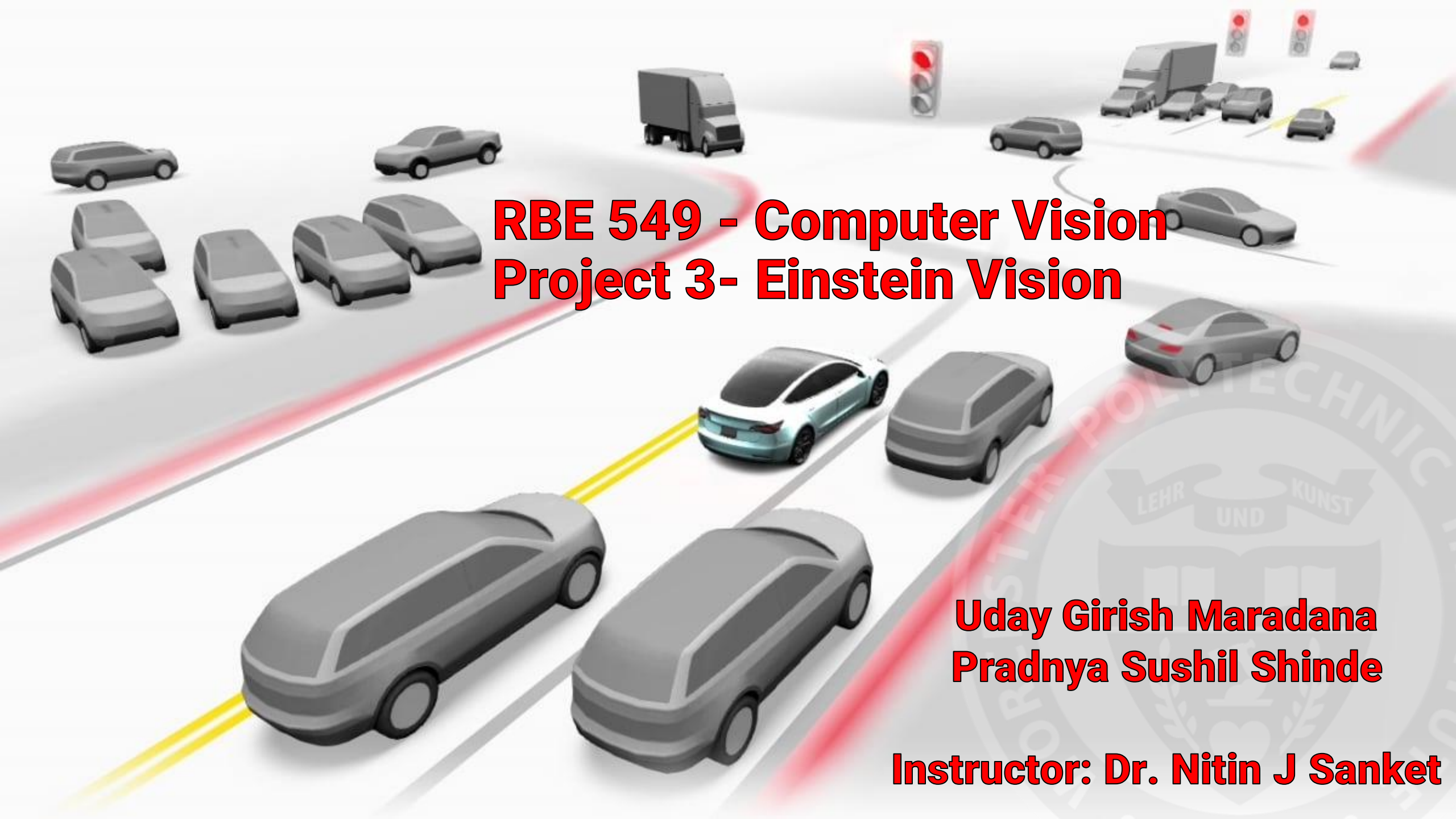


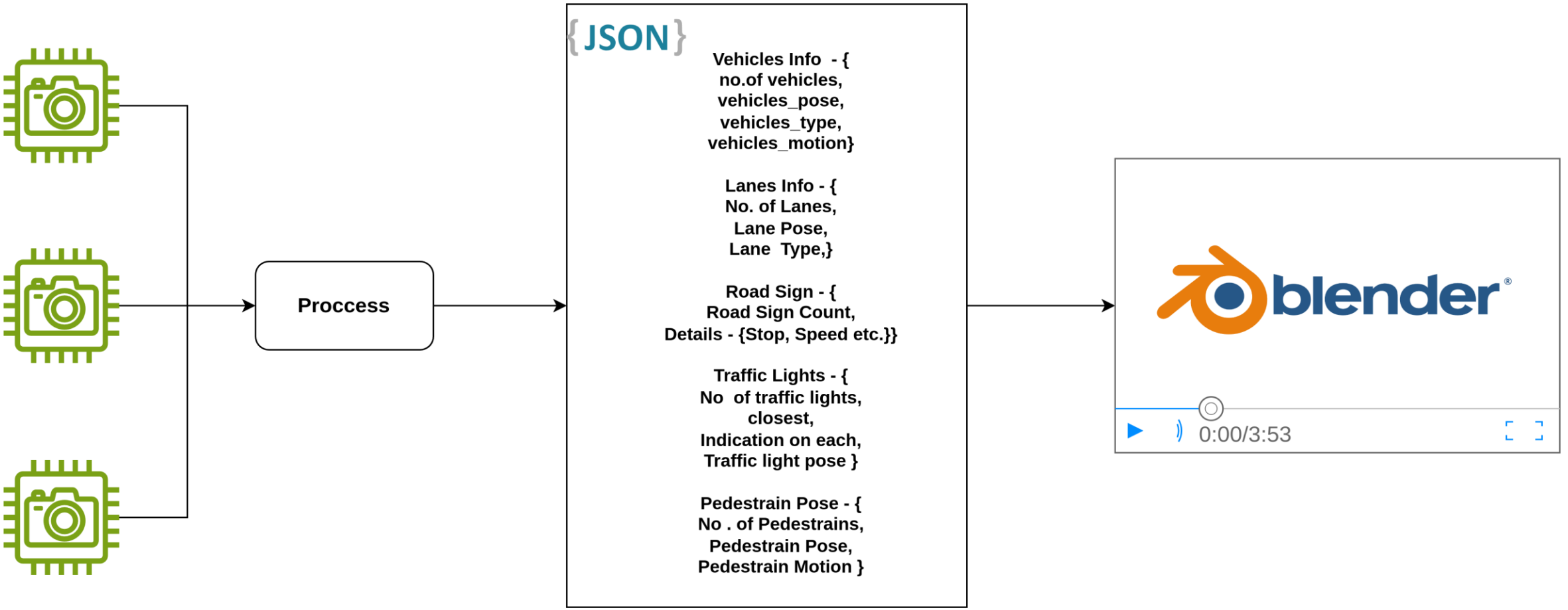
RBE 549 - Computer Vision
Project 3- Einstein Vision

Uday Girish Maradana
Pradnya Sushil Shinde

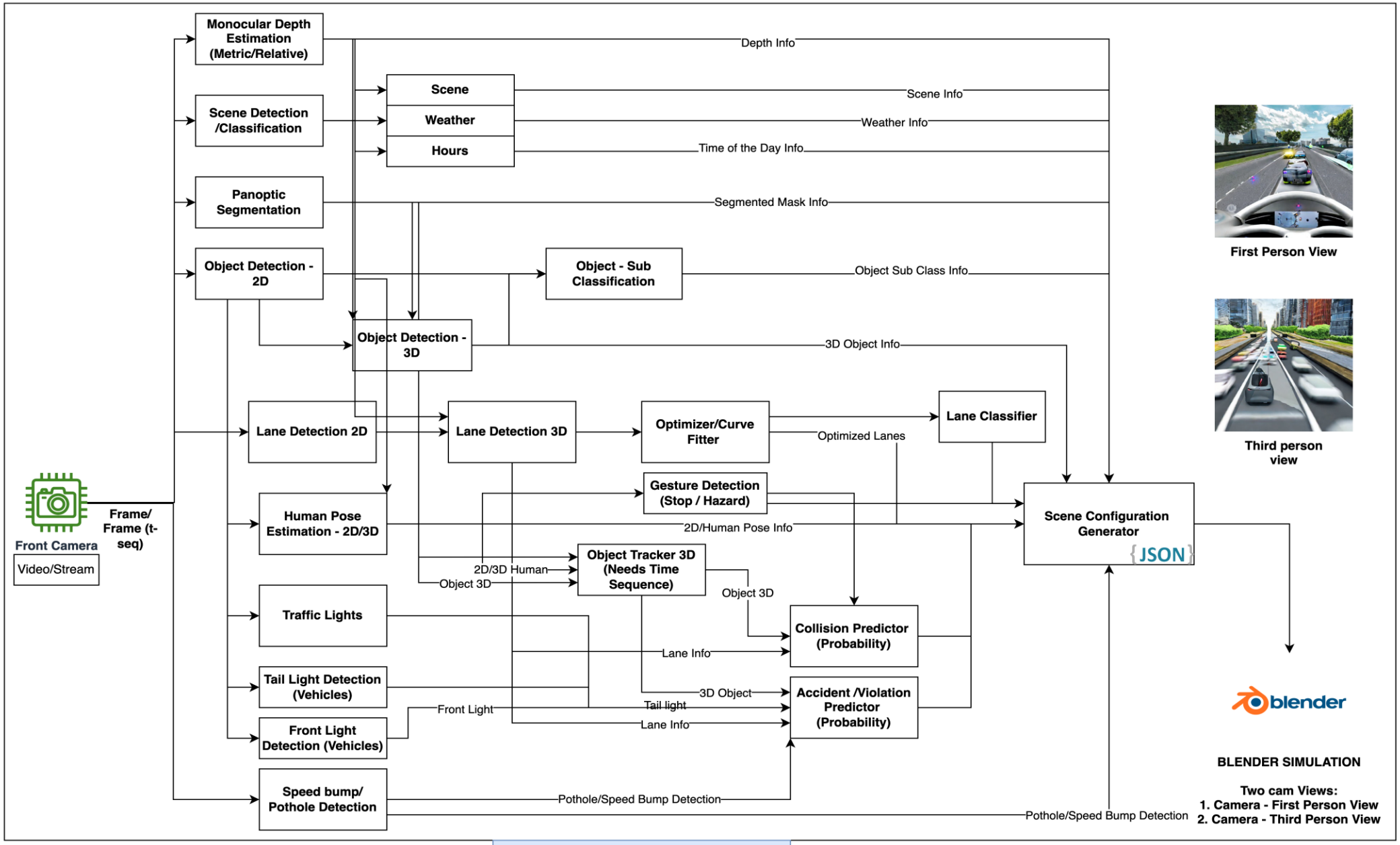
Instructor: Dr. Nitin J Sanket

Introduction

The aim of this project is to understand the concepts behind how to extract valuable information from a scene (Camera Stream) to gain insights and build an overall visualization system.

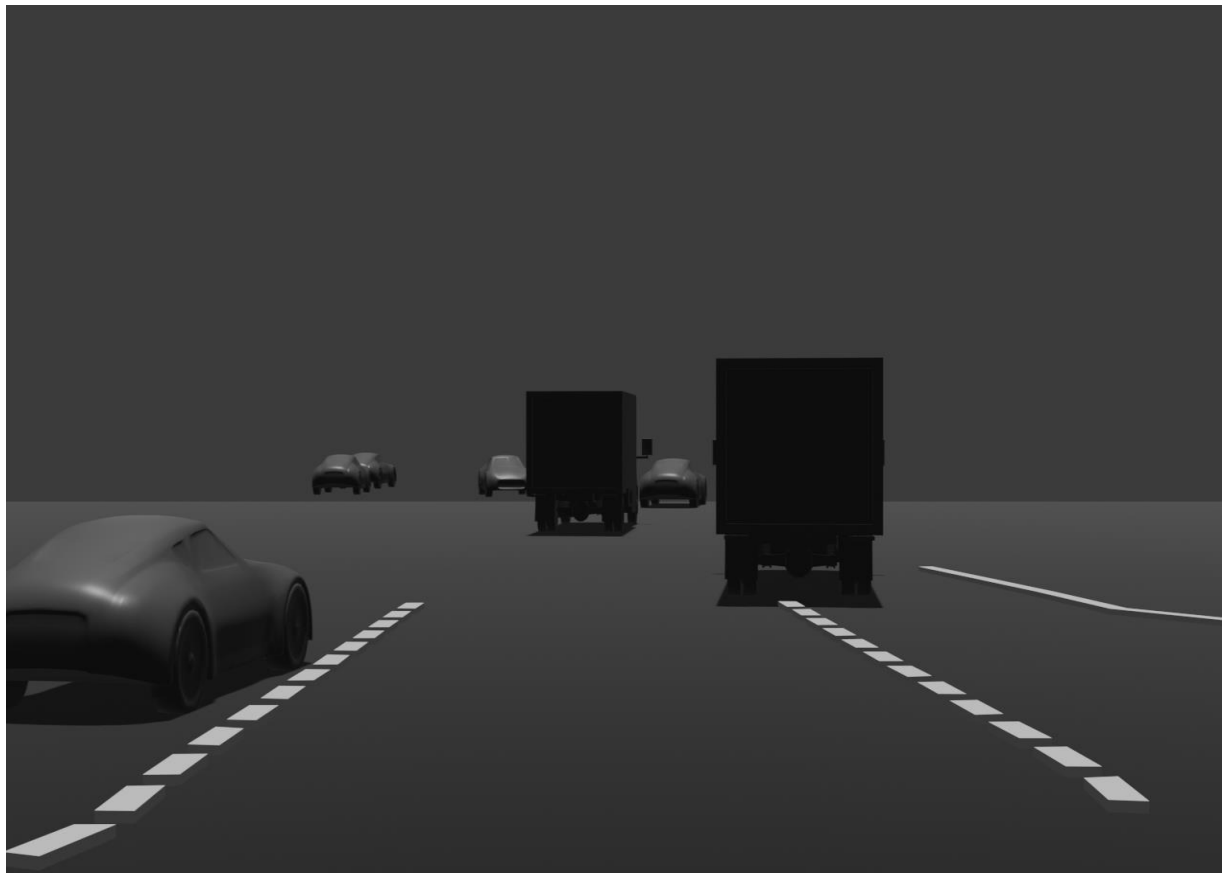


Initial thoughts ?



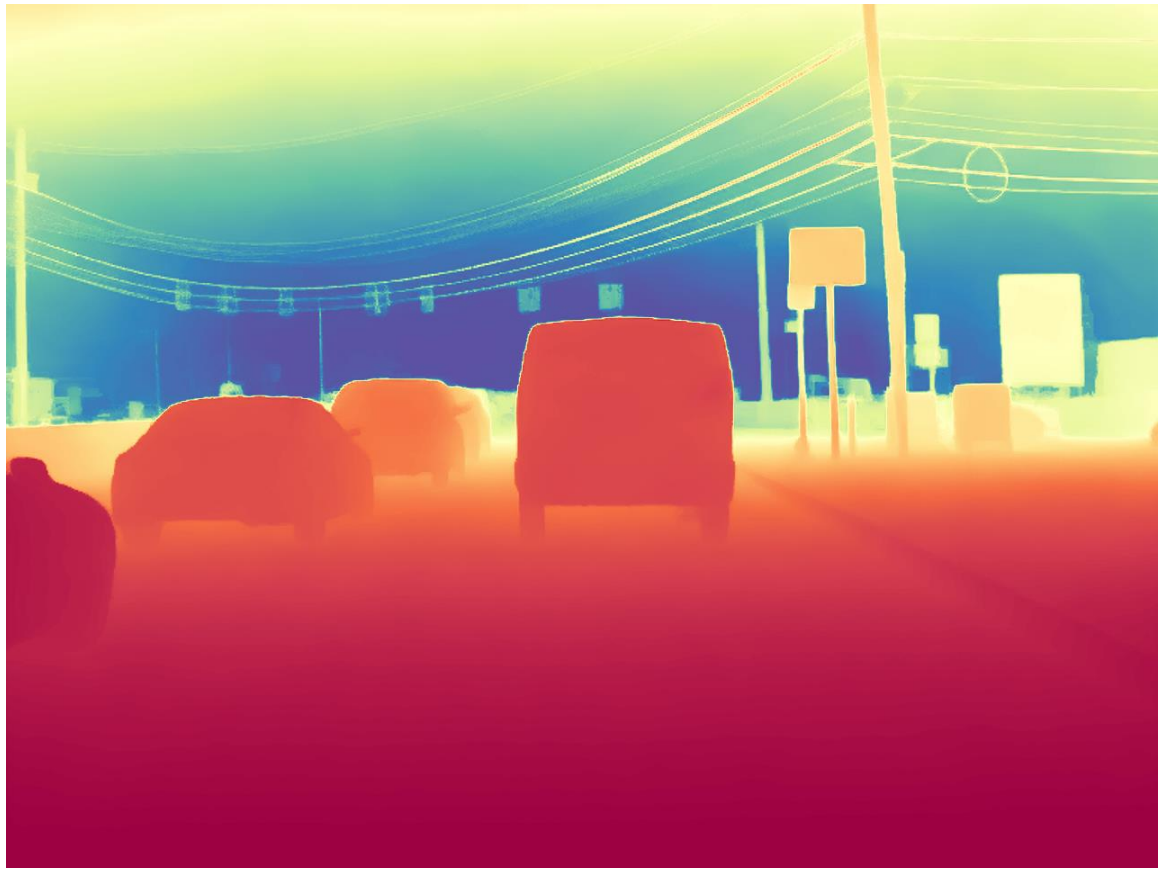
Phase II - (Advanced Single Video Pipeline)

Lane Detection – CLRer Net + Bezier Curves



H. Honda and Y. Uchida, 'CLRerNet: Improving Confidence of Lane Detection with LaneIoU', *arXiv [cs.CV]*. 2023.

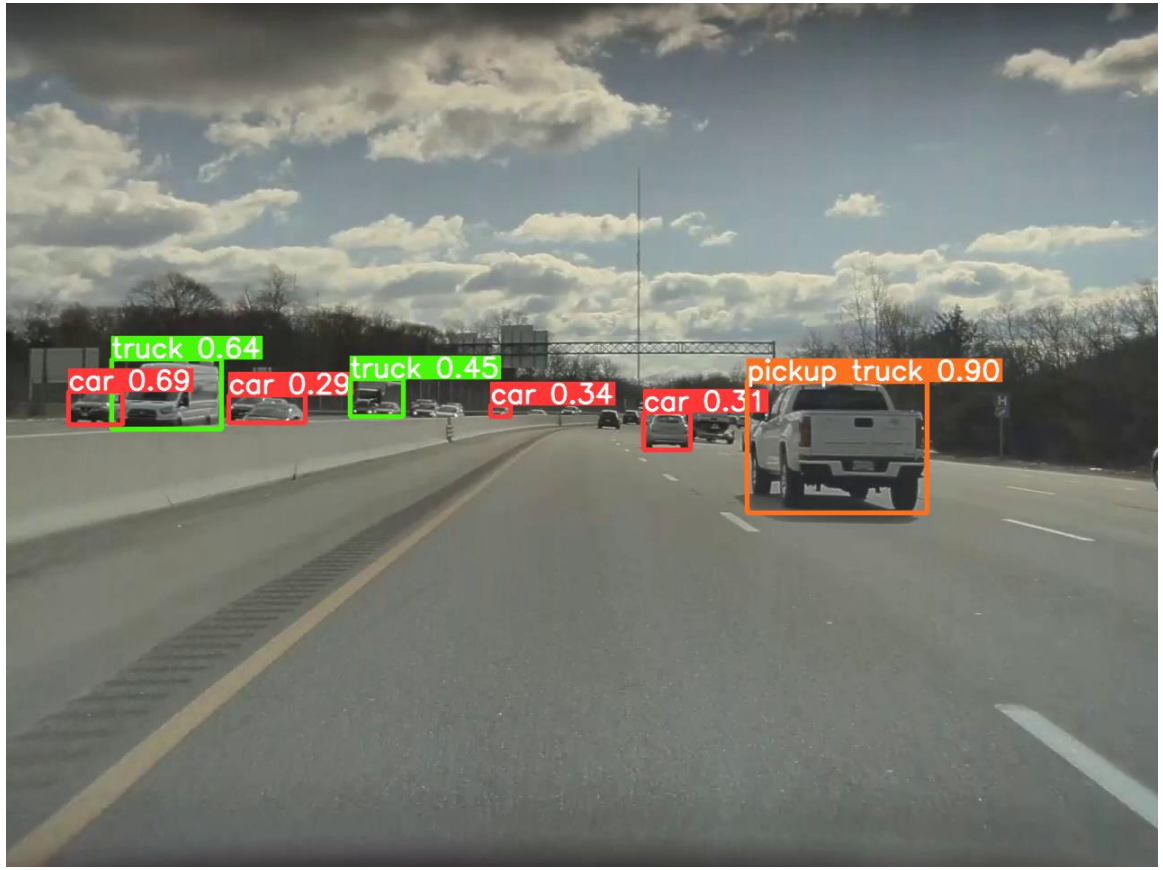
Depth Estimation – Marigold/ZoeDepth

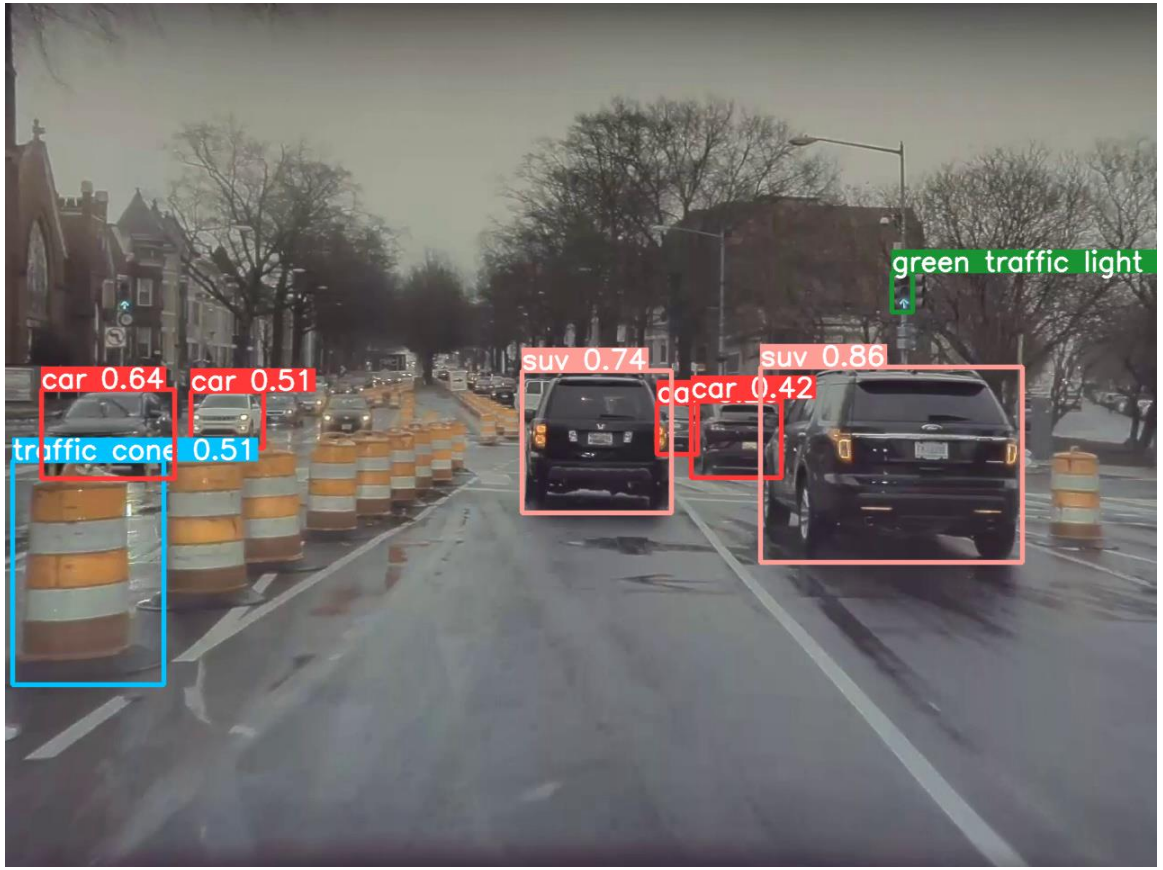


B. Ke, A. Obukhov, S. Huang, N. Metzger, R. C. Daudt, and K. Schindler, 'Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation', *arXiv [cs.CV]*. 2024.

S. F. Bhat, R. Birkl, D. Wofk, P. Wonka, and M. Müller, 'ZoeDepth: Zero-shot Transfer by Combining Relative and Metric Depth', arXiv [cs.CV]. 2023.

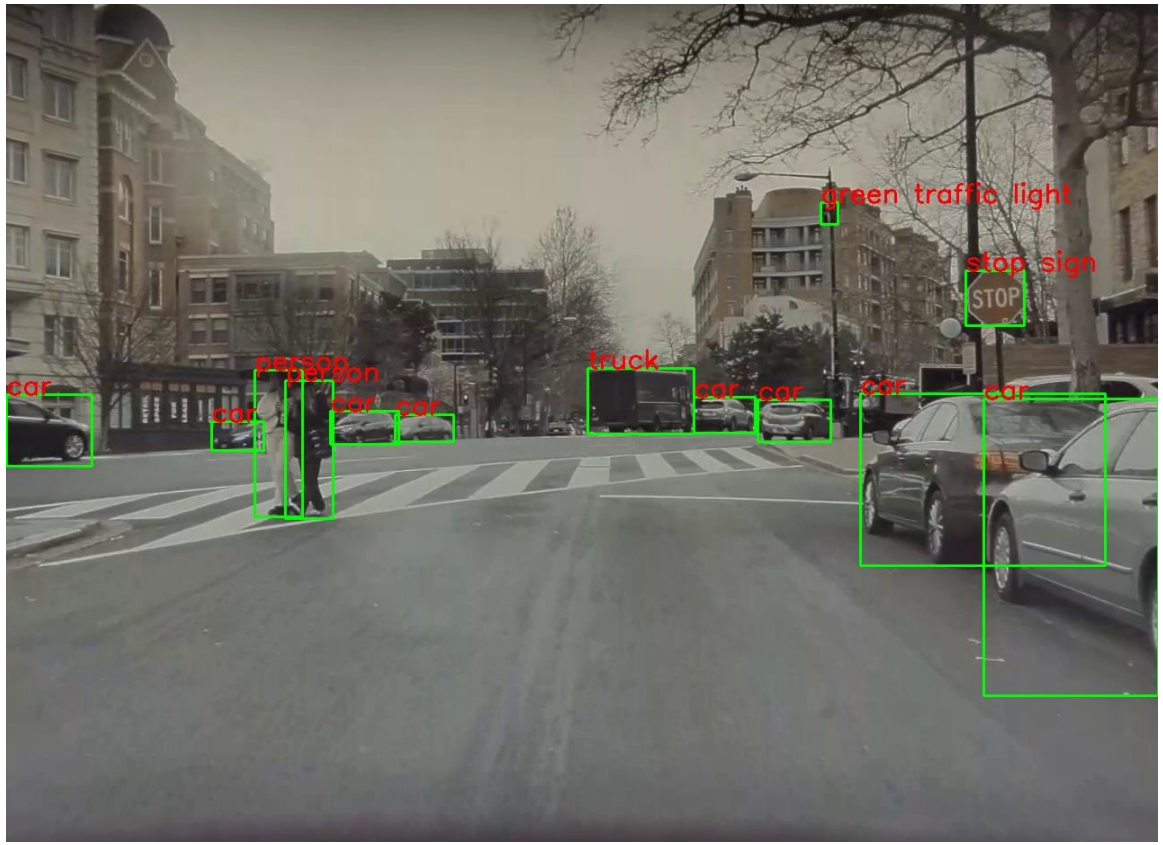
Object Detection 2D – Yolo World

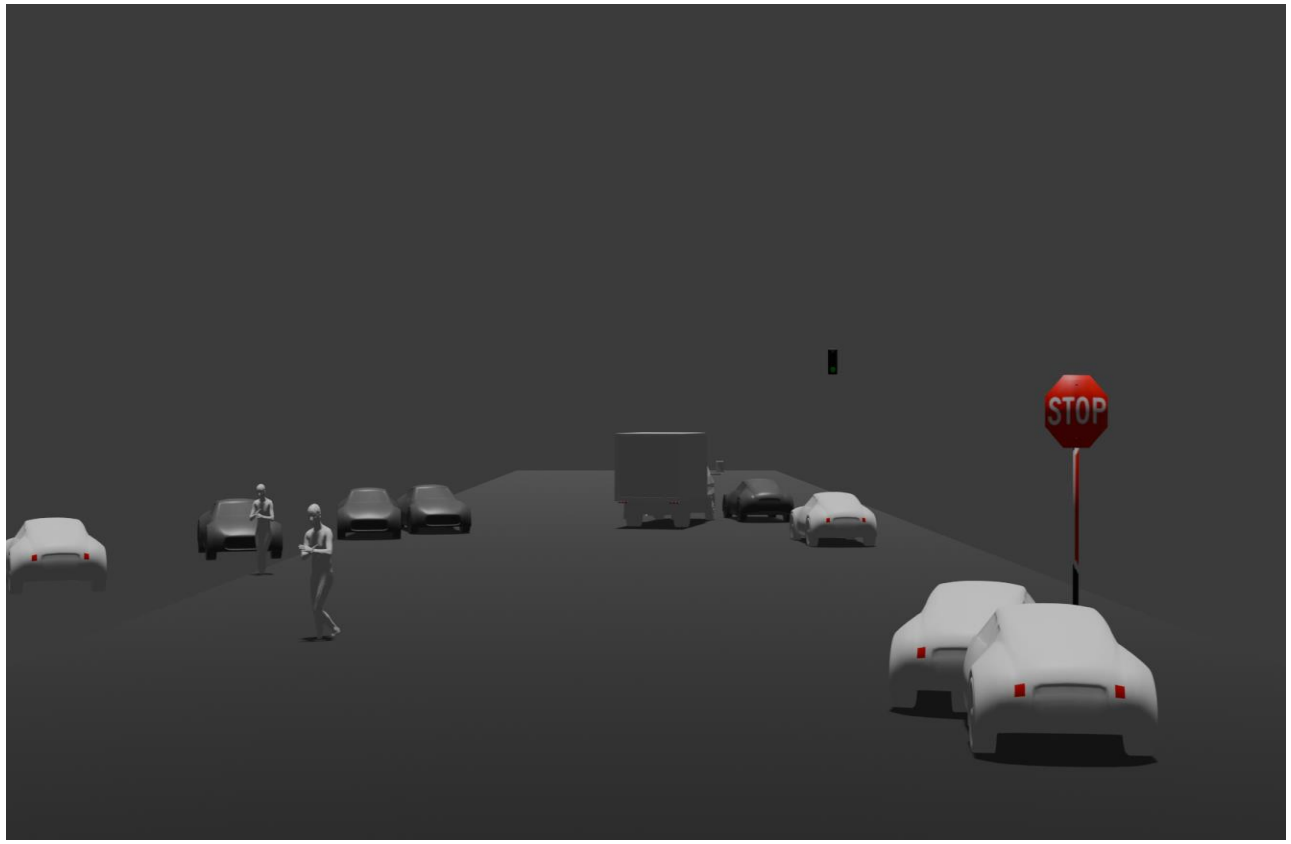




T. Cheng, L. Song, Y. Ge, W. Liu, X. Wang, and Y. Shan, 'YOLO-World: Real-Time Open-Vocabulary Object Detection', *arXiv [cs.CV]*. 2024.

Object Detection 2D – Yolo World





T. Cheng, L. Song, Y. Ge, W. Liu, X. Wang, and Y. Shan, 'YOLO-World: Real-Time Open-Vocabulary Object Detection', *arXiv [cs.CV]*. 2024.

Object Detection 3D – Yolo 3D

A. Mousavian, D. Anguelov, J. Flynn, and J. Kosecka, '3D Bounding Box Estimation Using Deep Learning and Geometry', *arXiv [cs.CV]*. 2017.

https://github.com/ruhyadi/YOLO3D

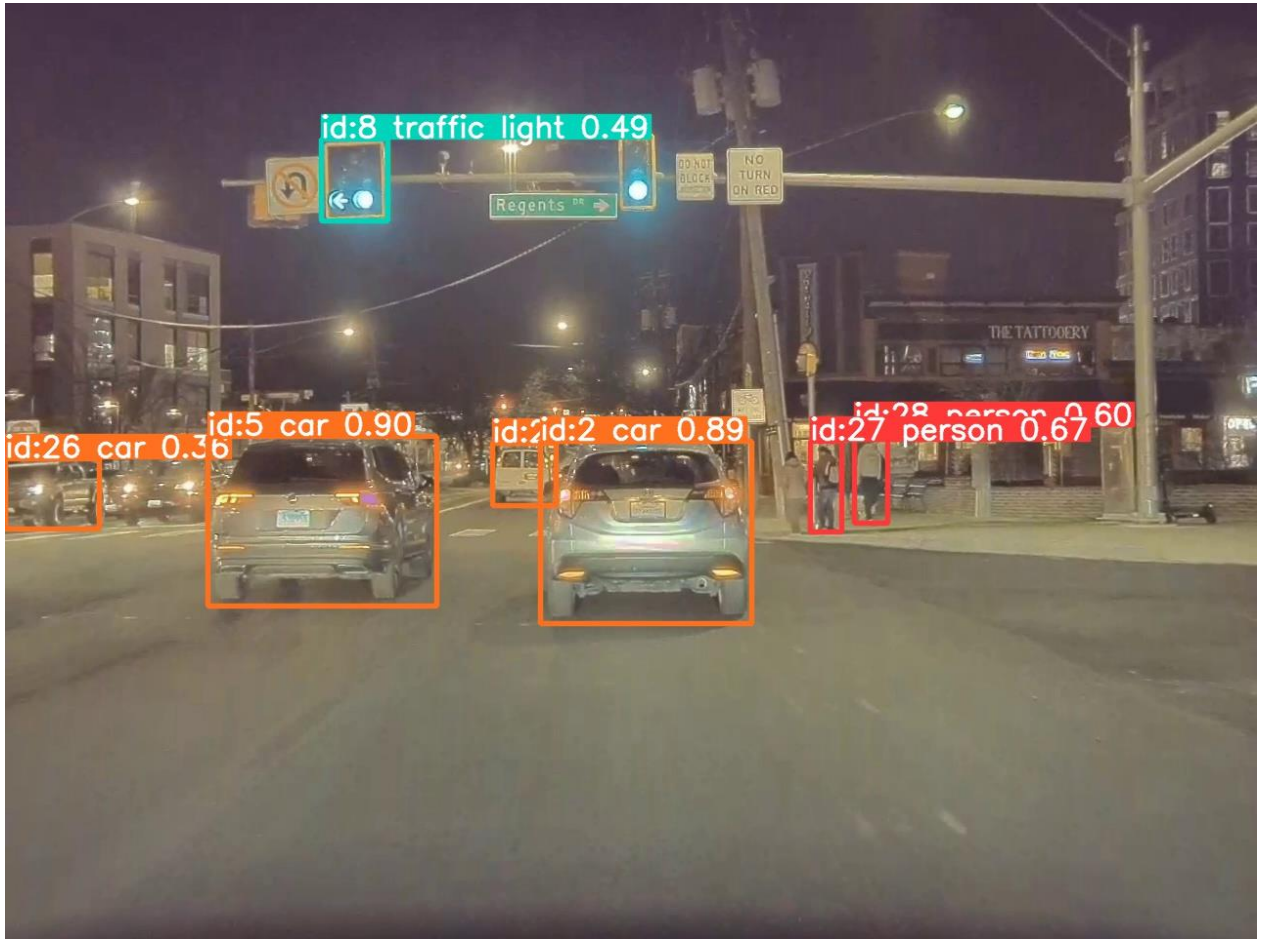
Road Markings

PyMAF

A. Mousavian, D. Anguelov, J. Flynn, and J. Kosecka, '3D Bounding Box Estimation Using Deep Learning and Geometry', *arXiv [cs.CV]*. 2017.

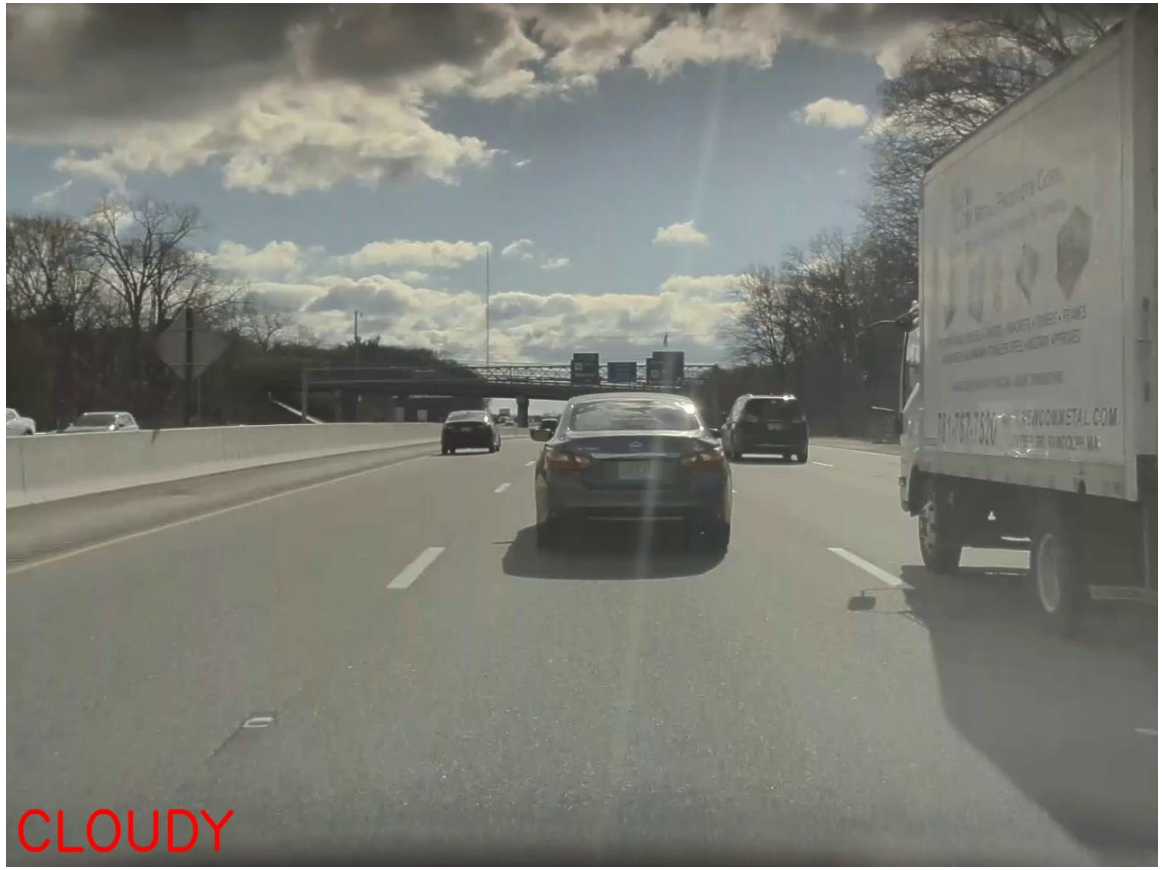
H. Zhang *et al.*, 'PyMAF-X: Towards Well-Aligned Full-Body Model Regression From Monocular Images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–16, 2023.

Multiple Object Tracking

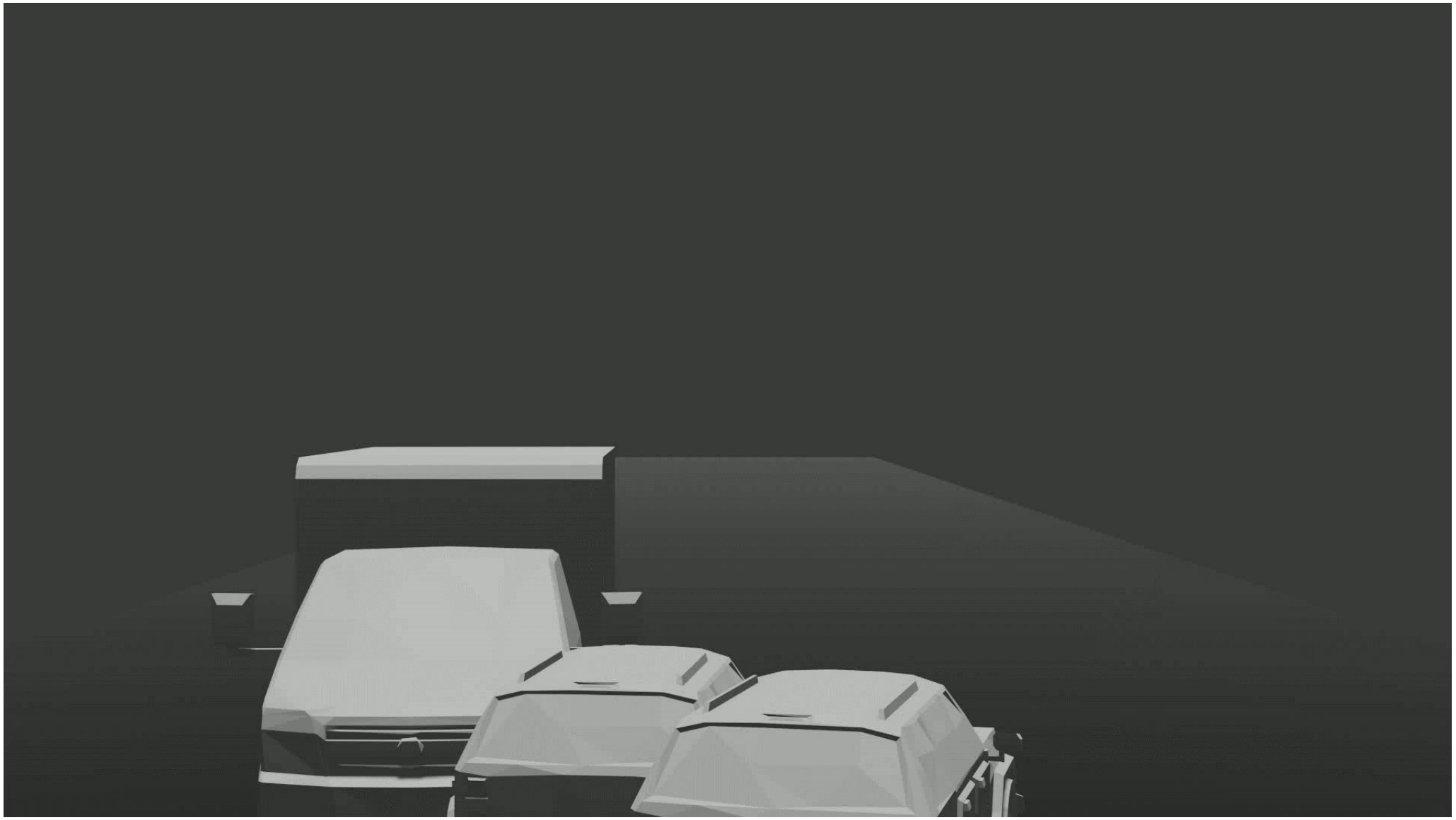


Y. Zhang *et al.*, 'ByteTrack: Multi-Object Tracking by Associating Every Detection Box', *arXiv [cs.CV]*. 2022.

Scene Classification

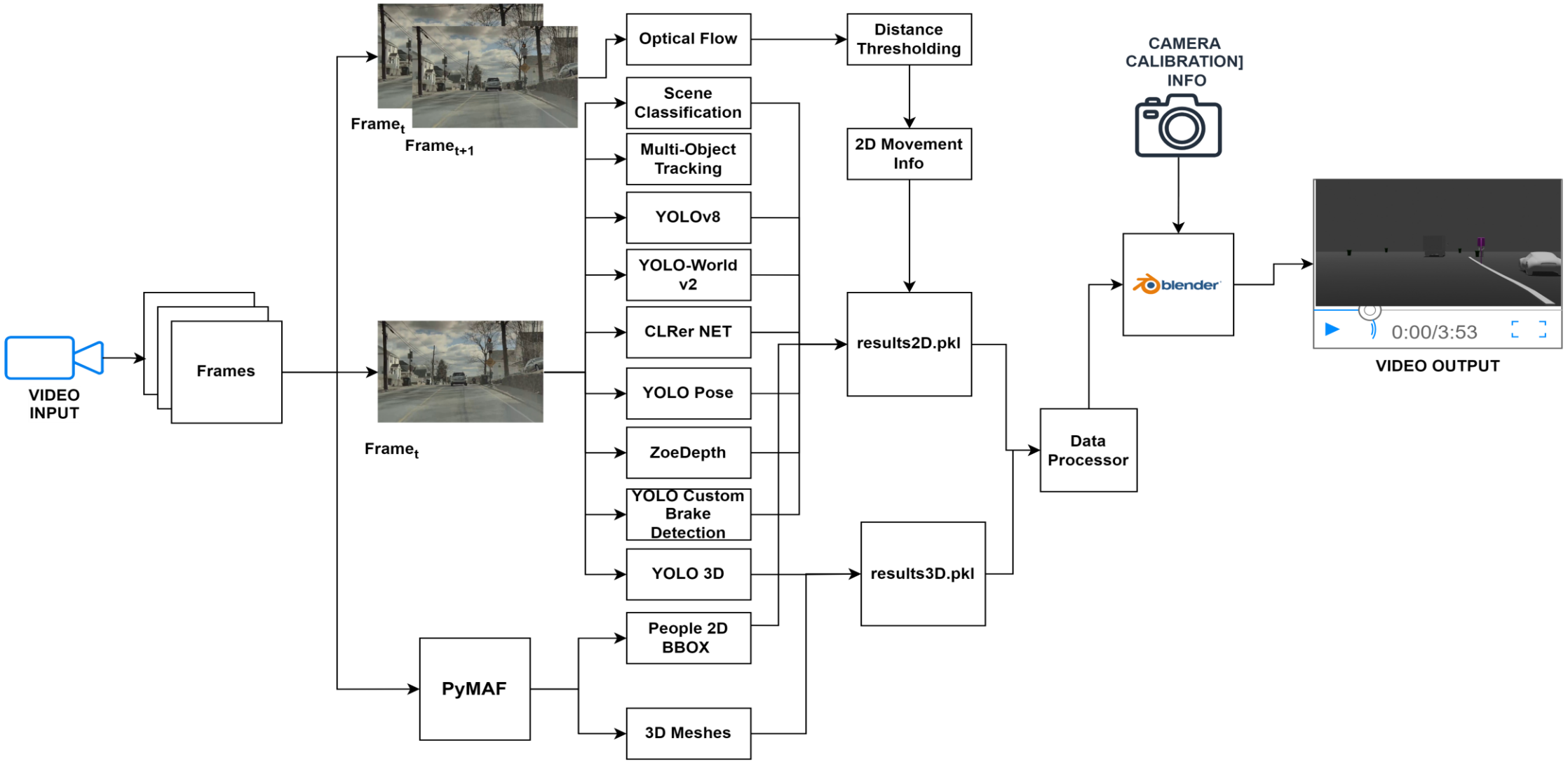


Video Output – Sample (Not that Great!)



Pipeline Overview

The following diagram shows an overview of the Final Perception & Rendering Pipeline.



Pipeline Overview – Output Structure

Each object final Structure:

{

'3d_world_coords': list(tuple(float, float, float))

'bbox_2d': list(tuple(float,float))

'class_name': string

'orientation': tuple(float, float, float)

'scale': tuple(float,float, float)

'score': float (Yolo World Det – Probability)

'state_label': string

'avg_velocity': float

'track_id': int

'pose_path': string (.obj file location)

}

Challenges

- Depth Inaccuracy leading to miscalculation of Object's 3D Position thereby we can see sometimes object randomly getting positioned in the simulation.
- Yolo 3D mispredictions leading to scale and orientation issues.
- PyMAF mesh Object - Sometimes the fit of the mesh is wrong. This was somehow seems solved from PyMaf SIMPLX but haven't tested it.
- Lane Classification is not very Robust.

Future Work Possible

- Better version of Depth should be used. - Marigold Depth
- Use of PyMAF SIMPLX Version.
- Yolo 3D Backend Updation and Training with Yolo World or V2
- Lane 3D based classifier - Something like Persformer (Tried – But seems the weights are wrong)
- Road Markings Robust Classification

Discussion

This is heavy ! So how does real Autonomous Vehicles does this?

We wrote , what we read from several places please scan the QR !
Mostly the concept is Multi headed Net or HydraNet some call those type of networks.
Watch Andrej Karpathy Talks on youtube!

SCAN ME

https://tinyurl.com/karpathytalks

Thank you !!