Memento: Object Detection and Tracking for Memory Recall

Thanuja Kumar
Bhavana Bandam
Narendra Kovi
Prasad Shinde

Overview

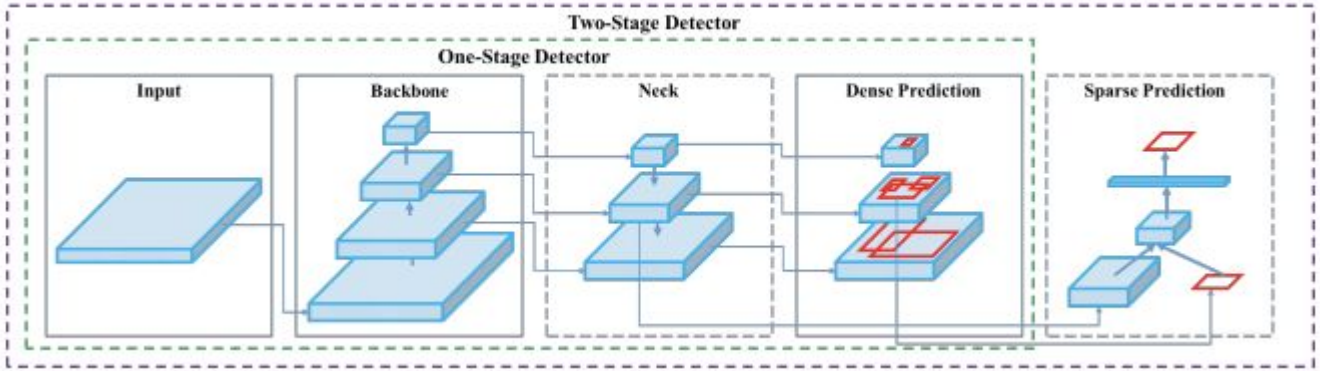
- Aim is to develop an object detection and tracking system to aid in memory recall for people suffering from amnesia
- Uses an object detection algorithm based on deep learning
- Detected objects last position will be saved which can be queried by the user later.

Steps

- Object detection
- Object tracking
- Object recall

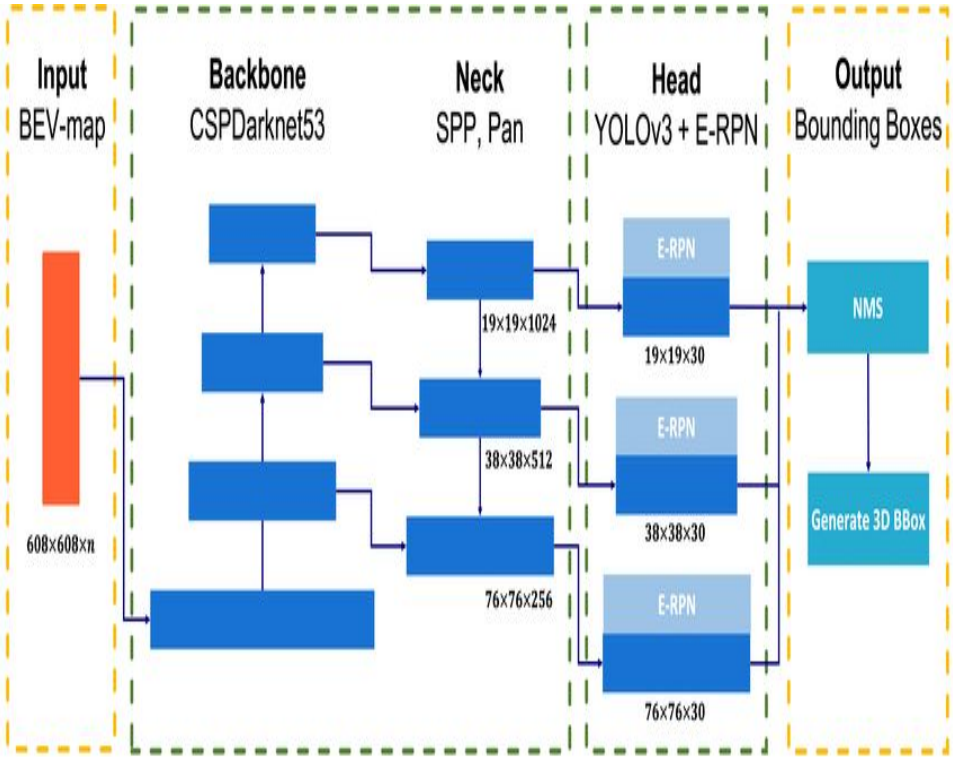
Object Detection using YOLO

- You Only Look Once (YOLO) is a One stage detection method
- YOLO v4 is a deep convolutional neural network(CNN) that takes an image as input and outputs a set of bounding boxes around the detected objects.



YOLOv4

- **Preprocessing**: The input image is resized.
- **Forward Pass**: It consists backbone network, neck network, and head network.
- **Bounding Box Prediction**: The head network outputs a set of bounding boxes and corresponding class probabilities for each object detected in the image.
- **Post-processing**: Involves filter out low-confidence detections and non-maximum suppression is applied to remove redundant bounding boxes.
- **Output**: The final output of YOLOv4 is a set of bounding boxes and class probabilities for each object detected in the input image.



YOLOv4

- We used cv2.dnn model called readFromDarkNet which takes input parameters of cfg and weights file.
- **Dataset**: The pretrained model uses “Coco Dataset”
- **yolov4.cfg** is a configuration file that defines the architecture of the YOLOv4 neural network.
- **yolov4.weights** contains the trained learned weights on “Coco Dataset” that enable the network to identify and localize objects within images and videos

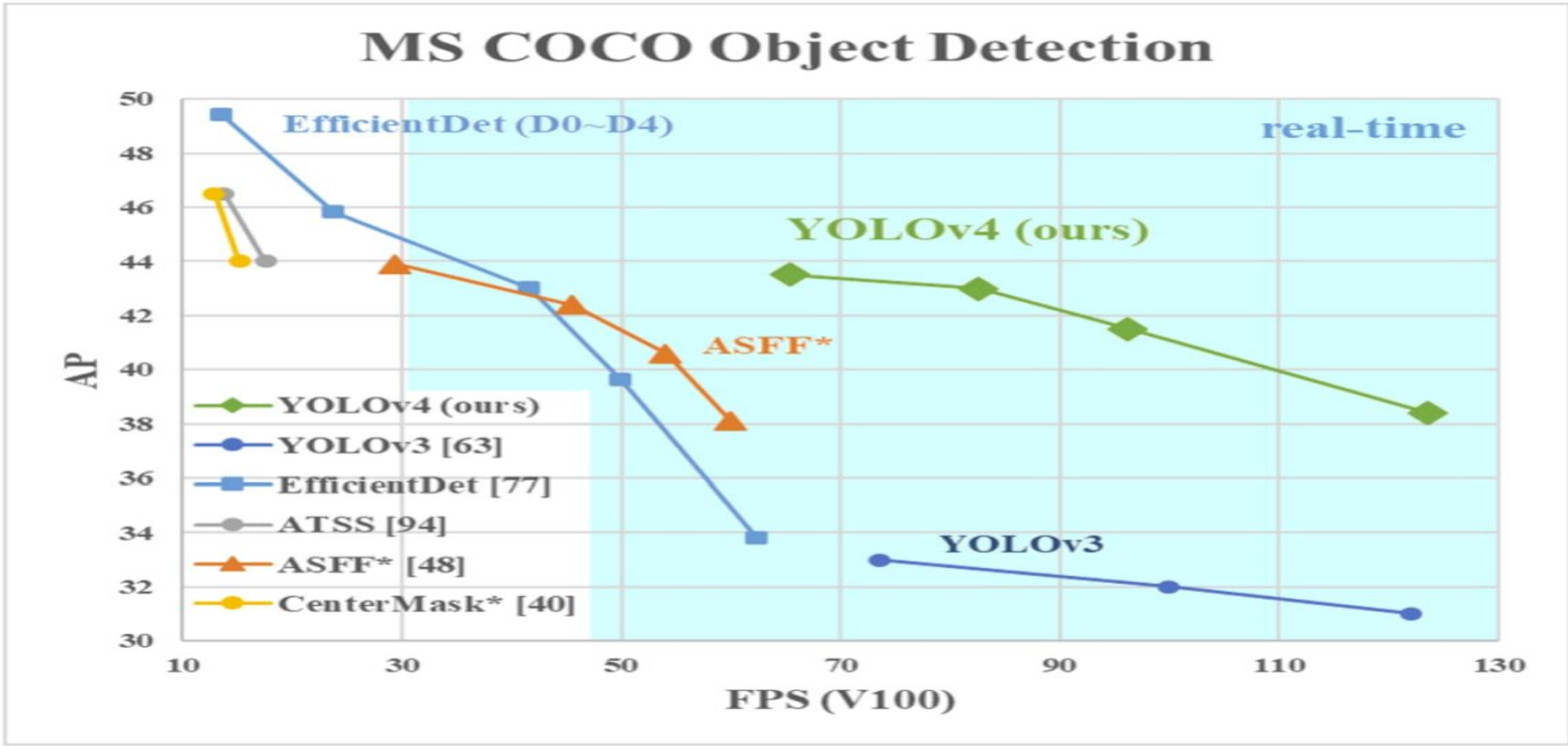


Figure 1: Comparison of the proposed YOLOv4 and other state-of-the-art object detectors. YOLOv4 runs twice faster than EfficientDet with comparable performance. Improves YOLOv3's AP and FPS by 10% and 12%, respectively.

Object tracking using kalman filter

- Kalman filtering is a mathematical algorithm that uses a series of measurements to estimate the state of a system.
- The Kalman filter is composed of two main steps: prediction and update.
- In the prediction step, the filter predicts the next state of the system based on its current state and a set of motion models.
- In the update step, the filter uses a set of measurements to update its estimate of the state of the system.

Object Recall

- Locations of the tracked objects are saved in memory.
- Location of the objects will always be upto date.
- When the user queries for the object location, it outputs the location in natural language description format with respect to other objects.
- The coordinates of the bounding-box of all the tracked objects are compared to determine the distance between them, which helps to predict where the object is located.


chair: 0.98

Metrics

- Overall Accuracy: 0.41 (40.74%)
- Overall Precision: 0.58 (57.89%)
- Overall Recall: 0.58 (57.89%)

Output:

- keyboard is to the right of laptop
- keyboard is below laptop
- Handbag is to the left of chair
- Handbag is above the chair

Predicted:

chair|right|below|laptop
chair|right|below|bottle
laptop|left|above|chair
laptop|right|below|bottle
bottle|left|above|chair
bottle|left|above|laptop

Ground truth:

chair|right|below|bottle
chair|right|below|laptop
laptop|left|above|chair
laptop|left|-|bottle
bottle|left|above|chair
bottle|right|-|laptop

Future work

- YOLOv4 has a limitation on detecting smaller objects accurately. Potential Solution is using “EfficientDet” Model.
- NLP Model for Object Recall

Questions?