Single-Agent Motion Prediction on the Waymo Open Motion Dataset

Gaurav Mitra
Georgia Institute of Technology
Atlanta, United States
mitragaurav10@gmail.com

I. INTRODUCTION

Autonomous vehicles must accurately predict the future trajectories of nearby agents for safe navigation. Trajectory forecasting models aim to anticipate where other vehicles or pedestrians will move in the next few seconds. This project developed a neural network model to predict a single agent's motion 8 seconds into the future given only 1 second of past trajectory data. I used the Waymo Open Motion Dataset (WOMD), a large-scale public dataset for motion forecasting in self-driving scenarios to train and evaluate a baseline model.

II. DATASET AND TASK

In WOMD, each sequence provides 1 second of history (10 frames) and 8 seconds of future (80 frames) at 10 Hz sampling. I utilized a subset of the dataset (20 TFRecord files for training/validation and 2 for testing). The prediction task was: given an agent's past 10 (x, y) positions, forecast the next 80 positions. I focused on one target agent at a time without any map or other agent context, essentially assuming motion continues in a similar manner unless the model learns to infer changes. To simplify learning, each trajectory was normalized to a local frame by translating the last observed position to the origin and rotating the coordinates so the agent's last heading points along the x-axis.

III. MODEL ARCHITECTURE

I implemented a custom neural network called **ConvMLP** for trajectory prediction. The ConvMLP architecture consists of two components:

1. **Convolutional Encoder:** Two 1D convolution layers (64 filters, kernel size 3, causal padding; ReLU activation) process the 10-step input sequence, capturing local motion trends. Causal padding ensures each timestep's features incorporate only past data (no leakage from future steps).
2. **MLP Decoder:** The encoder output is flattened and fed into a Dense layer (128 units, ReLU), then an output Dense layer produces 160 outputs representing 80 future (x, y) points. These are reshaped into an (80, 2) trajectory output.

This simple feed-forward architecture is efficient but limited. It essentially compresses the past trajectory into a single latent representation and assumes the future can be derived from that alone. It does not explicitly model complex temporal dependencies, and it lacks any scene context or interaction modeling.

IV. RESULTS AND DISCUSSION

I trained the ConvMLP model for 10 epochs using an Adam optimizer and mean squared error loss (MSE). Predictions Ire evaluated with standard metrics: Average Displacement Error (ADE) and Final Displacement Error (FDE). The model's accuracy was poor: it largely just extrapolated each agent's future motion as a continuation of its past motion. This yielded reasonable predictions when an agent continued straight at constant speed, but if the agent's behavior changed (e.g. turning or braking) the model did not anticipate it. The predicted trajectory would jump around the continuation of the past path, showing the model's attempt to generalize to changes in motion. Consequently, the error metrics (ADE/FDE) were high, especially in scenarios with significant trajectory changes.

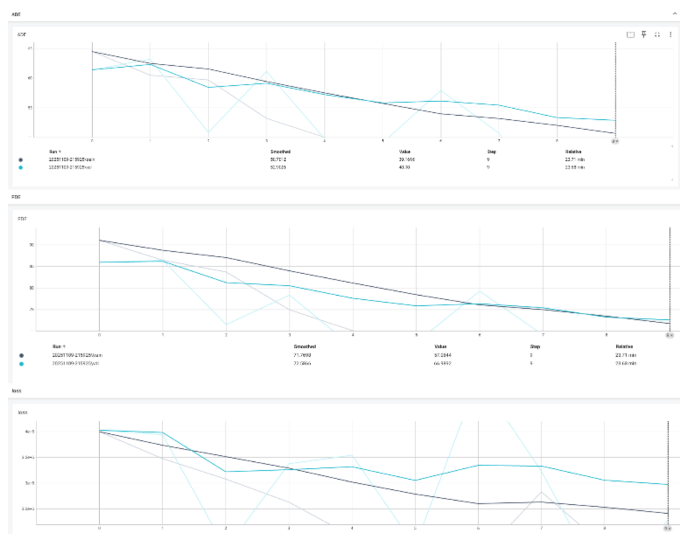

Fig. 1a-c. 10 Epoch Training metrics for ADE, FDE, and MSE, respectively.

V. CONCLUSION

This project demonstrated a baseline ConvMLP model for single-agent motion prediction on the Waymo Open Motion Dataset. The task of forecasting 8 seconds of future positions from only 1 second of past data is very challenging, and our simple ConvMLP architecture underperformed. It mainly produced straight-line extrapolations and struggled with sudden trajectory changes, underscoring the limitations of such a simplistic model for real-world motion forecasting. In practice, more advanced sequence models (e.g. recurrent or transformer networks) and incorporation of map or interaction context are needed to reliably predict complex agent behaviors. This work was intended to serve as a naive baseline and personal introduction into the data in the WOMD. My current focus is on integrating much more of the dataset and utilizing Google Compute Engine to run future larger models next to the dataset. This would remove the need for downloading the data.

REFERENCES

[1] Ettinger, S., Cheng, S., Caine, B., Liu, C., Zhao, H., Pradhan, S., Chai, Y., Sapp, B., Qi, C. R., Zhou, Y., Yang, Z., Chouard, A., Sun, P., Ngiam, J., Vasudevan, V., McCauley, A., Shlens, J., & Anguelov, D. (2021). *Large Scale Interactive Motion Forecasting for Autonomous Driving: The Waymo Open Motion Dataset.* In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 9710–9719).

[2] Chen, K., Ge, R., Qiu, H., Ai-Rfou, R., Qi, C. R., Zhou, X., Yang, Z., Ettinger, S., Sun, P., Leng, Z., Mustafa, M., Bogun, I., Wang, W., Tan, M., & Anguelov, D. (2024). *WOMD-LiDAR: Raw Sensor Dataset Benchmark for Motion Forecasting.* In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA).