Hierarchical Deep Reinforcement Learning through Scene Decomposition for Autonomous Urban Driving

Peggy (Yuchun) Wang, Maxime Bouton, Mykel J. Kochenderfer

SISL Stanford Intelligent Systems Laboratory

Problem

- Decision Making Under Uncertainty in Autonomous Driving Scenarios is hard
- Rule Based Methods are often used, but cannot scale or generalize
- Deep Reinforcement Learning (DRL) methods has been used on specific scenarios and are more robust, but would need to be trained on every single scenario. Still not able to generalize to unseen scenes and expensive to compute.

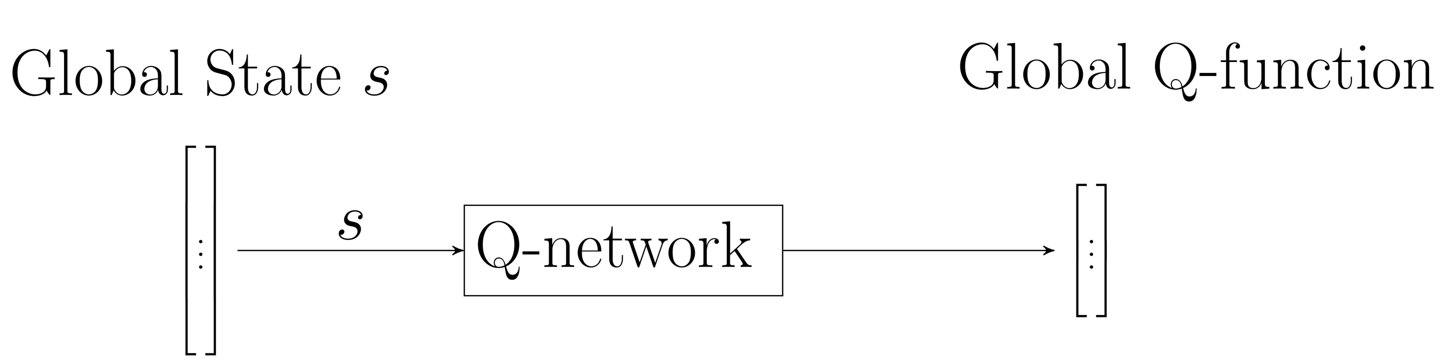
Approach

1. Use DRL to train micro-policies on micro-scenarios
2. Decompose a complex scenario into micro-scenarios
3. Fuse Q value functions of micro-scenarios to get Q function approximation of complex scenario
4. Extract policy from fused Q value function

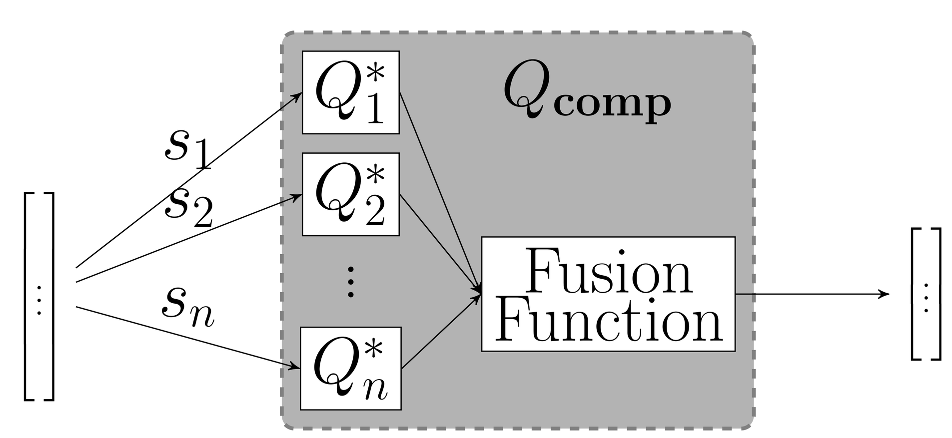
Utility Decomposition

$$\tilde{Q}^*_{comp}(s,a) = \tilde{Q}^*_1(s,a) + \tilde{Q}^*_2(s,a)$$

$$\tilde{\pi}^*_{comp}(s) = argmax_a \tilde{Q}^*_{comp}(s,a)$$



a. Regular Q-network

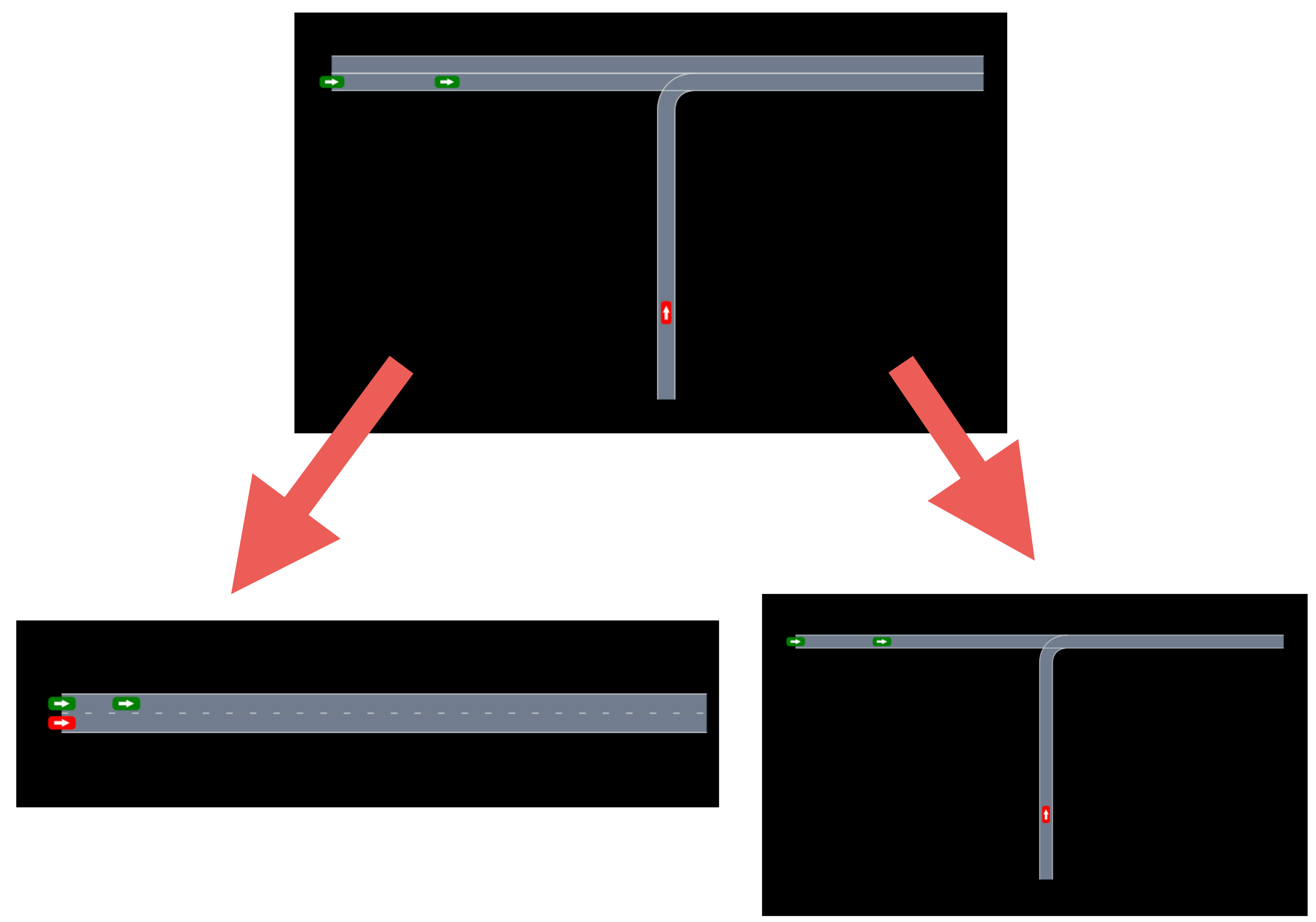


b. Q-Decomposition Network

Acknowledgments

A tremendous thank you to my mentor Maxime Bouton for all his help and input, as well as conceiving the idea of this interesting project. A tremendous thanks as well to Professor Mykel Kochenderfer for allowing me to work on this independent project in conjunction with CS191W and providing help and mentorship.

Scene Decomposition



Visualization of Policies

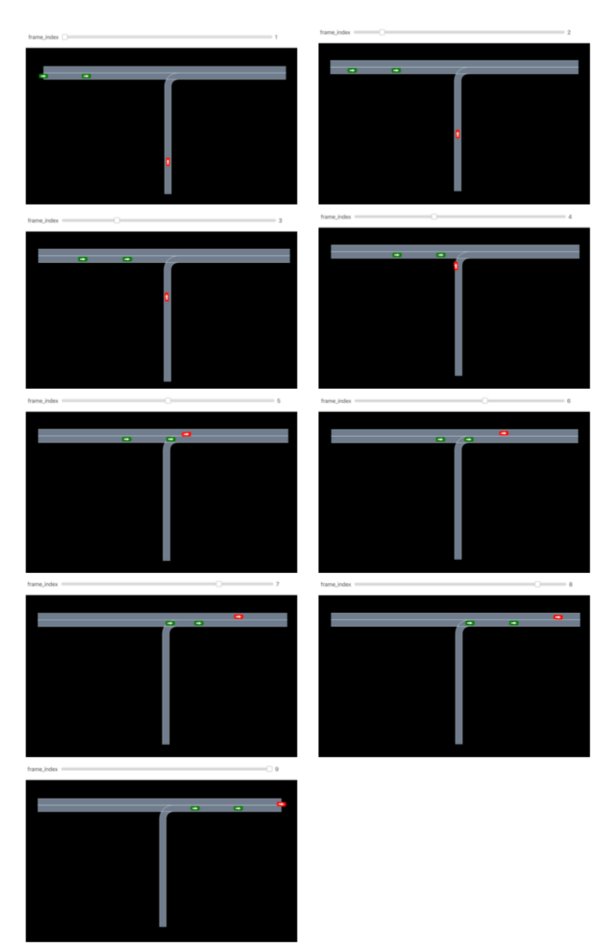


Figure 2: Visualization of the Q-decomposition policy. Starting from top left frame 1 to bottom left frame 9.

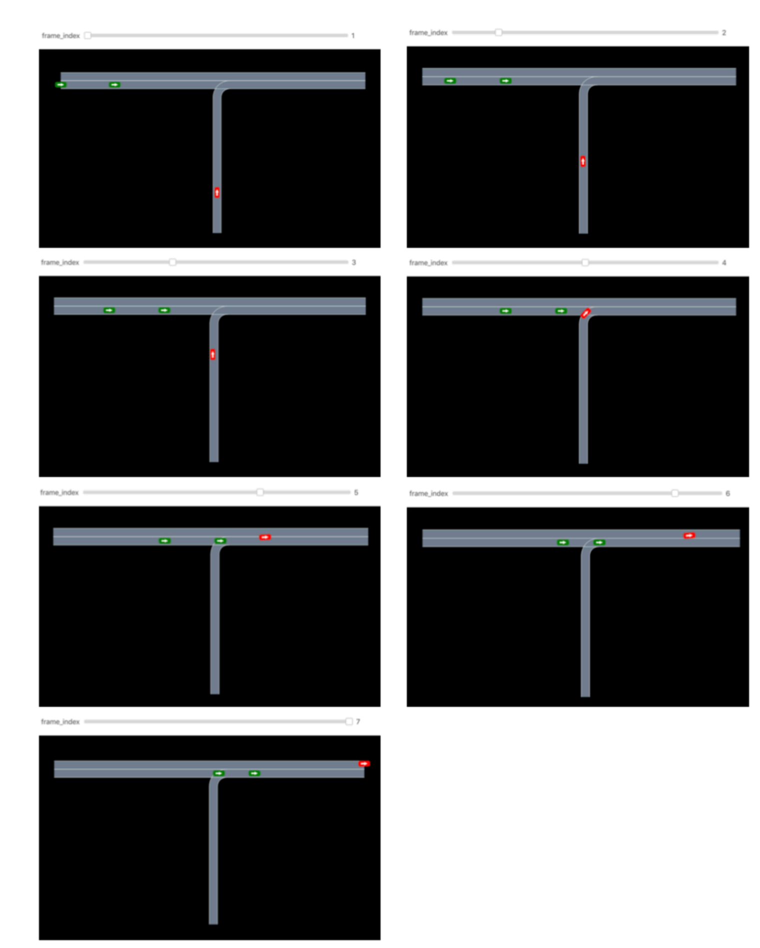


Figure 1: Baseline Policy, starting from top left frame 1 to bottom left frame 7

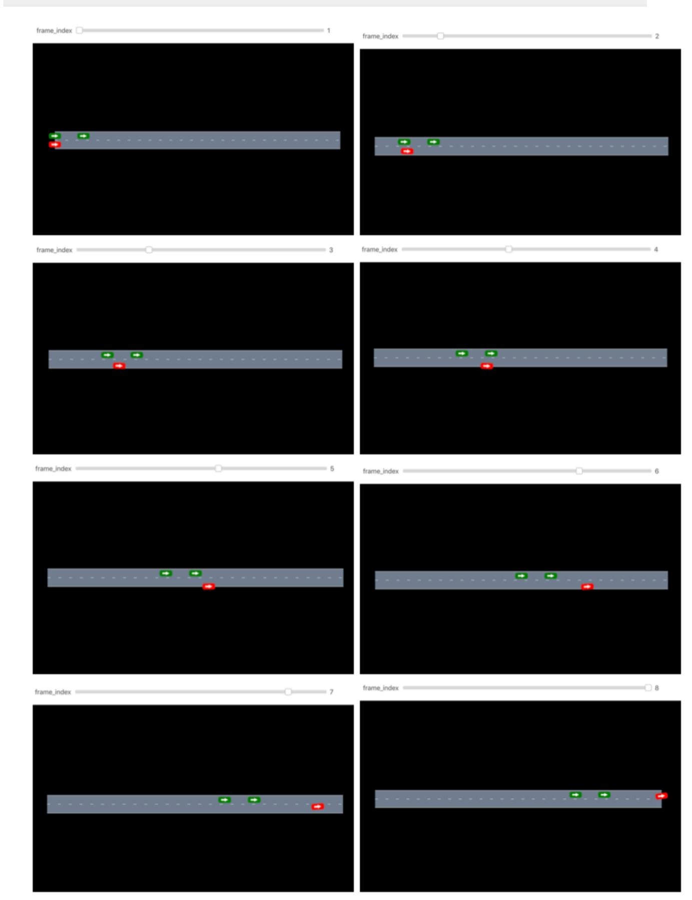


Figure 3: Visualization of the left lane-change micro-policy. Starting from top left frame 1 to bottom right frame 8.

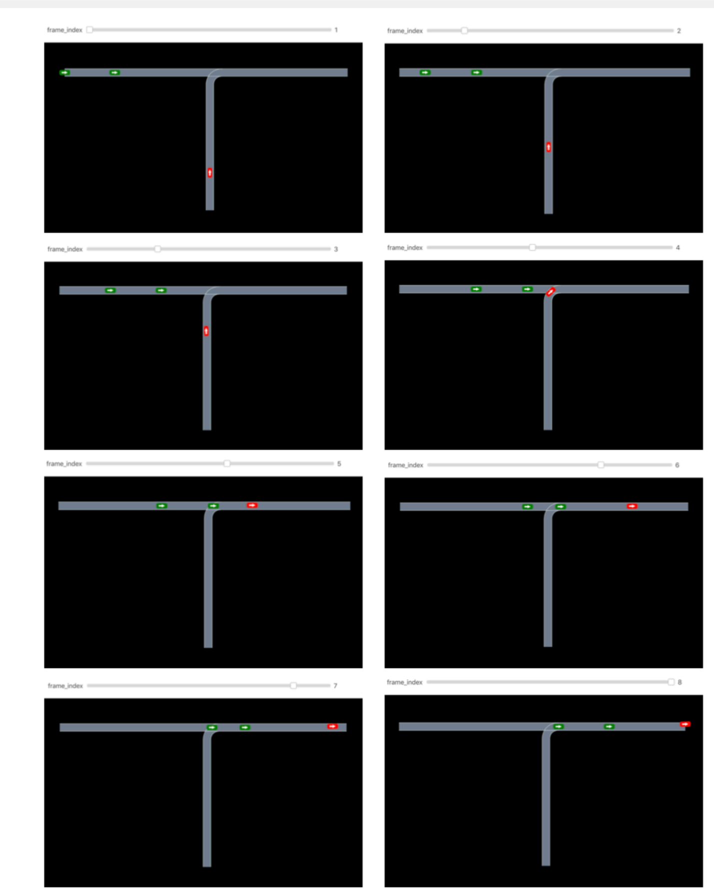
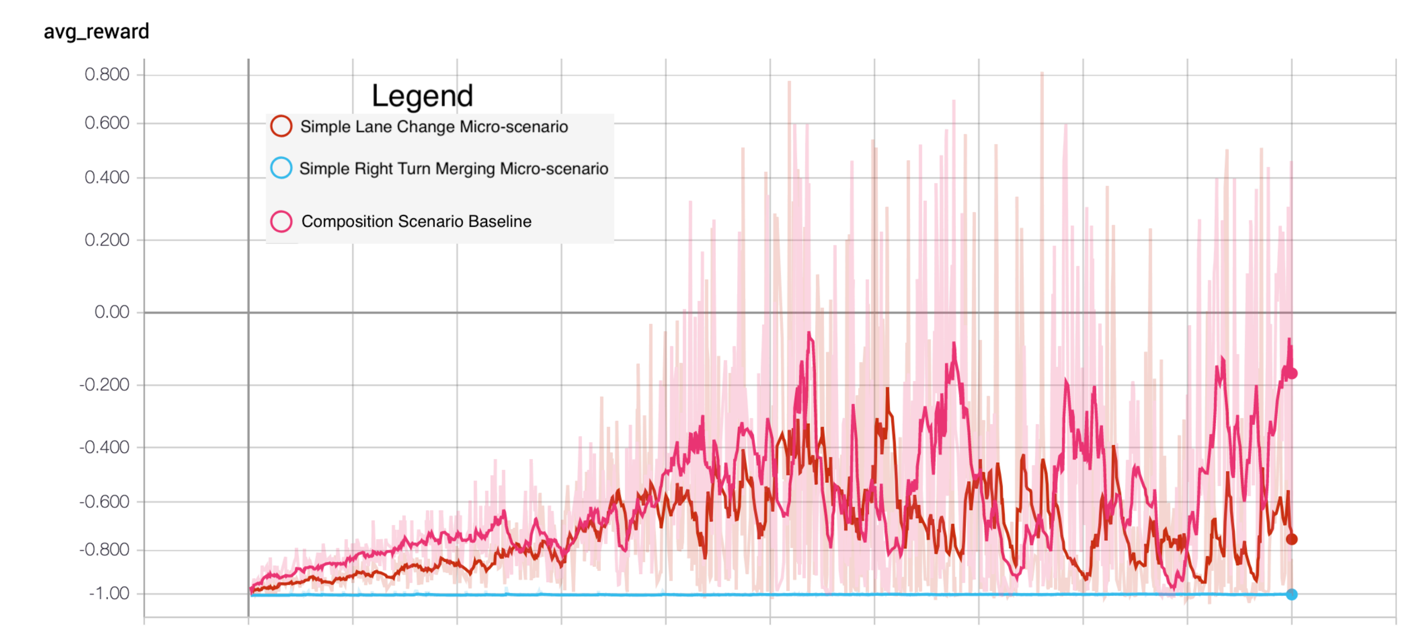
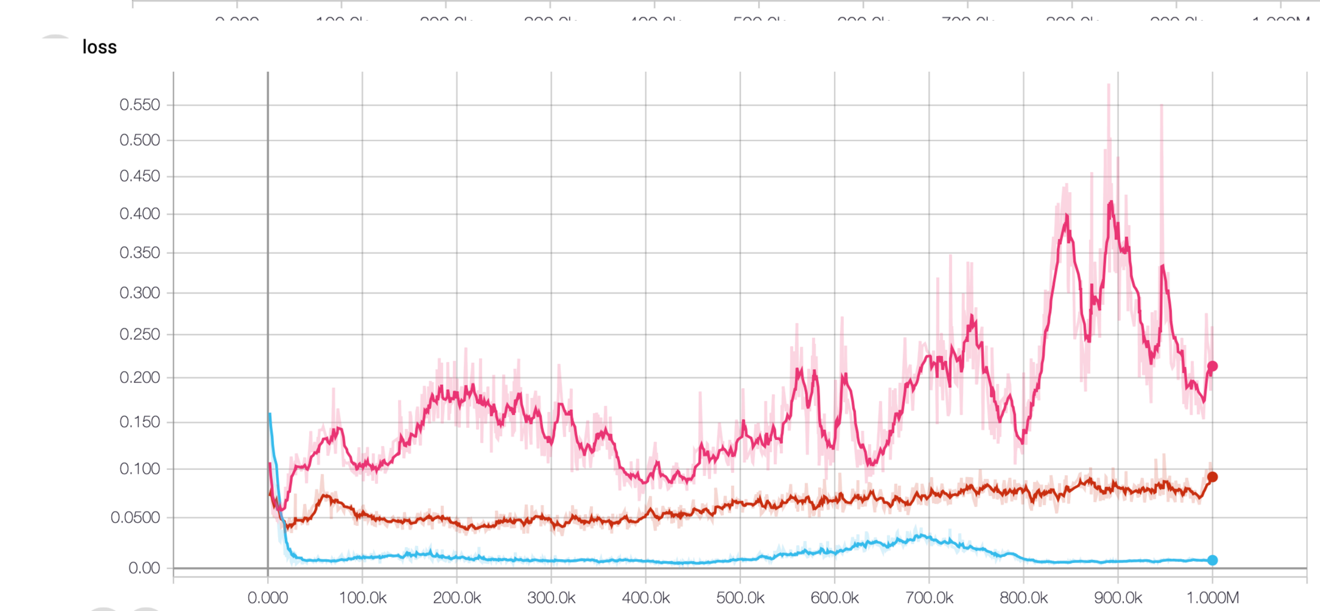


Figure 4: Visualization of the right-turn merge micro-policy. Starting from top left frame 1 to bottom right frame 8.

Results

Policy Name	Evaluation Reward	Timesteps
Baseline DRL	0.973	7
Q-Decomposition	0.968	9
Lane Change	0.960	8
Right Turn Merge	0.970	8





Conclusion

- Advantage
  - Ability to generalize from a set of micro-scenarios
  - More computationally efficient than DRL
  - Good performance, approximation is close to optimal policy
- Limitation
  - Not optimal policy, approximation only
  - Relies on an accurate scene decomposition function

Future Work

- Compose a "city-level" policy based on many micro-policies
- Investigate an efficient scene decomposition algorithm that is able to automatically decompose a high-level scene into a micro-scenario with efficiency and high degrees of accuracy
- Develop a formalism for state decomposition for urban driving and investigate efficient state representation
- Investigate how to these policies will interact with different driver models, stochastic worlds, and multiple agents