Vision-Based End-to-End Trajectory Prediction for Autonomous Vehicles using the Waymo Open Dataset

Anushka Tiwari, Computer Systems Engineering
Mentor: Prof. Bharatesh Chakravarthi
School of Computing and Augmented Intelligence

Introduction

Autonomous vehicles must predict where surrounding agents will move next to navigate safely. Current deep learning models operate as black boxes, it is unknown which visual features (lane markings, vehicles, traffic signals) most influence these predictions. This research investigates whether task-specific attention mechanisms can improve trajectory prediction by focusing on semantically meaningful road regions, using real-world driving data from Waymo's autonomous fleet.

Method

Dataset: Waymo Open Dataset Motion v1.3.1 -20 tfrecord files, 292,834 agent trajectory sequences extracted from real urban driving scenarios

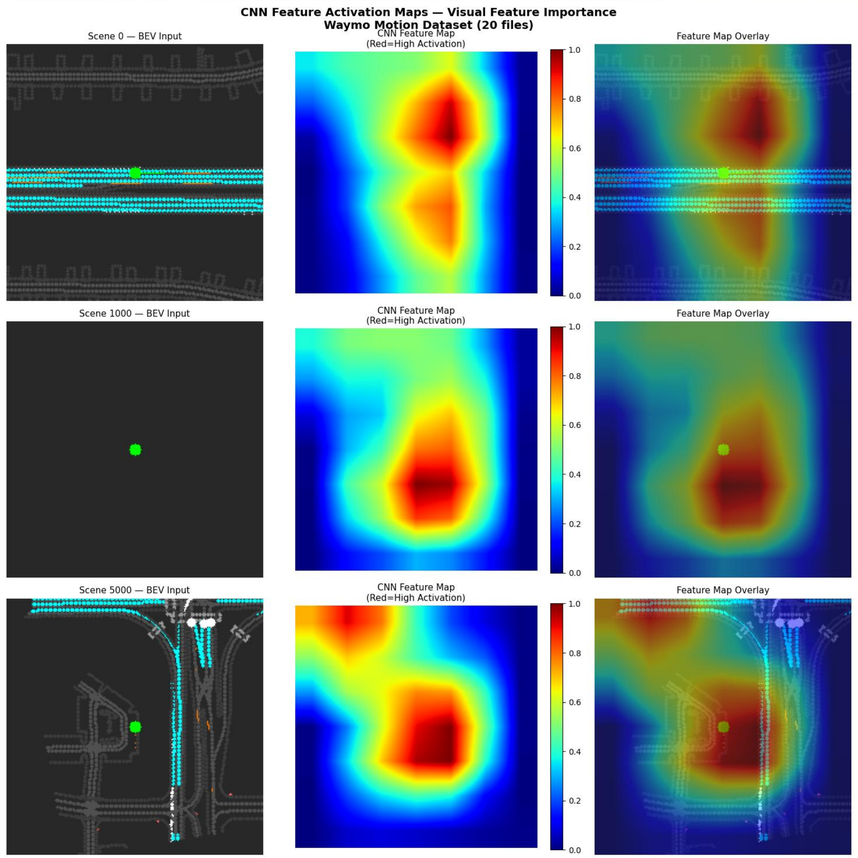
Data Pipeline:

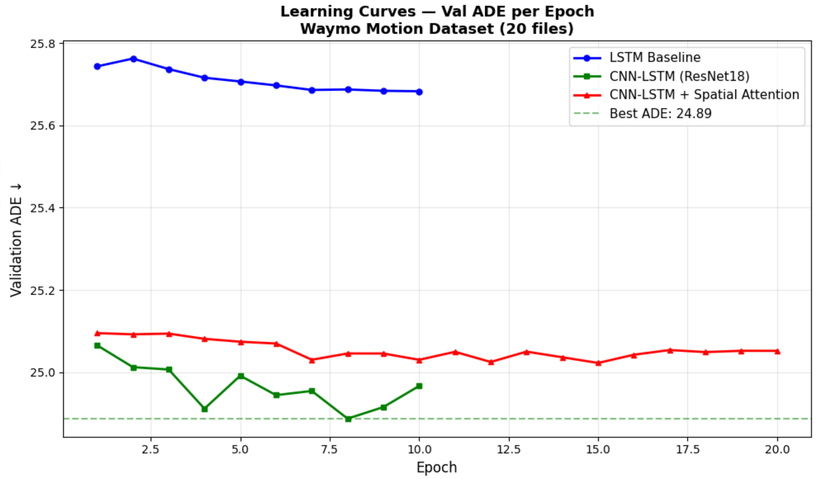
Parsed raw tfrecord files using TensorFlow

Extracted agent states: 10 past timesteps (x, y, vx, vy) → 80 future waypoints

Rendered 50,000 Bird's Eye View (BEV) scene images (224×224) from map features including lane markings, road geometry, and traffic light states

Training: Google Colab Pro (A100 GPU), PyTorch



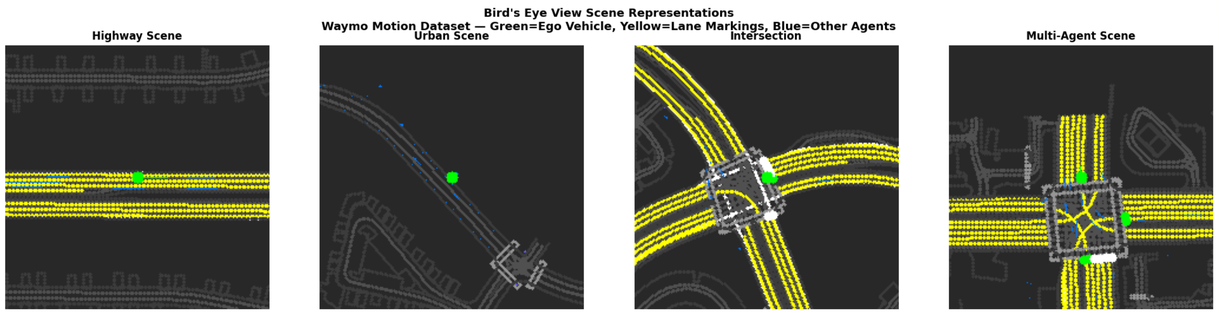


Conclusion

Visual scene context improves trajectory prediction over motion-only baselines. Freezing the ResNet18 backbone preserves ImageNet features and yields the best overall ADE (24.89). The attention mechanism shows promise but requires larger training data to learn semantically meaningful spatial priors. CNN feature maps demonstrate the model focuses on lane markings and road geometry - consistent with driving-domain expectations.

Future Work

- Scale training to full Waymo dataset (1,000 files, ~7M sequences)
- Implement explicit lane segmentation masks as attention supervision signals.
- Evaluate under occlusion, adverse weather, and long-tail scenarios
- Compare against Trajectron++ and ViP3D benchmarks
- Extend to Waymo Perception dataset for raw camera-based prediction



Tools & Technology

Python, PyTorch, TensorFlow, Google Colab Pro (A100 GPU), Google Cloud Storage, Waymo Open Dataset Motion v1.3.1, OpenCV, Matplotlib

Reference:

Sun, P. et al. (2020). Scalability in Perception for Autonomous Driving: Waymo Open Dataset. *CVPR*. - Salzmann, T. et al. (2021). Trajectron++. *ECCV*. Hu, Y. et al. (2023). ViP3D. *CVPR*.

FURI

Ira A. Fulton Schools of Engineering
Arizona State University