Semi automated annotation tool towards object detection and tracking in videos

Shyam Desai (01FE21BEC110),Shridhar Naragund (01FE21BEC116), Vinayak Nayak (01FE21BEC305)

Under the guidance of
Uma Mudenagudi, Ramesh Ashok Tabib

KLE Technological University, Vidyanagar, Hubballi-580031, Karnataka, India

January 29, 2024

Overview

- Introduction
- Motivation
- Literature Survey
- Problem Statement and Objectives
- Approach / Solution
- Dataset Analysis / Description
- Experimental Results/ Simulation / Observation
- Contributions
- Plan of Action and Conclusion
- References

Introduction towards Semi automated annotation tool

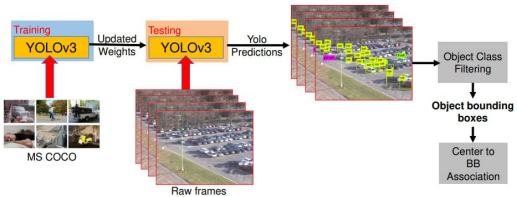


Figure: **Object Detection** involves identifying and locating objects within an image. They generate bounding boxes around the objects and may also classify them into specific categories

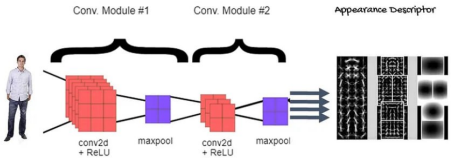


Figure: **Deep Appearance descriptor(object tracking)** involves identifying and locating objects within an image. They generate bounding boxes around the objects and may also classify them into specific categories.

- **Human Intervention:** Human reviews the automated detections and tracks, correcting any inaccuracies, and making adjustments where necessary. Human intervention is particularly valuable in complex scenes, where automated methods may struggle due to occlusions, lighting changes, or ambiguous situations.

Motivation

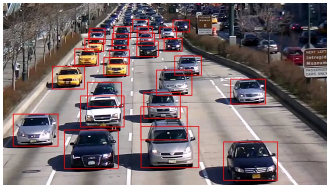


Figure: Object detection

Challenging Real-World Scenarios:

- In many real-world situations, object detection and tracking can be exceptionally challenging due to factors such as occlusions, complex backgrounds, lighting variations, and diverse object types.
- Fully automated systems often struggle to provide accurate and reliable results under above conditions.

Enhanced Accuracy and Reliability: The primary motivation is to improve the accuracy and reliability of object detection and tracking.

Literature Survey

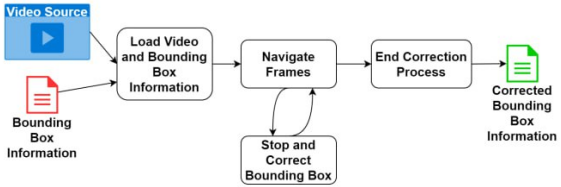


Figure: Annotation Tool

- Literature highlights the revolutionary impact of YOLO in real-time object detection, especially its grid-based approach for simultaneous predictions, which significantly improves speed and efficiency.[1]
- Various studies delve into the advancements in object tracking, with focus on methods like Optical Flow and Mean Shift for improving detection continuity, and more sophisticated techniques like SORT and DeepSORT for identity-aware tracking[2]

Literature Survey

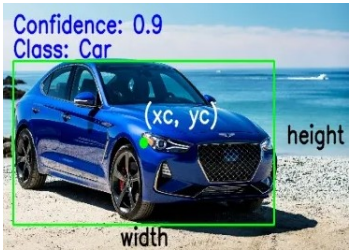


Figure: Bounding Box

Literature emphasizes the synergy between traditional computer vision methods (OpenCV, Optical Flow, Mean Shift) and deep learning techniques (YOLO, DeepSORT) to achieve a balance between real-time efficiency and tracking accuracy.[3]

Problem Statement and Objectives

Problem Statement

Semi automated annotation tool towards object detection and tracking in videos

Objectives

- Employ detection and tracking algorithms to detect and track objects (bounding boxes will act like annotations) within video frames.
- Develop annotation tool towards object detection and tracking
- Integrate the detection and tracking algorithm to annotation tool.
- Develop logic to hand-correct annotations in case of errors in annotations.

FUNCTIONAL BLOCK DIAGRAM

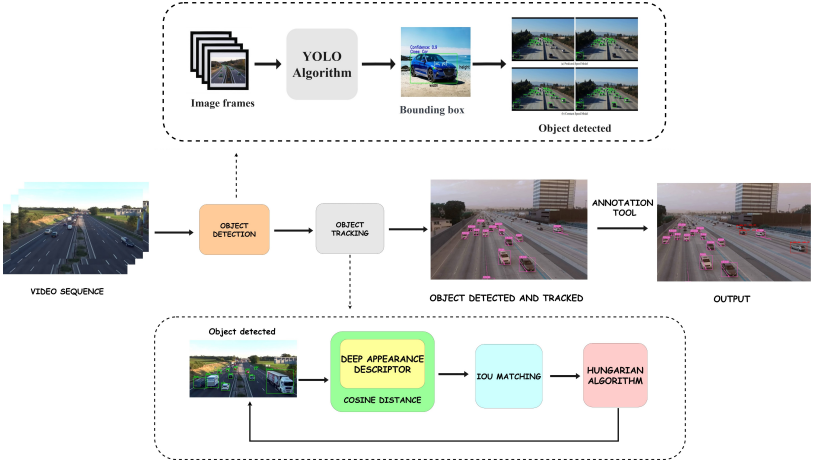


Figure: Block diagram framework for detection,tracking and annotation tool

Dataset Analysis / Description

- The COCO dataset is a large-scale dataset designed for object detection, segmentation, and captioning tasks.
- It contains images from a wide range of categories, and it is widely used in the computer vision community for benchmarking and training purposes.
- The dataset is labeled with bounding boxes around object instances, and each instance is associated with a specific category.including common objects such as person, car, dog, cat, and more.
- The COCO dataset is well-suited for training object detection models like YOLO because it provides a diverse set of images with multiple objects in various contexts.
- This diversity helps the model generalize well to different scenarios and object appearances.

Experimental Results

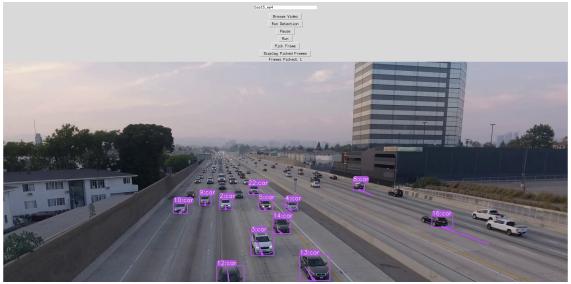


Figure: Annotation tool main window

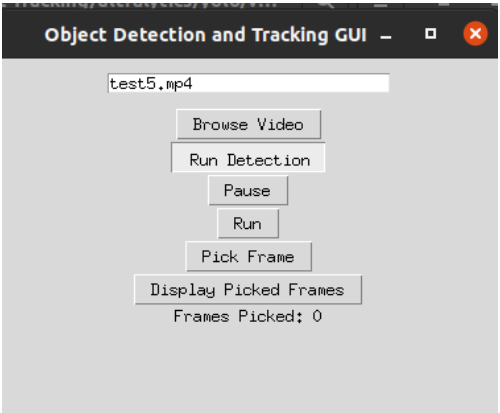


Figure: GUI Interface

Experimental Results

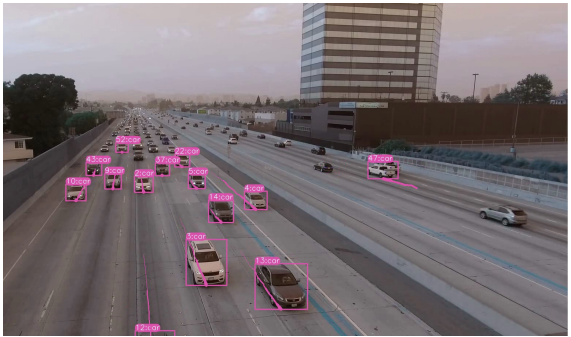


Figure: Before Annotation

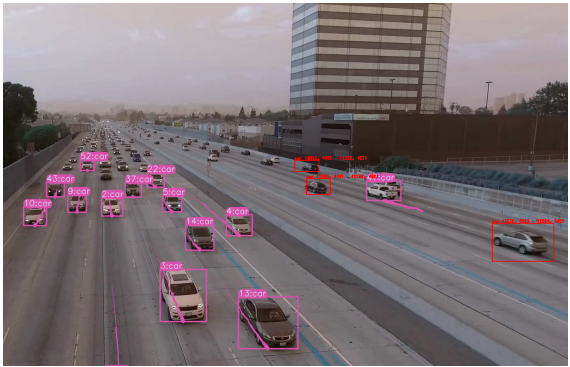
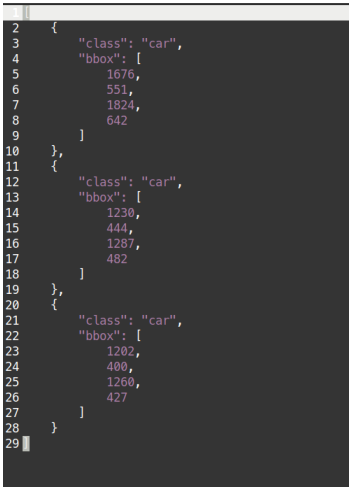


Figure: After Annotation

Experimental Results



```
[
    {
        "class": "car",
        "bbox": [
            1676,
            551,
            1824,
            642
        ]
    },
    {
        "class": "car",
        "bbox": [
            1230,
            444,
            1287,
            482
        ]
    },
    {
        "class": "car",
        "bbox": [
            1202,
            400,
            1260,
            427
        ]
    }
```

Figure: Coordinates of bounding box after annotation in .json format

Conclusions

- AI and machine learning are rapidly advancing disciplines with significant global impact
- YOLO algorithm stands out for efficient object detection by dividing images into grid cells and predicting bounding boxes
- Deep appearance descriptor encodes object information for similarity comparison, utilizing cosine distance matrices
- IOU matching and the Hungarian algorithm contribute to effective object tracking
- An annotation tool is employed to refine model accuracy, enhancing overall performance

References

- Al-Shakarji, Noor M., et al. "Semi-automatic system for rapid annotation of moving objects in surveillance videos using deep detection and multi-object tracking techniques." 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). IEEE, 2020.
- Gu, Chuang, and Ming-Chieh Lee. "Semiautomatic segmentation and tracking of semantic video objects." IEEE Transactions on Circuits and Systems for Video Technology 8.5 (1998): 572-584.
- Al-Shakarji, Noor M., et al. "Semi-automatic system for rapid annotation of moving objects in surveillance videos using deep detection and multi-object tracking techniques." 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). IEEE, 2020.
- Bewley, Alex, et al. "Simple online and realtime tracking." 2016 IEEE international conference on image processing (ICIP). IEEE, 2016.
- Du, Yunhao, et al. "Strongsort: Make deepsort great again." IEEE Transactions on Multimedia (2023).

Thank You