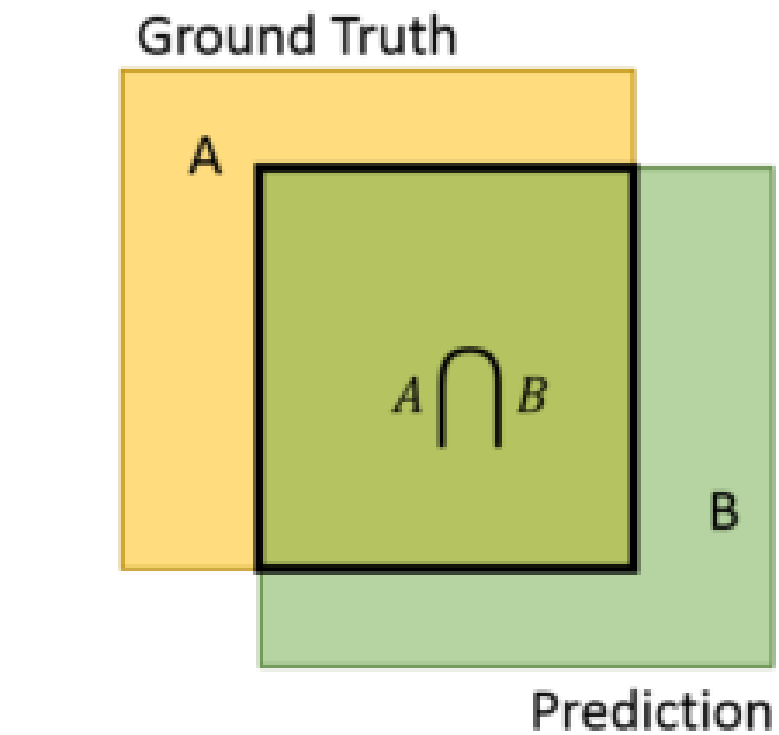
$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$




 Ground truth  
 Prediction

$$IoU = \frac{\text{area of overlap}}{\text{area of union}}$$



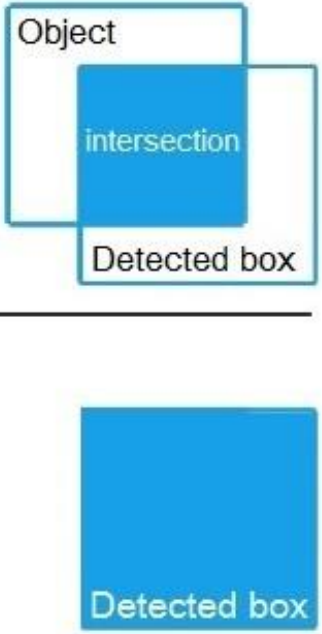


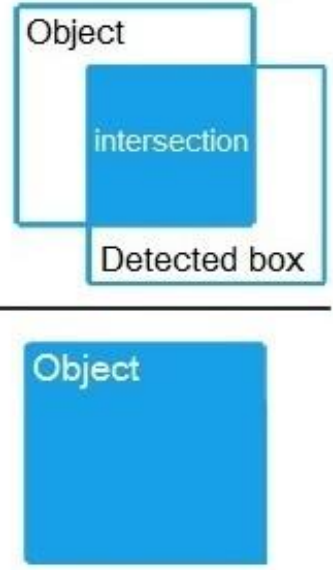
$$IOU_{AB} = \frac{A \cap B}{A + B - A \cap B}$$

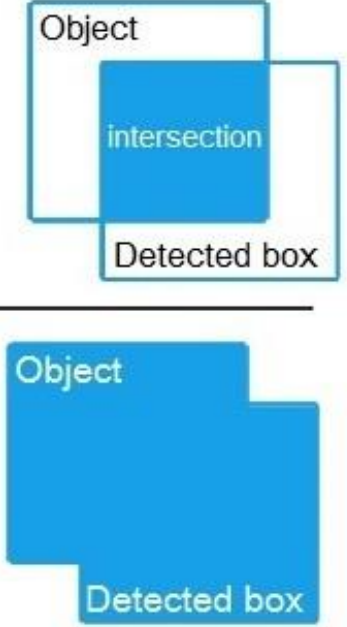
If  $IOU_{AB} > \text{threshold}$ , B is True Positive

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{Precision} = \frac{\text{Area of Intersection}}{\text{Area of Detected box}}$$


$$\text{Recall} = \frac{\text{Area of Intersection}}{\text{Area of Object}}$$


$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$




# AP (Average Precision)

用一个简单的例子来演示平均精度（AP）的计算。假设数据集中总共有5个苹果。我们收集模型为苹果作的所有预测，并根据预测的置信水平（从最高到最低）对其进行排名。第二列表示预测是否正确。如果它与 ground truth 匹配并且  $IoU \geq 0.5$ ，则是正确的。

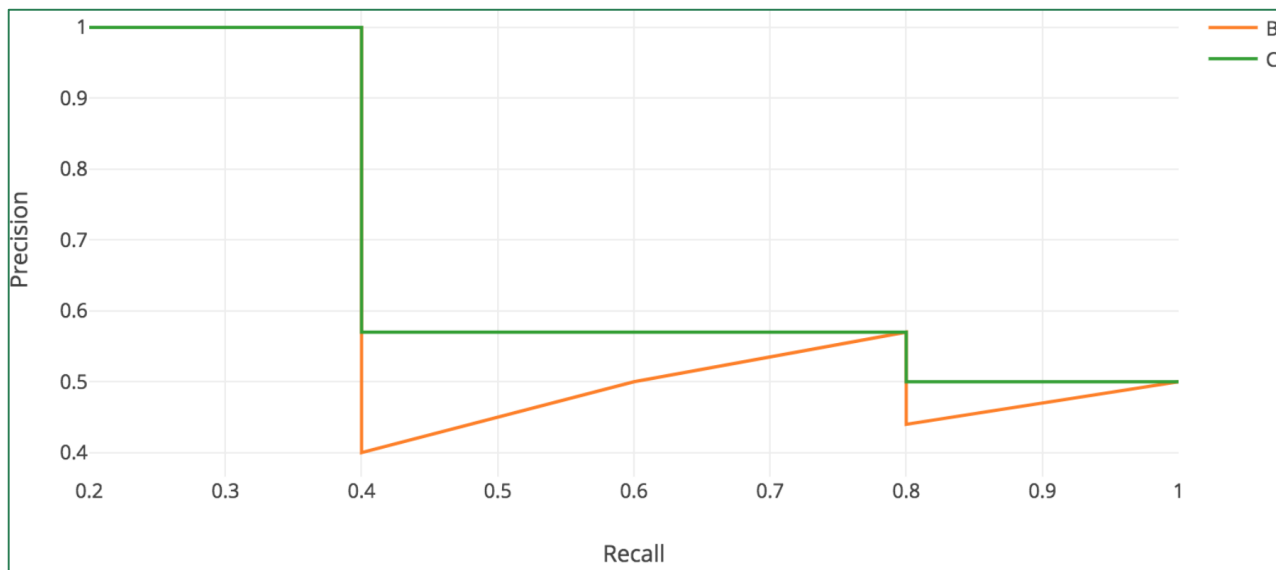
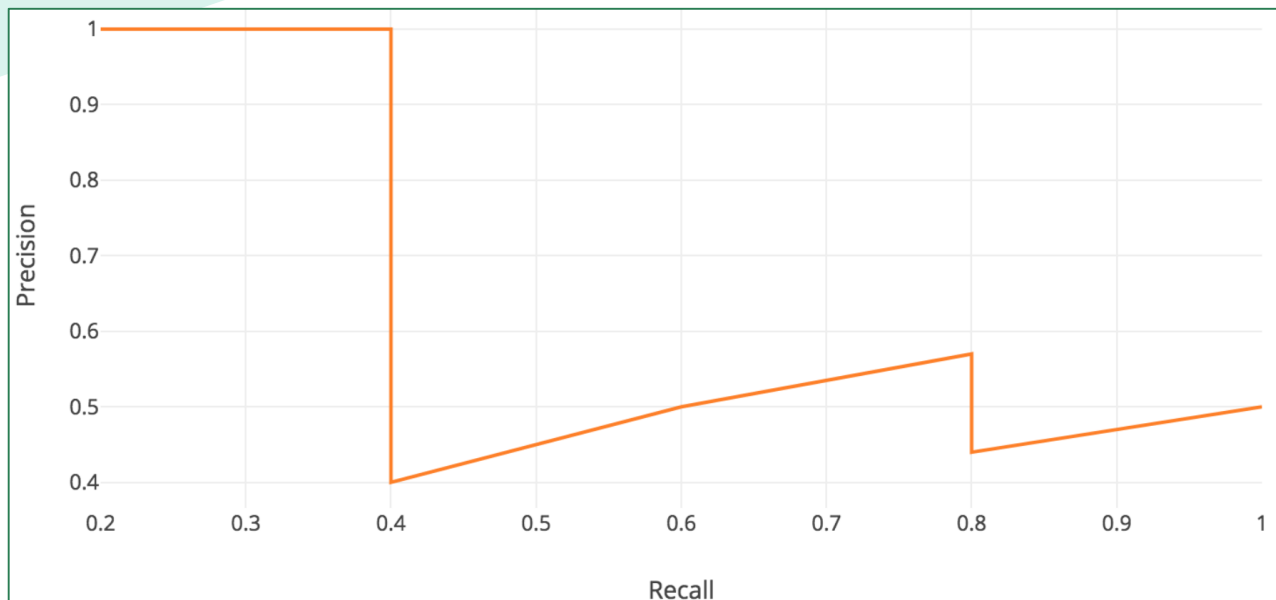
Rank	Correct?	Precision	Recall
1	True	1.0	0.2
2	True	1.0	0.4
3	False	0.67	0.4
4	False	0.5	0.4
5	False	0.4	0.4
6	True	0.5	0.6
7	True	0.57	0.8
8	False	0.5	0.8
9	False	0.44	0.8
10	True	0.5	1.0

Let's compute the precision and recall value for the row with rank #3.

**Precision** is the proportion of TP out of  $(TP+FP) = 2/3 = 0.67$ .

**Recall** is the proportion of TP out of the possible positives  $= 2/5 = 0.4$ .

Recall随着我们包含更多预测而增加，但Precision会上下波动。



AP在概念上可以视为在precision-recall graph（橙色线）下方的区域。

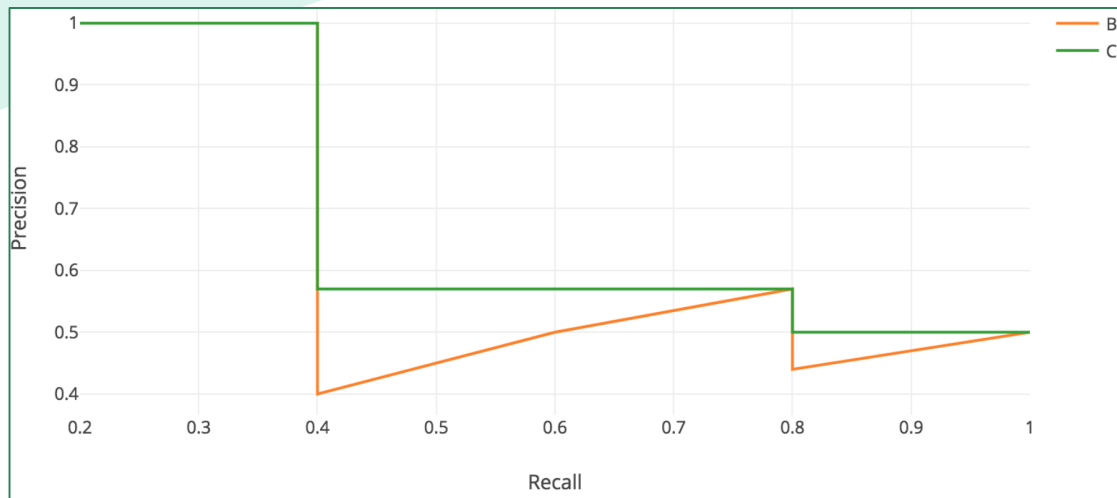
首先通过平滑锯齿形图形来近似这样的计算

We plot the graph with recall value at 0, 0.1, 0.2, ..., 0.9 and 1.0 and we replace the precision value with the maximum precision for any recall  $\geq \tilde{r}$

$$p_{interp}(r) = \max_{\tilde{r} \geq r} p(\tilde{r})$$

AP（平均精度）计算为这11个recall级别的最大精度的平均值：  
11-point interpolated average precision

$$AP = \frac{1}{11} \times (AP_r(0) + AP_r(0.1) + \dots + AP_r(1.0))$$



这近似于找到绿色曲线下的总面积并将其除以11。下面是更精确的定义：

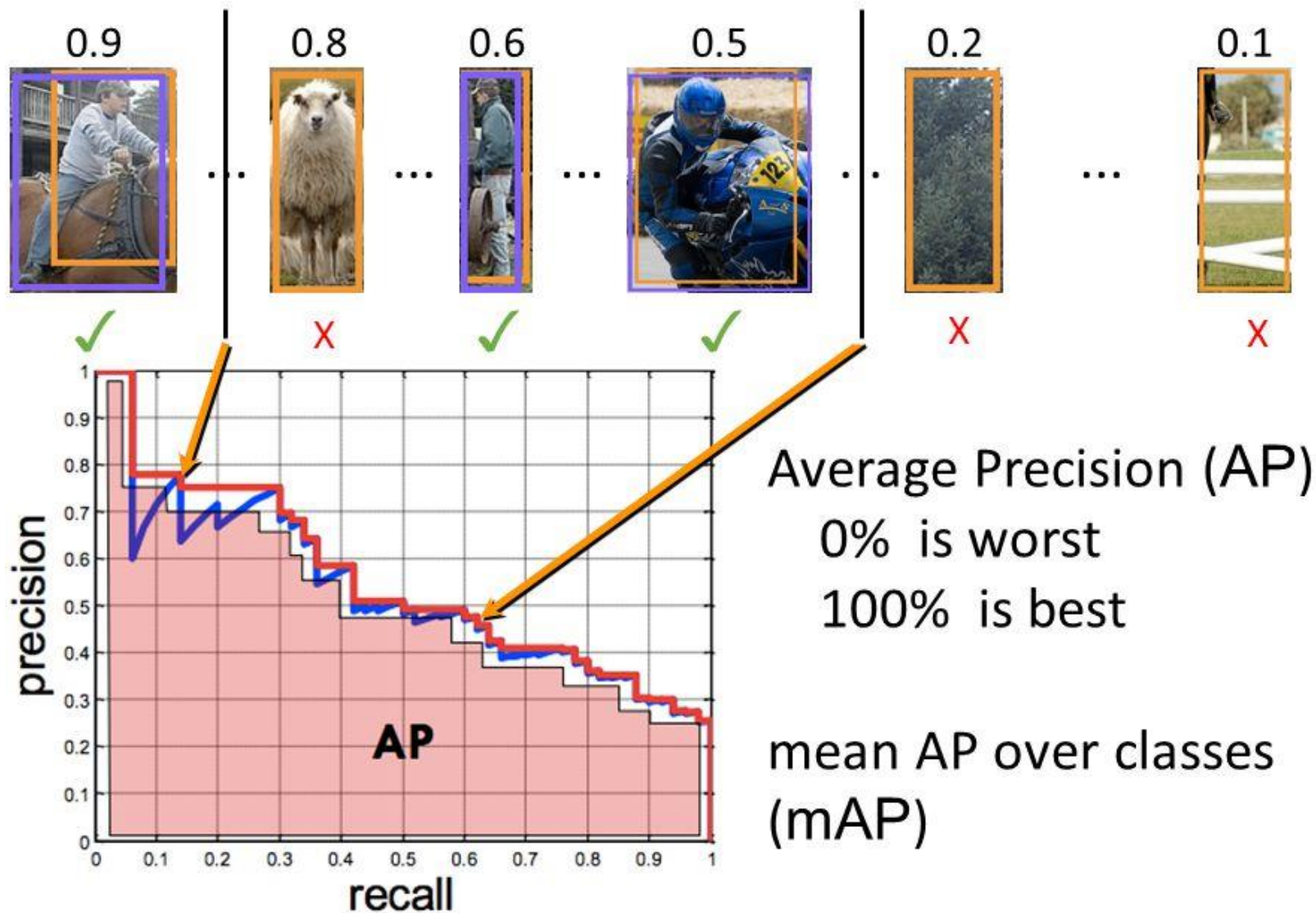
$$\begin{aligned} AP &= \frac{1}{11} \sum_{r \in \{0.0, \dots, 1.0\}} AP_r \\ &= \frac{1}{11} \sum_{r \in \{0.0, \dots, 1.0\}} p_{interp}(r) \end{aligned}$$

where

$$p_{interp}(r) = \max_{\tilde{r} \geq r} p(\tilde{r})$$

在本例中  $AP = (5 \times 1.0 + 4 \times 0.57 + 2 \times 0.5)/11$

- AP衡量的是学习出来的模型在每个类别上的好坏
- mAP衡量的是学出的模型在所有类别上的好坏。mAP就是取所有类别上AP的平均值。



## AP (Average Precision) in PASCAL VOC challenge

- 对于PASCAL VOC挑战, 如果 $IoU > 0.5$ , 则预测为正。但是, 如果检测到同一目标的多个检测, 则视第一个检测为正, 而视其余检测为负。
- PASCAL VOC CHALLENGE自2010年后换了一种计算方法。新的计算方法假设这N个样本中有M个正例, 那么我们会得到M个recall值 ( $1/M, 2/M, \dots, M/M$ ) ,对于每个recall值r, 我们可以计算出对应 ( $r' > r$ ) 的最大precision, 然后对这M个precision值取平均即得到最后的AP值。



# COCO mAP

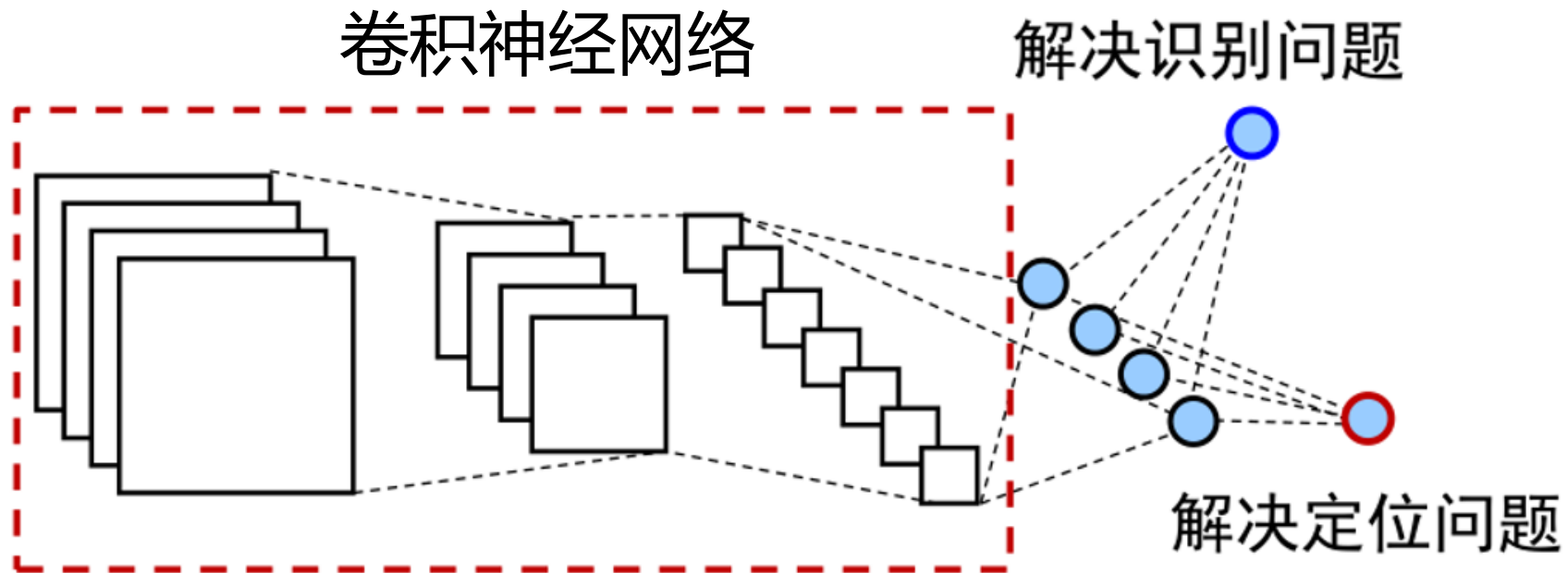
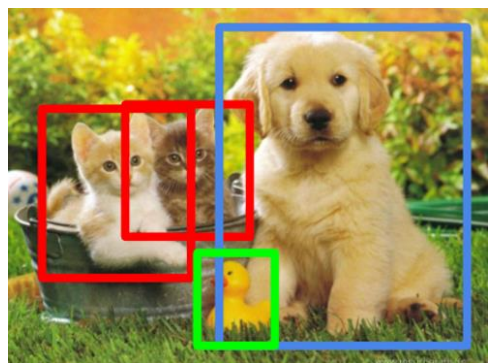
For COCO, AP is the **average over multiple IoU** (the minimum IoU to consider a positive match).

**AP@[.5:.95]** corresponds to the average AP for IoU from 0.5 to 0.95 with a step size of 0.05.

For the COCO competition, AP is the average over 10 IoU levels on 80 categories  
(AP@[.50:.05:.95]: start from 0.5 to 0.95 with a step size of 0.05).

**mAP@.75** means the mAP with IoU=0.75.

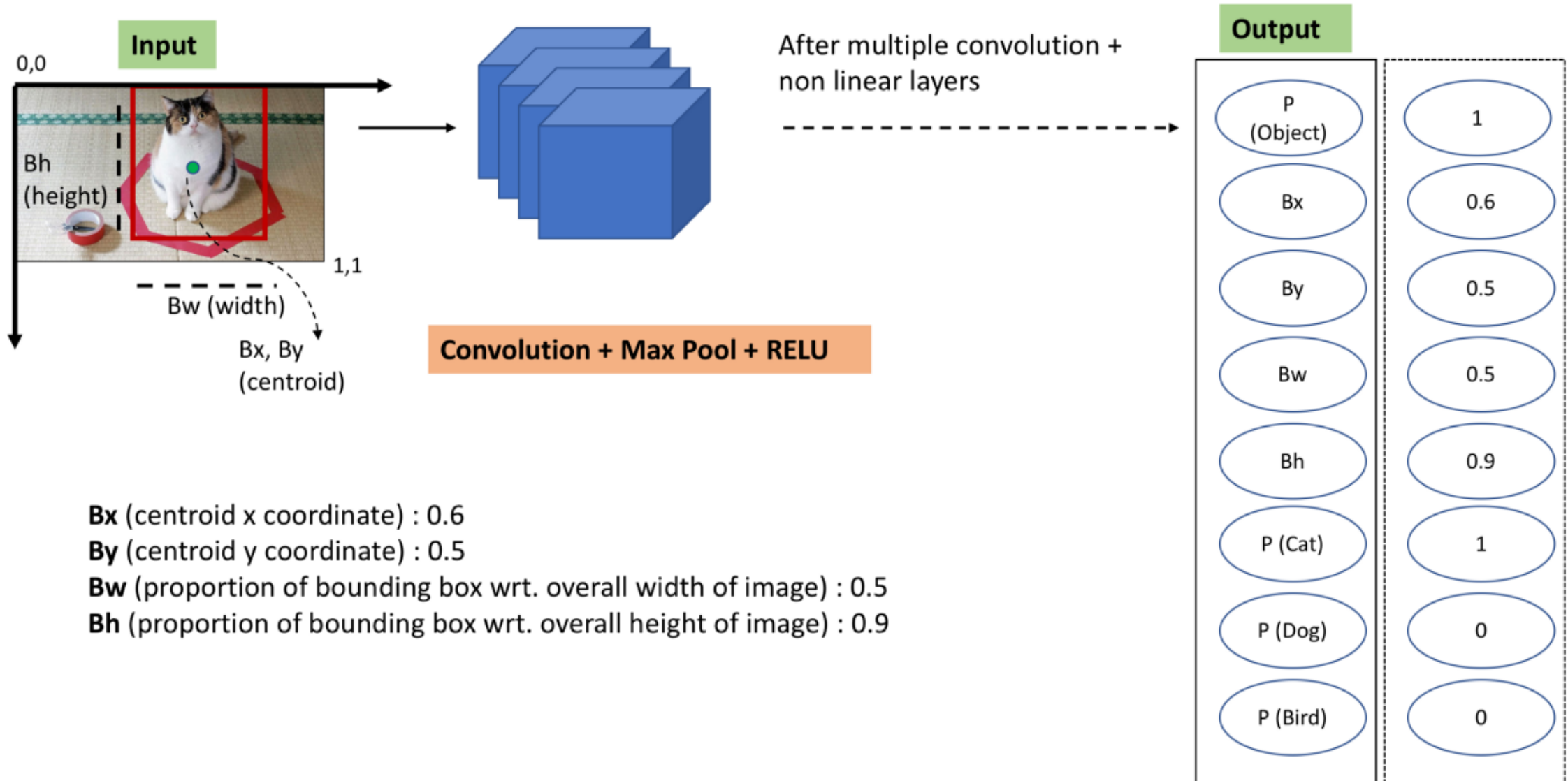
# 基于深度学习实现目标检测



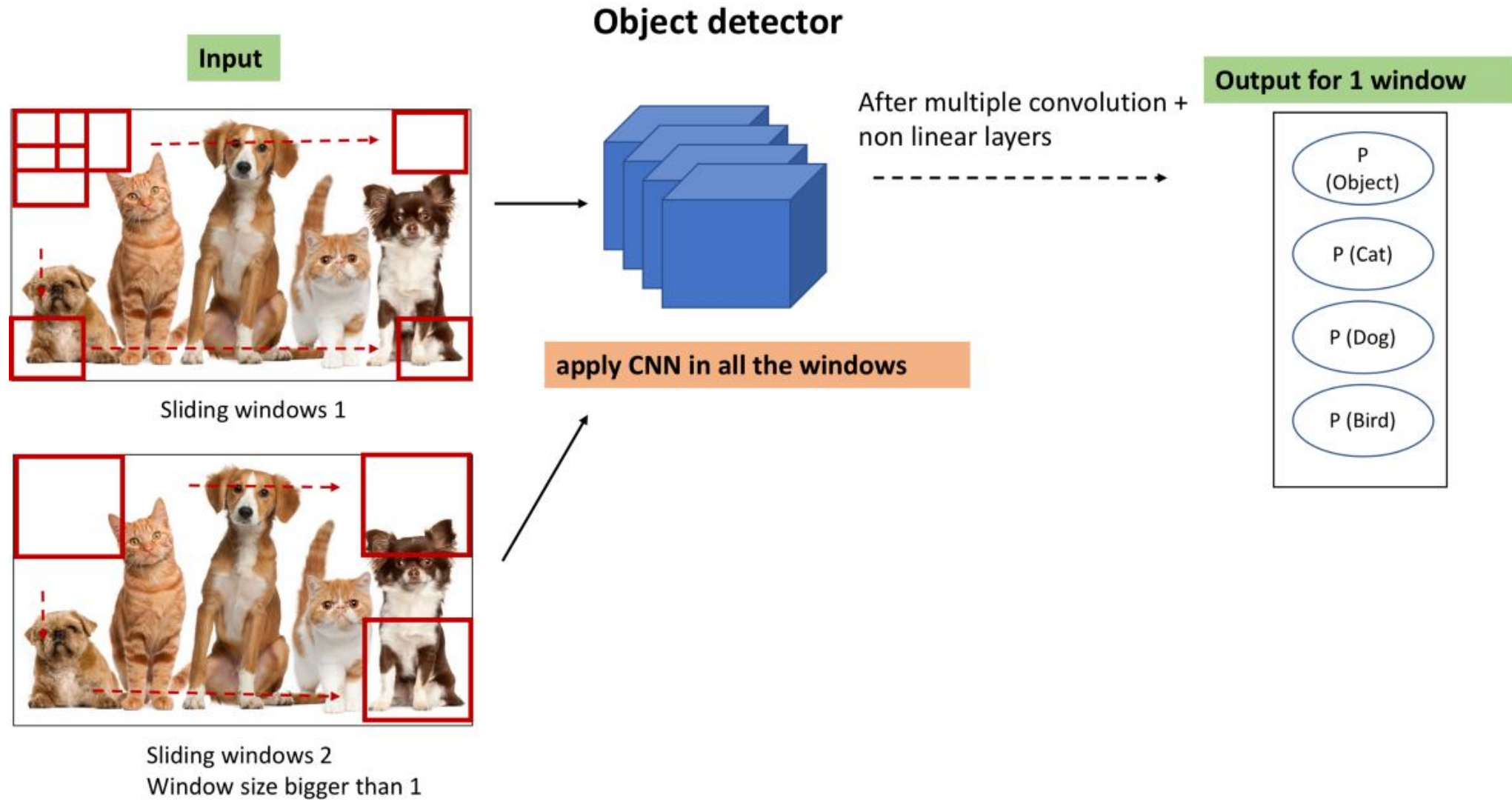
- 识别问题：输出目标的类别
- 定位问题：输出目标的位置

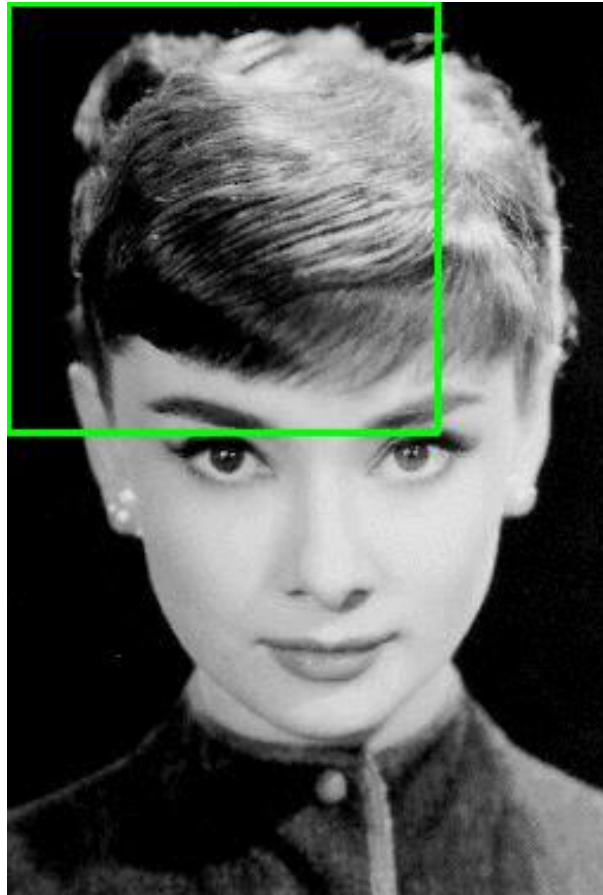
# Object classification and localization

Cat/ Dog/ Bird classifier and localizer



# Multiple objects detection and localization



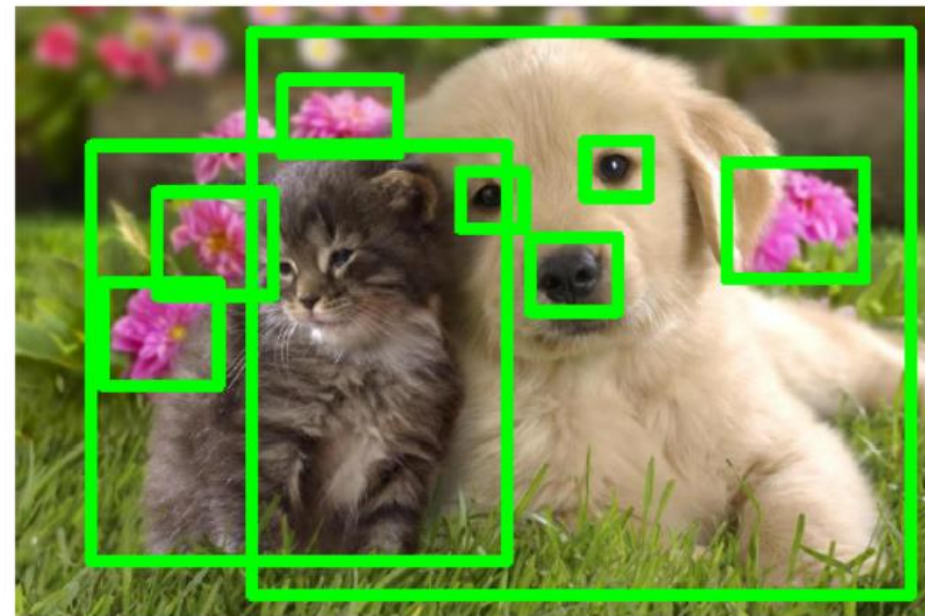


An example of the sliding window approach.

Each of the bounding boxes will be used as a region of interest (ROI).



## 区域建议 (Region Proposals)



- 基于建议框的方法 (Proposal-based; Two-Stage;)  
R-CNN, SPP-Net, Fast R-CNN, Faster R-CNN, FPN
- 免建议框方法 (Proposal-free; One-Stage; Single-Shot)  
**YOLO**, SSD, DSSD, RetinaNet

## Proposal-based

