Predicting Motion of Vulnerable Road Users using High-Definition Maps and Efficient ConvNets

Fang-Chieh Chou, Tsung-Han Lin, Henggang Cui, Vladan Radosavljevic, Thi Nguyen, Tzu-Kuo Huang, Matthew Niedoba, Jeff Schneider, Nemanja Djuric

Abstract
Following detection and tracking of traffic actors, prediction of their future motion is the next critical component of a self-driving vehicle (SDV), allowing the SDV to move safely and efficiently in its environment. This is particularly important when it comes to vulnerable road users (VRUs), such as pedestrians and bicyclists. We present a deep learning method for predicting VRU movement where we rasterize high-definition maps and actor’s surroundings into bird’s-eye view image used as input to convolutional networks. In addition, we propose a fast architecture suitable for real-time inference, and present an ablation study of rasterization choices.

Methods & Results
This work consists of two parts. First we propose a fast CNN architecture suitable for running in real-time onboard the SDV. Secondly we present an ablation study of various rasterization settings.

Fast CNN architecture:
We performed simple modifications on the MobileNet-V2 architecture to speed up the inference. Operations with high memory access costs are removed or reordered. The resulted architecture (FastMobileNet) is significantly faster on batched GPU inference (batch size = 32) than the original MobileNet-V2, with comparable accuracy.

Methods & Results
This work consists of two parts. First we propose a fast CNN architecture suitable for running in real-time onboard the SDV. Secondly we present an ablation study of various rasterization settings.

Fast CNN architecture:
We performed simple modifications on the MobileNet-V2 architecture to speed up the inference. Operations with high memory access costs are removed or reordered. The resulted architecture (FastMobileNet) is significantly faster on batched GPU inference (batch size = 32) than the original MobileNet-V2, with comparable accuracy.

Ablation Study:
We performed ablation studies on several rasterization choices of our input image representation.
- Raster frame rotation: whether we rotate the frame so that the object’s heading point ro north.
- Raster pixel resolution: spatial size of the pixel; larger resolution covers larger context but loses fine details.
- Lane direction: lane direction is encoded as a rainbow color for each lane segment.
- Traffic light: traffic lights as colored circles at intersections; uncrossable crosswalks are also colored green.
- Learned colors: instead of using manually picked colors, we let the network learn the color of each layer.
- Model pretraining: instead of training from scratch, we initialize VRU model with model trained on vehicle data.

Ablation Study:
We performed ablation studies on several rasterization choices of our input image representation.
- Raster frame rotation: whether we rotate the frame so that the object’s heading point ro north.
- Raster pixel resolution: spatial size of the pixel; larger resolution covers larger context but loses fine details.
- Lane direction: lane direction is encoded as a rainbow color for each lane segment.
- Traffic light: traffic lights as colored circles at intersections; uncrossable crosswalks are also colored green.
- Learned colors: instead of using manually picked colors, we let the network learn the color of each layer.
- Model pretraining: instead of training from scratch, we initialize VRU model with model trained on vehicle data.

Conclusion
We presented an efficient and effective solution to motion prediction of VRU actors. This is a critical problem in autonomous driving, as such actors have higher risk of injury and are less predictable since they may change behavior faster than vehicles. We applied recently proposed rasterization technique to generate raster images of actors’ surroundings encoding their context, used as input to deep CNN trained to predict actor trajectory. Moreover, we proposed a fast architecture suitable for real-time operations, and finally presented a detailed ablation study of various rasterization choices.