YOU LOOK ONLY ONCE (YOLO)

A Hands-on Introduction to the State of the Art Realtime Object Detection Deep Learning Model

Bradley Erickson

MD PhD

Pouria Rouzrokh

MD MPH MHPE

Radiology Informatics Laboratory
Mayo Clinic

What to Expect

Our goal for today:

To learn about an object detection deep learning model, called “YOLO”!

What we cannot do in 60 minutes:

To discuss every single line of notebook code / basics of DL programming!

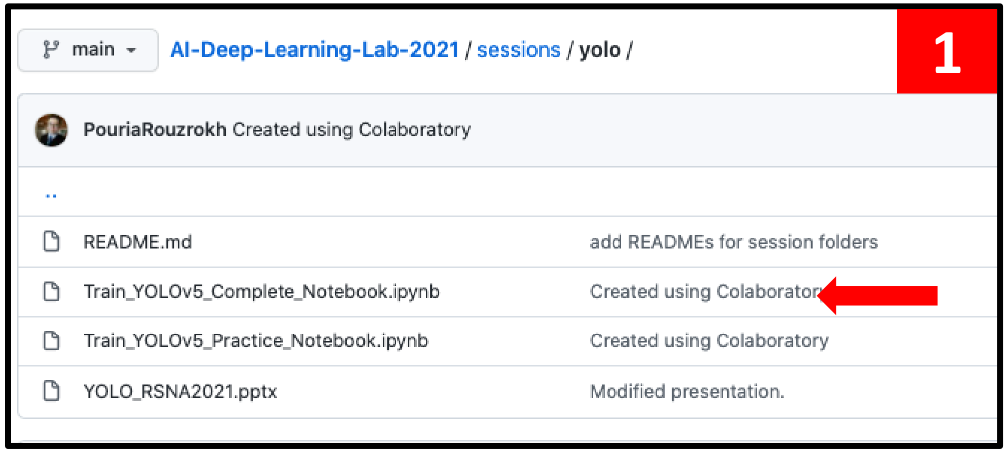
- The notebook will be accessible to you with full explanation in addition to the code.
- Ultralytics documentations include all the code details for implementing YOLO.

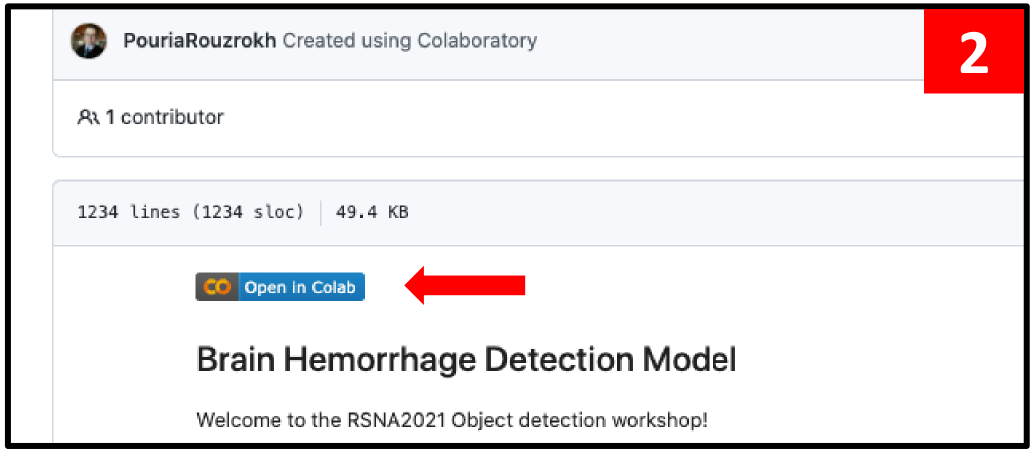
What we can do in 60 minutes:

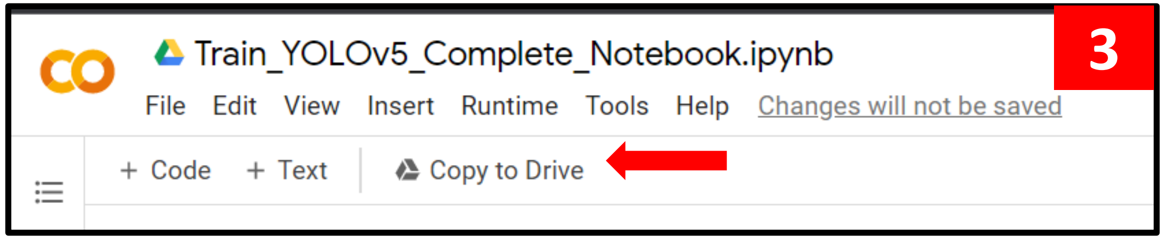
- To talk about the **big picture** of YOLO and how to implement YOLOv5.
- To talk about **handy tips** for obtaining a better performance from YOLOv5!

Let's Open Our Notebook!

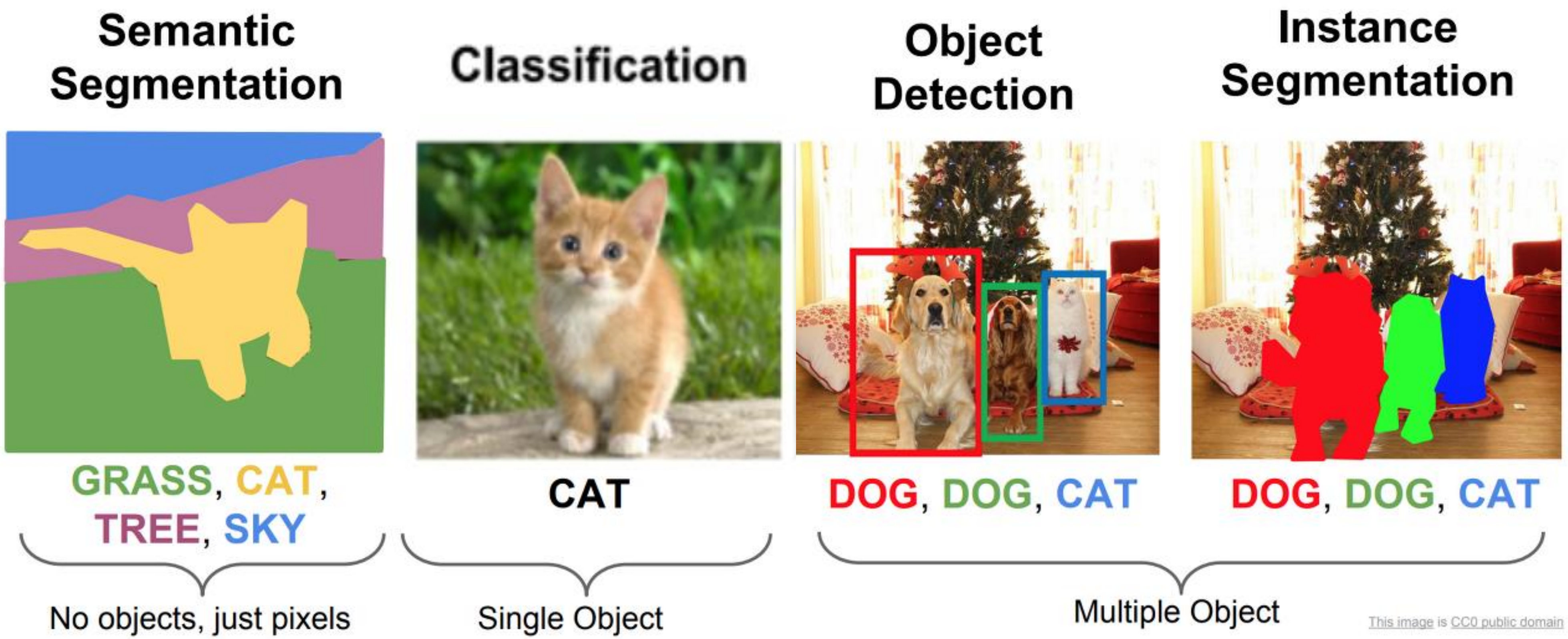
https://tinyurl.com/28bp6nm8



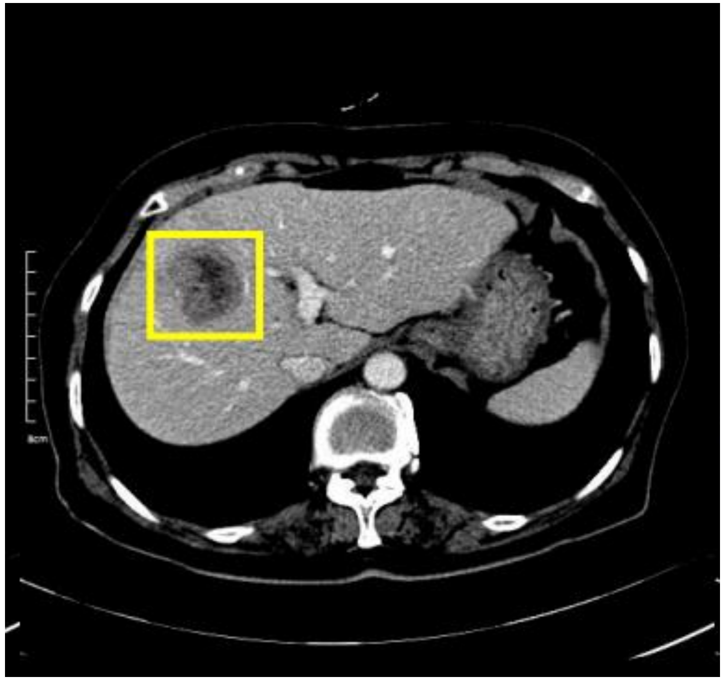


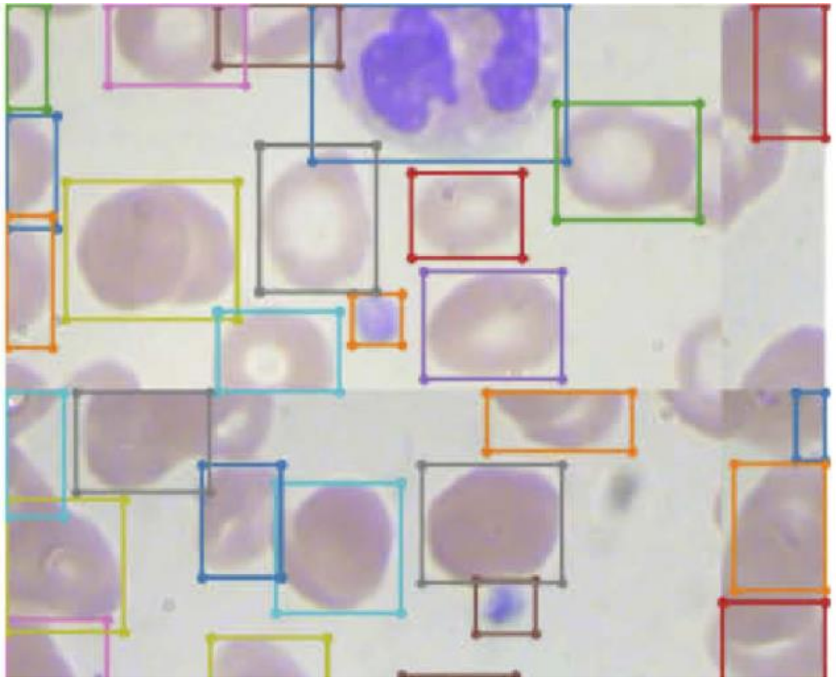


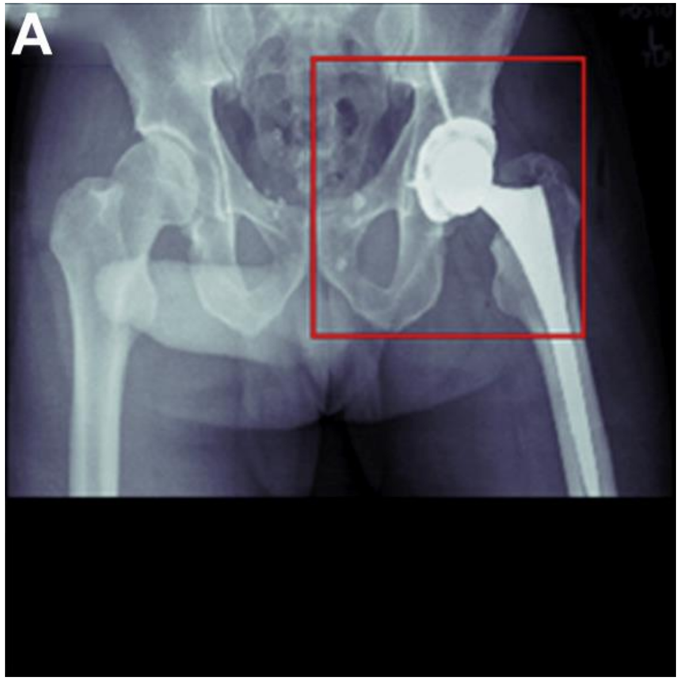
Defining Object Detection



Medical Applications of Object Detection



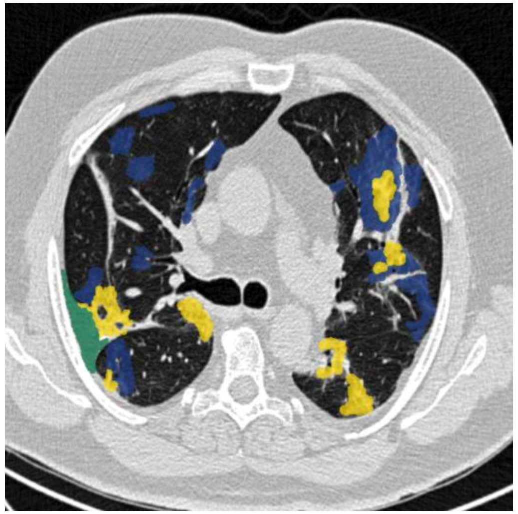


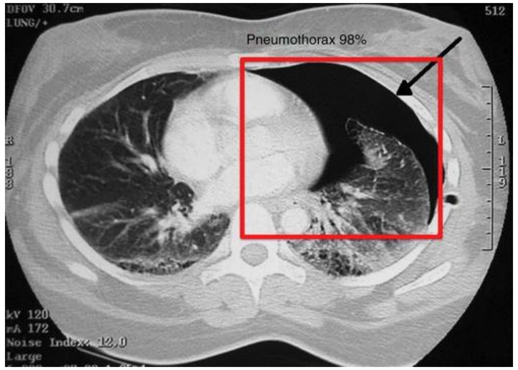


Sources:
- Liver Lesion Detection from Weakly-labeled Multi-phase CT Volumes with a Grouped Single Shot MultiBox Detector
- Improved detection performance in blood cell count by an attention-guided deep learning method
- Deep Learning Artificial Intelligence Model for Assessment of Hip Dislocation Risk Following Primary Total Hip Arthroplasty From Postoperative Radiographs

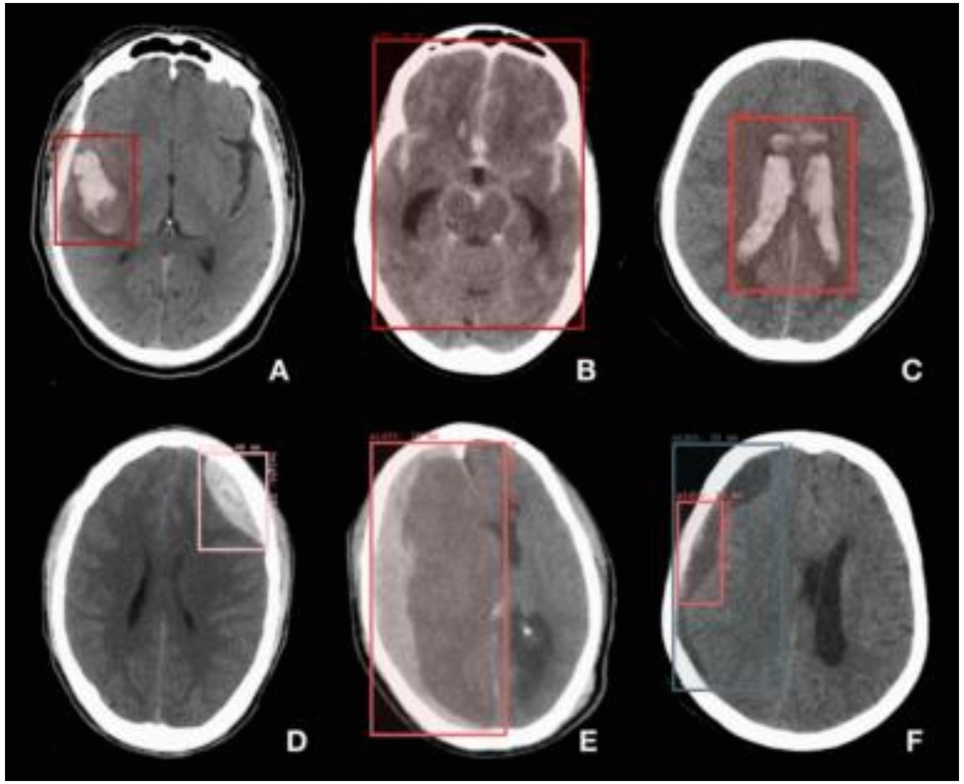
Object Detection vs. Object Segmentation

Object Detection	Semantic Segmentation
Does patch (bounding box)-level predictions.	Does pixel-level prediction.
Does not identify the exact borders or volume of an object.	Can identify the exact borders (and volume) of an object.
Less tedious manual labeling: Simply draw a bounding box!	More tedious manual labeling: Manually segment each pixel of interest!





What We Do Today!



To train a deep learning model to detect brain hemorrhage lesion on Head CT scans!

For our data, we use the publicly available CQ500 Head CT-scan dataset of patients with brain hemorrhage.

For our labels, we use the bounding box annotation on CQ500 dataset by Physionet.

Our Notebook Workflow

- **Part 0: Setting up the working directory**
- **Part 1: Extracting CT images from DICOM files**
- **Part 2: Collecting and visualizing the annotated bounding box labels**
- **Part 3: Data splitting and setting up the training and validation sets**
- **Part 4: Downloading the YOLO model and configuring it**
- **Part 5: View and modify the hyperparameters (optional)**
- **Part 6: Training**
- **Part 7: Inference**

80% of code!

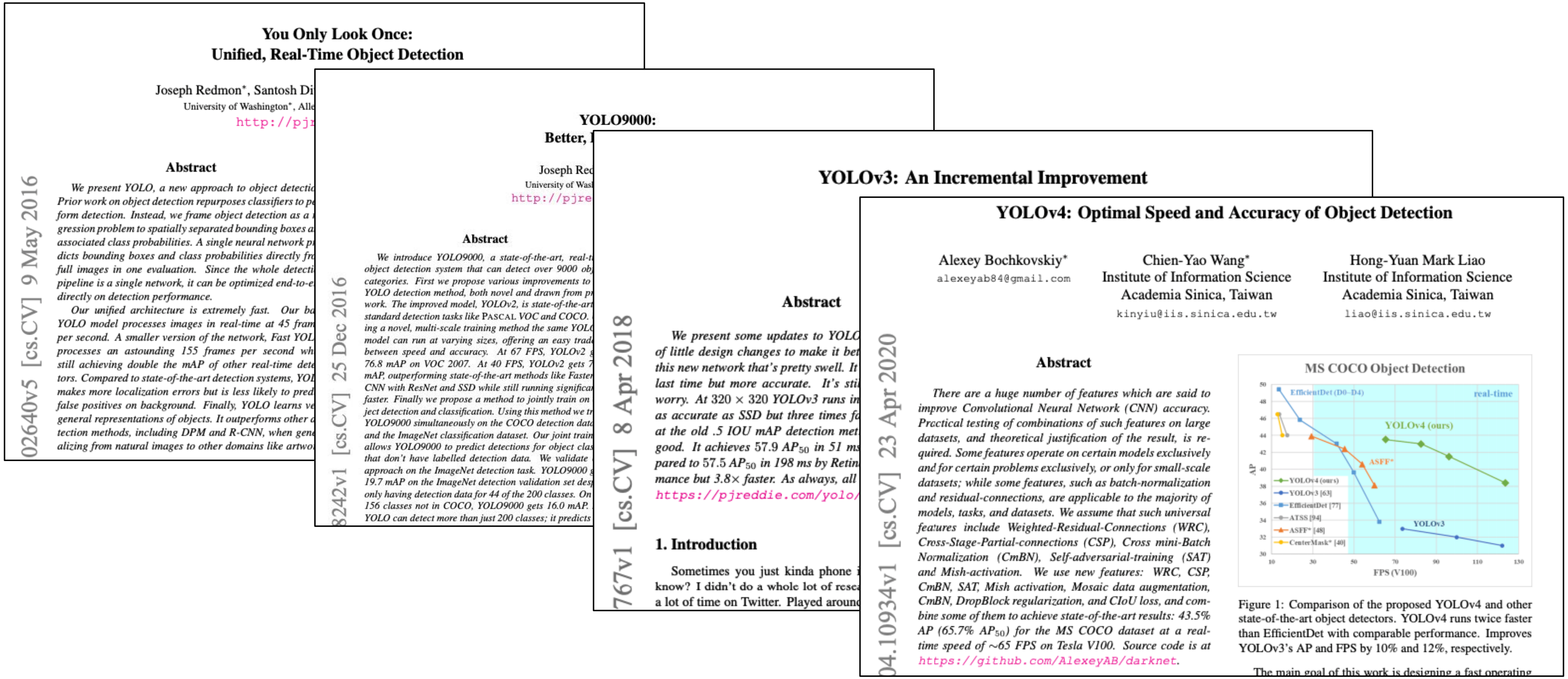
20% of code!

You Look Only Once (YOLO)

- The original YOLO model was the first object detection network to combine the problem of drawing bounding boxes and identifying class labels **in a single pass of training** (different than two-stage models like Faster RCNN)

- YOLO models are often **very fast and very small** – often making them faster to train and easier to deploy, especially to compute-limited edge devices.

You Look Only Once (YOLO)



You Only Look Once:
Unified, Real-Time Object Detection

Joseph Redmon*, Santosh Di
University of Washington*, Alle
http://pj

02640v5 [cs.CV] 9 May 2016

Abstract

*We present YOLO, a new approach to object detecti
Prior work on object detection repurposes classifiers to p
form detection. Instead, we frame object detection as a
gression problem to spatially separated bounding boxes a
associated class probabilities. A single neural network p
dicts bounding boxes and class probabilities directly fr
full images in one evaluation. Since the whole detecti
pipeline is a single network, it can be optimized end-to-e
directly on detection performance.*

*Our unified architecture is extremely fast. Our ba
YOLO model processes images in real-time at 45 fram
per second. A smaller version of the network, Fast YOL
processes an astounding 155 frames per second wh
still achieving double the mAP of other real-time dete
tors. Compared to state-of-the-art detection systems, YOL
makes more localization errors but is less likely to pred
false positives on background. Finally, YOLO learns ve
general representations of objects. It outperforms other d
tection methods, including DPM and R-CNN, when gene
alizing from natural images to other domains like artwo*

YOLO9000:
Better, I

Joseph Re
University of Was
http://pjre

8242v1 [cs.CV] 25 Dec 2016

Abstract

*We introduce YOLO9000, a state-of-the-art, real-t
object detection system that can detect over 9000 ob
categories. First we propose various improvements to
YOLO detection method, both novel and drawn from p
work. The improved model, YOLOv2, is state-of-the-ar
standard detection tasks like PASCAL VOC and COCO.
ing a novel, multi-scale training method the same YOL
model can run at varying sizes, offering an easy trad
between speed and accuracy. At 67 FPS, YOLOv2 g
76.8 mAP on VOC 2007. At 40 FPS, YOLOv2 gets 7
mAP, outperforming state-of-the-art methods like Faster
CNN with ResNet and SSD while still running significa
faster. Finally we propose a method to jointly train on
ject detection and classification. Using this method we t
YOLO9000 simultaneously on the COCO detection dat
and the ImageNet classification dataset. Our joint trai
allows YOLO9000 to predict detections for object cla
that don't have labelled detection data. We validate
approach on the ImageNet detection task. YOLO9000
19.7 mAP on the ImageNet detection validation set des
only having detection data for 44 of the 200 classes. On
156 classes not in COCO, YOLO9000 gets 16.0 mAP.
YOLO can detect more than just 200 classes; it predicts*

YOLOv3: An Incremental Improvement

767v1 [cs.CV] 8 Apr 2018

Abstract

*We present some updates to YOLO
of little design changes to make it bet
this new network that's pretty swell. It
last time but more accurate. It's sti
worry. At 320 × 320 YOLOv3 runs in
as accurate as SSD but three times fa
at the old .5 IOU mAP detection met
good. It achieves 57.9 AP_{50} in 51 ms
pared to 57.5 AP_{50} in 198 ms by Retin
mance but 3.8× faster. As always, all
https://pjreddie.com/yolo/*

1. Introduction

Sometimes you just kinda phone i
know? I didn't do a whole lot of resea
a lot of time on Twitter. Played aroun

YOLOv4: Optimal Speed and Accuracy of Object Detection

Alexey Bochkovskiy*
alexeyab84@gmail.com

Chien-Yao Wang*
Institute of Information Science
Academia Sinica, Taiwan
kinyiu@iis.sinica.edu.tw

Hong-Yuan Mark Liao
Institute of Information Science
Academia Sinica, Taiwan
liao@iis.sinica.edu.tw

04.10934v1 [cs.CV] 23 Apr 2020

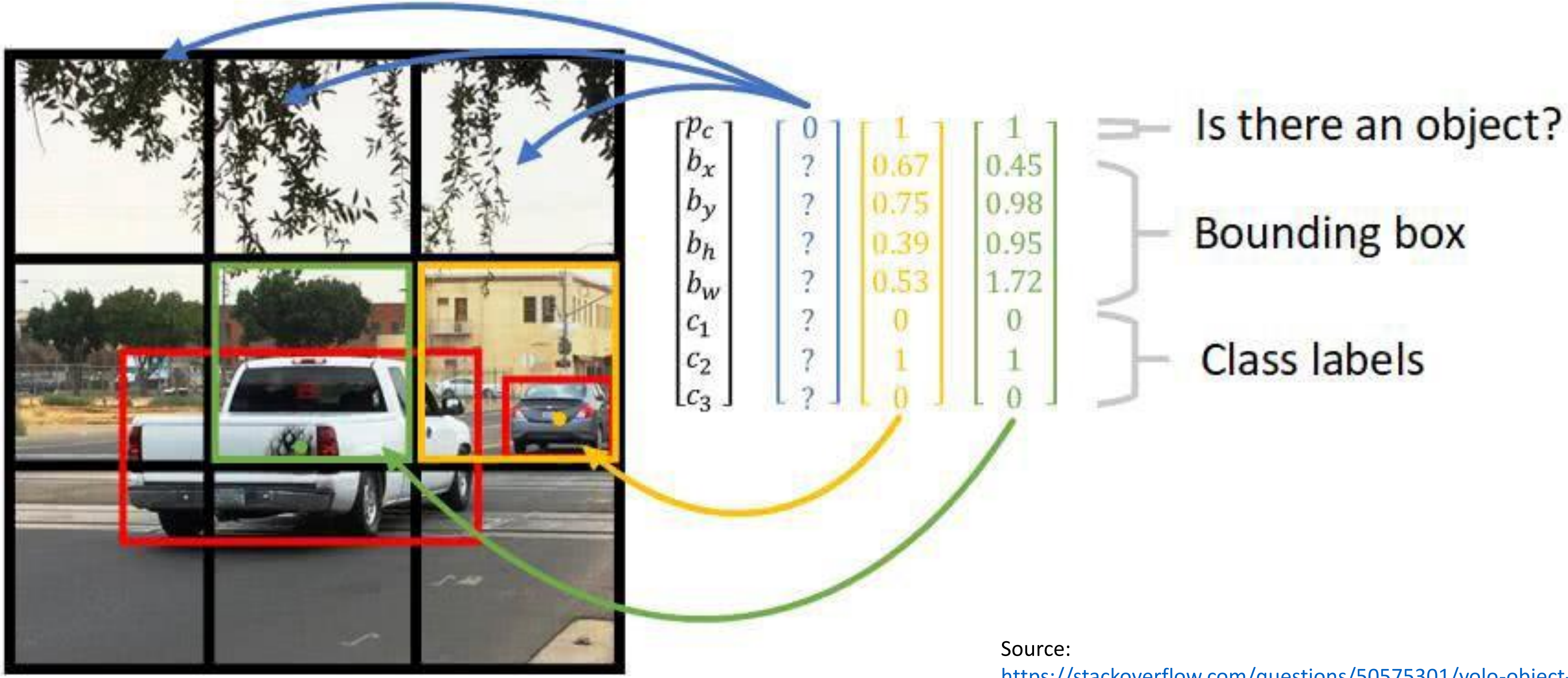
Abstract

There are a huge number of features which are said to improve Convolutional Neural Network (CNN) accuracy. Practical testing of combinations of such features on large datasets, and theoretical justification of the result, is required. Some features operate on certain models exclusively and for certain problems exclusively, or only for small-scale datasets; while some features, such as batch-normalization and residual-connections, are applicable to the majority of models, tasks, and datasets. We assume that such universal features include Weighted-Residual-Connections (WRC), Cross-Stage-Partial-connections (CSP), Cross mini-Batch Normalization (CmBN), Self-adversarial-training (SAT) and Mish-activation. We use new features: WRC, CSP, CmBN, SAT, Mish activation, Mosaic data augmentation, CmBN, DropBlock regularization, and CIoU loss, and combine some of them to achieve state-of-the-art results: 43.5% AP (65.7% AP_{50}) for the MS COCO dataset at a realtime speed of ~65 FPS on Tesla V100. Source code is at https://github.com/AlexeyAB/darknet.

Figure 1: Comparison of the proposed YOLOv4 and other state-of-the-art object detectors. YOLOv4 runs twice faster than EfficientDet with comparable performance. Improves YOLOv3's AP and FPS by 10% and 12%, respectively.

The main goal of this work is designing a fast operating

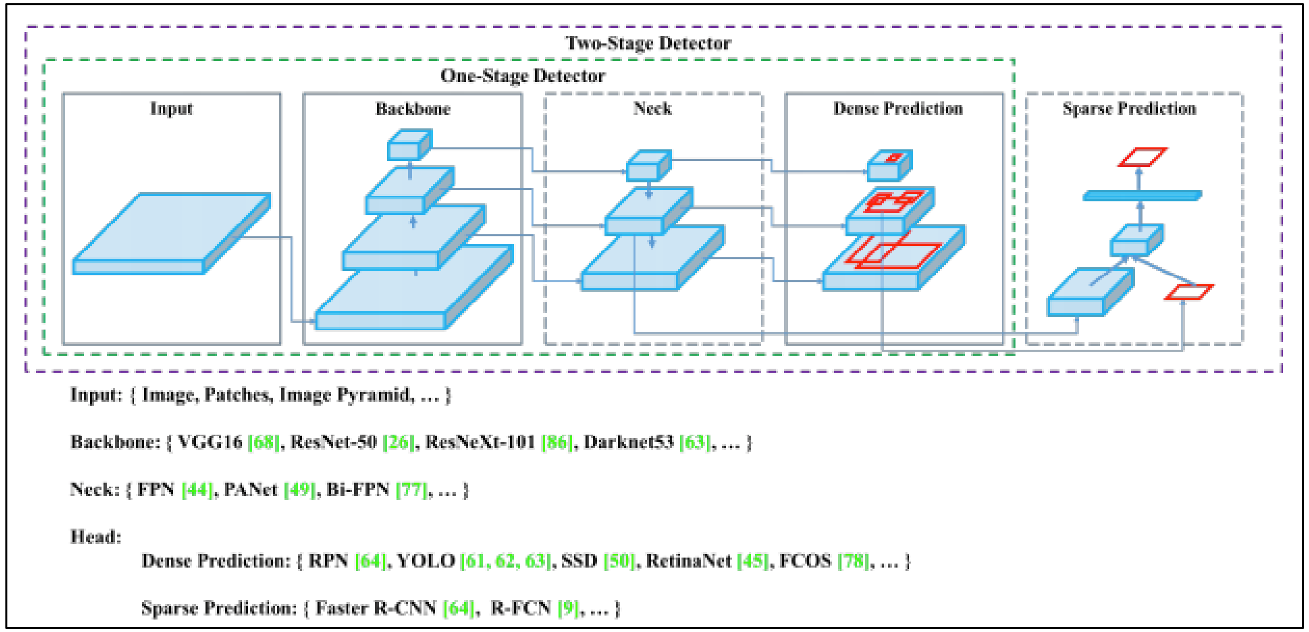
Fundamental Idea of YOLO



Source:
https://stackoverflow.com/questions/50575301/yolo-object-detection-how-does-the-algorithm-predict-bounding-boxes-larger-than

Well, YOLO has evolved a lot!

- **New versions of YOLO use advanced architectures, data augmentation techniques, and training strategies! Please refer to their papers for understanding the novelties.**



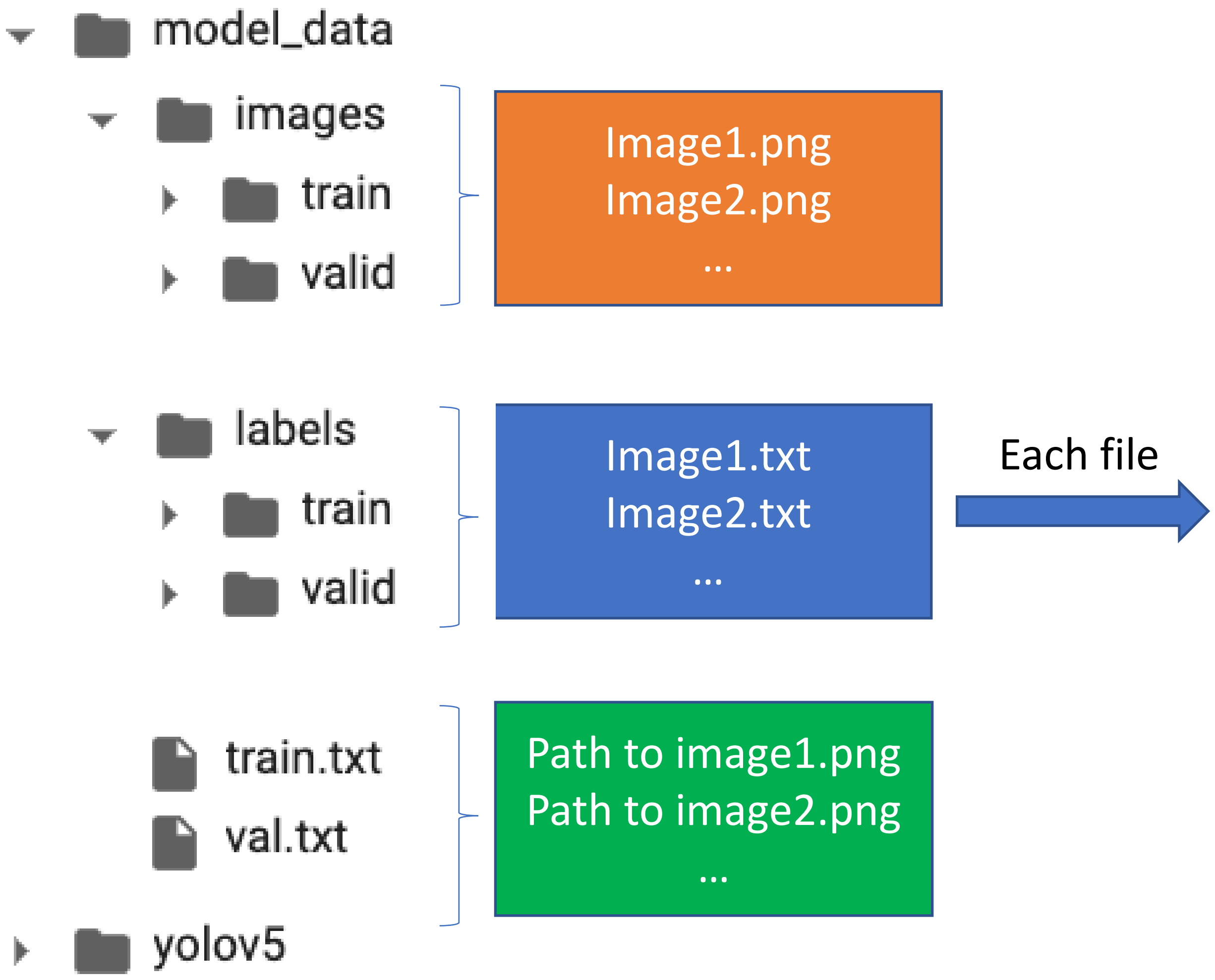
YOLOv5

- One of the latest versions of YOLO.
- No scientific manuscript released yet.
- Based on YOLOv4 + some technical improvements.
- Based on **Pytorch** (as opposed to Darknet), and therefore significantly smaller than YOLOv4.
- Easy implementation for custom training and helpful documentation/ tutorials by **Ultralytics**.

Big Picture of Training a YOLOv5 Model

- Ultralytics has implemented YOLOv5 as a ready to train package, i.e., training and inference with it take only 1 to 2 lines of code!
- However, most of your time will be spent on collecting and organizing your data as YOLOv5 expects it to be.
- When your data is ready, you should clone the YOLOv5 model directory from the Ultralytics GitHub page, and then configure it to know your data.
- You can select different size variants of YOLOv5, with or without transfer learning from the models trained on ImageNet.

Data Organization

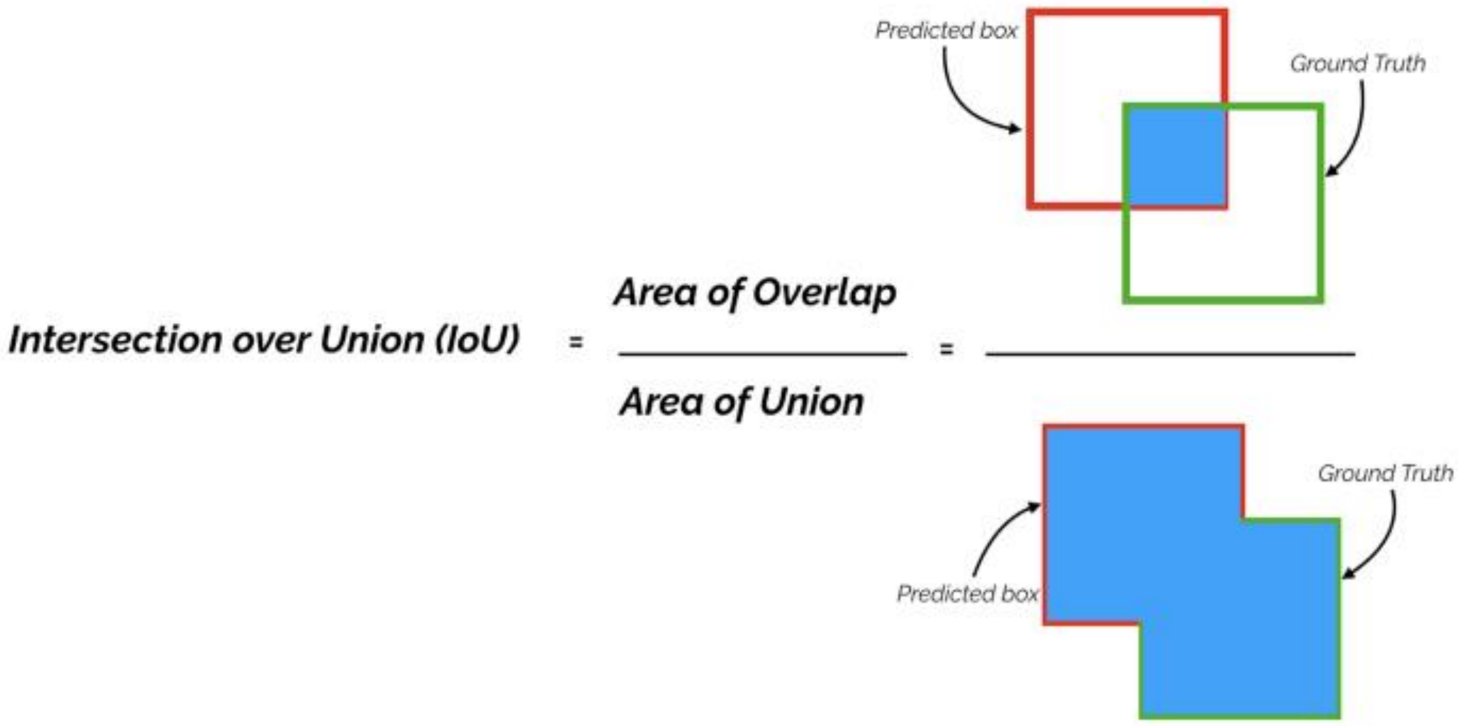


- One row per object
- Each row is `class x_center y_center width height` format.
- Box coordinates must be in **normalized xywh** format (from 0 - 1). If your boxes are in pixels, divide `x_center` and `width` by image width, and `y_center` and `height` by image height.
- Class numbers are zero-indexed (start from 0).

Inside Image1.txt

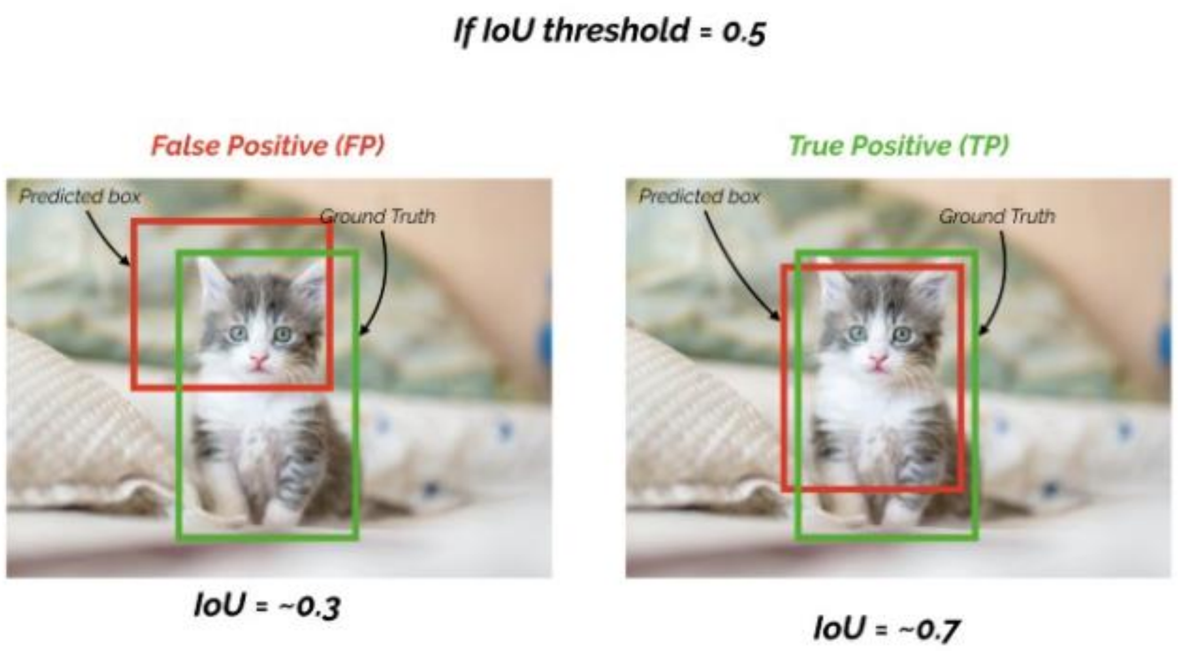
0 0.21 0.45 0.18 0.23
1 0.6 0.74 0.2 0.1

Intersection over Union (IOU)



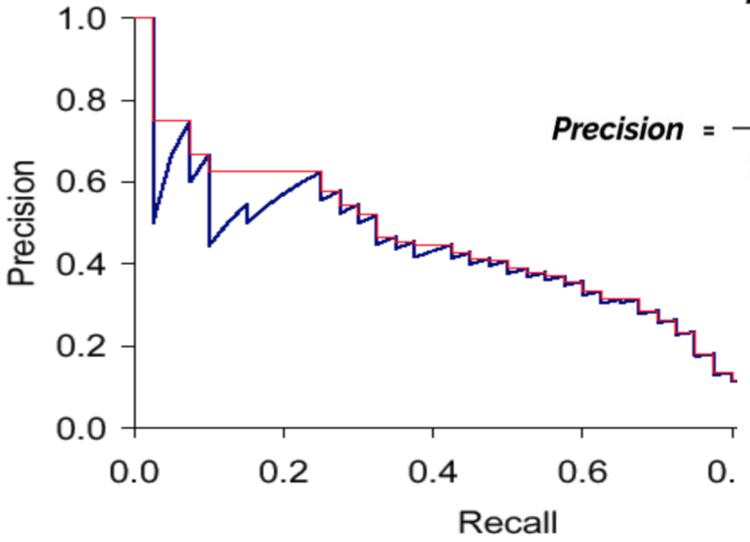
Source: https://towardsdatascience.com/map-mean-average-precision-might-confuse-you-5956f1bfa9e2

Mean Average Precision (mAP)



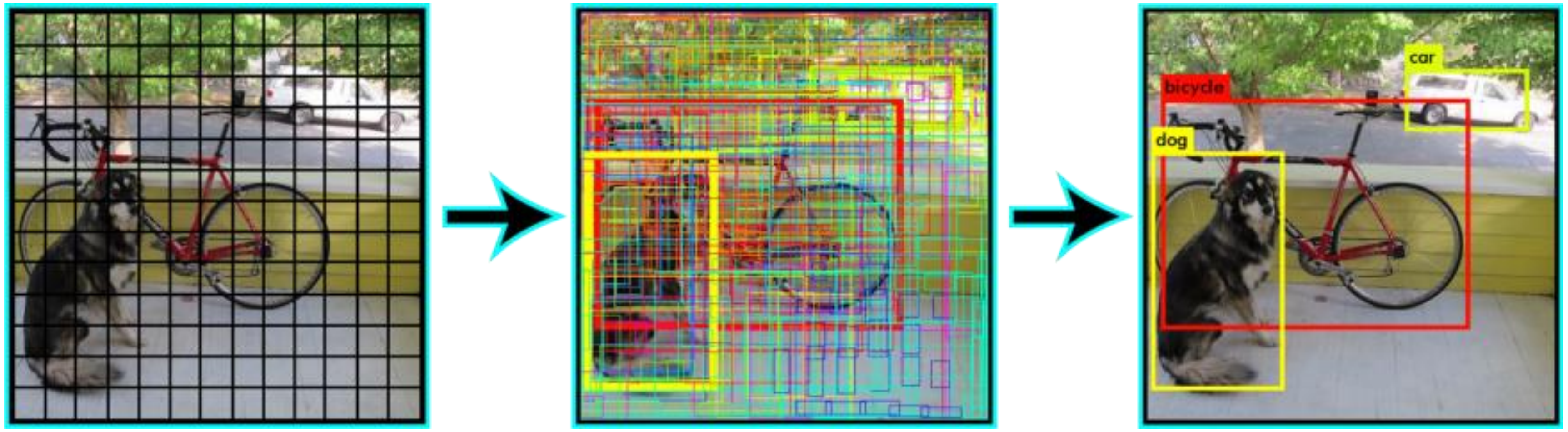
$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$



The area under the Precision-Recall curve for an object detection model is called the "Average Precision" of that model. If you take the mean of these "Average Precision" scores across all classes which the model predicts, you will achieve a metric for evaluating your model: mean Average Precision (mAP)!

Non-max Suppression



Non-Max Suppression:

To select the best bounding box, from the multiple predicted bounding boxes, these object detection algorithms use non-max suppression.
In this technique, if two bounding boxes have an IoU over a certain threshold (e.g., 60%), the bounding box with the lower confidence score will be discarded.

Source: https://www.pyimagesearch.com/2018/11/12/yolo-object-detection-with-opencv/

Thank you for your attention!