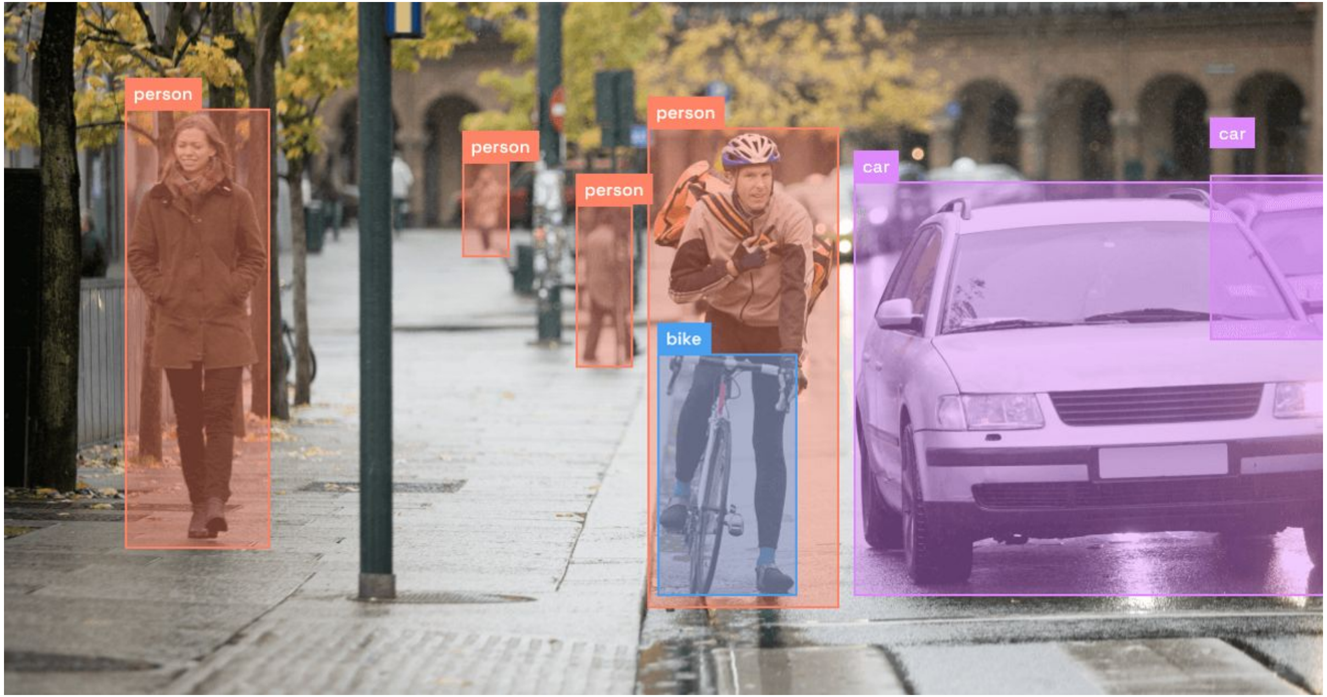
CSE4261: Neural Network and Deep Learning

Lecture: 25.06.2025

Sangeeta Biswas, Ph.D.
Associate Professor,
University of Rajshahi, Rajshahi-6205, Bangladesh

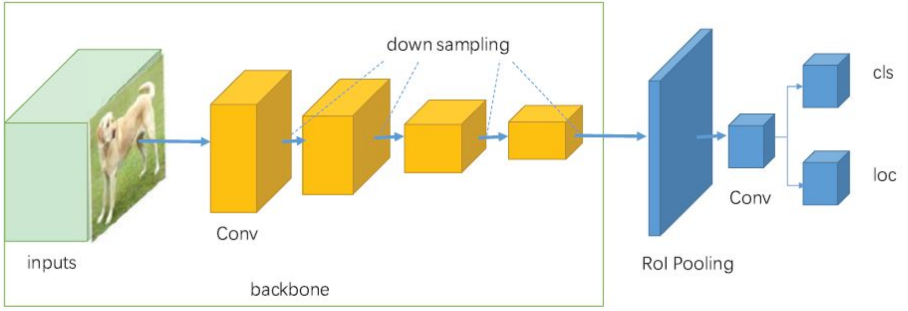
Object Detection



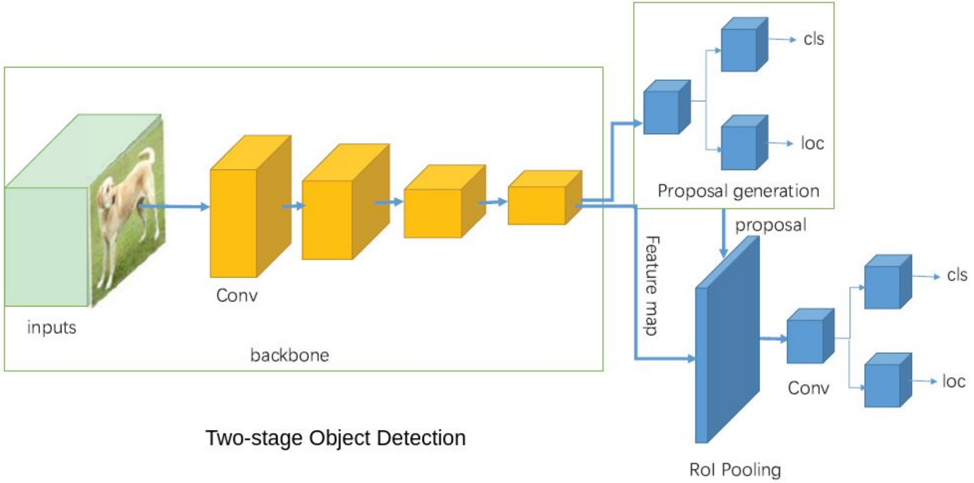
Types of Object Detector

Depending on how many times input image is seen by the object detector system, it is divided into two categories:

1. One Stage Object Detector:
   a. YOLO, SSD
2. Two-Stage Object Detector:
   a. R-CNN, Fast R-CNN, Faster R-CNN



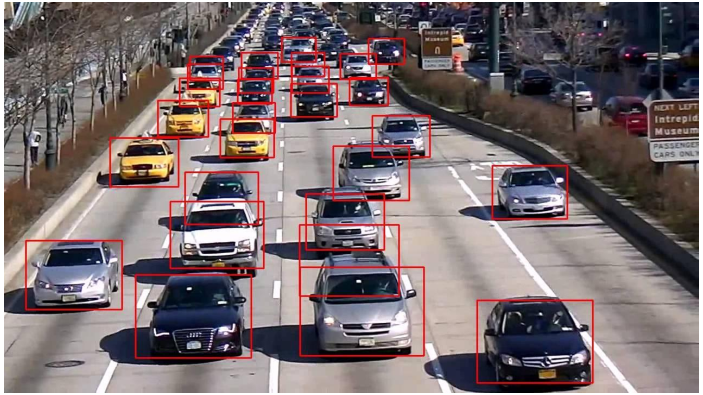
One-stage Object Detection



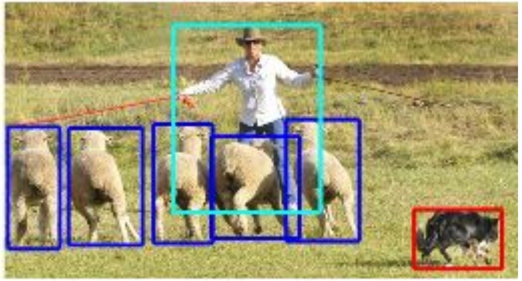
Two-stage Object Detection

One-Stage vs. Two-Stage Object Detector

- One-stage object detection is better suited for speedy real-time applications



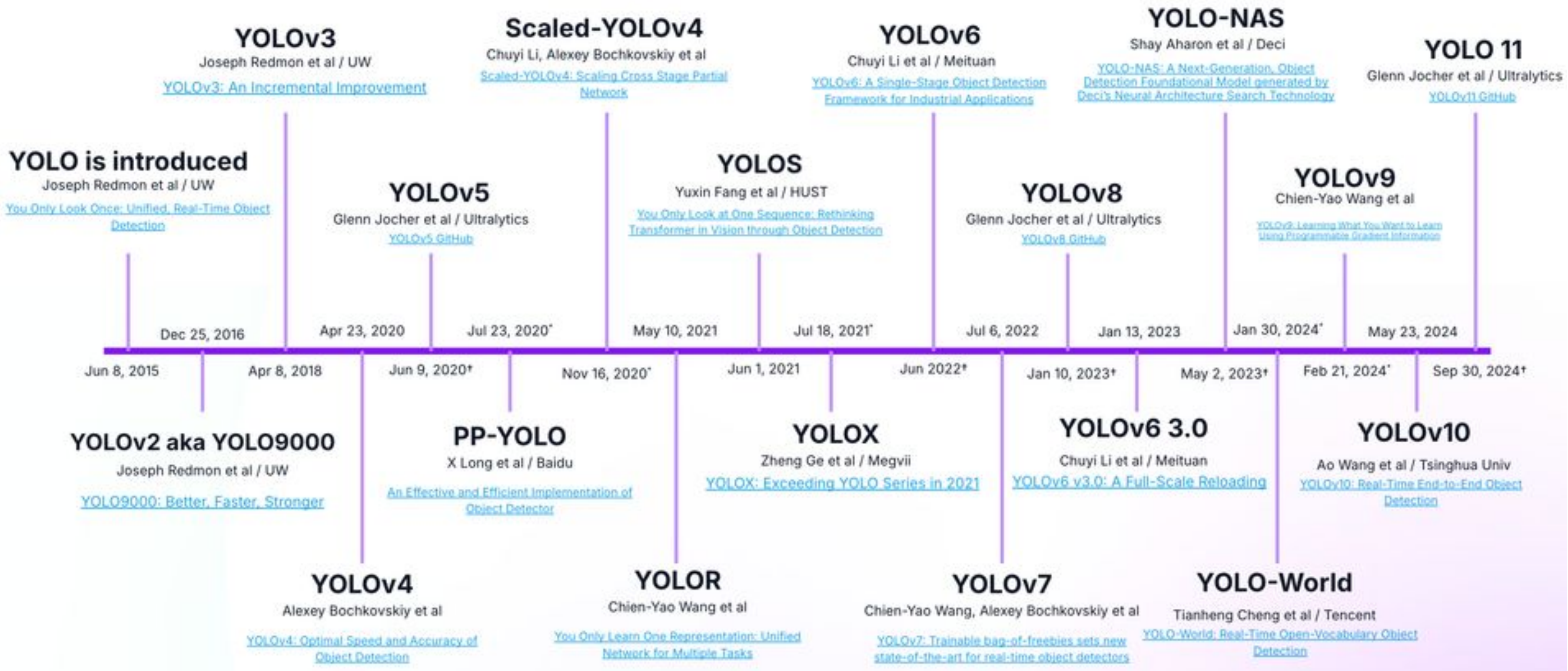
- Two-stage object detection is better for applications where accuracy is more important than speed.
- Two-stage is more computationally expensive than one-stage.



YOLO

- **Y**ou **O**nly **L**ook **O**nce
- It is a one-stage object detector
- It is the most popular **real-time** object detection system.
- It processes images in one time. Therefore, it is faster than two-stage object detection systems such as Faster R-CNN.
  - YOLOv8 has a GPU latency of 1.3ms, while Faster R-CNN has a GPU latency of 54ms.
  - YOLOv3 can operate in real time with a frames per second (FPS) that is more than eight times faster than Faster R-CNN.

Different Versions of YOLO



Jun 8, 2015
YOLO is introduced
Joseph Redmon et al / UW
You Only Look Once: Unified, Real-Time Object Detection

Dec 25, 2016
YOLOv2 aka YOLO9000
Joseph Redmon et al / UW
YOLO9000: Better, Faster, Stronger

Apr 8, 2018
YOLOv3
Joseph Redmon et al / UW
YOLOv3: An Incremental Improvement

Apr 23, 2020
YOLOv4
Alexey Bochkovskiy et al
YOLOv4: Optimal Speed and Accuracy of Object Detection

Jun 9, 2020†
YOLOv5
Glenn Jocher et al / Ultralytics
YOLOv5 GitHub

Jul 23, 2020*
PP-YOLO
X Long et al / Baidu
An Effective and Efficient Implementation of Object Detector

Nov 16, 2020*
Scaled-YOLOv4
Chuyi Li, Alexey Bochkovskiy et al
Scaled-YOLOv4: Scaling Cross Stage Partial Network

May 10, 2021
YOLOR
Chien-Yao Wang et al
You Only Learn One Representation: Unified Network for Multiple Tasks

Jun 1, 2021
YOLOS
Yuxin Fang et al / HUST
You Only Look at One Sequence: Rethinking Transformer in Vision through Object Detection

Jul 18, 2021*
YOLOX
Zheng Ge et al / Megvii
YOLOX: Exceeding YOLO Series in 2021

Jun 2022†
YOLOv6
Chuyi Li et al / Meituan
YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications

Jul 6, 2022
YOLOv7
Chien-Yao Wang, Alexey Bochkovskiy et al
YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors

Jan 10, 2023†
YOLOv8
Glenn Jocher et al / Ultralytics
YOLOv8 GitHub

Jan 13, 2023
YOLOv6 3.0
Chuyi Li et al / Meituan
YOLOv6 v3.0: A Full-Scale Reloading

May 2, 2023†
YOLO-NAS
Shay Aharon et al / Deci
YOLO-NAS: A Next-Generation, Object Detection Foundational Model generated by Deci's Neural Architecture Search Technology

Jan 30, 2024*
YOLO-World
Tianheng Cheng et al / Tencent
YOLO-World: Real-Time Open-Vocabulary Object Detection

Feb 21, 2024*
YOLOv9
Chien-Yao Wang et al
YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information

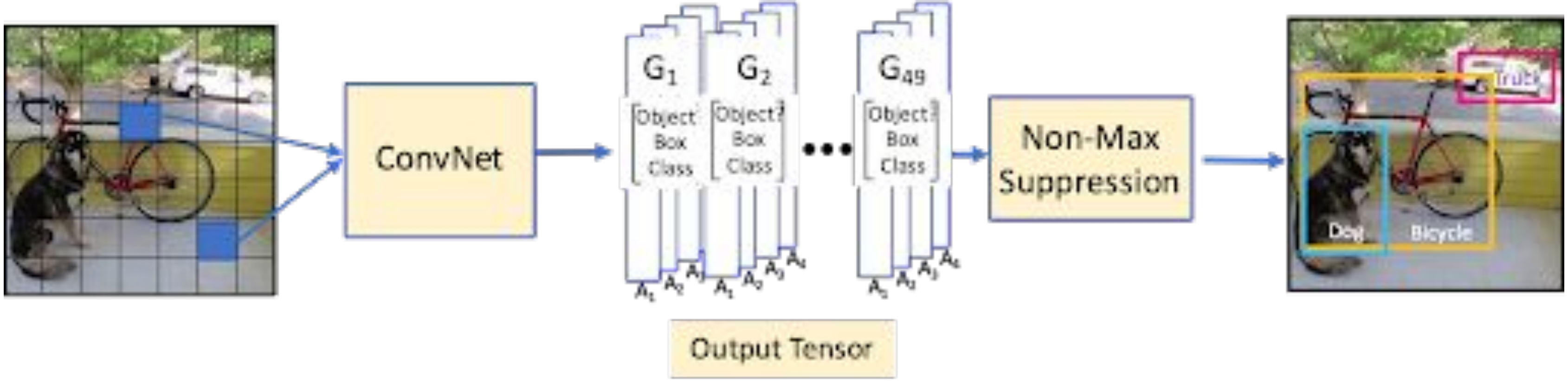
May 23, 2024
YOLOv10
Ao Wang et al / Tsinghua Univ
YOLOv10: Real-Time End-to-End Object Detection

Sep 30, 2024†
YOLO 11
Glenn Jocher et al / Ultralytics
YOLOv11 GitHub

On Feb, 2025, YOLOv12: https://docs.ultralytics.com/models/yolo12/

How YOLO Works

- Instead of using two separate stages for box prediction and category prediction like R-CNN, YOLO employs a single network to simultaneously predict bounding boxes and object categories in a single stage.



How YOLO WOrks

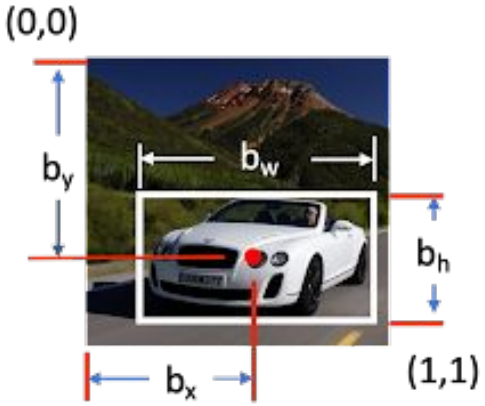
YOLO follows the following steps:

- Step-1: Divide the original image into NxN grid cells of equal shape
- Step-2: Determine the bounding boxes corresponding to all the objects in the image.
- Step-3: Keep relevant grid-box candidates based on Intersection over Union (IoU) value
- Step-4: Keep only the boxes with the highest probability detection score using Non-Max Suppression (NMS) technique

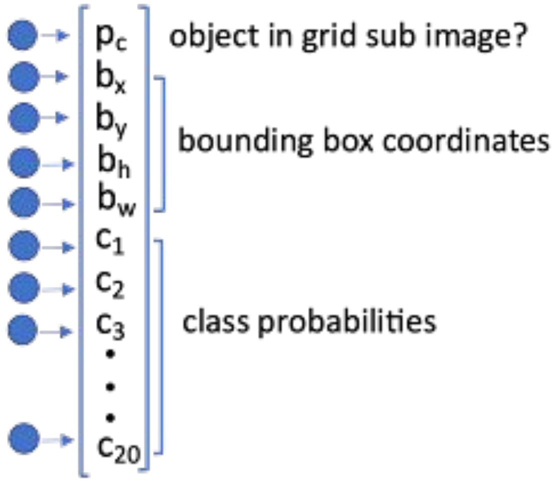
Output Vector of YOLO Model

size_of_output_vector = (5 + size_of_one_hot_vector) * number of objects per grid

Here,
size_of_output_vector= (5+20)*1 = 25



Grid cell
bounding box coordinates

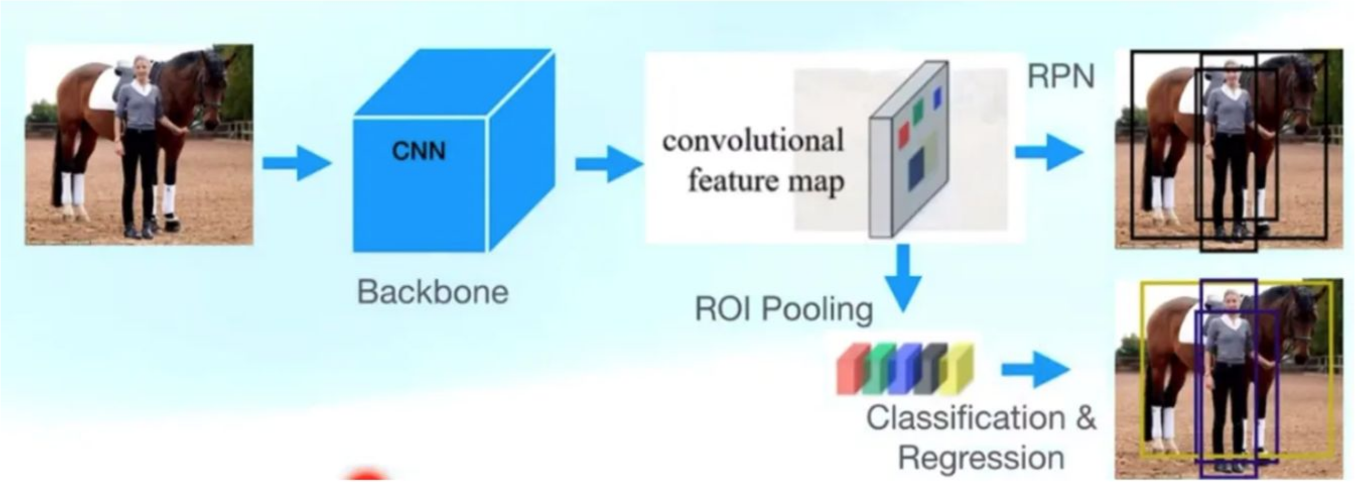


ConvNet output vector

Two Stage Object Detector

Stage-1: Generate bounding boxes to locate potential objects (i.e., region proposals) in an image

Stage-2: Classify bounding boxes to determine the object’s class.



Cons of R-CNN

- Feature extraction and classification need to be done for 2000 region proposals, which are computationally intensive and time-consuming
- Cannot be implemented for large datasets or real-time scenarios.

If a surveillance camera that needs to quickly identify multiple objects, then R-CNN or other two-stage approaches may struggle to provide fast results due to its sequential nature, potentially causing delays in critical applications.