MIXTURE OF EXPERTS MODEL FOR TRAJECTORY PREDICTION

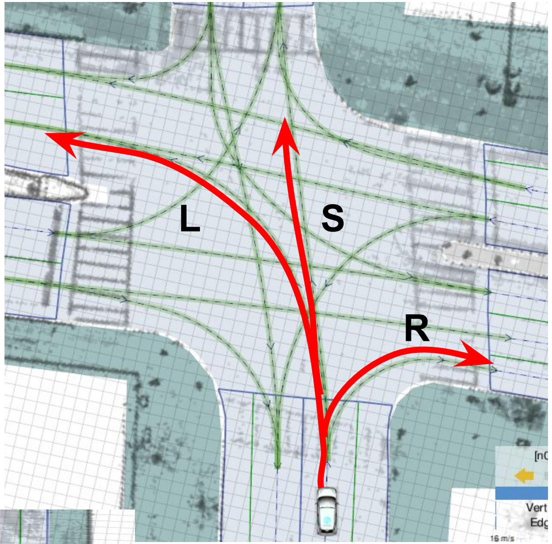
PULKIT KATDARE

15 AUGUST, 2019

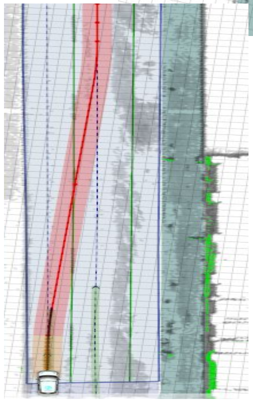
APTIV

PREDICTION TEAM

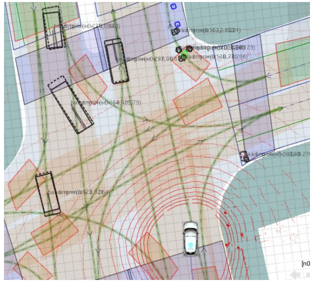
- Model possible future states of the world
- Planner uses to enable safe and natural driving by better anticipating interactions with others
- Flexible models support different behaviors in multiple cities/countries (Boston, SG, Vegas)



Lane Target Prediction



Lane Changing



Pass/yield behavior

OVERVIEW

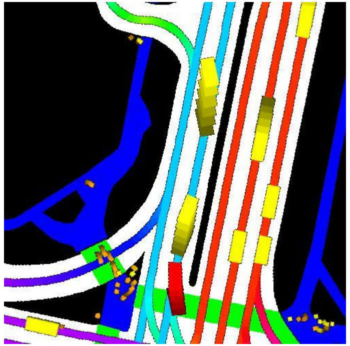
- We use different kind of models based on
  - Physics
  - Micro-planners
  - Machine learning

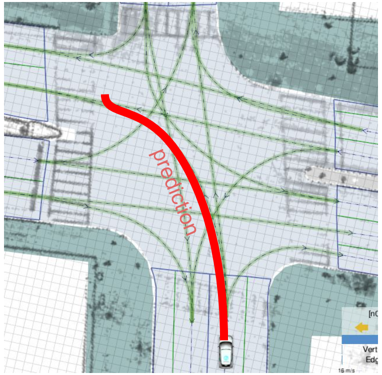
- Each of these models are generally well-suited for different kinds of motion

- Can we leverage these models to hopefully improve our results?

APTIV

TRAJECTORY PREDICTION MODEL





• APTIV •

TOWARDS MULTIPLE PREDICTIONS

- Impossible to predict exact trajectories for vehicles

- Is it possible to leverage the models in our repository along with their probability to improve prediction

- Could help planning team develop risk-aware algorithms

APTIV

RELATED WORK

1. Multimodal Trajectory Predictions for Autonomous Driving using Deep Convolutional Networks (Cui et al.)
2. Adaptive Mixture of Local Experts (Jacobs et al.)

- [1]: FCN for multi-modal prediction
- [2]: Mixture of Expert based architecture

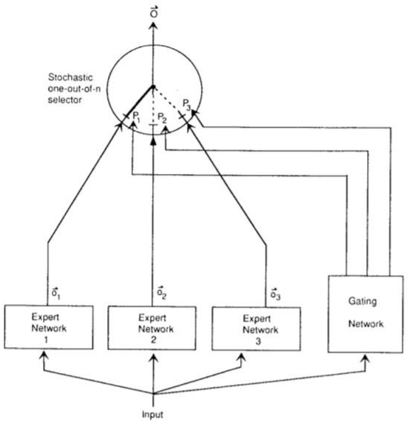
$$m^* = \underset{m \in \{1,\ldots,M\}}{\arg\min} \; dist(\boldsymbol{\tau}_{ij}, \tilde{\boldsymbol{\tau}}_{imj}). \tag{3}$$

$$\mathcal{L}_{ij}^{class} = -\sum_{m=1}^{M} I_{m=m^*} \log p_{im}, \tag{5}$$

Probability For the mth mode

Ground Truth For jth mode

Predicted Trajectory for mth mode



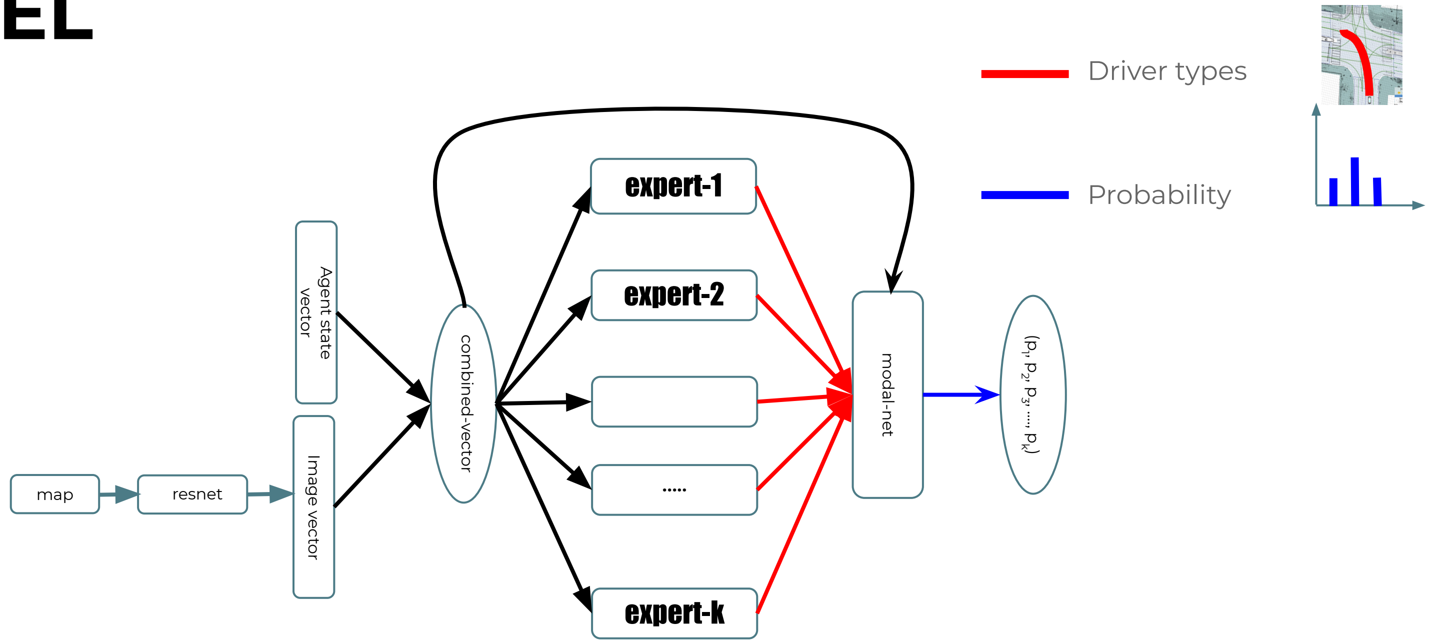
APTIV

APPROACHES

1. END-TO-END MIXTURE-OF-EXPERTS MODEL

2. FROZEN MOE MODEL

END-TO-END MIXTURE-OF-EXPERTS (MOE) MODEL



APTIV

END-TO-END MODEL

$$\mathcal{L}(Z, \tau) = \sum_{i=1}^{k} p_i L(E_k(Z), \tilde{\tau})$$

Z: input to the network

τ : ground truth

E_k(.): output by a specific expert-k

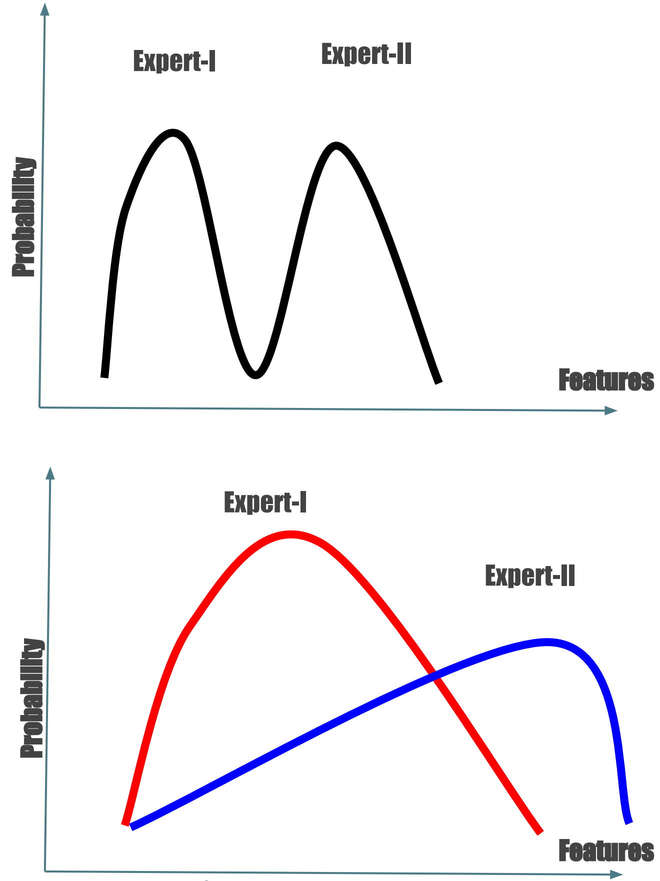
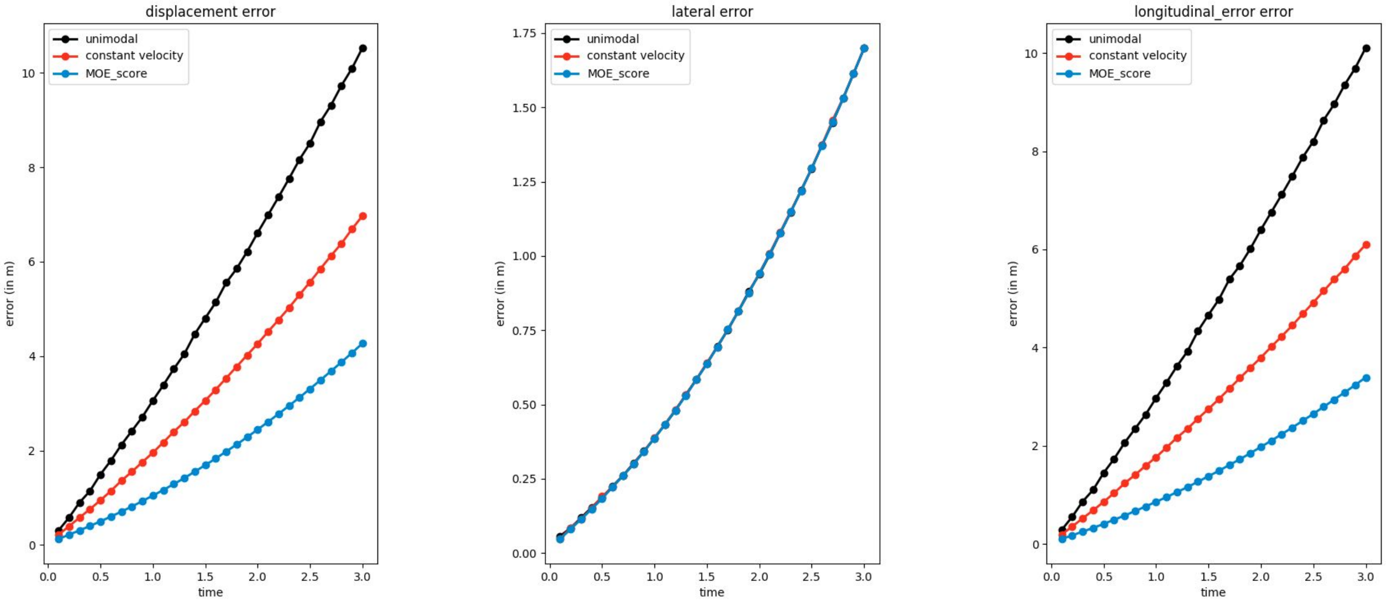


Illustration of Mode Collapse

1. Adaptive Mixture of Local Experts (Jacobs et al.)

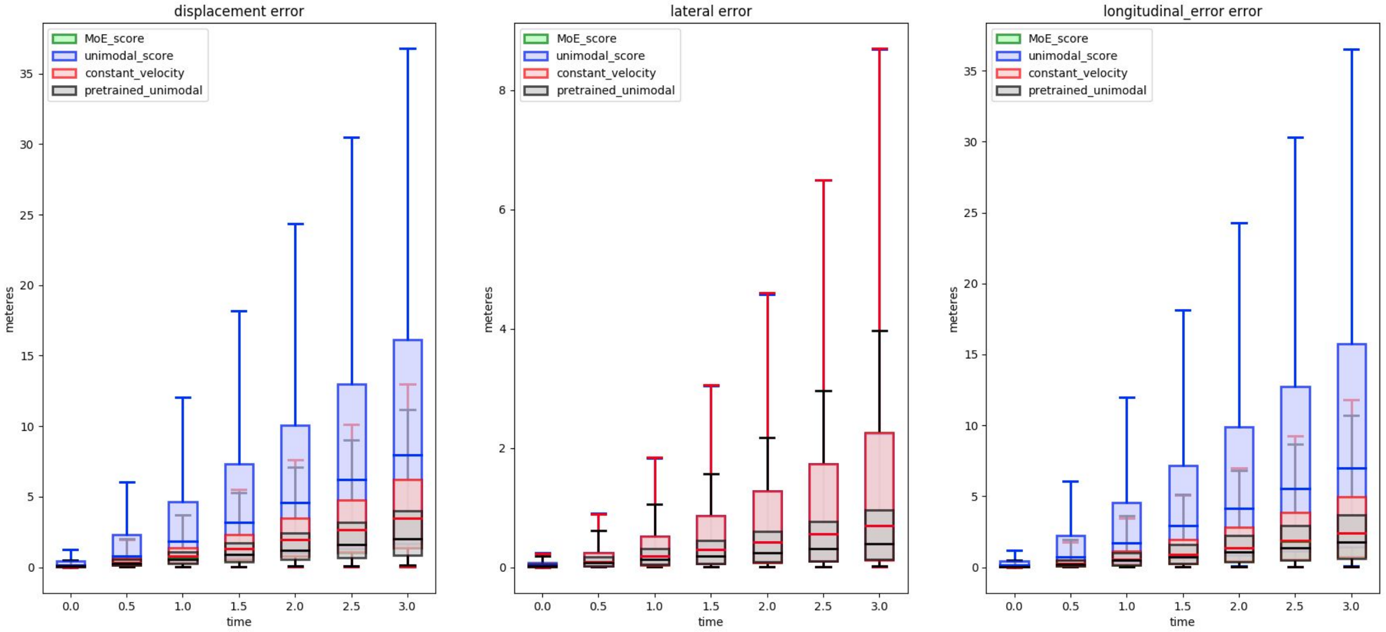
APTIV

EXPERTS: CONSTANT VELOCITY, VEHICLES (TRAIN TRACKS: 300K VEHICLES)



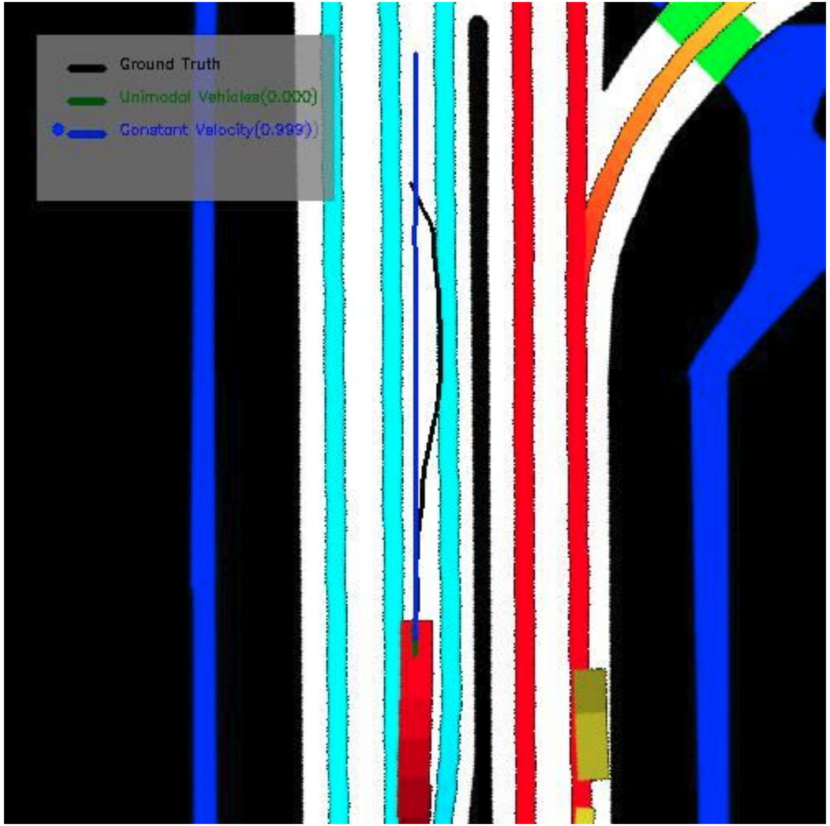
APTIV

EXPERTS: CONSTANT VELOCITY, VEHICLES (TRAIN TRACKS: 300K VEHICLES)



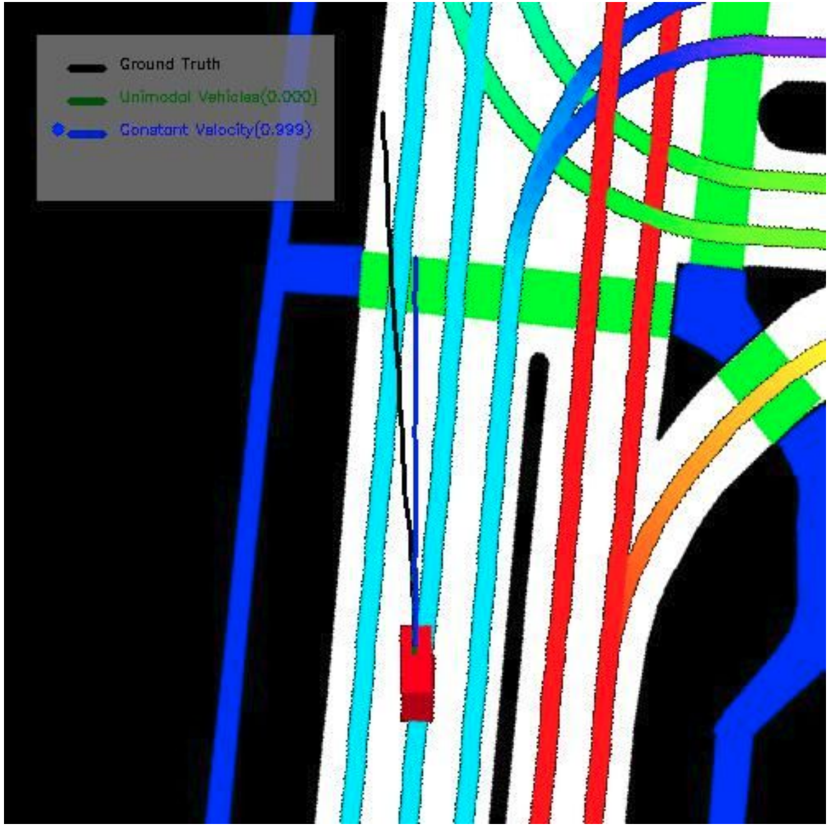
APTIV

RESULTS



APTIV

RESULTS



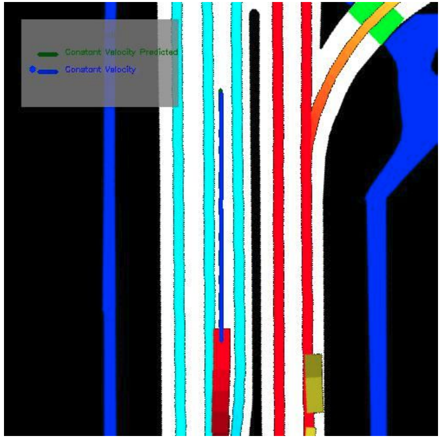
APTIV

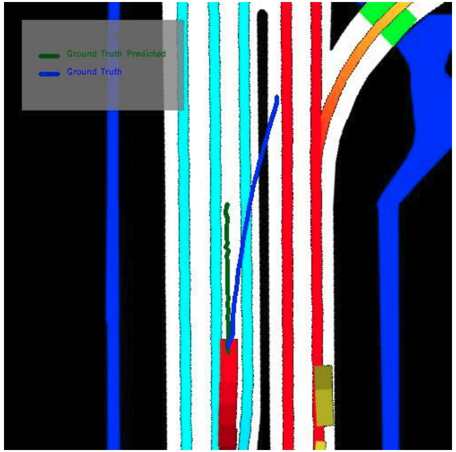
CONCLUSION

- Individual experts perform worse than the unimodal ones

- Mode collapse
  - Constant velocity is the dominant prediction
  - Model complexity of two models are different

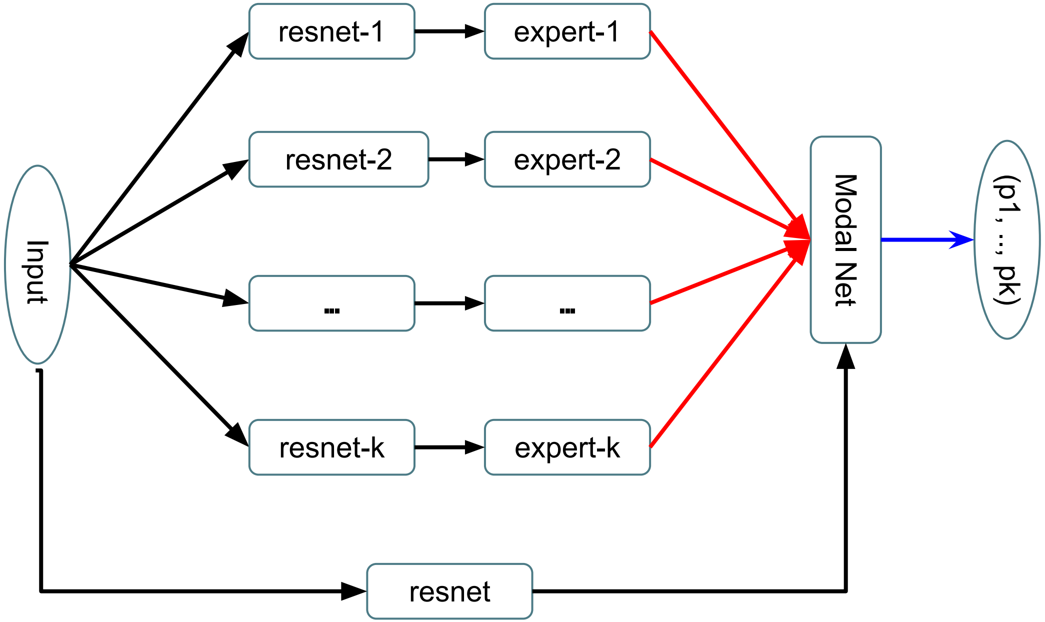
- Leverage existing models to learn a probability distributions over driver types



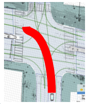


APTIV

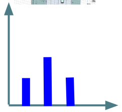
FROZEN MOE MODEL



Driver types

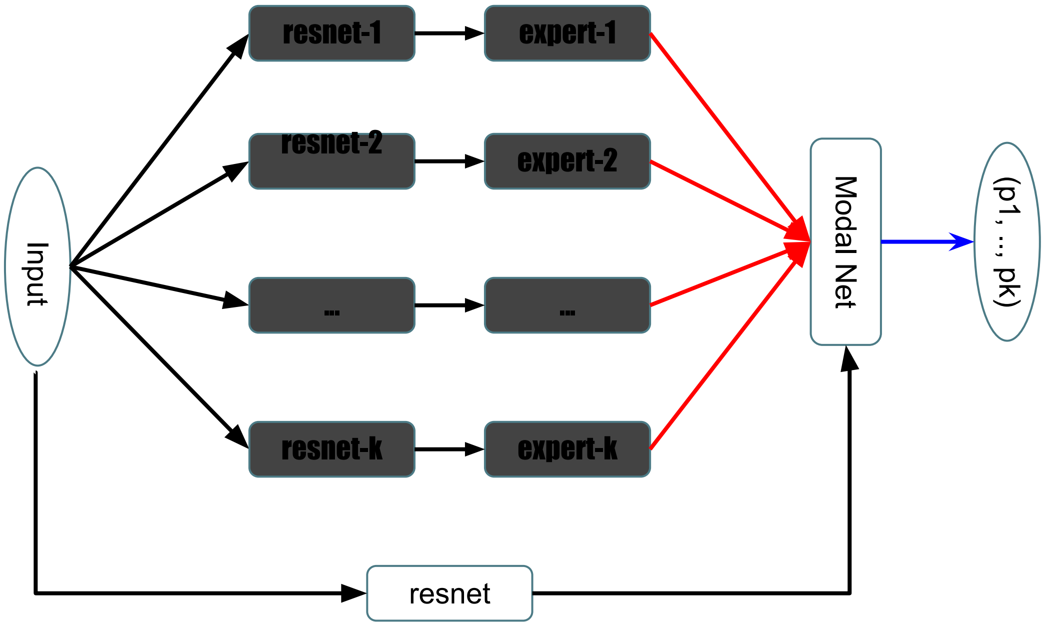


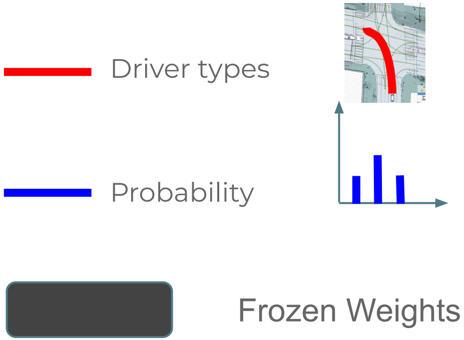
Probability



APTIV

FROZEN MOE MODEL





APTIV

TRAINING THE DISCRIMINATOR

Predicted Trajectory for ith mode

- Proposes a loss function just for discriminator training [1]

$$k^* = \arg min_{i \in \{1,2,..,k\}} dist(\tau_{ij}, \tilde{\tau}_j)$$

Ground Truth

$$\mathcal{L}_{ij}^{class} = - \sum_{i \in \{1,2,\ldots,k\}} I_{i=i^*} \log(p_{im})$$

Probability For the mth mode

- In our particular case, the dist function is a smooth L1-norm

APTIV

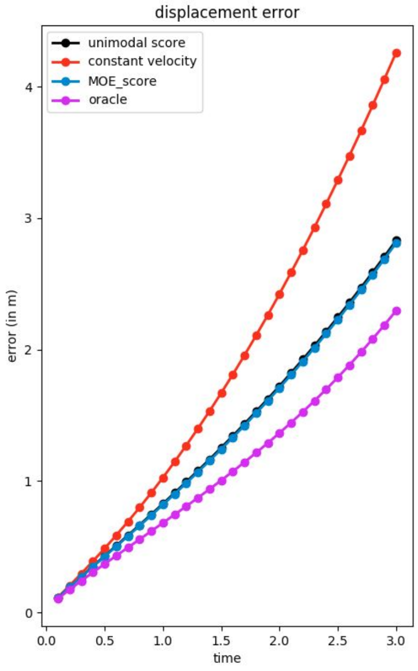
EXPERIMENTS

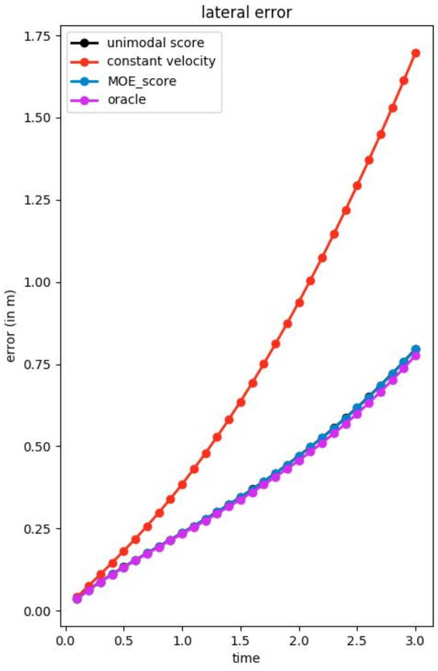
1. Experts: Constant Velocity, Unimodal
   Train Tracks: 300K, 3 Million
   Training Data: **Vehicle Tracks**

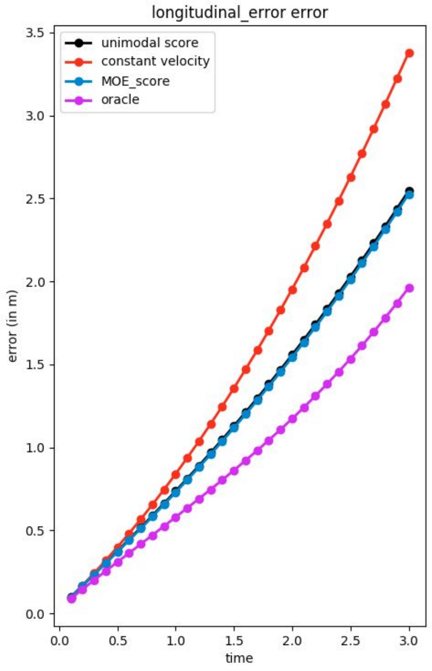
2. Experts: Vehicles, Pedestrian, Cyclists, Constant Velocity
   Train Tracks: 600k, 3 Million, 5 Million
   Training Data: **Vehicle, Pedestrian and Cyclists Tracks**

3. Experts: Pedestrians, Vehicles, Unclassified
   Train Tracks: 6 Million
   Training Data: **Pedestrians, Vehicles, Unclassified Tracks**

EXPERTS: CONSTANT VELOCITY, VEHICLES (TRAIN TRACKS: 300K)

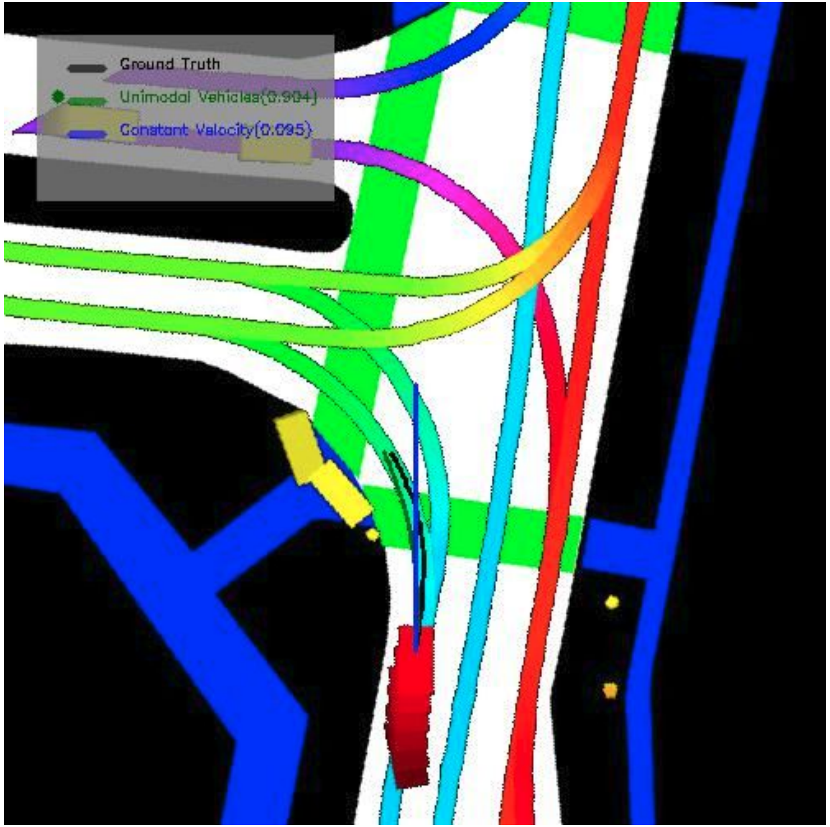






RESULTS (TRAIN TRACKS: 300K)

Unimodal > Constant Velocity
Predicted: Unimodal ✓

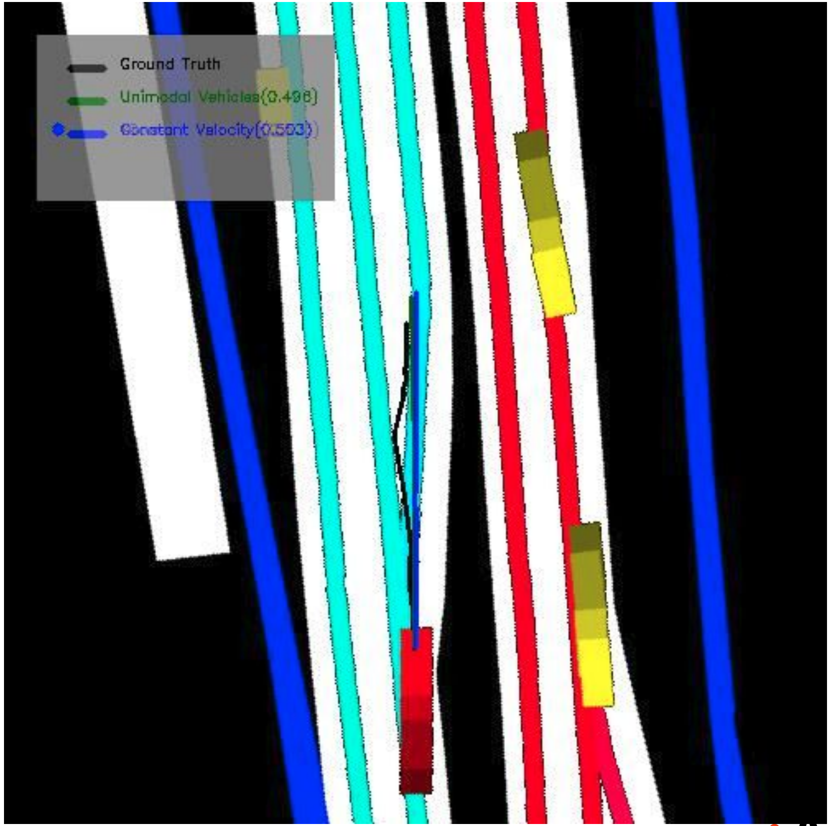


APTIV

RESULTS (TRAIN TRACKS: 300K)

Constant Velocity > Unimodal

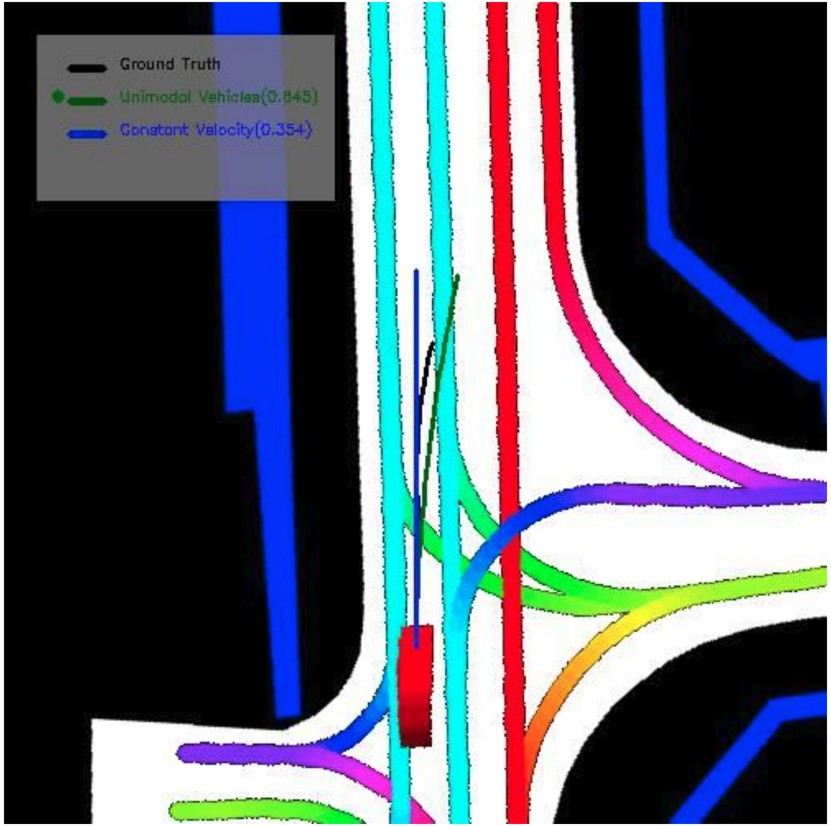
Predicted: Constant Velocity ✓



RESULTS (TRAIN TRACKS: 300K)

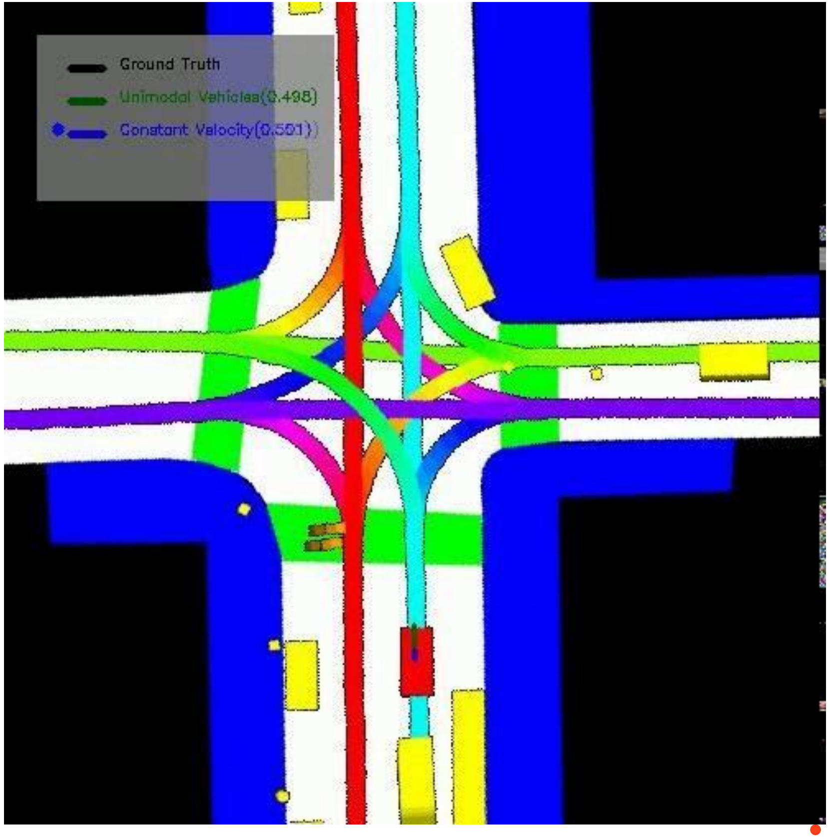
Constant Velocity > Unimodal

Predicted: Unimodal ×



APTIV

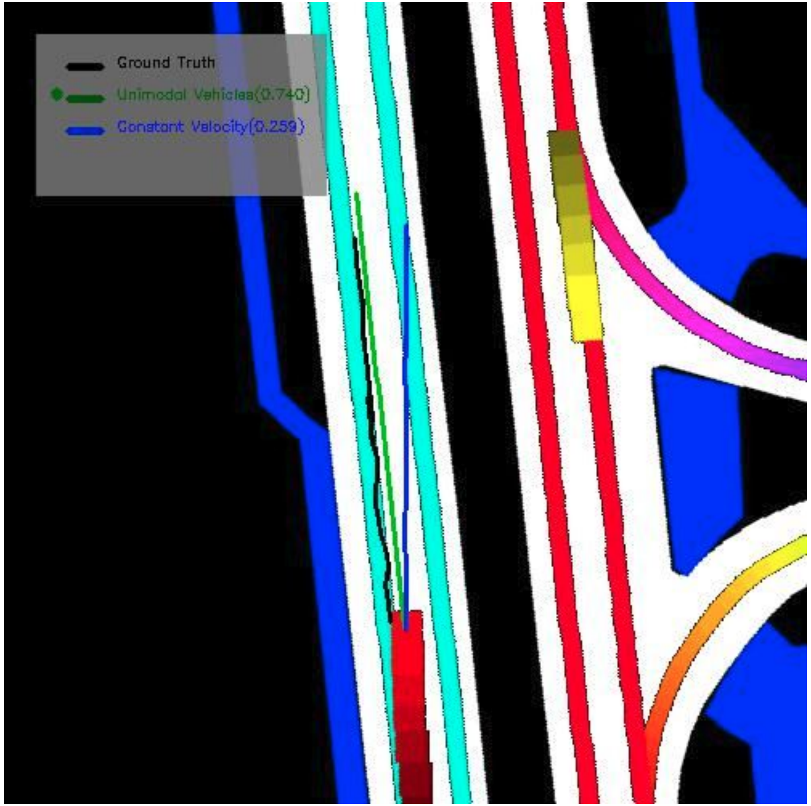
VISUALISATION (TRAIN TRACKS: 300K)



APTIV
23

RESULTS (TRAIN TRACKS: 3MIL)

Unimodal > Constant Velocity
Predicted: Unimodal ✓



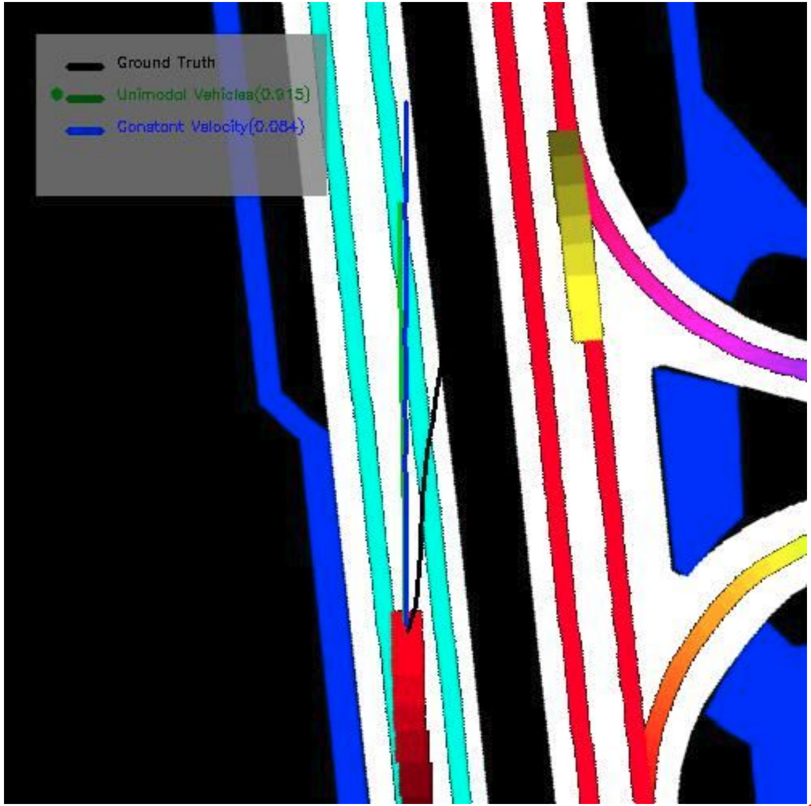
APTIV

24

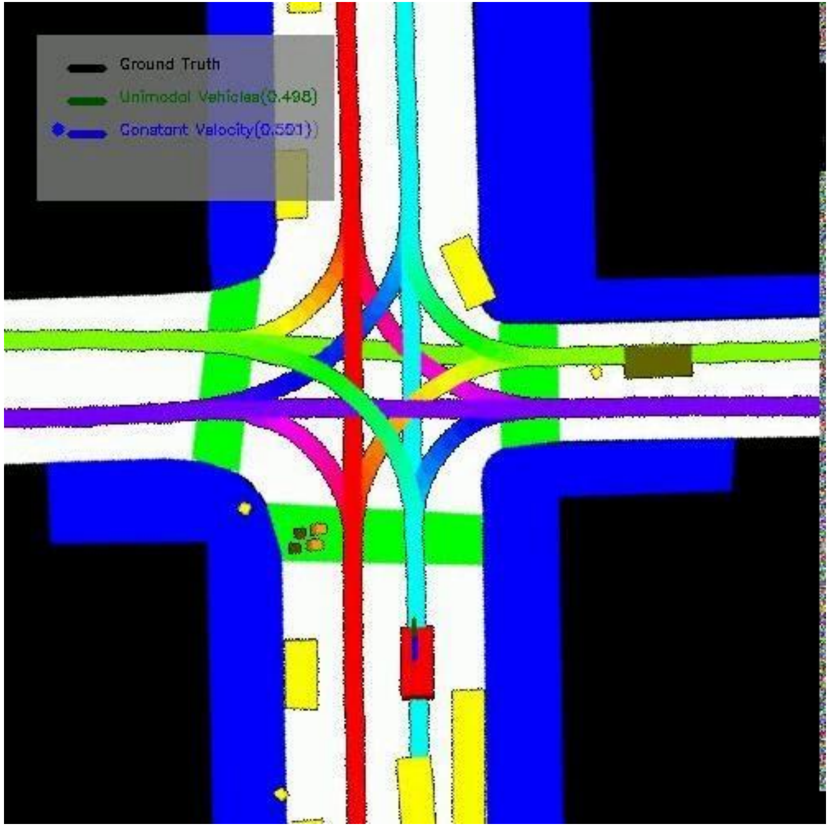
RESULTS (TRAIN TRACKS: 3MIL)

Constant Velocity > Unimodal

Predicted: Constant Velocity ✓



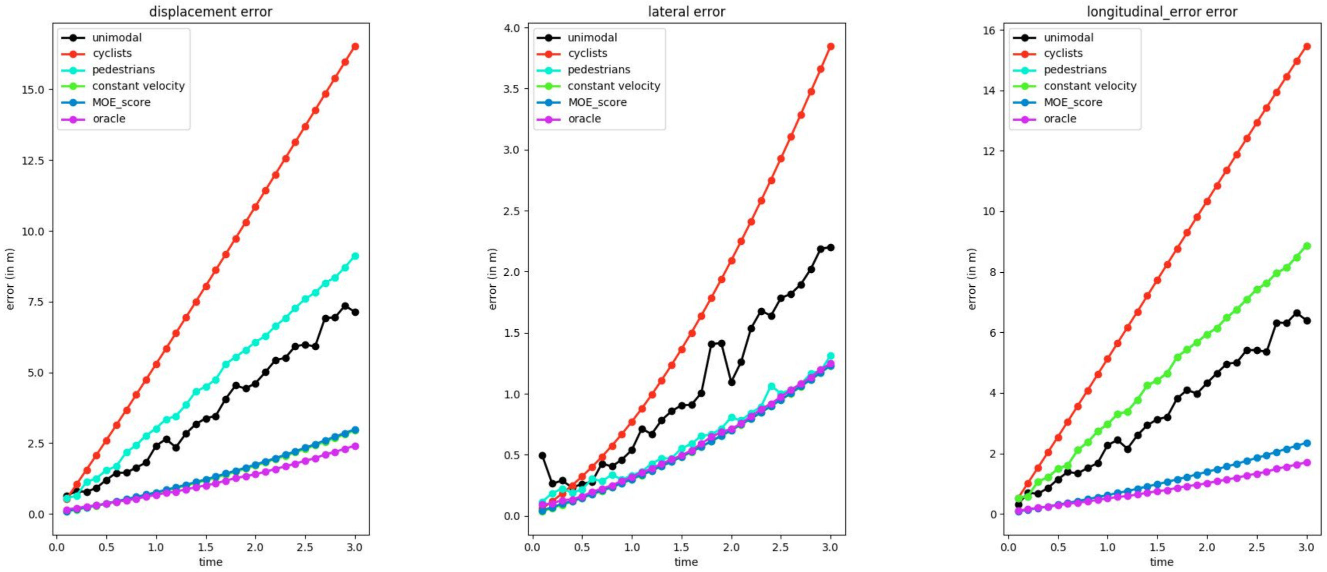
VISUALISATION (TRAIN TRACKS: 3MIL)



APTIV

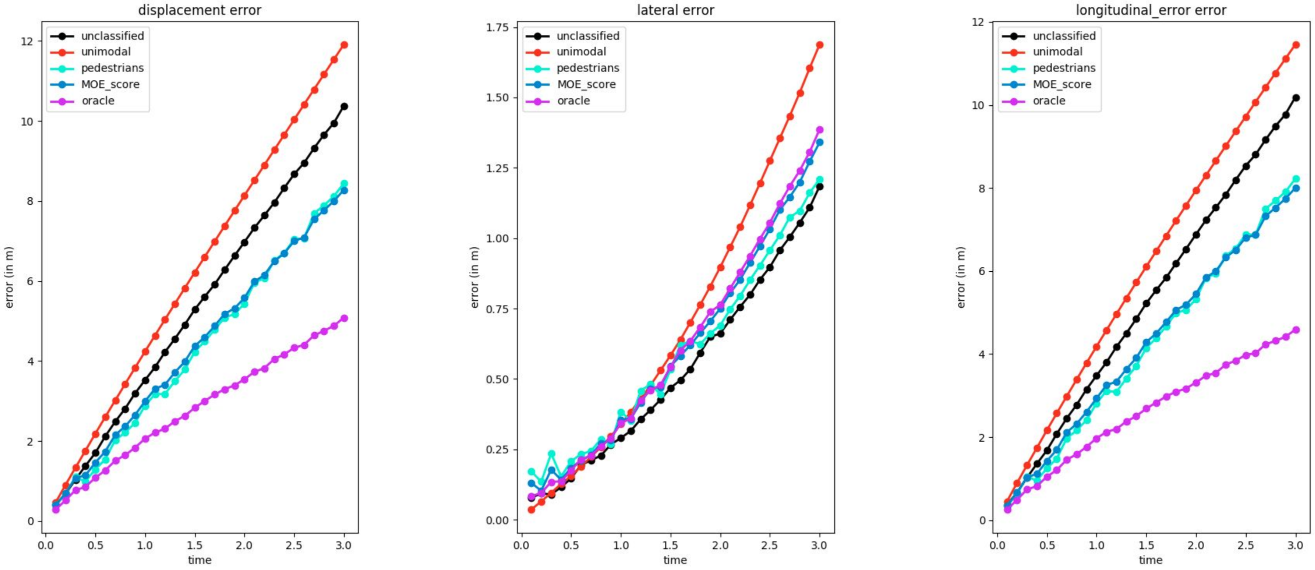
26

EXPERTS: VEHICLES, PEDESTRIAN, CYCLISTS, CONSTANT VELOCITY (TRAIN TRACKS: 5MIL)



Other experiments: 600k, 3 million

EXPERTS: VEHICLES, PEDESTRIAN, UNCLASSIFIED (TRAIN TRACKS: 6MIL)



APTIV

28

CONCLUSION

- End-to-End Model does not provide improvement over Vehicles Model
- Frozen MoE Model (Experts: Constant Velocity, Unimodal Vehicles) does not provide improvement over Unimodal Vehicles model
- Frozen MoE Model (Experts: Unimodal Vehicles, Unimodal Pedestrians, Unimodal Unclassified) provides average score values similar to the Pedestrian Model, and better scores than the Unimodal Vehicles and Unimodal Unclassified Models (computed on all tracks)
- Future work:
    - Hyper-parameter optimisation
    - Distance function based on angle between trajectories instead of smooth L1
    - Different Architecture

APTIV

29

IMPACT

- Mixture of experts model automatically chooses expert for each track
  - Eliminates the need to have different model for different types of tracks
- Created initial architecture for such an approach that can be used as a building block for further development

ACKNOWLEDGEMENT

- Corina (Advisor)

- Eric
- Freddy
- Tung Phan
- Sang Uk Lee

QUESTIONS?