CV / VLMs

Unit 5: State-of-the-Art Object Detection Techniques

TIL-AI
TODAY I LEARNED-AI

5.2.1

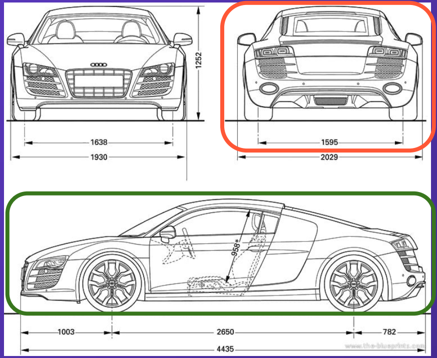
Anchor-Free Object Detection

CenterNet and FCOS transformer models

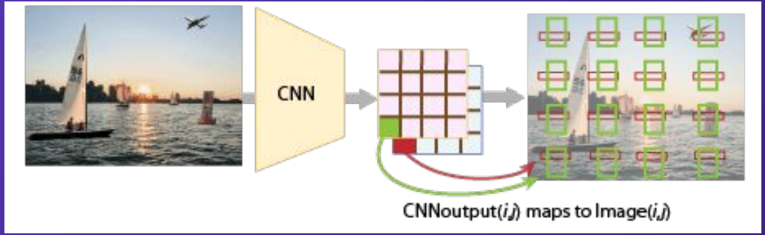
What is an Anchor Box? (recap)

- An anchor box is responsible for predicting an object class

- To effectively detect objects, a model has to try all possible anchor boxes for every grid cell



Anchor Box: a vehicle would be 1:1 (square) when looking from the front or rear, but 2:1 (rectangular) when viewed from the side.



Two anchor boxes (ships and airplane) are used to make two predictions per grid in the image.

Anchor-Free vs. Anchor-Based OD

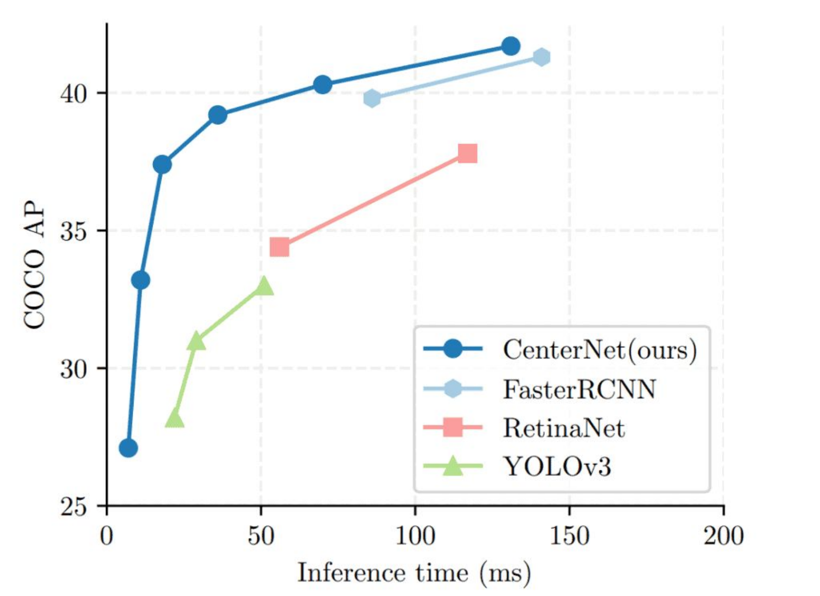
While anchor-free object detection methods have equivalent accuracy to anchor-based methods, they offer several unique advantages:

1. **Simplified anchor identification**: Anchor-free methods eliminate the need to identify suitable anchors, which can be a complex problem due to factors like object physical properties, perspectives, and hyperparameters.

2. **Reduced computational complexity**: Anchor-based methods require more anchor boxes to improve accuracy, leading to more complex architectures and calculations. Anchor-free methods avoid this complexity.

3. **Improved generalizability**: Anchor-free object detection is more generalizable and can be extended to other tasks like keypoint detection and 3D object detection.

Speed vs Accuracy Plots



CenterNet latency vs mAP

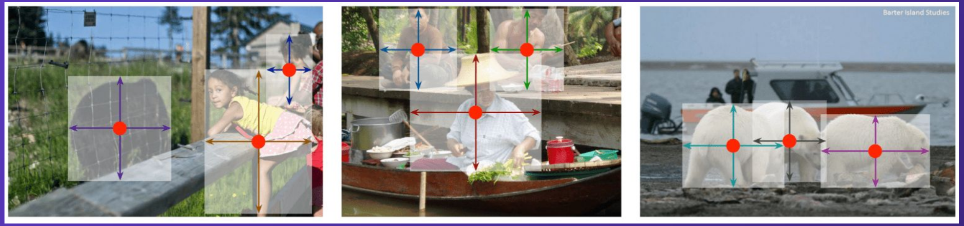
Adapted from Paper:Object as Points

CenterNet: Objects as Points - Anchor Free Object Detection Explained (learnopencv.com)

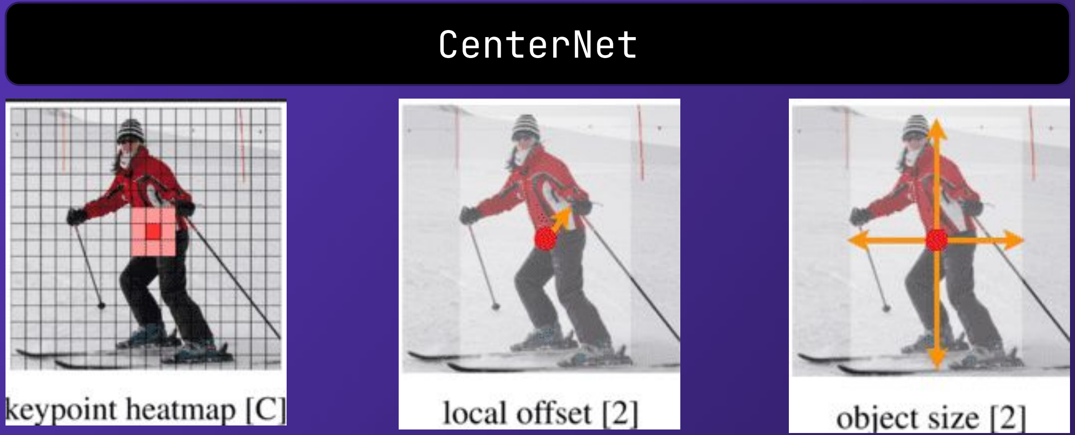
Introduction to CenterNet

In CenterNet, an object is represented by the center point (key-points) of its bounding box, which is crucial for its localization.

- **Keypoint heatmap:**
  A set of heatmaps that would predict the likelihood of a pixel being the keypoint.
- **Local Offset:**
  Offset to improve key-points precision
- **Bounding box size:**
  Predicting width and height of the bounding box



CenterNet: Objects as points.



CenterNet

CenterNet Components

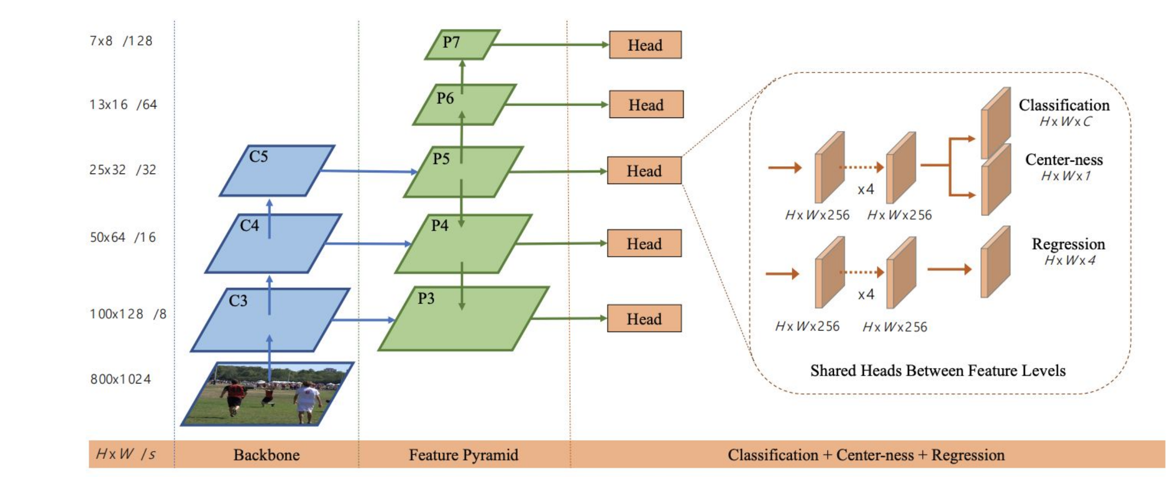
CenterNet: Objects as Points - Anchor Free Object Detection Explained (learnopencv.com)

Fully Convolutional One-Stage Object Detection (FCOS) Introduction

- FCOS (Fully Convolutional One-Stage Object Detection) uses a pixel-wise prediction for object detection.

- It utilizes FCN (Fully Convolutional Networks for semantic segmentations)

- Consist of 3 stages
  - Backbone
  - Feature Pyramid
  - Head
    - Classification loss
    - Center-ness loss
    - Regression loss



(top) network architecture of FCOS

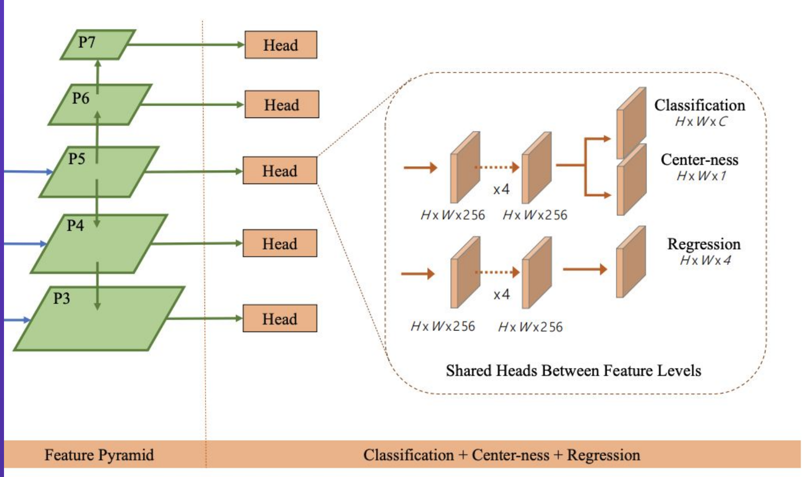
FCOS- Anchor Free Object Detection Explained | LearnOpenCV #

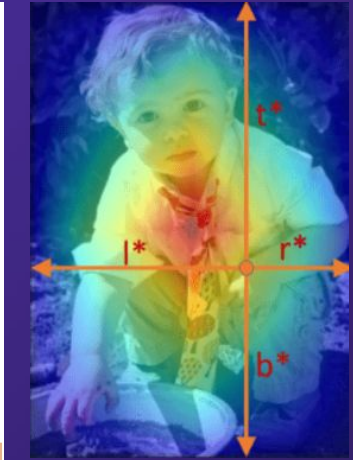
Fully Convolutional One-Stage Object Detection (FCOS) Introduction (cont.)

- Classification
  - Predicting object classifications per spatial location.
  - Focal loss - modified standard cross entropy criterion

- Center-ness
  - Center-ness close to the bounding box center.
  - BCE (binary cross-entropy error/log) loss.

- Regression
  - Distance from center to edges (left, top, right, bottom)
  - IoU loss- which measures the overlap between the predicted and ground truth bounding boxes.





(top) network architecture of FCOS