University of Information Technology, VNU-HCM

CS117.O21.KHTN – Computational Thinking

GROUP 3

VEHICLE COUNTING IN REGION OF INTEREST FOR HCM ROADWAY

Instructor

PhD. Ngo Duc Thanh

thanhnd@uit.edu.vn

Team Members

Hoang Ngoc Quan (Leader)	Le Trong Dai Truong	Nguyen Nhat Minh
22521178	22521576	21521135

INTRODUCTION

Traffic congestion is a major issue in urban areas, and counting vehicles accurately can provide valuable data for traffic management and planning. However, current methods for vehicle counting in roadways often involve manual human observation or the use of fixed sensors, which can be time-consuming, expensive, and limited in scope. Addressing these challenges, we propose an efficient and accurate method to count vehicles in a specific region of interest for a roadway in Ho Chi Minh City using AI, Computer Vision, and video analysis techniques.

Given a video or image footage of a roadway in Ho Chi Minh City, the task is to develop a method that can count accurately the number of vehicles within a defined Region of Interest (ROI).

Input/Output

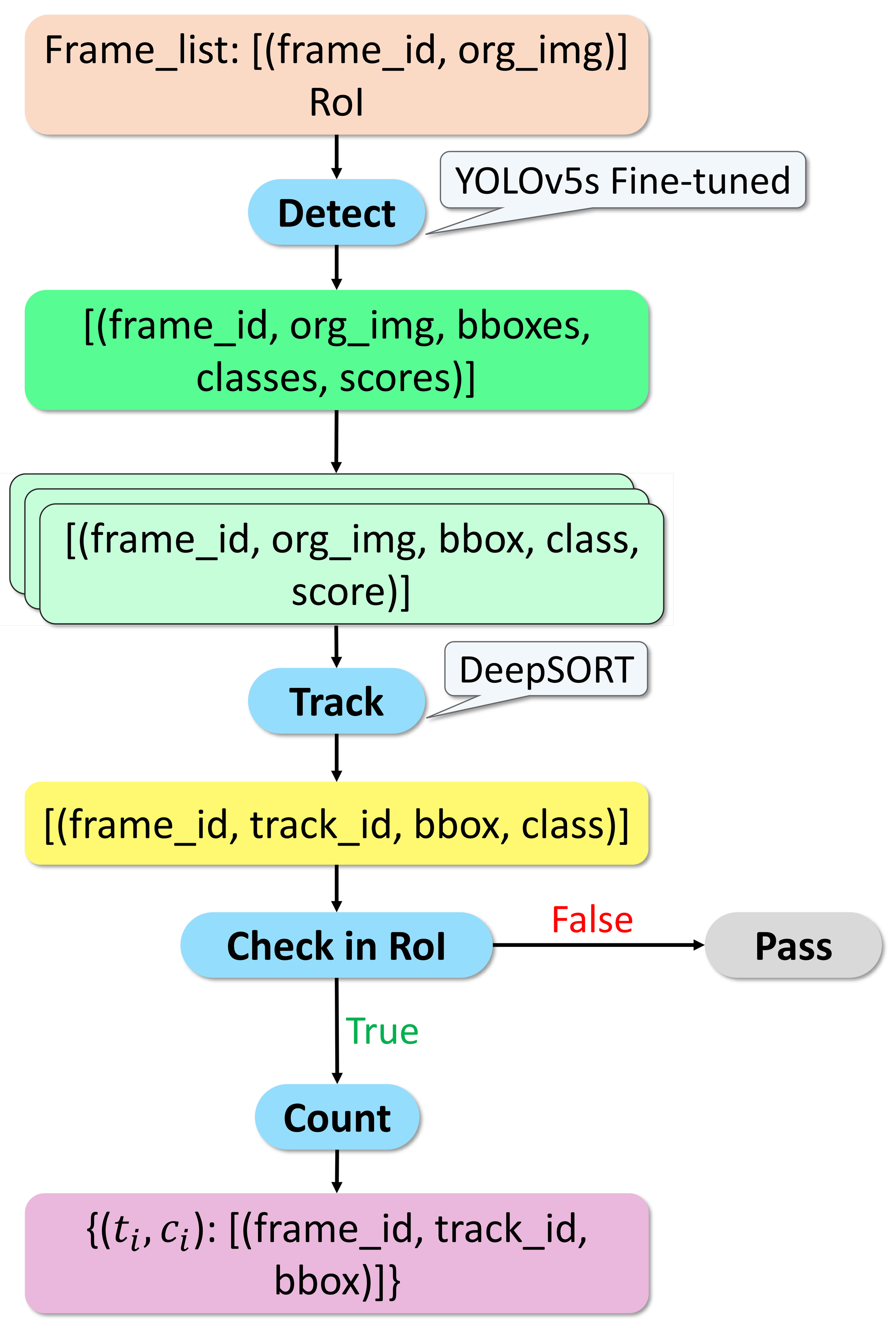
Input:

- A training dataset $\mathcal{D} = \{(image_i, \{bb_j, cls_j\}_i)\}_{i=0}^{N}$ with $cls_j \in \mathcal{L} = \{l_1, \dots, l_M\}$
- A list of interested vehicle types $\mathcal{T} = \{t_1, \dots, t_C\}$ defined by user, $\mathcal{T} \subset \mathcal{L}$
- A video X is needed to detect and count the number of interested vehicle types appearing in RoI of the video. RoI is a polygon $P = (x_1, y_1, \dots, x_V, y_V)$ with V vertices, (x_i, y_i) is coordinate of i-th vertex, defined by user to specify an area within video X **where** the vehicles will be counted.

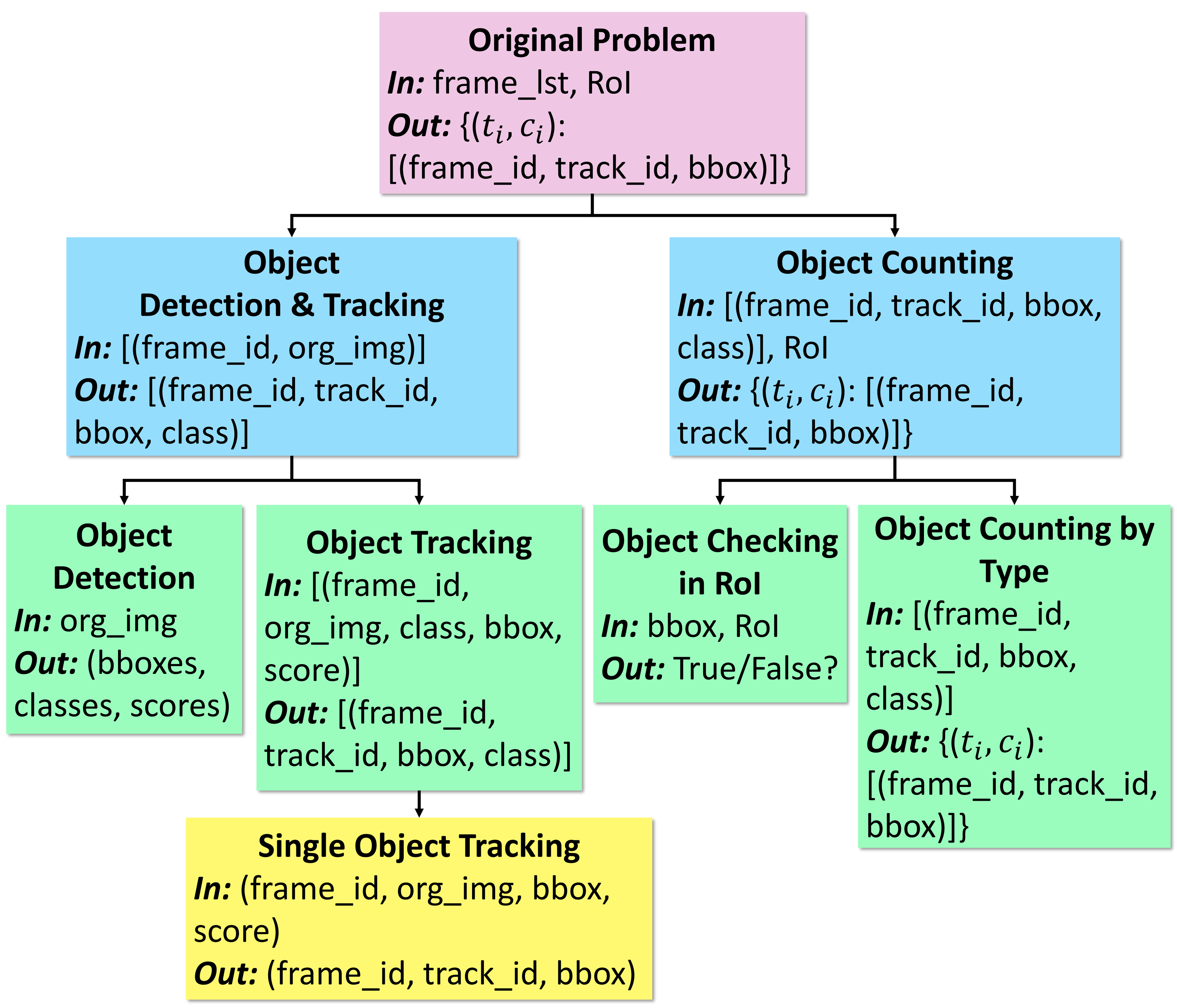
Output:

- A detail information file containing: The number of vehicles appearing in the RoI for each type in $\mathcal{T}$, bounding boxes, classes and IDs of objects of interest within each frame of the video X.

Algorithm



Decomposition Tree



Evaluation

Dataset:

AI Challenge HCMC 2020 (for both training and testing)

Reasons:

- Captured from different traffic cameras looking from above at the target area roadway in HCM City
- Diverse cases from different vehicle types, traffic densities, and weather conditions

Metrics:

FPS (Frame per Second): is the number of frames per second of video, or the number of frames that a system can process per second.

mAP (mean Average Precision): computes the mean AP value for all classes. AP computes the Average Precision value for Recall values over 0 to 1. It is a popular metric to evaluate performance of Object Detection models. Steps to compute:

- Determine IoU threshold to decide whether the predicted bounding box (bbox) is TP or FP. Predicted bbox has IoU with ground truth bbox equal or greater than IoU threshold is TP, and vice versa.
- Sort predicted bboxes in descending order of their confident scores.
- Compute Precision and Recall at each level of sorted bboxes list.

$$Precision = \frac{TP}{TP + FP}, Recal = \frac{TP}{num.of\ ground\ truth\ bbox}$$

- Draw Precision-Recall Curve based on calculated Precision and Recall values.
- AP is the area under the Precision-Recall Curve, calculate this value by methods: 11-point Interpolation, Area Under Curve – AUC, Interpolated AP

mMCR (mean Miss Counting Rate): MCR represents the proportion of actual objects that the model fails to detect over total actual objects. mMCR is mean of MCR of all classes, MCR can be calculated following this formula:

$$\mathrm{MCR} = \frac{|\mathrm{num.\,of\ actual\ objects} - \mathrm{num.\,of\ detected\ objects}|}{\mathrm{num.\,of\ actual\ objects}}$$

Constraints/Requirements

Constraints:

Camera: fixed position and angle, unobstructed view for clear images. Be able to capture clear images day and night, with HD resolution (720x1080 px), record video at maximum speed of 60 FPS. Regular lens cleaning to prevent dust and dirt accumulation. Utilize existing traffic surveillance cameras if possible.

Vehicle types: are user-defined and limited to common transports in Viet Nam.

RoI: is a user-defined polygon specifying the area for vehicle counting.

Requirements:

Processing speed: real-time or close to real-time video processing at a minimum speed of 10 FPS, minimal required hardware: 8GB RAM, 2.0 GHz CPU, 2 GB GPU.

Reliability: Precise identification of vehicles of different sizes, shapes, orientations within the RoI, ensuring an overall detection accuracy exceeding 90% on testing set.

Maintains a Miss Counting Rate: under 5% at good conditions (good lighting, no obstruction), under 10% at average complex conditions (light rain, thin fog/dust, no obstruction), under 20% at extremely complex conditions (heavy rain, dense fog/dust, small obstruction small insects such as mosquitoes, flies, moths)) on testing set.

Conclusion

- By applying computational thinking, we found an acceptable solution to the problem posed. In practice, the model can run real-time on a device hardware configuration: 8GB RAM, 2.0GHz CPU, 2GB VRAM GPU (NVIDIA MX250).
- However, the issue of its reliability in use still needs to be tested and evaluated further, before the method is applied in practice.