Introduction to Computer Vision

Presented by: Besma Guesmi

Software Engineer Specializing in Computer Vision

"AI is the new electricity. Just as electricity transformed numerous industries starting 100 years ago, AI is now poised to do the same."

–Andrew Ng

Co-Founder @Coursea

OUR Agenda

Intro to Computer Vision

Intro to Object detection

YOLO

Let’s practice

What is Computer Vision?

01

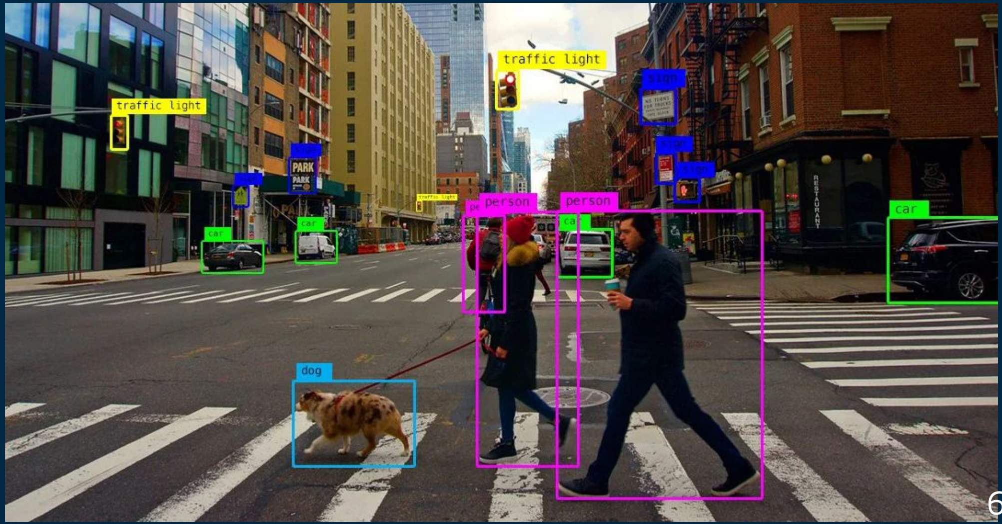
What is Computer Vision?

- A field of Computer Science focuses on analysing and processing visual images or videos intelligently like human
- Enable machines to learn and understand the images at pixel level through training and validation
- Machines retrieve visual information, handle it and interpret the result by getting training from special ML Software algorithms.

What is Computer Vision?

Make computers understand images and video.

- ➢ What kind of scene?
- ➢ Where are the cars?
- ➢ How far is the building?



How Computer Vision Works?

02

How Computer Vision works?

Computer Vision works in three basics steps:

- ➢ **Acquiring an image (image capture)**
- ➢ **Preprocessing the image (noise reduction, image enhancement…)**
- ➢ **Understanding the image (Classification, detection, segmentation)**

How Computer Vision works?

- CV algorithms that we use today are based on pattern recognition
- We train computers on massive amount of visual data - computer process the image, label objects on them, and find patterns on those images.

Computer Vision Applications

03

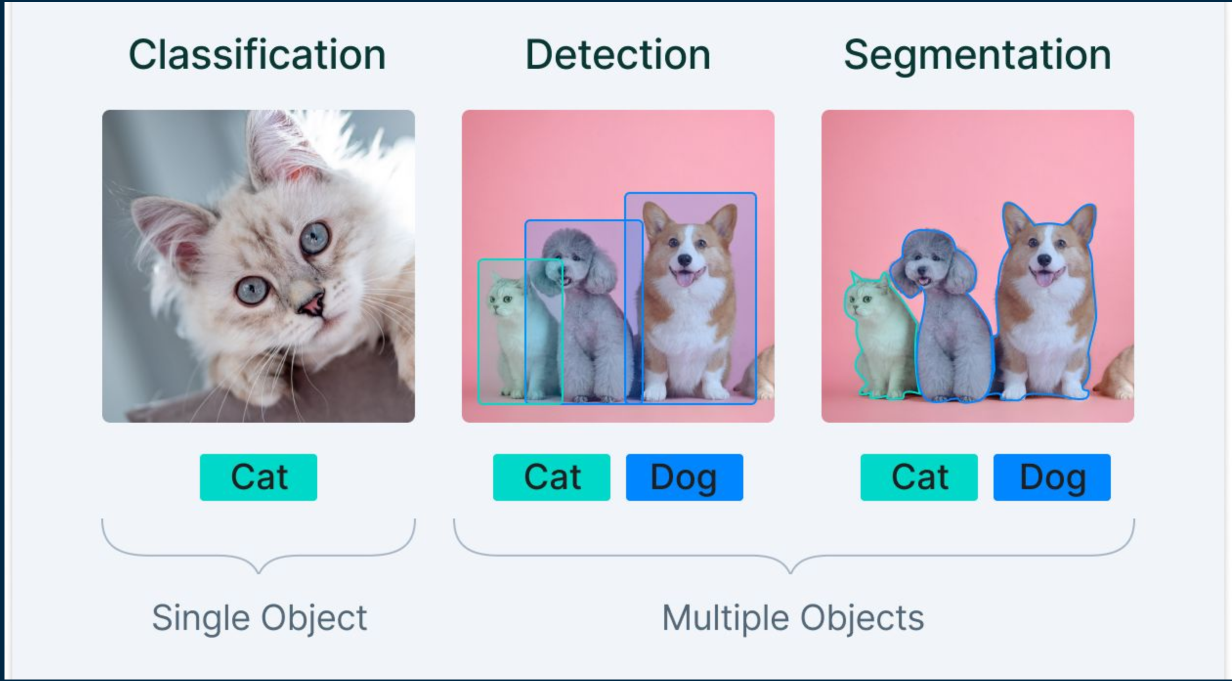
Computer Vision applications

- CV is one of the ML (Machine Learning) area where core concepts are already being integrated into major products that we use every day, like:
  - Self driving cars
  - Facial recognition
  - Healthcare
  - etc

Computer Vision Tasks

04

Computer Vision Tasks



Object detection

05

Object detection

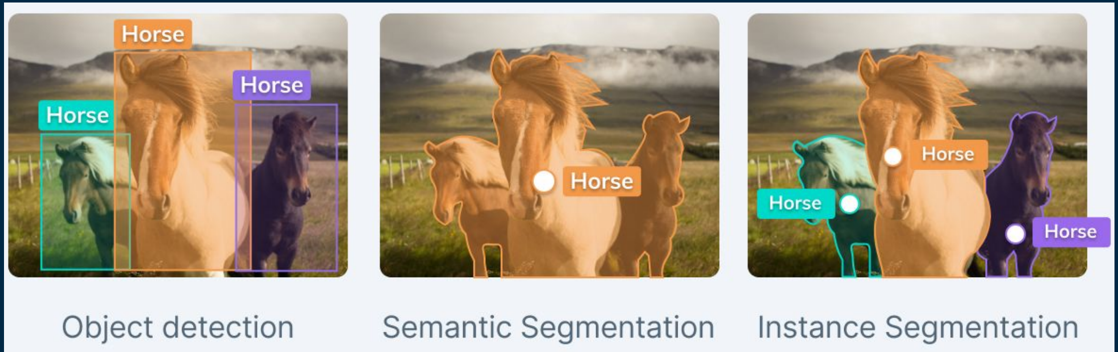
- Object detection is a computer vision task that involves identifying and locating objects within an image or video.
- Object detection comes down to drawing bounding boxes around detected objects, which allow us to locate them in a given scene.

Object detection vs. image classification

- **Image classification** assigns a single label to an entire image using a classifier, but it doesn't pinpoint where the labelled object is.
- **Object detection** goes further by not only classifying objects, but also drawing bounding boxes around them to precisely locate them in the image.

Object detection vs image segmentation

- **Semantic segmentation** marks all pixels of that class, but doesn't outline object boundaries.
- **Object detection** doesn't segment objects, but precisely locates them with bounding boxes.
- **Instance segmentation** combines both by first detecting objects and then segmenting them within the detected boxes, yielding distinct regions for each instance.



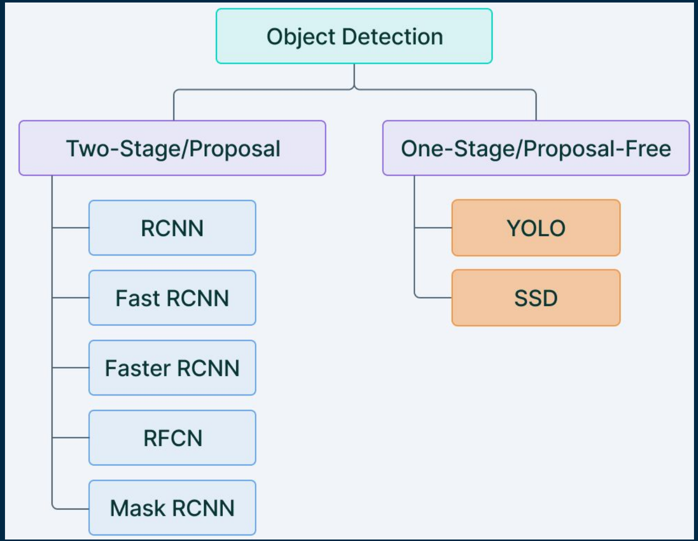
How does object detection work ?

05

How does object detection work?

Object detection generally is categorized into 2 stages:

- ➢ **Single-stage object detectors.**
- ➢ **Two-stage object detectors.**



Single Object detection – YOLO

05

Single Object detection – YOLO

- ➢ A single-stage detector removes the ROI extraction process and directly classifies and regresses the candidate anchor boxes.
- ➢ YOLO (You Only Look Once3): Uses fewer anchor boxes and divides the input image into an S x S grid for regression and classification, developed using darknet neural networks.
- ➢ YOLOv4 (YOLOv3 upgrade) works by breaking the object detection task into two pieces, regression to identify object positioning via bounding boxes and classification to determine the object's class.
- ➢ YOLOv5 is an improved version of YOLOv4 with a mosaic augmentation technique for increasing the general performance of YOLOv4.

Data Preparation – YOLO

05

Data Annotation

The data annotation tools for object detection allow the user to draw bounding boxes around the objects

- ➢ **LabelImg**
- ➢ **VGG Image Annotator (VIA)**
- ➢ **LabelMe**
- ➢ **Roboflow**
- ➢ **CVAT (Computer Vision Annotation Tool)**
- ➢ **Labelbox**

Data Augmentation

Data augmentation techniques are used to artificially increase the size of the dataset.

- ➢ **Rotation: Rotating images by various degrees to simulate different viewpoints.**
- ➢ **Flip: Flipping images horizontally or vertically to create mirror images.**
- ➢ **Scaling: Resizing images to different scales to simulate objects at varying distances.**
- ➢ **Translation: Shifting images horizontally or vertically to simulate object movement.**
- ➢ **Crop and Pad: Cropping and padding images to focus on specific regions or change aspect ratios.**

Data Preparation

Data preprocessing is a crucial step in preparing raw data for machine learning algorithms

- **Data Cleaning: Handling missing data / Outlier detection**
- **Data Normalization**
- **Data Imbalance Handling**
- **Data splitting**
- **Noise reduction**

YOLO – Training

06

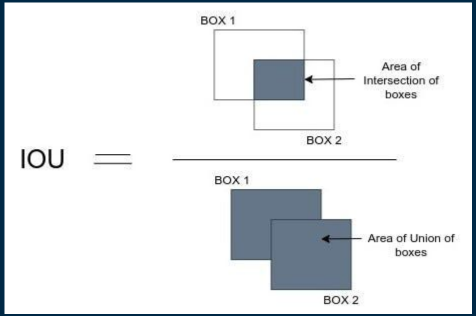
YOLO – Metrics

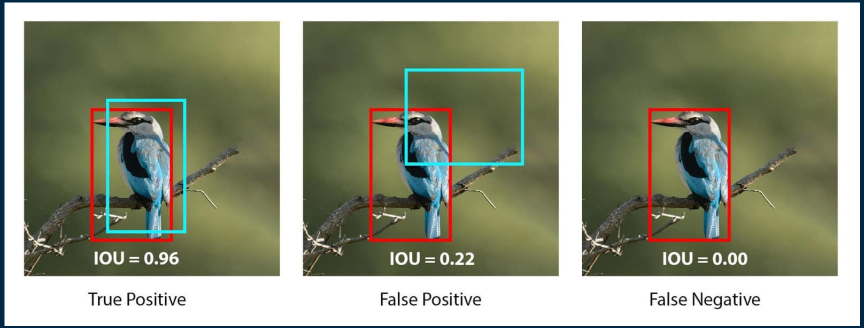
06

YOLO – Metrics

Metrics help quantify how well the model is identifying and localizing objects within images

- **Intersection over Union (IoU): measures the overlap between predicted bounding boxes and ground truth bounding boxes. It is calculated as the area of overlap divided by the area of union between the two boxes. A higher IoU indicates better localization accuracy.**





YOLO – Metrics

- **Precision (P): it measures the accuracy of positive predictions made by the model. It is calculated as the number of true positive detections (correctly identified instances) divided by the total number of positive predictions (true positives + false positives).**
- **Recall (R) : measures the ability of the model to find all the relevant instances of the object class in the dataset. It is calculated as the number of true positive detections divided by the total number of actual instances (true positives + false negatives).**

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{\#\text{ ground truths}}$$

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

YOLO – Metrics

- **mAP50 (Mean Average Precision at IoU 0.5): This metric calculates the average precision (AP) for each object class at an Intersection over Union (IoU) threshold of 0.5 and then takes the mean across all classes. An IoU of 0.5 means that a predicted bounding box is considered correct if it overlaps with the ground truth bounding box by at least 50%.**
- **mAP50-95 (Mean Average Precision from IoU 0.5 to 0.95): This is a variation of mAP that considers the average precision across a range of IoU thresholds, from 0.5 to 0.95, typically in increments of 0.05.**

Any questions?

bessmagsm@gmail.com

THANK YOU!

[Besma Guesmi]