

Perception

⇒ Common perception Suite:

- (Stereo) Camera
- LiDAR
- Radar

⇒ Other Sensors:

- Ultrasonic (near range)
- GPS
- IMU
- Odometer

⇒ Sensor fusion: Integrate information from multiple sensors.

* Perception Tasks

- Classification
- Detection
- Semantic Segmentation
- Instance Segmentation
- { Panoptic Segmentation }

* Anatomy of an Object Detector

Input: RGB Image

Output:

- bounding boxes defined by (x, y, w, h) or (c_x, c_y, w, h) or (x_1, y_1, x_2, y_2)
- Confidence Scores in $[0, 1]$

⇒ General Approach:

① Extract regions (Sliding Window Approach)

② Classify & Score regions

③ Keep high Scoring regions

(Non-maximum Suppression)
(NMS)

↓
(Intersection-over-Union)
(IoU)

⇒ Object Detection Datasets

- * Pascal VOC 2012
- * ImageNet 2014
- * MS COCO 2014
- * LVIS 2019

⇒ Greedy Matching for Comparing Predicted bounding box with ground truth bounding box.

* Traditional Object Detectors

Haar features

Histogram of
Gradients
(HOG)

Deformable Part
Model
(DPM)

⇒ Main innovations: better features

★ Modern Object Detectors

⇒ Rely mainly on Convolutional Neural Network (CNN)

⇒ Paradigms for Object Detection:

1. Two-stage approaches:

↳ Extract first regions that are then classified.

2. Single Stage:

↳ Directly produce classified bounding boxes.

★ Deep Learning Frameworks

⇒ All operations must be implemented using GPU

TensorFlow

PyTorch

⇒ Selective search instead of Sliding window

