Objects recognition

Objects detection

Skoltech

Ekaterina Trofimova, Senior Specialist, Center for Applied AI (Skoltech),
Junior Research Fellow, Higher School of Economics

Course Structure

CLASSIC COMPUTER VISION PROBLEMS

- Computer vision introduction
- Image classification. Transfer learning
- Image objects recognition
- Image objects segmentation
- Homework

ADVANCED COMPUTER VISION PROBLEMS

- Image generation: AE and VAE
- Image generation: GAN
- Style transferring
- Homework
- Image reconstruction
- Video processing

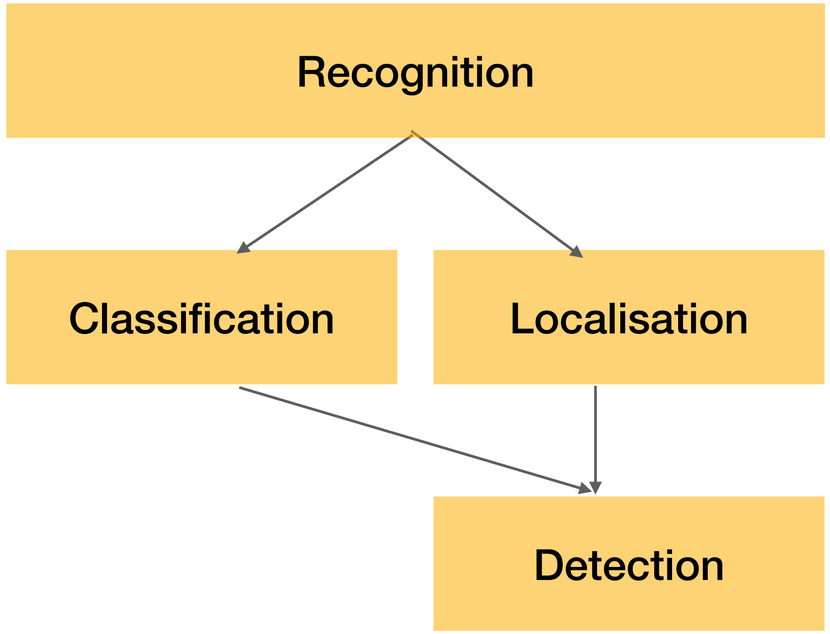
Skoltech

Lecture outline

- Difference between classification, localisation and detection
- Object recognition application
- Main network architectures
- Recognition metrics

Skoltech

Objects recognition tasks



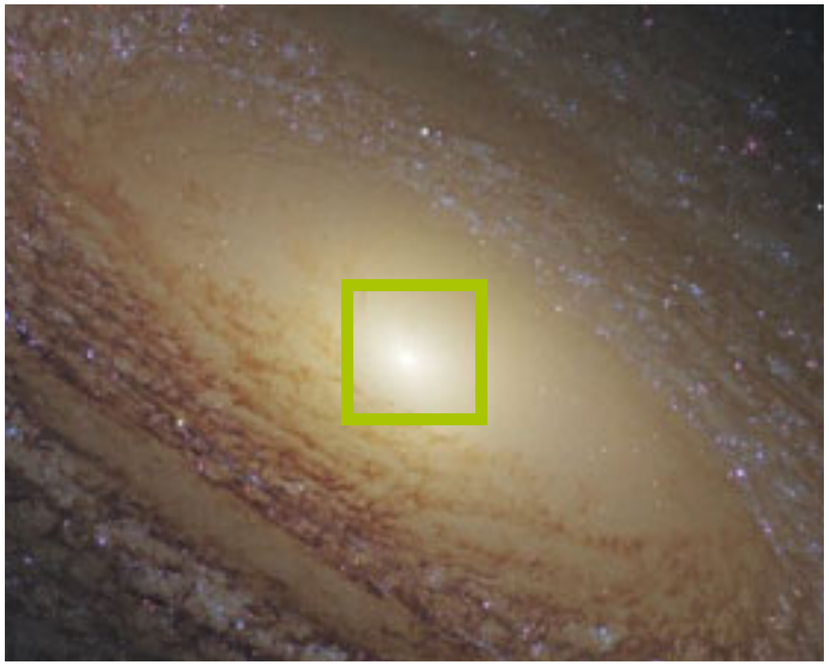
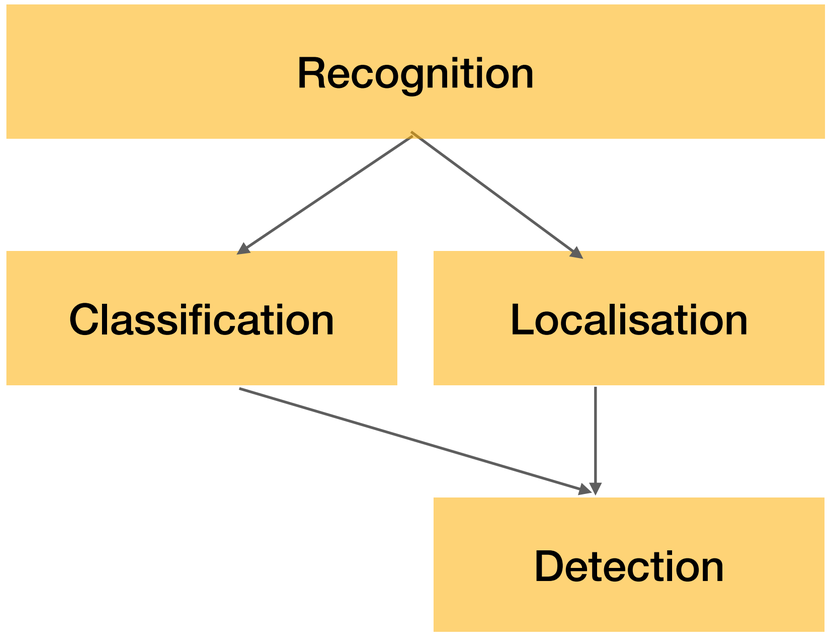
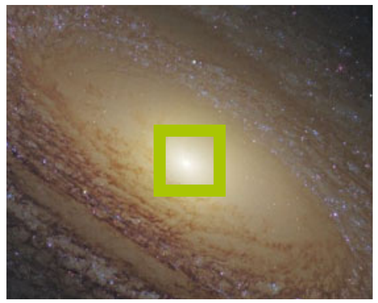


Figure 1. Localised star example.
Source: https://astrobackyard.com/types-of-galaxies

Skoltech

Objects recognition tasks

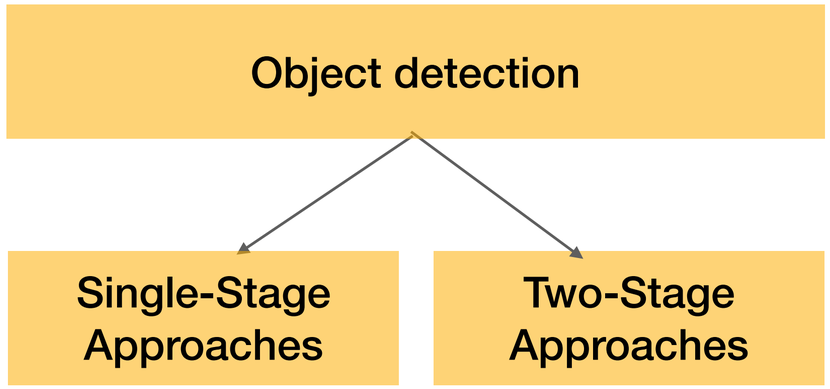




- In **classification algorithms**, the last layer produces a class label or the probability value of an object belonging to a particular class.
- **Localization algorithms**, on the contrary, predict 4 real numbers, that is, they solve the regression problem.
- If there are 2 objects in the image, we need 8 coordinates. If there are 5 objects - 20.

Skoltech

Objects detection methods



YOLO (You Only Look Once): YOLO is known for its real-time object detection capabilities. It divides the image into a grid and simultaneously predicts bounding boxes and class probabilities for each grid cell.

R-CNN (Region-based Convolutional Neural Networks): First generates region proposals using selective search and then extracts features from these regions using CNNs.

Object detectors solve the following two problems:

1. find an arbitrary number of objects (possibly even zero)
2. classify each object and estimate its size using a bounding box

Two-stage detectors divide these tasks into two stages. Single-stage - combine both tasks into one stage.

Regions With CNNs (R-CNN)

Ross Girshick et al. *Rich feature hierarchies for accurate object detection and semantic segmentation*. UC Berkeley. 2014

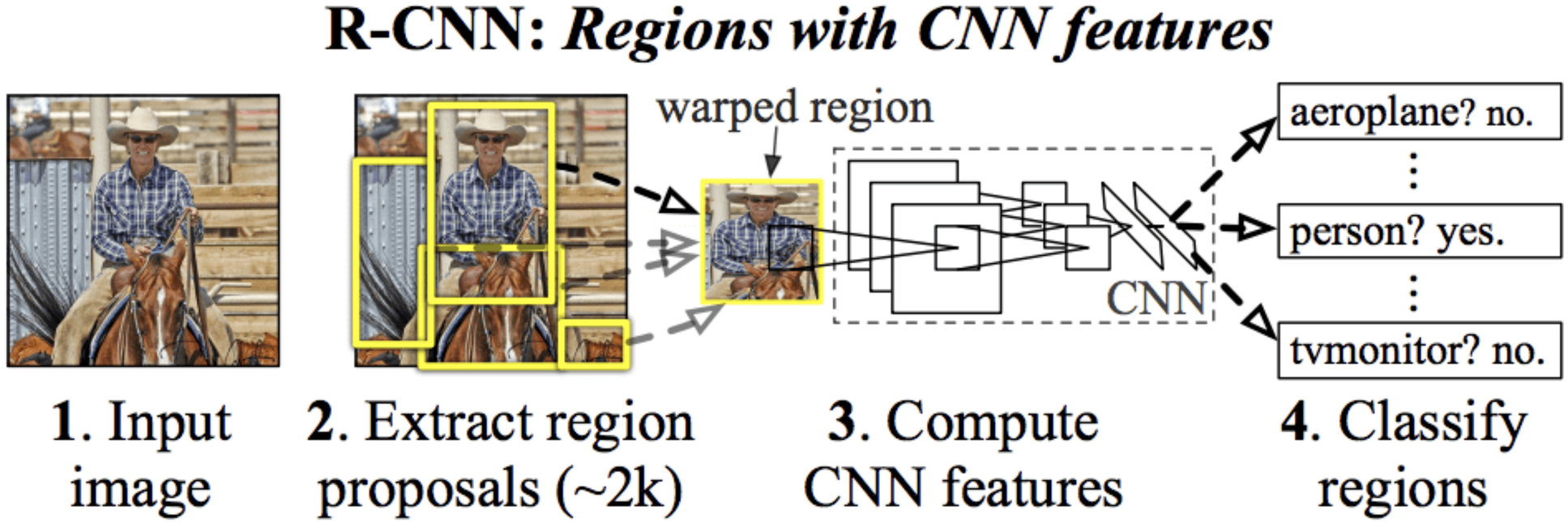


Figure 2. Object detection system overview. The system (1) takes an input image, (2) extracts around 2000 bottom-up region proposals, (3) computes features for each proposal using a large convolutional neural network (CNN), and then (4) classifies each region using class-specific linear SVMs.

Fast R-CNN

Ross Girshick. *Fast R-CNN*. UC Berkeley. 2015

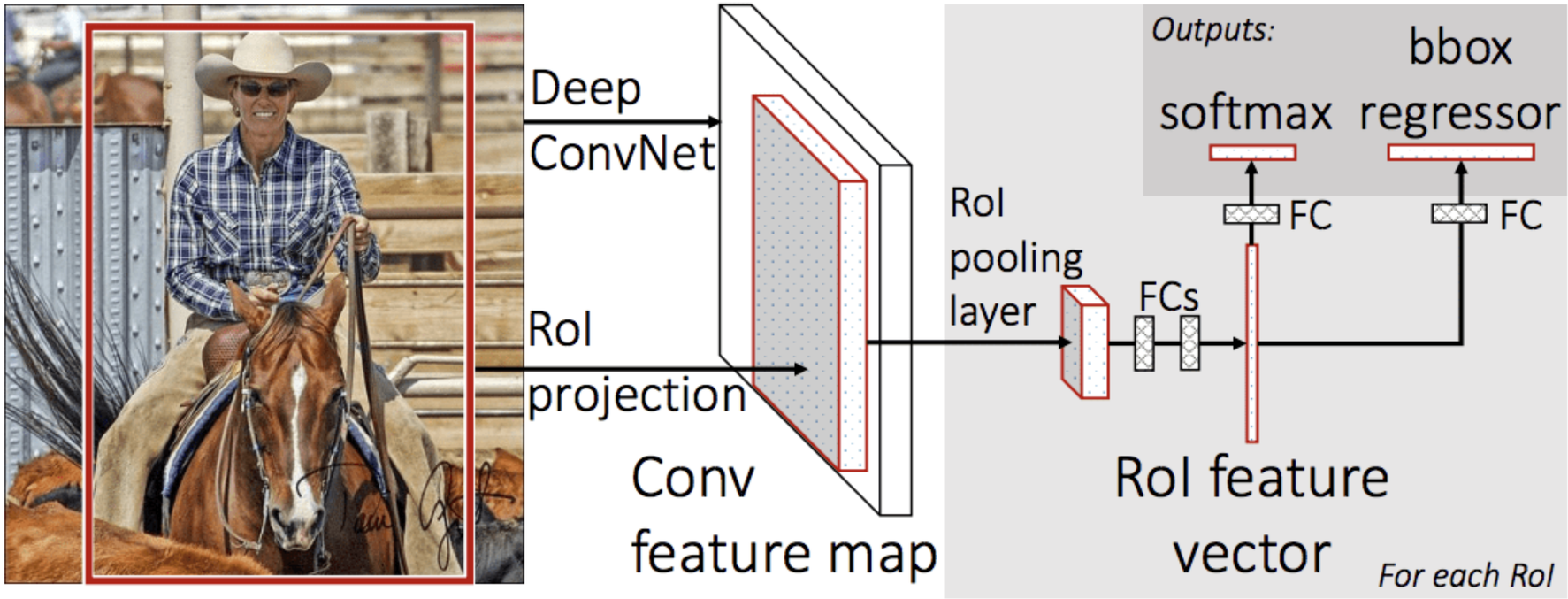


Figure 3. Fast R-CNN architecture. An input image and multi- ple regions of interest (RoIs) are input into a fully convolutional network. Each RoI is pooled into a fixed-size feature map and then mapped to a feature vector by fully connected layers (FCs). The network has two output vectors per RoI: softmax probabilities and per-class bounding-box regression offsets. The architecture is trained end-to-end with a multi-task loss.

Skoltech

Faster R-CNN

Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. 2016

Selective search → Region Proposal Networks (RPN)

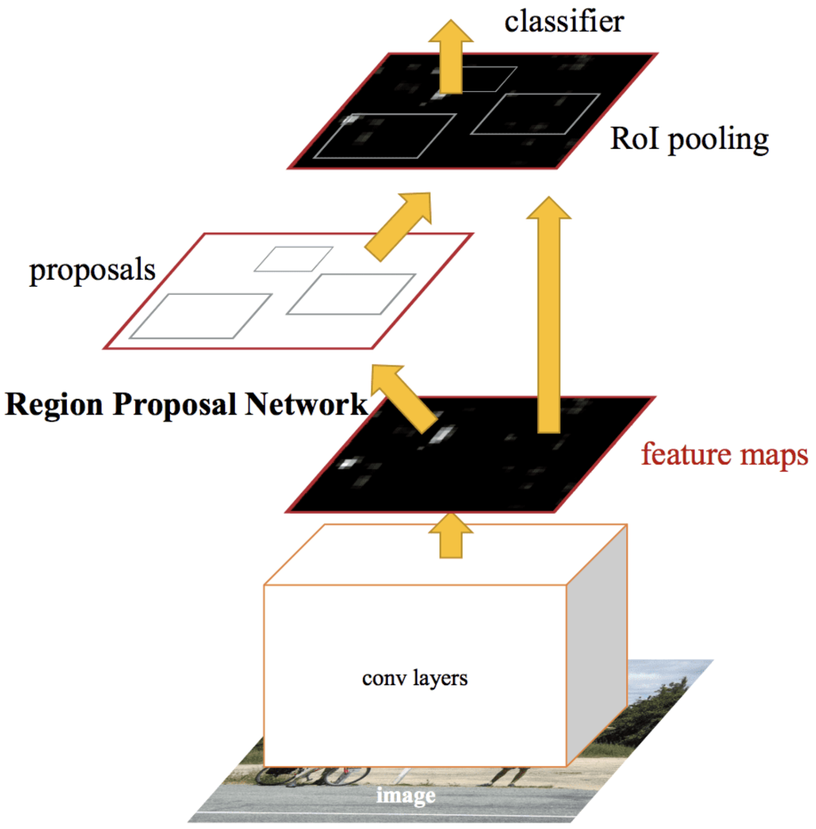


Figure 4. Faster R-CNN architecture.

YOLO

Joseph Redmon et al. *You Only Look Once: Unified, Real-Time Object Detection.* 2016

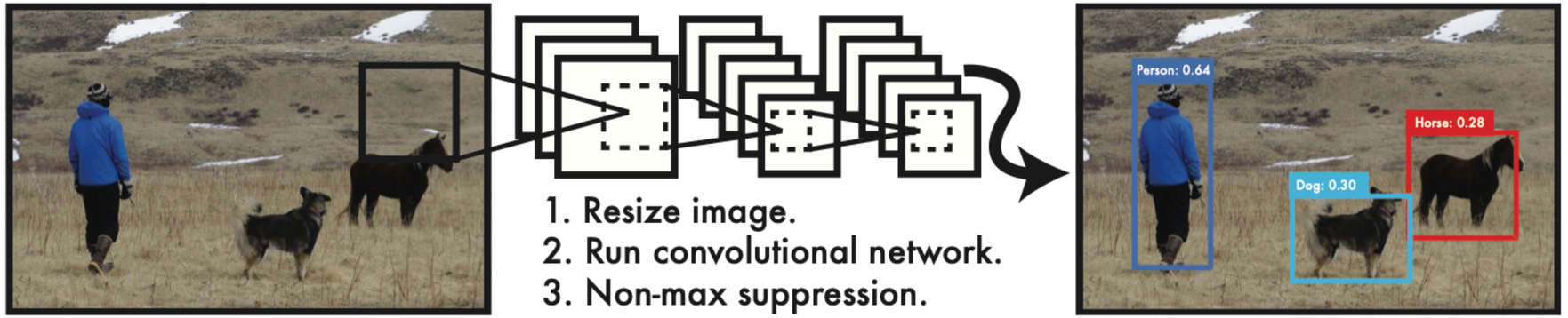


Figure 5. The YOLO Detection System. Processing images with YOLO is simple and straightforward. Our system (1) resizes the input image to 448 × 448, (2) runs a single convolutional net- work on the image, and (3) thresholds the resulting detections by the model's confidence.

Intersection over Union

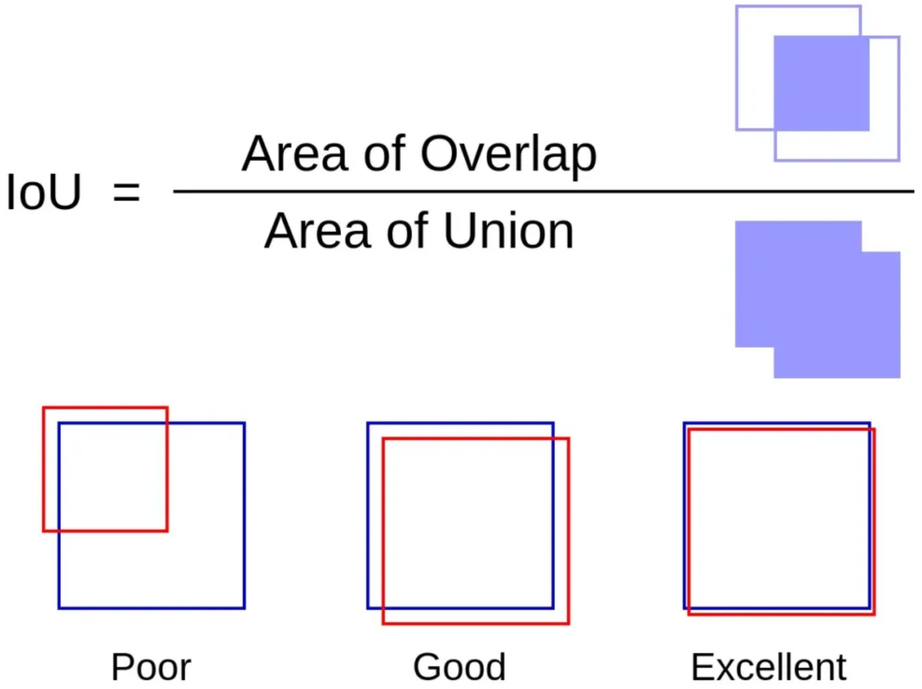


Figure 6. IoU calculation.
Source: https://idiotdeveloper.com/what-is-intersection-over-union-iou/

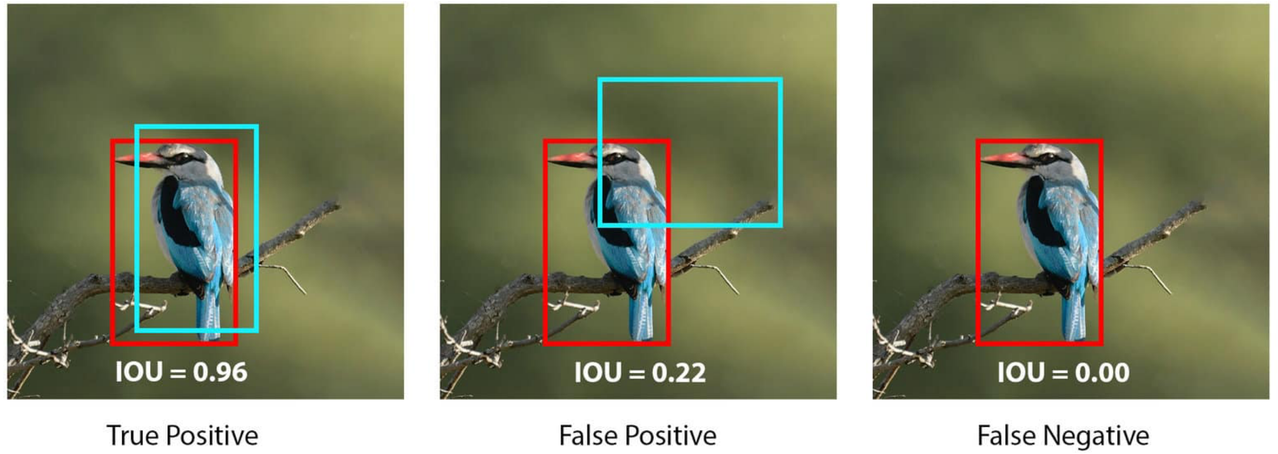


Figure 7. Different IoU illustration.
Source: https://learnopencv.com/intersection-over-union-iou-in-object-detection-and-segmentation/

Object detection summary

- Object detection is a fundamental computer vision task that involves identifying and localizing objects within images or video frames.
- It goes beyond object recognition by providing precise bounding box locations for detected objects, enabling spatial understanding.
- Object detection techniques can be categorized into two-stage and single-shot approaches, each with its trade-offs in accuracy and speed.
- Two-stage approaches, such as Faster R-CNN, involve region proposal and feature extraction, followed by classification and bounding box regression.
- Single-shot approaches, like YOLO, predict object classes and bounding boxes directly from fixed grid cells, achieving real-time performance.

Next time: Objects segmentation