ARTIFICIAL INTELLIGENCE IN TRANSPORTATION COMPETITION

PRELIMINARY DESIGN REPORT

Team Name: SkyVision AI
Team ID: 479586
Application ID: 1989598

CONTENTS

1. TEAM DIAGRAM

1.1 Team Summary

SkyVision AI team was established in 2024. We aim to contribute to the global AI landscape and create safer, more efficient transportation systems for the future.The SkyVision AI team consists of four undergraduate students. Team chart is presented in figure 1.

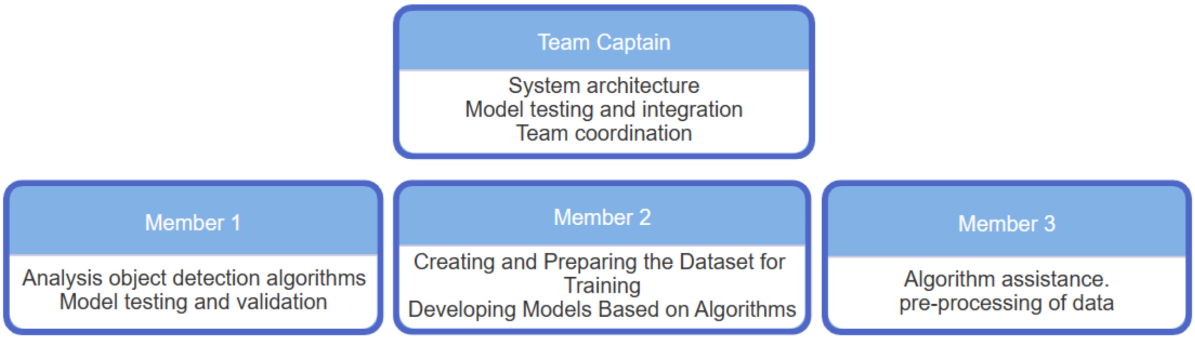


Figure 1: Team Chart of SkyVision AI Team.

2. CURRENT STATUS ASSESSMENT OF THE PROJECT

As of now, we have made progress in labeling and training part of the intended dataset for our model, but development is still ongoing. We are using YOLOv8 and DeepSort algorithms for object detection, and work on pose estimation is also in progress. Initially, we attempted training using an online GPU, but it was not successful. Consequently, we shifted to using a hardware GPU to improve efficiency.

Our current model demonstrates decent accuracy in detecting vehicles, but it faces challenges in accurately detecting humans. To address these issues, we continue to collect and label datasets, refine our training process, and focus on algorithm improvement and optimization. Our team members are actively working on these tasks.

3. ALGORITMS AND SYSTEM ARCHITECTURE

3.1. Datasets

3.1.1. Dataset Provided by the Committee:

- This dataset is provided directly by Teknofest for the current year as well as data from previous years.
- It serves as a foundational dataset for our project, allowing us to train and validate our model on real-world scenarios encountered during Teknofest events.

3.1.2. DOTA Dataset (Dataset of Object Detection in Aerial Images):

- DOTA [1] is a large-scale dataset specifically designed for object detection in aerial images.
- It has 1,793,658 object instances across 18 categories, annotated with oriented bounding boxes.
- The dataset includes 11,268 aerial images, providing diverse scenarios and object variations.

3.1.3. Stanford Aerial Vehicle Dataset:

- This dataset consists of 30,000 images created by Stanford University [2].
- It covers various angles and heights, making it suitable for our project's requirements.

3.1.4. VisDrone-Dataset:

- The VisDrone [3] dataset includes 288 video clips (261,908 frames) and 10,209 static images.
- Captured by different drone-mounted cameras, it covers a wide range of aspects:
    - Locations (from 14 different cities in China).
    - Environments (urban and rural).
    - Objects (pedestrians, vehicles, bicycles, etc.).

3.1.5. Additional Datasets:

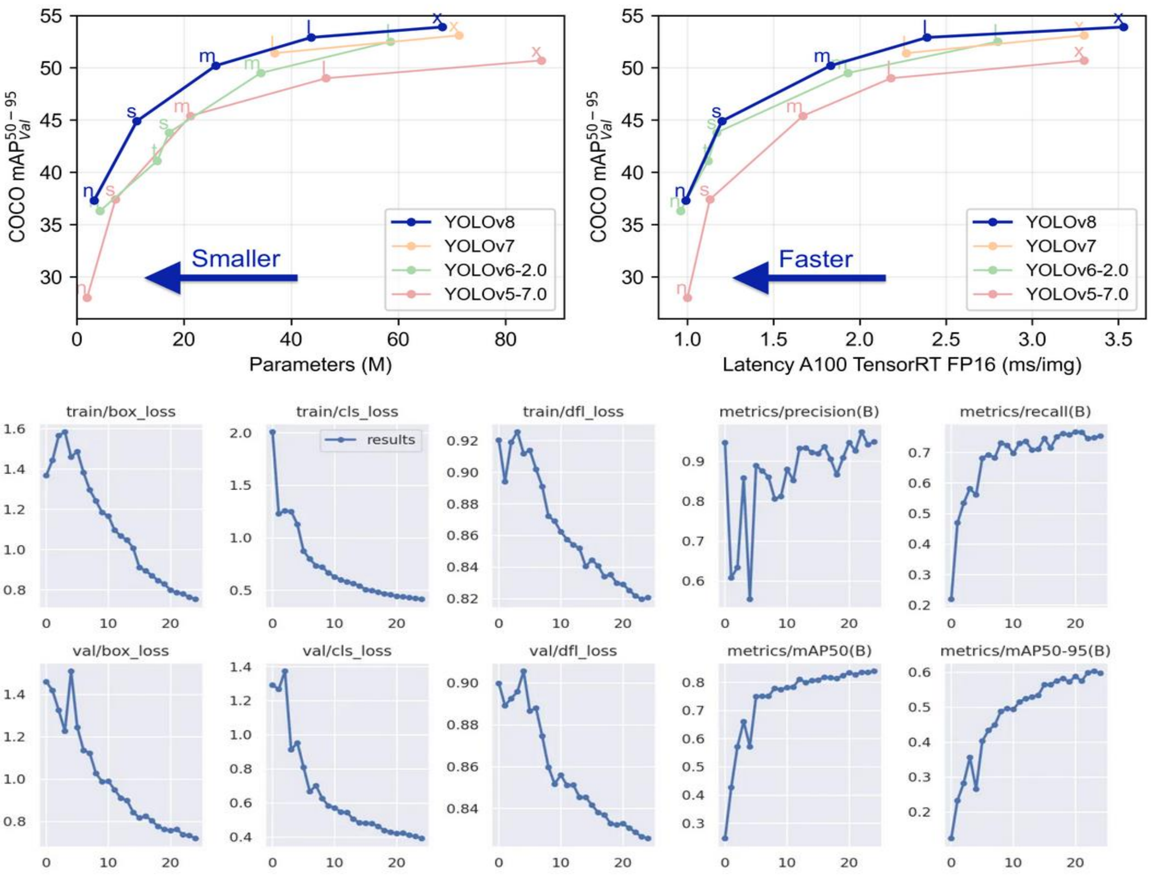
- As our project progresses, we may incorporate other datasets based on the algorithm's performance.
- We will ensure that the data distribution is corrected and aligned with our model's requirements.
- These datasets collectively provide a diverse set of scenarios, object types, and environmental conditions, enabling robust training and evaluation of our model.

3.2. Algorithms

We are considering the use of several algorithms based on their effectiveness and suitability for the task. For now, we are utilizing YOLO algorithms[4].

YOLO (You Only Look Once):

- YOLO is an object detection algorithm that employs a single neural network to predict bounding boxes and class probabilities for objects within an image.
- It stands out for its simultaneous detection and classification, making it efficient and real-time capable.
- YOLOv8, the latest iteration, is currently utilized in our system for object detection.
- We selected YOLOv8 after comparing it with previous versions using custom datasets, and it has demonstrated superior performance.



Key metrics tracked by YOLOv8

3.3. Flow Chart

Flowcharts created for the algorithms of the mission are given in Figure 3.3.1 and Figure 3.3.2.

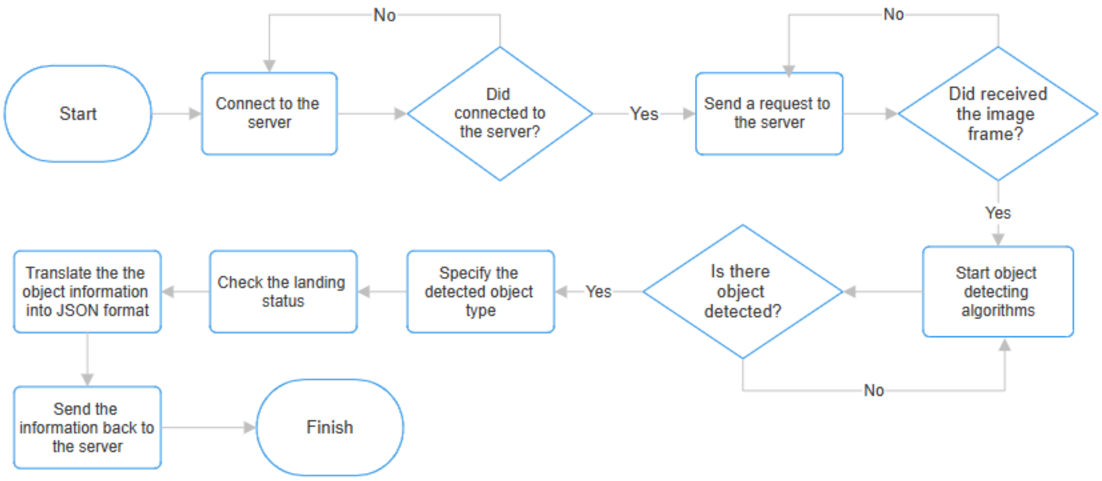


Figure: 3.3.1

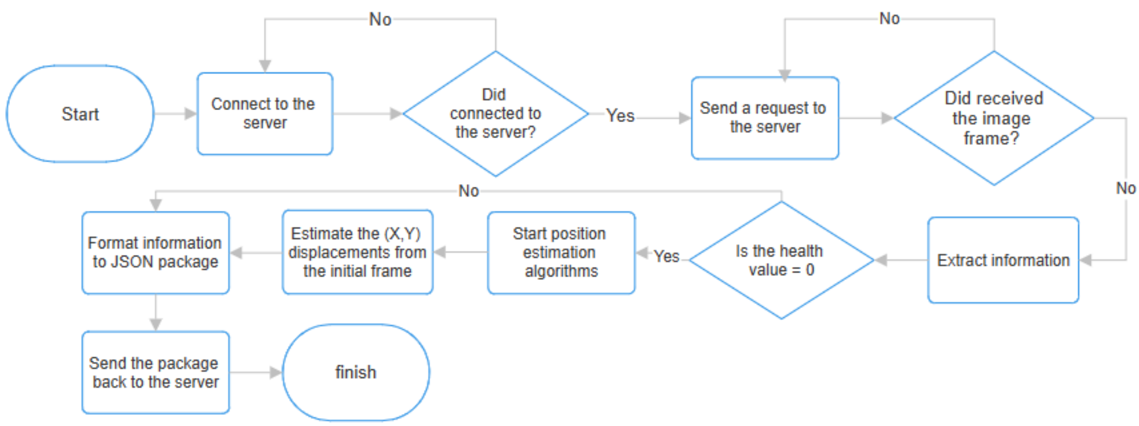


Figure: 3.3.2

4. ORIGINALITY

In our project, we strive to enhance the performance of our model by exploring additional methods that can make it faster and more accurate. Here are the key areas of focus:

1. Hybrid Tracking Approach: YOLOv8 and Deep Sort

- We propose a hybrid tracking structure that combines the strengths of two different libraries: YOLOv8 and Deep Sort [5]. Each library may have its own weaknesses, but by using both, we aim to mitigate these limitations. This approach promises more robust object tracking capabilities.

2. Augmenting Datasets with Parking Station Photos

- While our existing datasets provide a foundation, they may not be sufficient for certain tasks like Flying Ambulance Landing (FAL) and Flying Car Parking (FCP). To address this, we plan to supplement our datasets with additional photos, especially those covering parking stations. These new images will enrich our training data and improve model performance.

3. Larger Training Datasets for Neural Networks

- The Neural Network (NN) benefits from a larger sample size. Therefore, we are actively collecting more images to create an extensive dataset. A well-curated dataset ensures better generalization and flexibility for our algorithm.

4. Optimizing Neural Network Structure

- In coding, simplicity often leads to efficiency. To enhance our Neural Network, we focus on streamlining the code. By minimizing loops and selecting the most suitable libraries, we aim to create a concise yet powerful architecture.

5. PROJECT SCHEDULE

The working plan created according to time planning is given in Table 5.1.

Done

Will be Done

Table 5.1. Time planning table

To Do List	January	Febuary	March	April	May	June
Project planning	Done					
Collecting Dataset		Done	Done			
Labeling Dataset		Done	Done			
Training Dataset			Done			
Validating Dataset			Done			
Object Detection Algorithim Developing			Done	Done	Will be Done	Will be Done
Position Estimation Algorithim Developing			Done	Done	Will be Done	Will be Done
Begin to Develop a Working Model			Done	Done		
PDR				Done		
Finalizing Object Detection Algorithim					Will be Done	Will be Done
Finalizing Estimation Algorithim Algorithim					Will be Done	Will be Done
Finalizing Working Model					Will be Done	Will be Done
Online Contest Simulation					Will be Done	Will be Done

6. RESULTS AND RESEARCH

Our model, leveraging the YOLOv8 algorithm for detection and DeepSort for tracking, is capable of identifying intended objects and pinpointing their coordinates. This is achieved through the use of the OpenCV library and other fundamental Python libraries. The performance statistics of our model are as follows: mAP (Mean Average Precision): 83.9%, Precision: 87.6%, and Recall: 80.5%. Shown Figures 6.1 and Figure 6.2 for an illustrative example of the model's input and output.

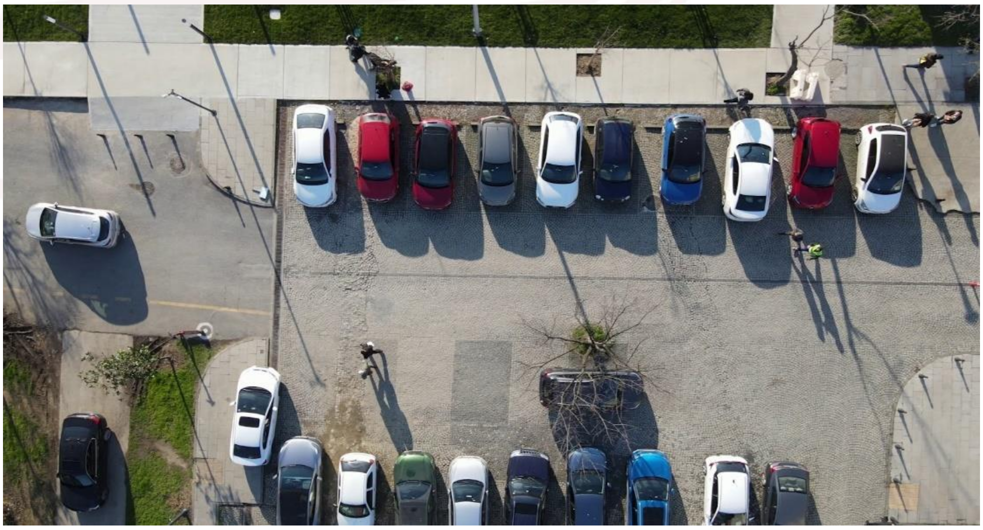


Figure: 6.1. input.

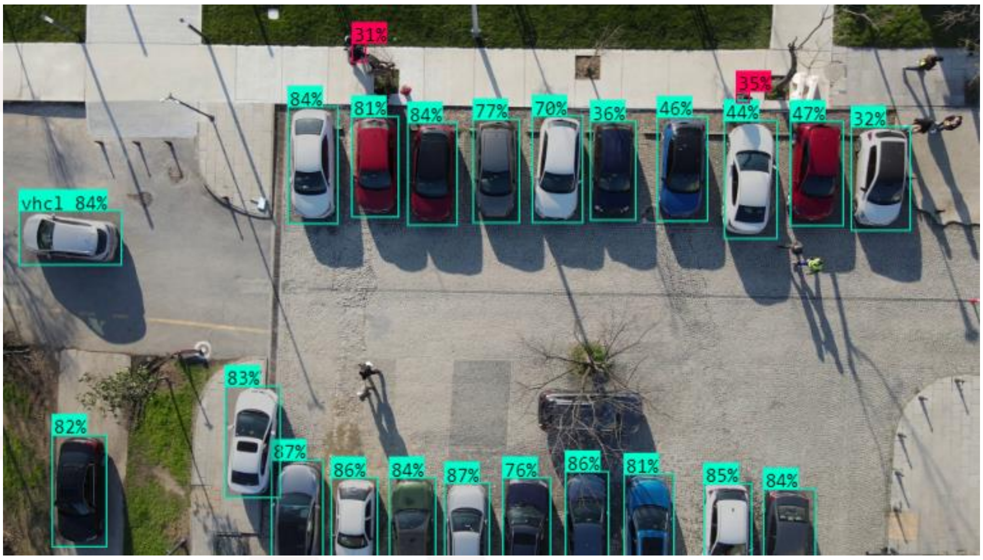


Figure: 6.2. Output.

Our model exhibits proficiency in detecting Vehicles, FAL, and FCP with commendable precision. However, it occasionally fails to detect Vehicles and confuses between FAL and FCP. Moreover, the model's performance in detecting humans is subpar. The reasons for these shortcomings include:

- The diminutive size of human objects within the frames.
- The scarcity of human objects in the dataset.
- The inability to recognize objects due to environmental factors and lack of variety.
- The model's failure due to insufficient memory for operation.

To address these issues, we propose the following solutions:

- Augment the frames using OpenCV to enhance the model's object detection capabilities.
- Optimize the algorithms to reduce memory consumption, thereby improving the model's efficiency.
- Enrich the dataset to increase its variety.

In general, our team is continuously working on improving the model to achieve the desired level of performance.

7. REFERENCE

[1] Ding, J., Xue, N., Long, Y., Xia, G., & Lu, Q. (2019, June). *Learning RoI Transformer for Detecting Oriented Objects in Aerial Images.* In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[2] A. Robicquet, A. Sadeghian, A. Alahi, S. Savarese, *Learning Social Etiquette: Human Trajectory Prediction In Crowded Scenes* in European Conference on Computer Vision (ECCV), 2016.

[3] Zhu, Pengfei, Wen, Longyin, Du, Dawei, Bian, Xiao, Fan, Heng, Hu, Qinghua, & Ling, Haibin. (2021). *Detection and Tracking Meet Drones Challenge.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11), 7380-7399.

[4] A. Nazir and M. A. Wani, "You Only Look Once - Object Detection Models: A Review," *in 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom),* New Delhi, India: IEEE, 2023.

[5] Nicolai Wojke, Alex Bewley, and Dietrich Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3645-3649, 2017, IEEE