Project 3

Einstein Vision

Niranjan Kumar Ilampooranan &
Thanikai Adhithiyan Shanmugam

Worcester Polytechnic Institute

April 11, 2024

Motivation

- Self-driving car - plethora of important data
- In need of compact presentation of data to user
- Neat UI/visualization goes a long way

Objective

- Detect important features from scene - dashcam footage
  - Lane markings
  - Vehicles in proximity
  - Traffic lights
  - Pedestrians
  - Status of each element
- Render on Blender

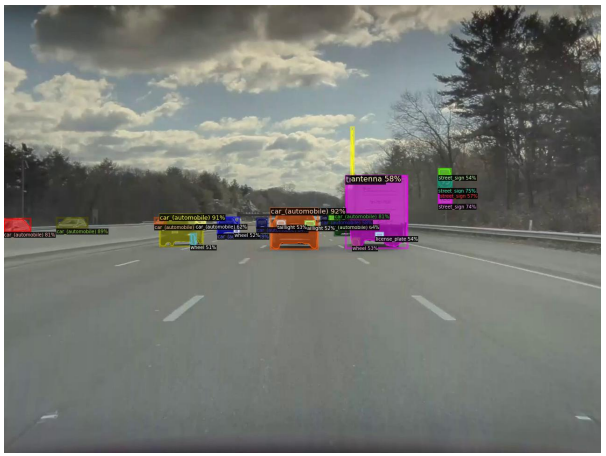
Object Detection - YOLOv9

- Objects Detected [1] - Cars, Trucks, Stop Sign, Pedestrians, SpeedBreakers, Traffic Lights
- Pre-trained weights - MSCOCO dataset, Custom weights

- Challenges - Lot of False Positives, Signs and Objects not detected.

Object Detection - Detic

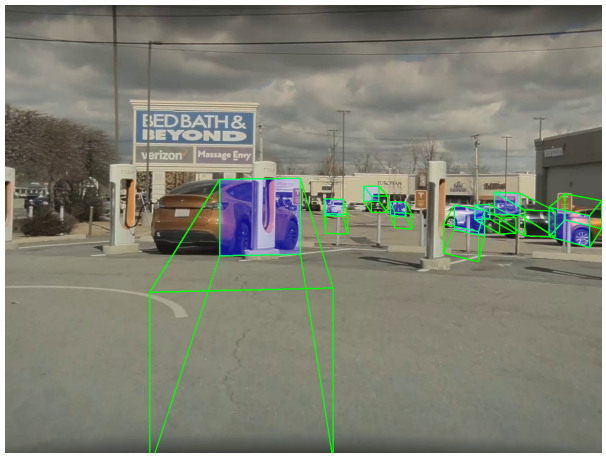
- Objects Detected [2] - 30,000 Classes !!! + Cars Classification
- Pre-trained weights - Open-vocabulary LVIS, COCO

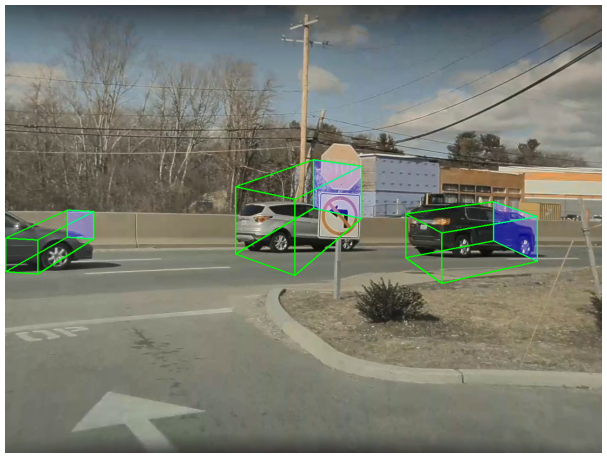


- Challenges - Low resolution on cars, Integrating with yolo3D provided bad results

Object Orientation - Yolo3D

- Detects 3D bounding boxes and yaw of vehicles [3]
- Uses yolov8(2D) and Regressor(ResNet) for 3d bounding box





- Challenges - Pre-trained results were really bad. No proper pre-trained network for the regressor.

Depth Estimation

Zoedepth [4]

- Pretrained Weights - NYU Depth and KITTI Dataset
- Scaling Ambiguity within each frame of a scene
- Works Poorly for fairly distant objects

Marigold [5]

- Uses Stable Diffusion models
- Marigold outperformed Zoedepth in every criterion.
- Scaling - set constant through each frame of every scene

Pedestrian Pose Estimation

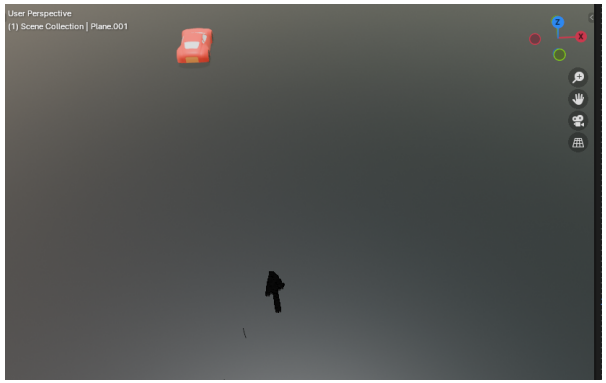
- I2L-MeshNet [6] for pose estimation of humans
- Image cropped - only human - sent to network
- Example output - rendered mesh imposed

- Network - cascaded PoseNet and MeshNet

Lane Detection

Mask R-CNN

- Model - Yolov8 , Dataset - Custom Dataset(Kaggle)



Blender visualization of lanes

- Points detected - Bezier curve control points
- Modified mesh to resemble shape of curve
- For arrow markings - points considered as interior of polygon mesh

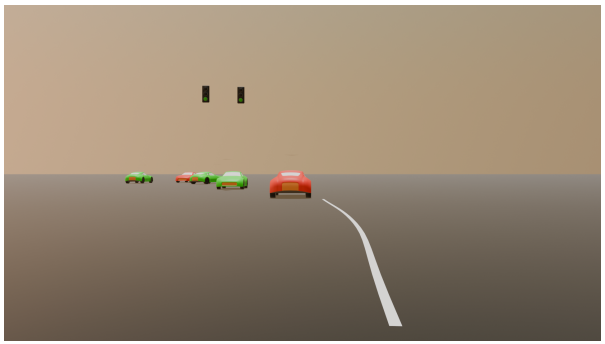
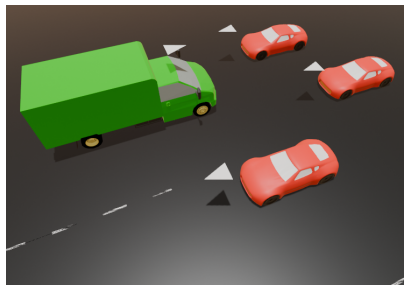
Optical Flow

RAFT [7]

- Provides Optical Flow image for every frame. (x and z direction)
- Pre-trained Weights - KITTI, SINTEL, Flying Things-3d

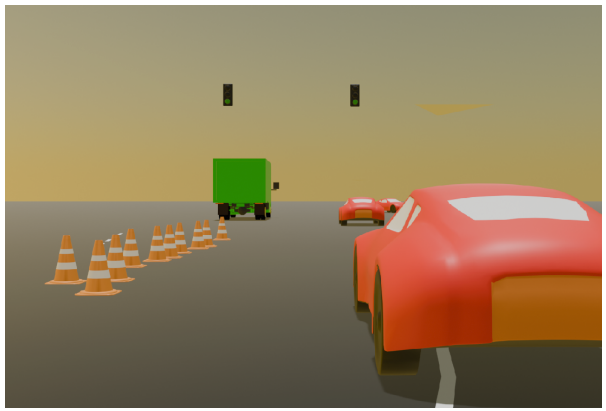
Classification

- CV2 for homography, SIFT features.
- Reprojected image1 correspondences on image2.
- Used Sampson distance, computed error and set threshold.
- Arrows above car showing predicted trajectory of vehicle.

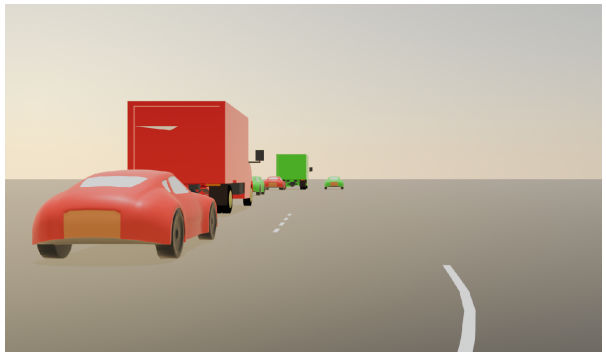




Footage vs Rendered Scene

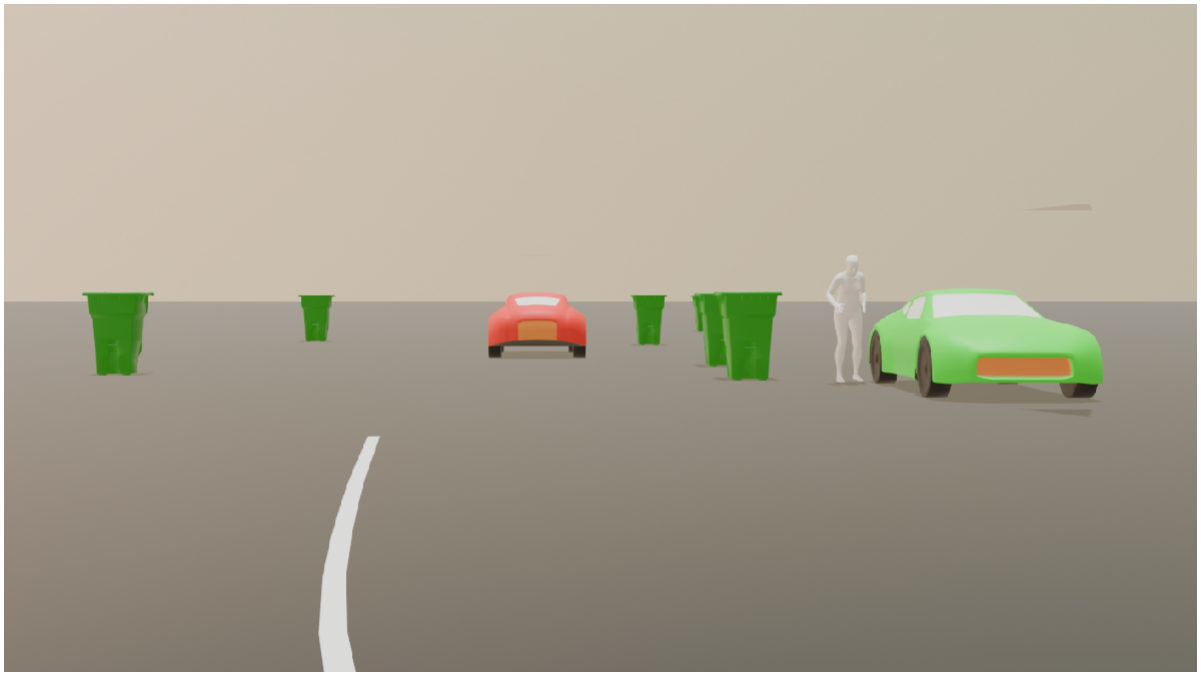
- Sample from scene 2



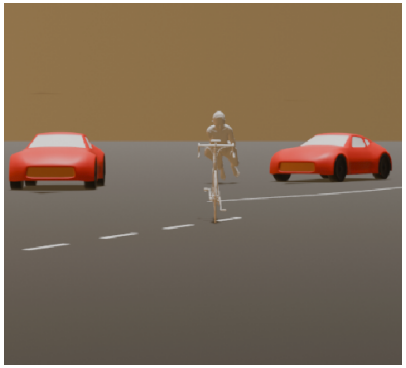
- Sample from scene 13



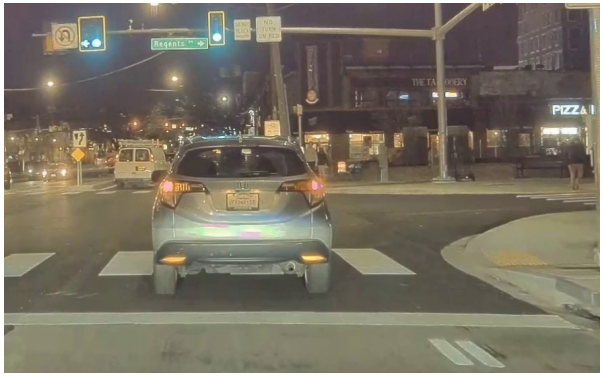
Result - Video



Challenges - Human Pose



- Poor results I2L -
  - when human only partially visible or in different orientations
  - when human too close to another human or even an object
  - indistinguishable colour of attire from background



[1] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," 2024.

[2] X. Zhou, R. Girdhar, A. Joulin, P. Krähenbühl, and I. Misra, "Detecting twenty-thousand classes using image-level supervision," in *ECCV*, 2022.

[3] A. Mousavian, D. Anguelov, J. Flynn, and J. Kosecka, "3d bounding box estimation using deep learning and geometry," 2017.

[4] S. F. Bhat, R. Birkl, D. Wofk, P. Wonka, and M. Müller, "Zoedepth: Zero-shot transfer by combining relative and metric depth," *arXiv preprint arXiv:2302.12288*, 2023.

[5] B. Ke, A. Obukhov, S. Huang, N. Metzger, R. C. Daudt, and K. Schindler, "Repurposing diffusion-based image generators for monocular depth estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[6] G. Moon and K. M. Lee, "I2l-meshnet: Image-to-lixel prediction network for accurate 3d human pose and mesh estimation from a single rgb image," 2020.

[7] Z. Teed and J. Deng, "Raft: Recurrent all-pairs field transforms for optical flow," 2020.

[8] H. Zhang and C. Ye, "Sampson distance: A new approach to improving visual-inertial odometry's accuracy," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 9184–9189, 2021.