End-to-End Planning for Autonomous Driving

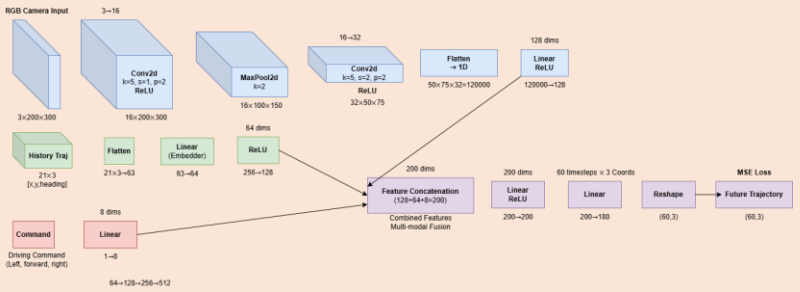
Sara Heidaripour, Aurelien Genin

Introduction

In this work, we developed three end-to-end deep learning-based planners that predict the future motion of the ego vehicle using different input modalities (e.g., past motion, RGB camera views, depth data, etc.). These planners are trained using simulation or real-world data.

Dataset: a curated subset of the nuPlan dataset with simulated sensors

Basic End-to-End Planner



We also tried using transformer encoder and decoder layers to account for the temporal dimension. However, the results turned out to be worse compared to the simpler method above.

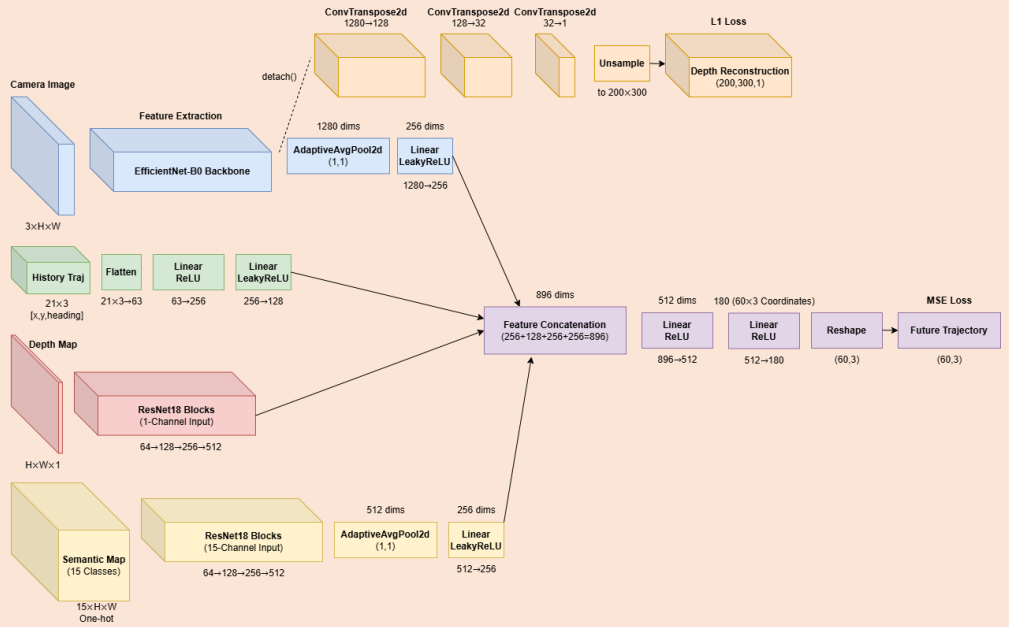
Inspiration for Perception-Aware Planning

Similarities to **CramNet**

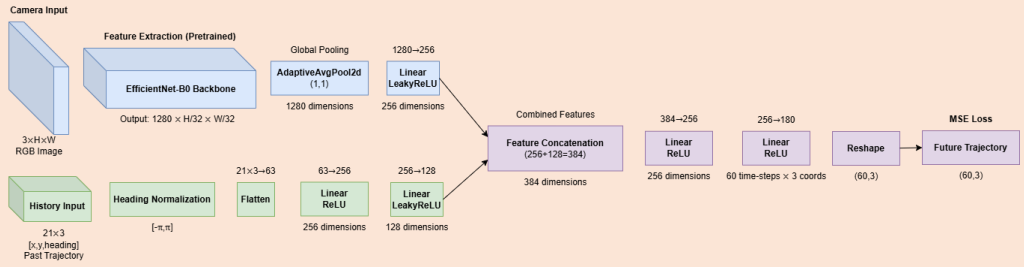
- Multi-Modal Feature Encoding
- Late Fusion
- Auxiliary Learning
- Intermediate Feature Processing

CramNet Architecture (by Waymo) Published in ECCV 2022

Perception-Aware Planning

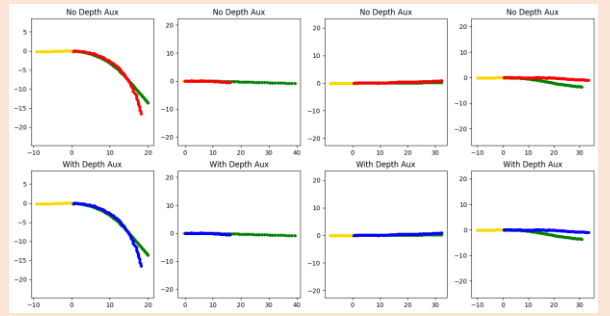


Sim-to-Real Generalization



The data being simulated was worsening the performance in the previous two phases due to the usage of pretrained backbones. In this phase, the real-world data works even better since the pretrained backbones are also trained on real-world data.

Challenge



The auxiliary task (depth map) did not appear to be effective enough in performance improvement. Possible reason might be its weight in the loss function. Further fine-tuning for landa may help improving the performance.

Results

Phase	ADE Public Test Set	ADE Private Test Set
1	2.11	2.00
2	1.60	1.65
3	1.55	-

Most Helpful Strategy: Using Pretrained Backbones (Last classification layers were removed)

Further Improvement: Fine-tuning for landa + Data Augmentation