Latitude AI 1: Detecting 3D Objects in Autonomous Driving

AI Studio Final Presentation

Break Through Tech Virtual Program @ Cornell Tech
December 6th, 2024

Introductions

Meet Our Team!

Salma Bhar
Case Western Reserve University

Natasha Kamara Martinez
Wellesley College

Jessamine Qu
Carnegie Mellon University

Sara Phondge
Rutgers University

Our AI Studio TA and Challenge Advisor

Kade Lin
AI Studio TA

Jhony Kaesemodel Pontes
Challenge Advisor

Presentation Agenda

1. Project Overview
2. Initial Approach and New Approach
3. Modeling and Evaluation
4. Final Thoughts
5. Questions?

AI Studio Project Overview

Project Overview

This project bridges the gap between 2D object detection and 3D scene understanding, essential for applications like self-driving cars and robotics. While 2D detection identifies object locations in images, it lacks depth information, crucial for safe navigation. Our goal is to enhance 2D detections with depth to approximate a 3D scene understanding.

Our Goals

1. Use an open-source 2D detection model on a single camera view from the nuScenes dataset.
2. Enhance the accuracy of object detection with depth information.
3. Contribute to the development of safer autonomous systems by addressing challenges in depth-aware object detection.

Business Impact

- **Enhanced Safety and Navigation**
- **Increased Depth Awareness**
- **Enhanced System Reliability**

Resources We Leveraged

- Google Colab
- NuScenes Dataset
- Detectron2
- DepthAnything
- Matplotlib, Numpy, Scipy

Background

Detectron2

2D Object Detection with Detectron2

- What it is?
- How we use it?
- Why it's useful?

Limitations of 2D Detections

- Lack of Depth Information?
- Why Depth Matters?

Data Understanding

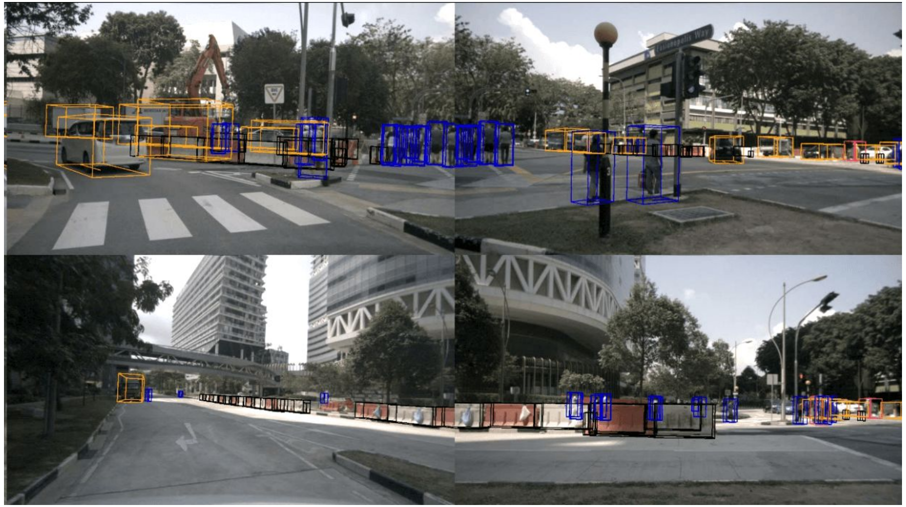
Our Initial Approach

Steps

1. Feature Extraction
2. Bird's Eye View Transformation
3. 3D Bounding Box Prediction
4. Training with Ground Truth

Why We Simplified

- Time Constraints
- Resource Limitations



Our Initial Approach: Lessons learned

- Set Realistic Expectations
- Recognize Resource Constraints
- Adapt to Uncertainty

Key Takeaway: This process wasn’t time wasted! It was an invaluable chance to strengthen our technical abilities and enhance teamwork.

Modeling

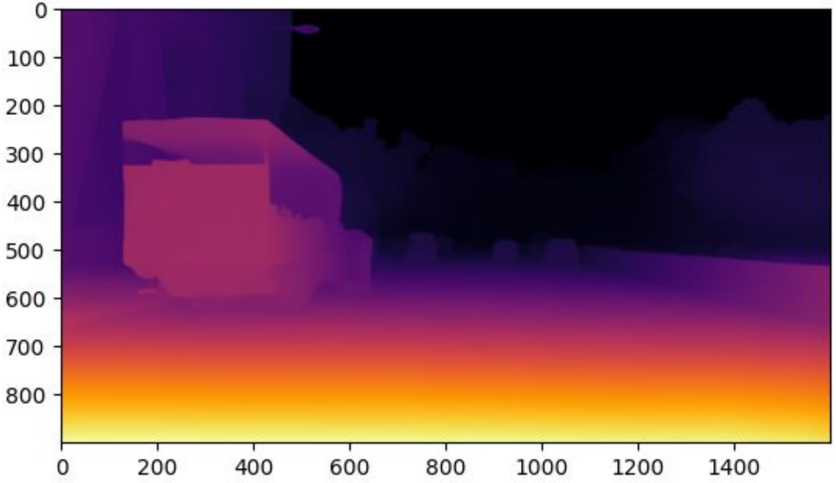
Our New Approach

1. Focus on Front Camera View

- Simplifies problem
- Reduces computational requirements.

2. Integrate Depth Estimation

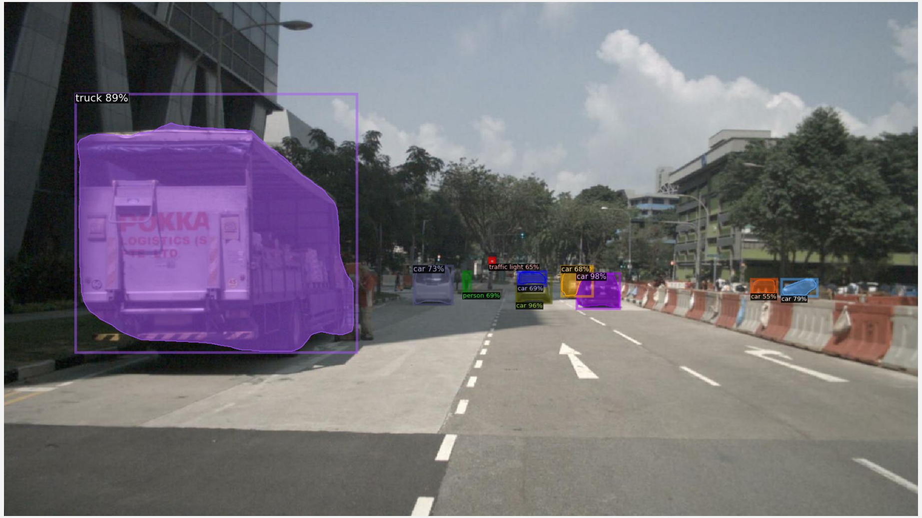
- Used DepthAnything to estimate depth from a single image
- Process: Ran DepthAnything to get a depth map, providing relative depth information for every pixel.
- Ran Detectron2 on the same image to get 2D bounding boxes around detected objects.



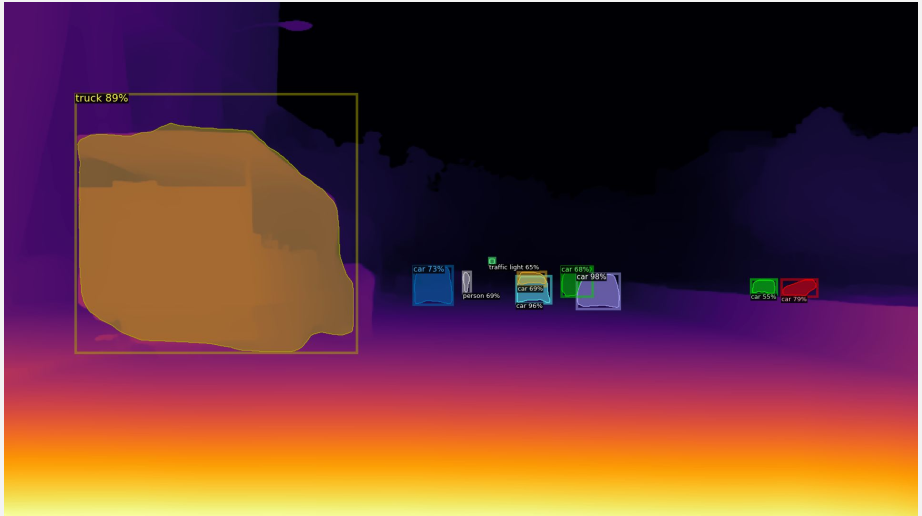
Depth map from DepthAnything

3. Combining Detections with Depth

- Instance Segmentation for Masks: Use Detectron2 to generate precise object segmentation masks.
- Depth Calculation within Masks: For each object, calculate median depth only within its mask, avoiding background interference and providing a more accurate distance estimate.



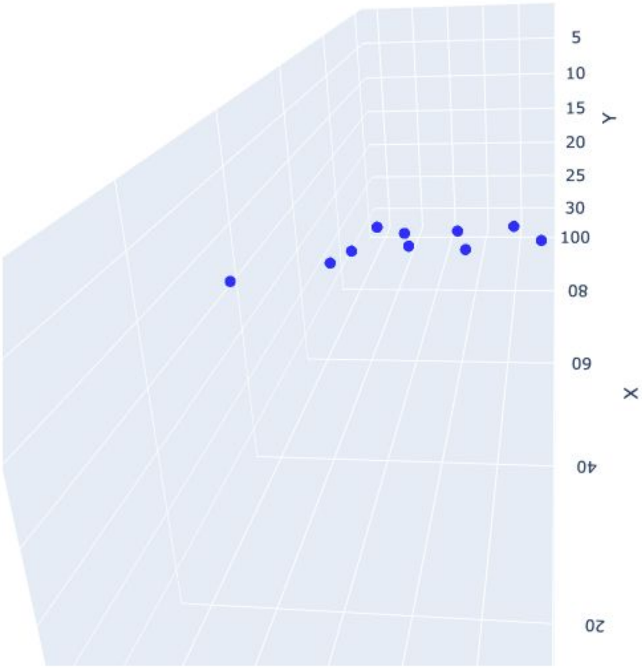
Object detections from Detectron2

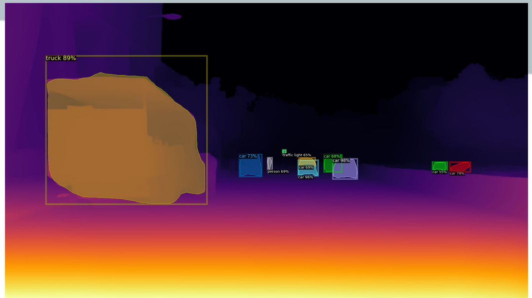


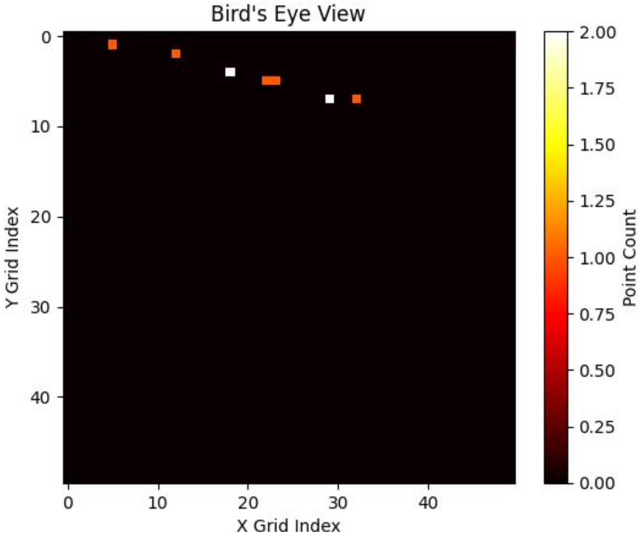
Bounding boxes and segmentation masks overlaid onto depth map

4. From 2D to 3D

- Back-Projection of Centers: Use depth data and nuScenes camera calibration to map 2D detections into 3D space.
- Average Object Sizes for Cuboids: Apply typical dimensions for each object class (e.g., cars, trucks).
- Orientation Estimation:
  - Basic: Assume objects face forward, aligned with camera view.
  - Enhanced: Use road context, lane markings, and vanishing points.

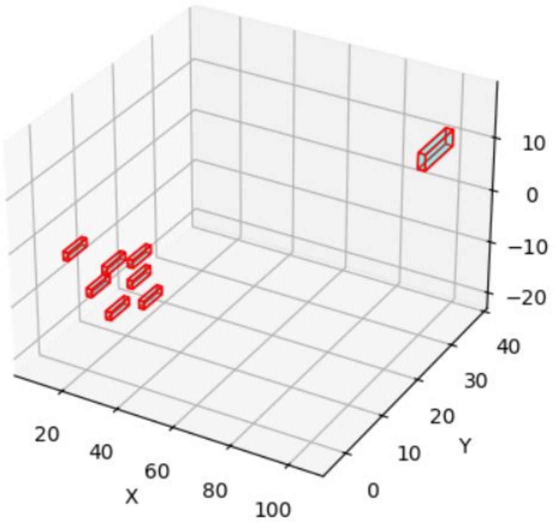




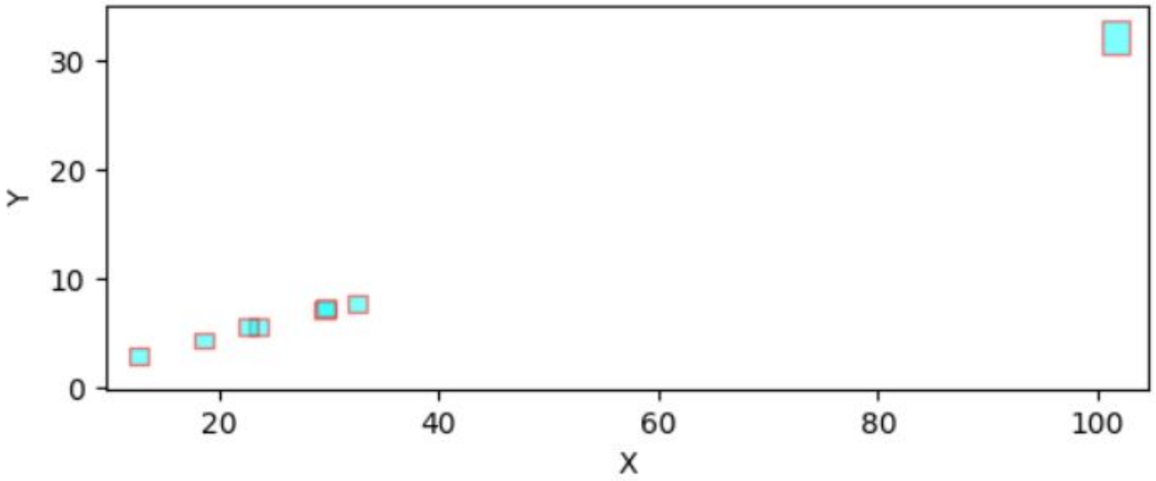


Center points in 3D Space

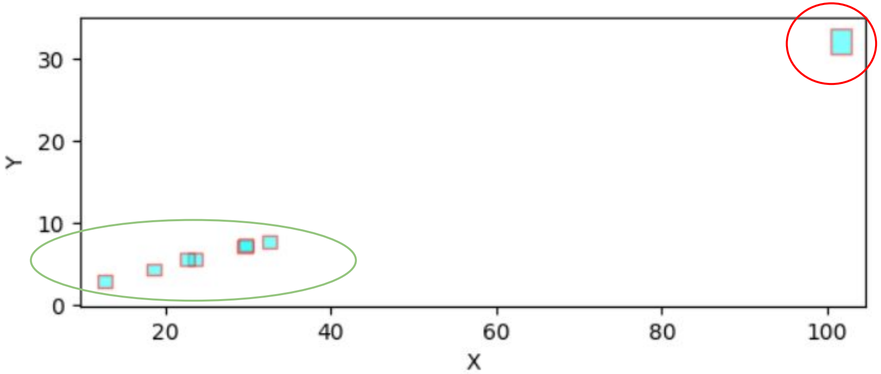
Center points in BEV



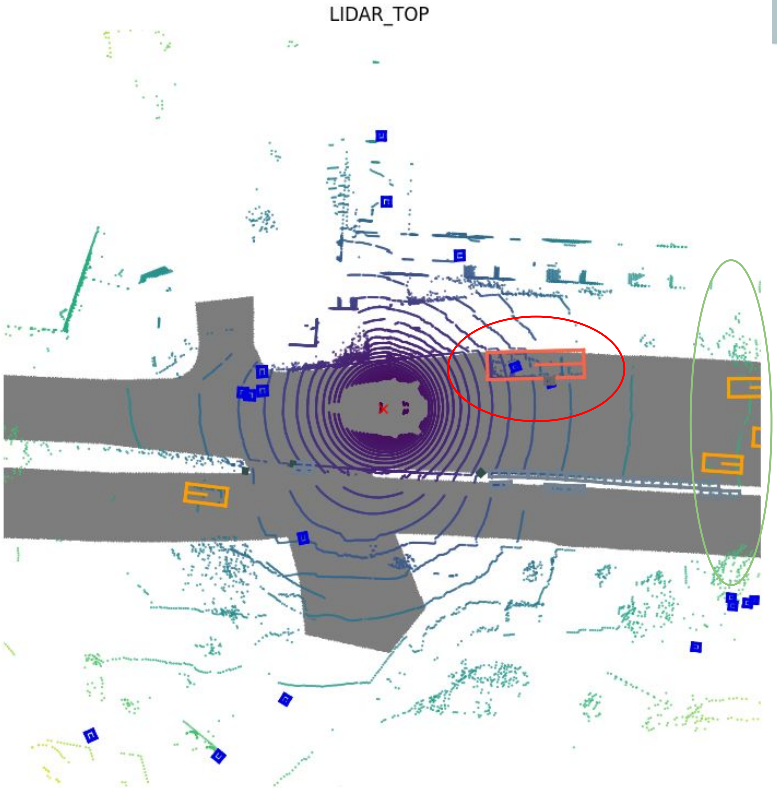
Cuboids in 3D Space



Visualization in Bird’s Eye View

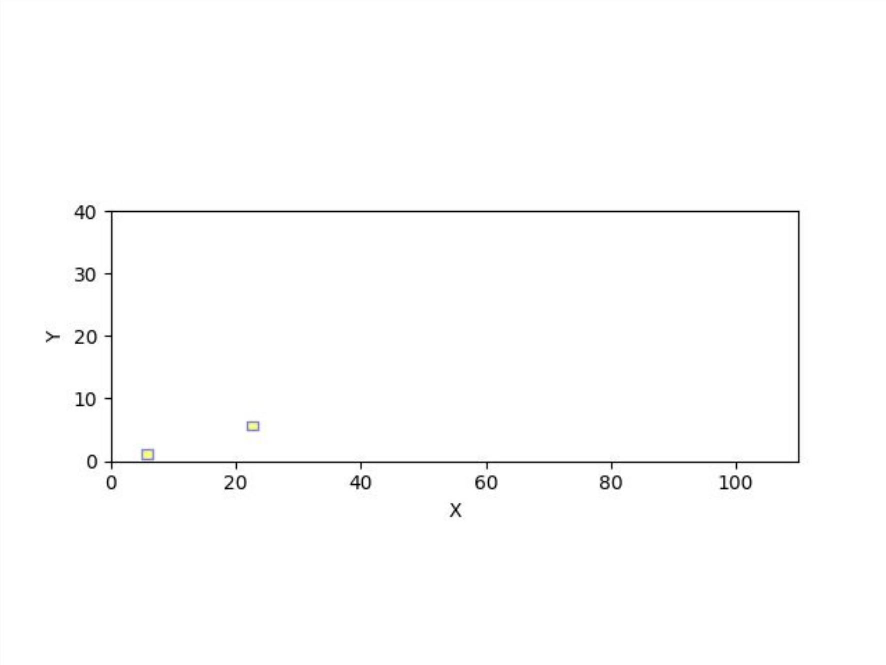


Final Detections in Bird's Eye View

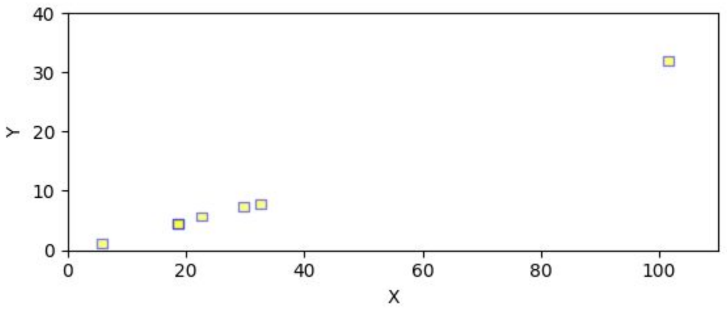


Ground Truth Lidar Bird's Eye View

Model Evaluation (Ground truth vs Cuboids) Scene 3 [GIF Animation]



Model Evaluation (Ground truth vs Cuboids) Scene 6 [GIF Animation]

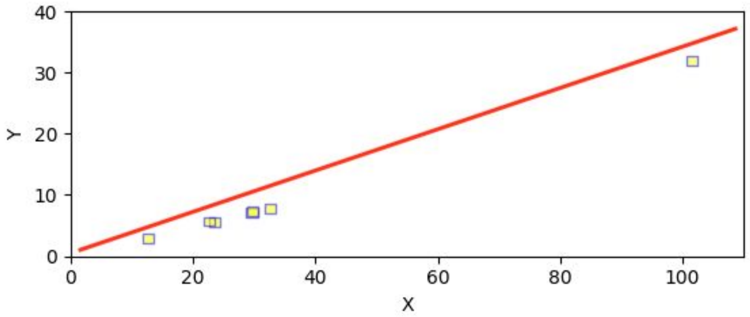


General Evaluation

Ultimately, our model did not perform as expected.

Why?

- Detections being inaccurate
- Issues with projection
- Depth estimation being off

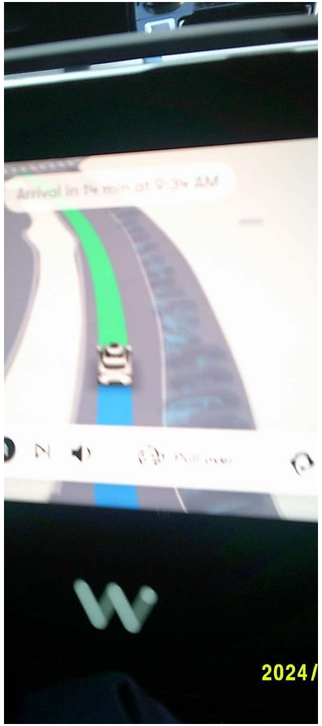


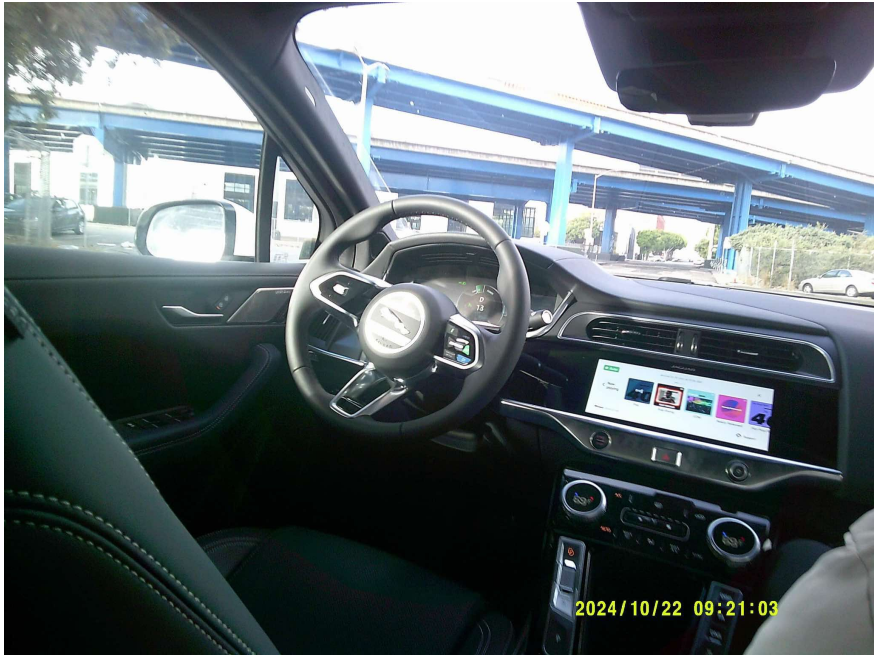
These issues would have taken some time to investigate. While we don't have a perfect model, we still learned a lot.

Final Thoughts

Experiencing the project firsthand

Salma’s Waymo Autonomous Robotaxi experience in San Francisco:





What We Learned

- We encountered real-world challenges, such as hardware limitations and computational demands, which highlighted the importance of resource management and efficiency in applied AI projects
- We deepened our understanding of machine learning algorithms.
- We learned the importance of clear communication and documentation
- We experienced the need to adapt our goals and methods as we learned more about the technical and practical aspects of the project
- Feedback from mentors and peers was invaluable

Potential Next Steps

- Expanding 3D detection on more categories (pedestrians, buses, traffic lights, etc...)
- If enough computational resources are available, revisit our initial approach i.e. a FastBEV-based approach
- Expanding the method on multiple cameras instead of a single camera
- Optimize depth estimation models for faster and more accurate depth predictions

Questions?