INFO 7375 - Neural Networks & AI

Home Work to Chapter - 16

Submitted By:

Abdul Haseeb Khan
NUID: 002844724
khan.abdulh@northeastern.edu

How Object Detection Works

Object detection is a computer vision task that combines **classification** (what objects are present) with **localization** (where they are located) to identify and locate multiple objects within an image simultaneously.

The Object Detection Pipeline

1. **Input**: Raw image (RGB pixels)
2. **Processing**: Neural network identifies objects and their locations
3. **Output**: List of (class label, bounding box, confidence score) for each detected object

Feature Extraction Backbone

The foundation of any detector is a convolutional neural network that extracts hierarchical features:

- **Early layers**: Detect edges, corners, textures
- **Middle layers**: Combine simple features into parts (wheels, faces, windows)
- **Deep layers**: Represent complete objects with semantic meaning
- **Common backbones**: ResNet, VGG, EfficientNet, MobileNet

Grid-Based Detection (YOLO Approach)

The image is divided into an S×S grid, where each cell:

1. Predicts B bounding boxes
2. Assigns confidence scores for boxes
3. Predicts class probabilities

Per Grid Cell Output:
- Bounding box coordinates (x, y, w, h)
- Objectness score (probability object exists)
- Class probabilities (for each possible class)

Training Process

- Data Preparation:
  - Annotated images with bounding boxes and class labels
  - Data augmentation (scaling, cropping, color jittering)
- Forward Pass:
  - Extract features through backbone
  - Generate predictions for all possible locations
- Matching Strategy:
  - Assign ground truth objects to predictions
  - Based on IoU overlap or center location
- Backpropagation:
  - Calculate loss for matched predictions

- Update weights to minimize combined loss

What is the meaning of the following terms: object detection, object tracking, occlusion, background clutter, object variability?

Object Detection: Identifying and locating objects within an image (answering "what" and “where").

Object Tracking: Following the movement of detected objects across video frames over time.

Occlusion: When objects are partially hidden or blocked by other objects in the scene.
Background Clutter: Complex or noisy backgrounds that make it difficult to distinguish objects from their surroundings.

Object Variability: Variations in object appearance due to different poses, scales, lighting conditions, or viewpoints.

What is an object bounding box do?

A bounding box is a rectangular frame that tightly encloses a detected object, defined by coordinates (typically x, y, width, height or corner points). It provides spatial localization of objects in an image.

What is the role of the loss function in object localization?

The loss function measures the error between predicted and ground-truth bounding boxes. It typically combines:

- Localization loss: Measures bounding box coordinate accuracy (often using MSE or IoU-based metrics)
- Classification loss: Measures object class prediction accuracy

What is facial landmark detection and how does it work?

Identifies specific facial keypoints (eyes, nose, mouth corners, etc.). It works by:

- Training CNNs to predict (x,y) coordinates of predefined facial points
- Using regression to output continuous coordinate values
- Often employing cascaded networks for refinement

What is convolutional sliding window and its role in object detection?

A technique that applies a CNN classifier to multiple overlapping regions of an image by sliding a fixed-size window across different positions and scales. While conceptually important, it's computationally expensive and largely replaced by more efficient methods.

Describe YOLO and SSD algorithms in object detection.

YOLO (You Only Look Once):

- Single-pass detection dividing image into grid cells
- Each cell predicts bounding boxes and class probabilities simultaneously
- Extremely fast but can struggle with small objects

SSD (Single Shot Detector):

- Uses multiple feature maps at different scales
- Predicts objects at various resolutions from different network layers
- Better at detecting objects of varying sizes than YOLO

What is non-mas suppression, how does it work, and why I is needed?

What it is: Post-processing technique to eliminate duplicate detections

How it works:

1. Sort bounding boxes by confidence score
2. Keep highest-scoring box
3. Remove boxes with high IoU (overlap) with kept box
4. Repeat until all boxes processed

Why needed: Object detectors often produce multiple overlapping predictions for the same object; NMS ensures each object is detected only once.