

Chapter 02. Image Localization 구현 및 성능 개선하기

# Image Localization

# Computer Vision Task

## Semantic Segmentation



GRASS, CAT,  
TREE, SKY

No objects, just pixels

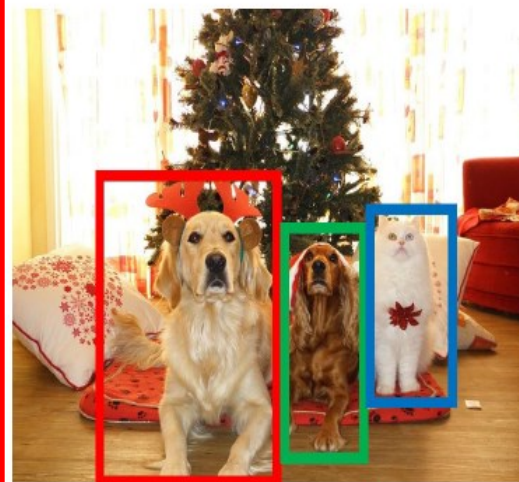
## Classification + Localization



CAT

Single Object

## Object Detection



DOG, DOG, CAT

Multiple Object

## Instance Segmentation



DOG, DOG, CAT

This image is CC0 public domain

# Classification & Localization

**Classification:** C classes

**Input:** Image

**Output:** Class label

**Evaluation metric:** Accuracy



CAT

**Localization:**

**Input:** Image

**Output:** Box in the image (x, y, w, h)

**Evaluation metric:** Intersection over Union

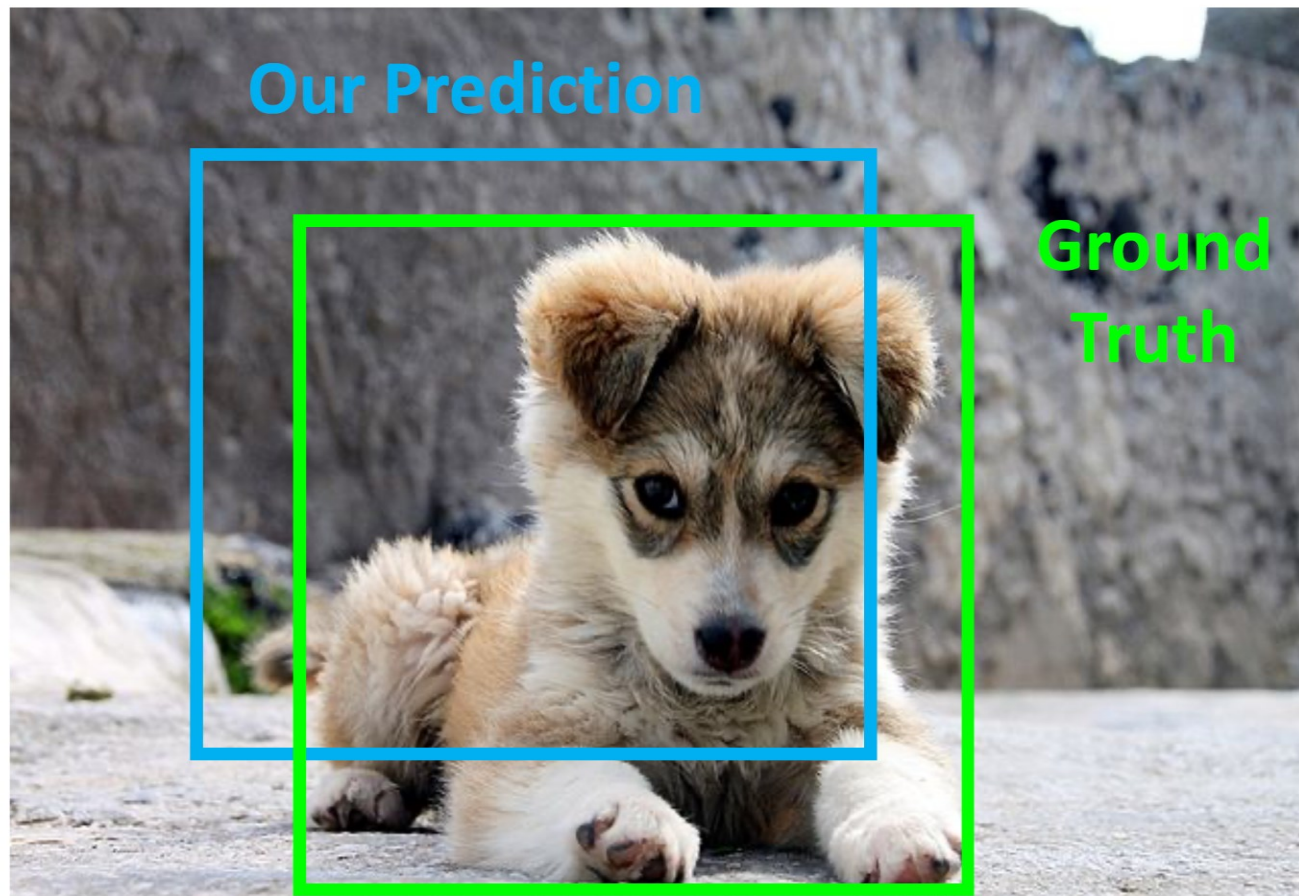


(x, y, w, h)

**Classification + Localization:** Do both

# Comparing Boxes: Intersection over Union (IoU)

How can we compare our prediction to the ground-truth box?



[Puppy image](#) is licensed under [CC-A 2.0 Generic license](#). Bounding boxes and text added by Justin Johnson.

Slide from Deep Learning for Computer Vision @Univ. of Michigan

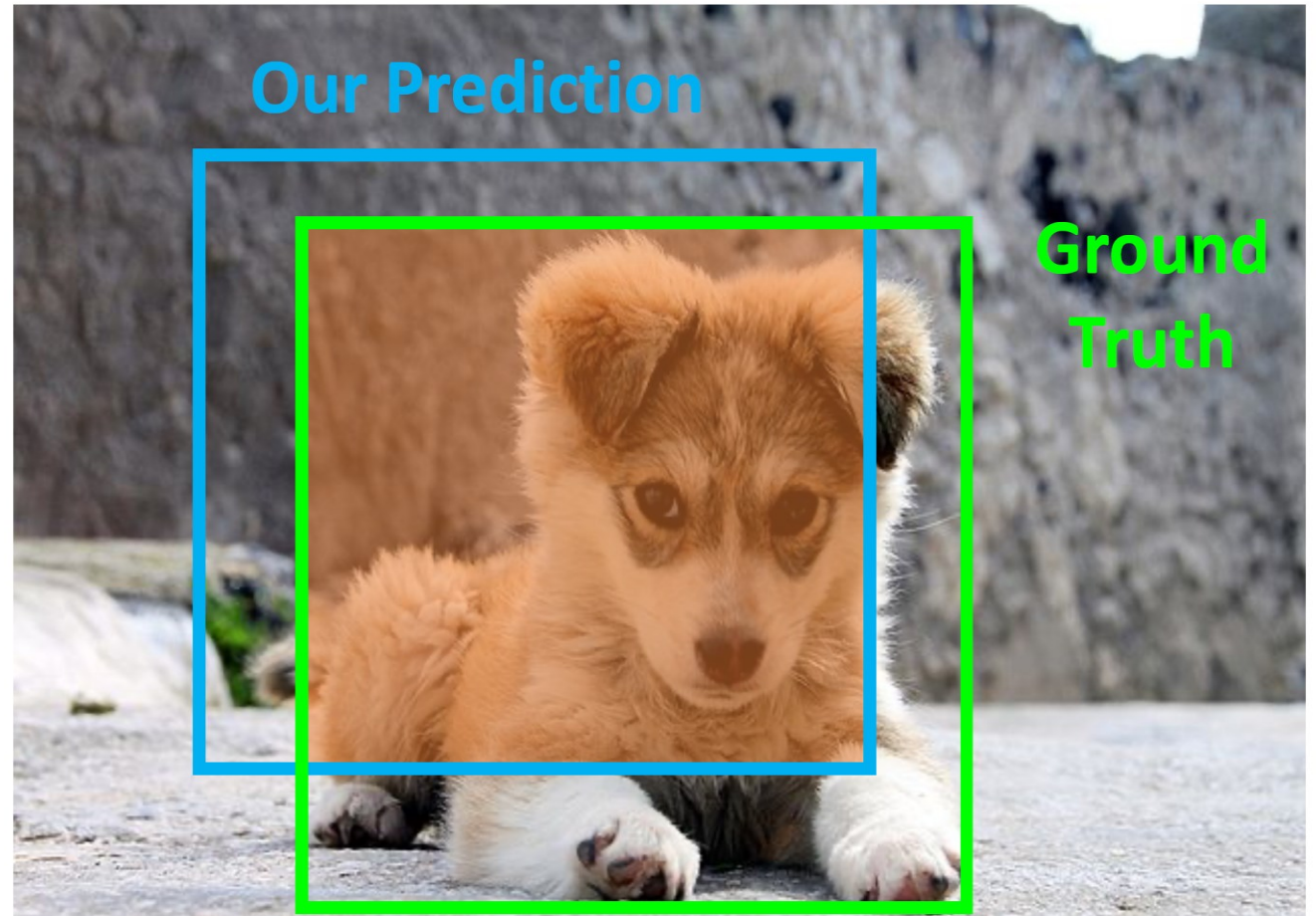


# Comparing Boxes: Intersection over Union (IoU)

How can we compare our prediction to the ground-truth box?

**Intersection over Union (IoU)**  
(Also called “Jaccard similarity” or “Jaccard index”):

$$\frac{\text{Area of Intersection}}{\text{Area of Union}}$$



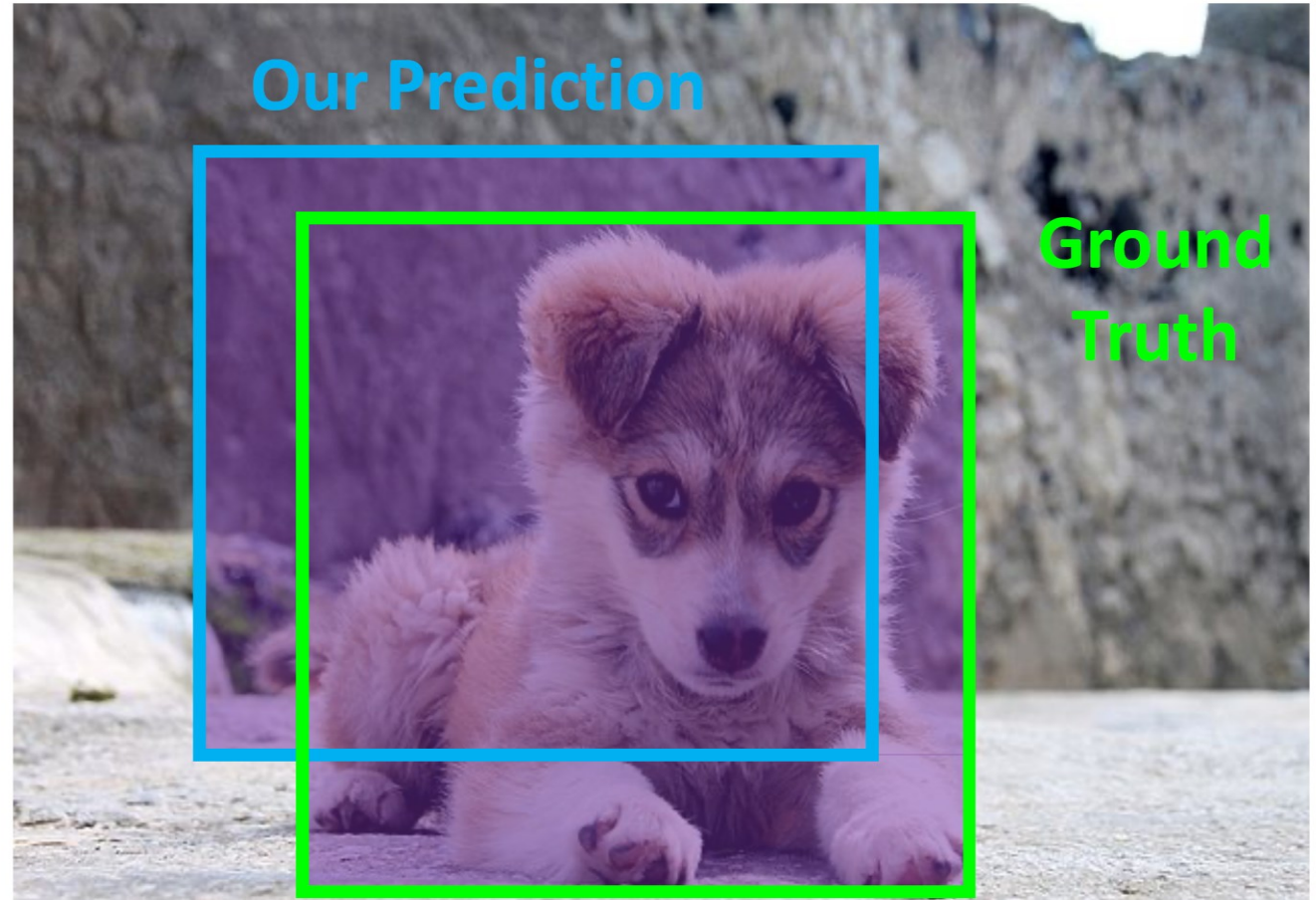
[Puppy image](#) is licensed under [CC-A 2.0 Generic license](#). Bounding boxes and text added by Justin Johnson.

# Comparing Boxes: Intersection over Union (IoU)

How can we compare our prediction to the ground-truth box?

**Intersection over Union (IoU)**  
(Also called “Jaccard similarity” or “Jaccard index”):

$$\frac{\text{Area of Intersection}}{\text{Area of Union}}$$



[Puppy image](#) is licensed under [CC-A 2.0 Generic license](#). Bounding boxes and text added by Justin Johnson.

Slide from Deep Learning for Computer Vision @Univ. of Michigan



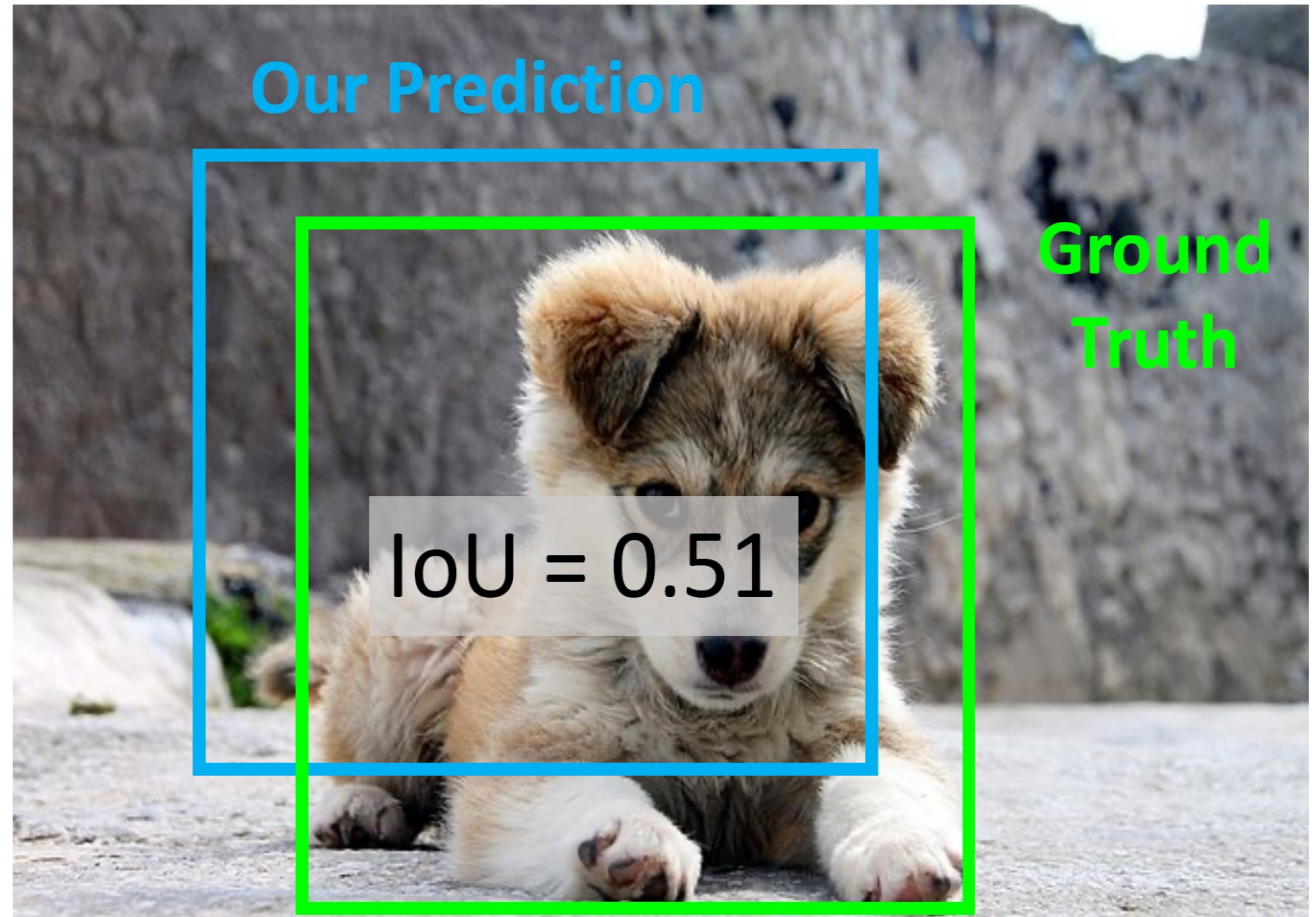
# Comparing Boxes: Intersection over Union (IoU)

How can we compare our prediction to the ground-truth box?

**Intersection over Union (IoU)**  
(Also called “Jaccard similarity” or “Jaccard index”):

$$\frac{\text{Area of Intersection}}{\text{Area of Union}}$$

$\text{IoU} > 0.5$  is “decent”



[Puppy image](#) is licensed under [CC-A 2.0 Generic license](#). Bounding boxes and text added by Justin Johnson.

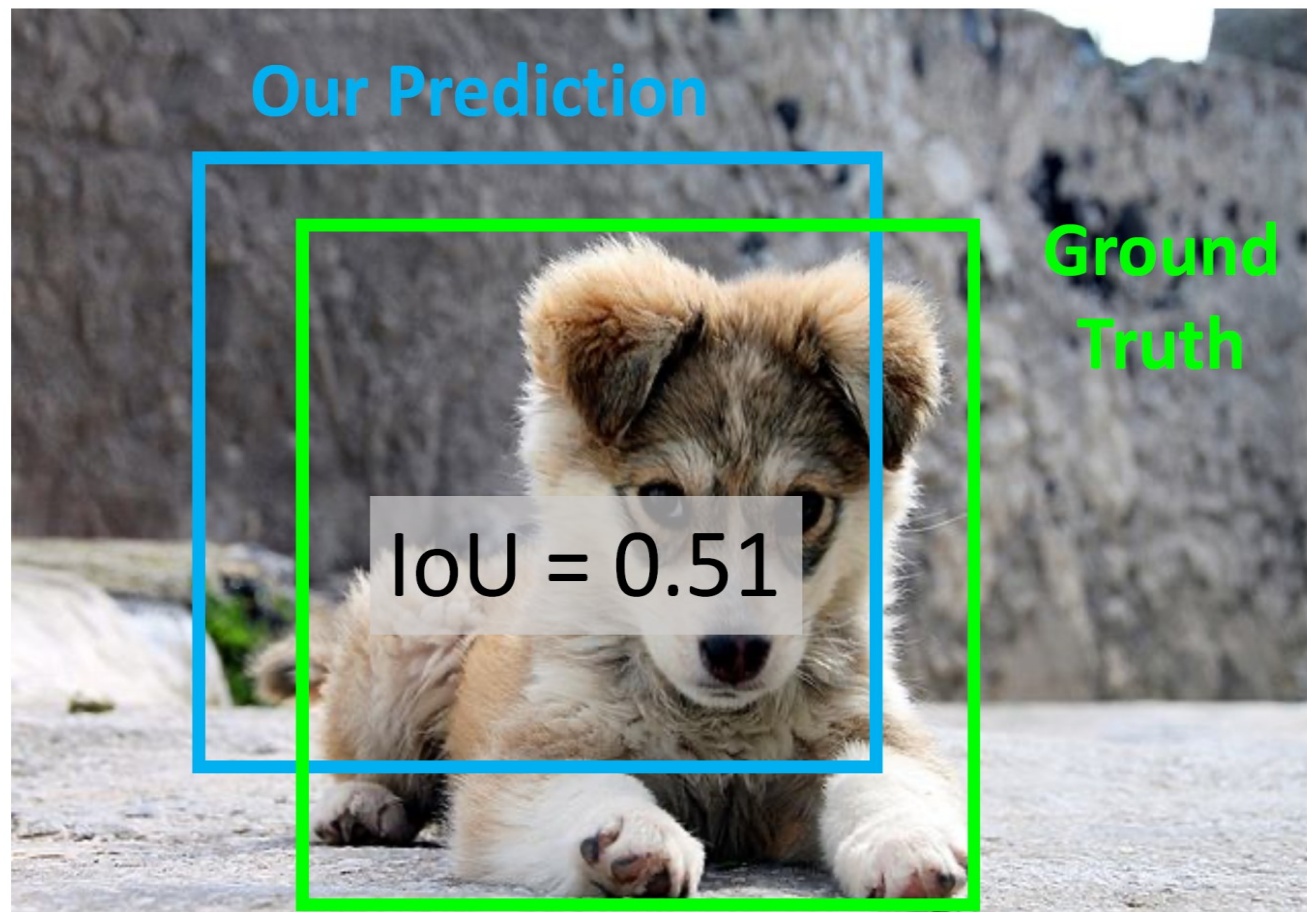
# Comparing Boxes: Intersection over Union (IoU)

How can we compare our prediction to the ground-truth box?

**Intersection over Union (IoU)**  
(Also called “Jaccard similarity” or “Jaccard index”):

$$\frac{\text{Area of Intersection}}{\text{Area of Union}}$$

$\text{IoU} > 0.5$  is “decent”



[Puppy image](#) is licensed under [CC-A 2.0 Generic license](#). Bounding boxes and text added by Justin Johnson.

Slide from Deep Learning for Computer Vision @Univ. of Michigan



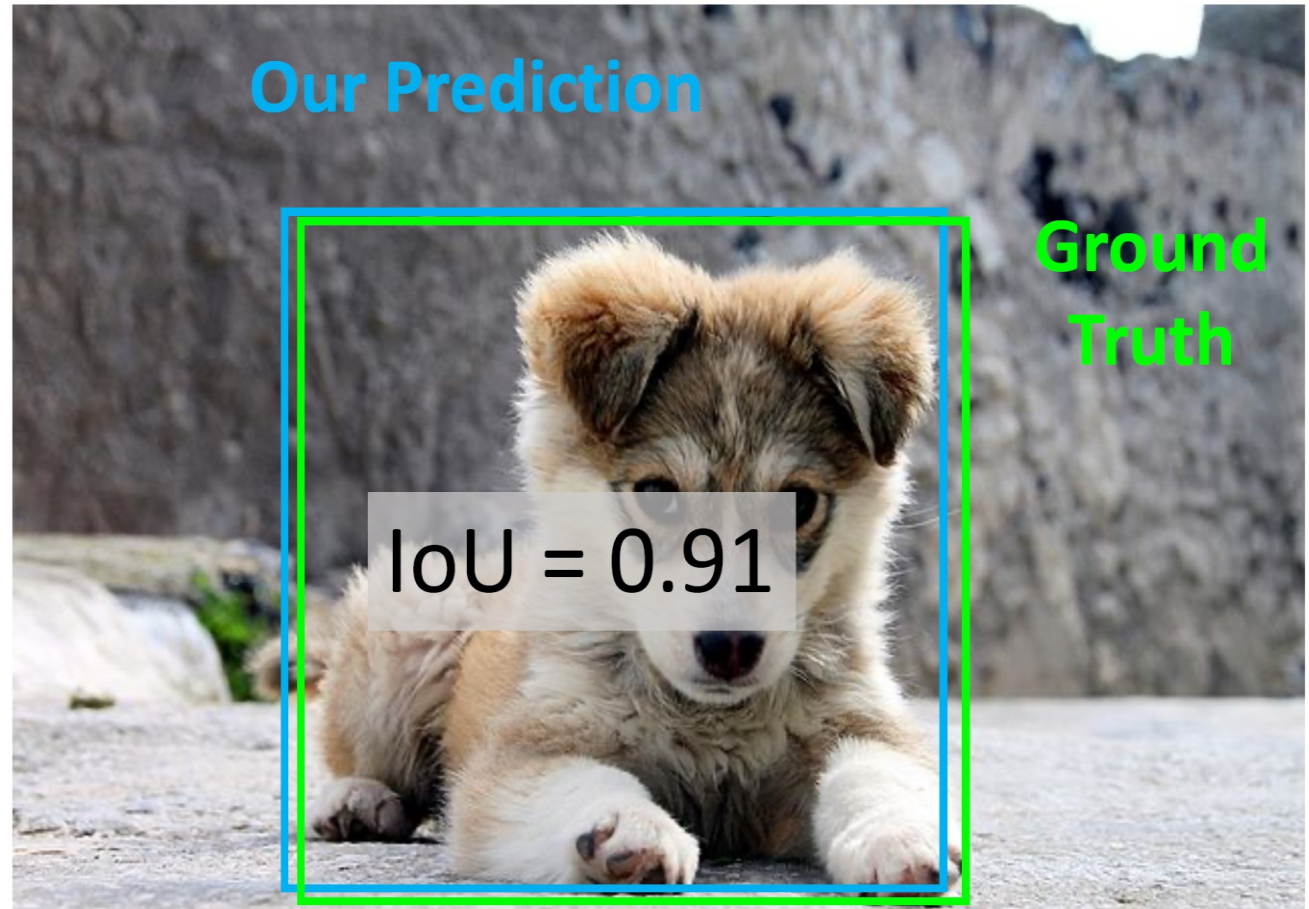
# Comparing Boxes: Intersection over Union (IoU)

How can we compare our prediction to the ground-truth box?

**Intersection over Union (IoU)**  
(Also called “Jaccard similarity” or “Jaccard index”):

$$\frac{\text{Area of Intersection}}{\text{Area of Union}}$$

IoU > 0.5 is “decent”,  
IoU > 0.7 is “pretty good”,  
IoU > 0.9 is “almost perfect”



[Puppy image](#) is licensed under [CC-A 2.0 Generic license](#). Bounding boxes and text added by Justin Johnson.

Slide from Deep Learning for Computer Vision @Univ. of Michigan

**Fully  
Connected:**  
4096 to 4

$(x, y, w, h)$

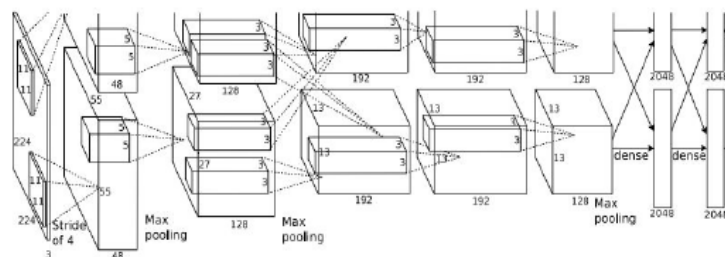
**Correct box:**  
(x', y', w', h')

Copyright **FASTCAMPUS** Corp. All Rights Reserved

# Classification + Localization



This image is CC0 public domain



Fully  
Connected:  
4096 to 1000

**Class Scores**

Cat: 0.9  
Dog: 0.05  
Car: 0.01  
...

Correct label:  
Cat

**Softmax  
Loss**

**Multitask Loss**

**+** → **Loss**

**Vector:**  
4096  
Fully  
Connected:  
4096 to 4

**Box**

**Coordinates**  
(x, y, w, h)

**L2 Loss**

Correct box:  
(x', y', w', h')

Treat localization as a  
regression problem!