A. Supplementary Material for the Active Markov Game Framework

Most existing MARL methods adopt Markov Games (and their extensions such as POMDPs and Dec-POMDPs) as the underlying modeling framework, sharing the same environment transition assumption: $s_{t+1} \sim P(s_{t+1} \mid s_t, a_t)$. Under this assumption, the role of a policy π is mainly to serve as an action-sampling mechanism and to induce experience distributions; once the current state and joint action are given, the policy itself no longer affects the environment transition. In standard MARL, the value function of agent i is typically defined as $V_i^{\pi_i,\pi_{-i}}(s) = \mathbb{E}\big[\sum_{t=0}^{\infty}\gamma^t r_i(s_t,a_t)\big]$, but during training opponent policies are continuously updated (i.e., $\pi_{-i}^{(k+1)} \neq \pi_{-i}^{(k)}$), which results in different Bellman operators being applied at different training iterations (e.g., $T_{\pi_{-i}^{(k)}}V^{(k)} \neq T_{\pi_{-i}^{(k+1)}}V^{(k+1)}$). Consequently, the learning target of the value function changes over time, which is commonly regarded as a primary source of non-stationarity in MARL. In contrast, AMG relaxes this *action-sufficiency* assumption at the modeling level by explicitly incorporating policies into the environment dynamics: $s_{t+1} \sim P_U(s_{t+1} \mid s_t, a_t, \pi_t)$. The policy profile π_t on which the transition depends may encode strategy-level, long-term behavioral characteristics that cannot be fully captured by a single-step action. Formally, define a policy equivalence class $\Pi(a_t, s_t) = \{\pi : \pi(\cdot \mid s_t) \Rightarrow a_t\}$. In standard Markov Games, for any $\pi^{(1)}, \pi^{(2)} \in \Pi(a_t, s_t)$, the induced transition distributions are identical; in AMG, however, we allow the existence of $\pi^{(1)}, \pi^{(2)} \in \Pi(a_t, s_t)$ such that $P_U(s_{t+1} \mid s_t, a_t, \pi^{(1)}) \neq P_U(s_{t+1} \mid s_t, a_t, \pi^{(2)})$, thereby capturing the influence of strategy-level differences on environment evolution. Equivalently, for analysis purposes, AMG can be viewed as defining an augmented state representation that includes the policy profile, under which the environment transition becomes stationary; based on this modeling choice, the value function of agent i can be defined over the augmented state space and satisfies a standard Bellman equation, e.g., $V_i(s) = \mathbb{E}_{a\sim\pi}[r_i(s,a) + \gamma\,\mathbb{E}_{s'\sim P}V_i(s')]$, allowing the value function to be defined on an augmented state space while preserving Bellman consistency.

B. Experimental Supplement

This supplementary material provides additional visualizations and experimental evidence that further support the findings presented in the main paper. The presented results highlight interactive behaviors, strategy variations, and the effects of reward design under the proposed Active Markov Game (AMG) framework. Specifically, we include a wide range of driving interactions at unsignalized four-way and T-shaped intersections, covering potential collision configurations, yielding behaviors, conservative decision-making patterns, and aggressive opponent strategies. Fig.9,10,11,12 illustrate the diversity of interactions in four-way intersections, showing how the ego agent adapts to different opponent behaviors and how realistic strategies naturally emerge through the co-evolutionary training process. Fig.13,14,15 further demonstrate the model's generalization capability in T-shaped intersections by visualizing potential conflict layouts, ego-vehicle yielding maneuvers, and opposing-vehicle yielding strategies. These results confirm that the proposed method generates diverse, context-aware, and realistic driving strategies in complex and previously unseen scenarios, clearly surpassing traditional rule-based baselines. Additionally, Fig.16,17 compare training outcomes under sparse rewards and shaped rewards, revealing that sparse rewards lead to unstable and slow learning, whereas shaped rewards provide denser feedback that substantially improves training stability and convergence speed. Overall, the supplementary materials demonstrate that the AMG framework effectively models policy-dependent environmental dynamics, that the co-evolutionary mechanism produces a rich set of adaptive interactive strategies, and that dense reward shaping plays a crucial role in enabling efficient learning in complex multi-agent environments.

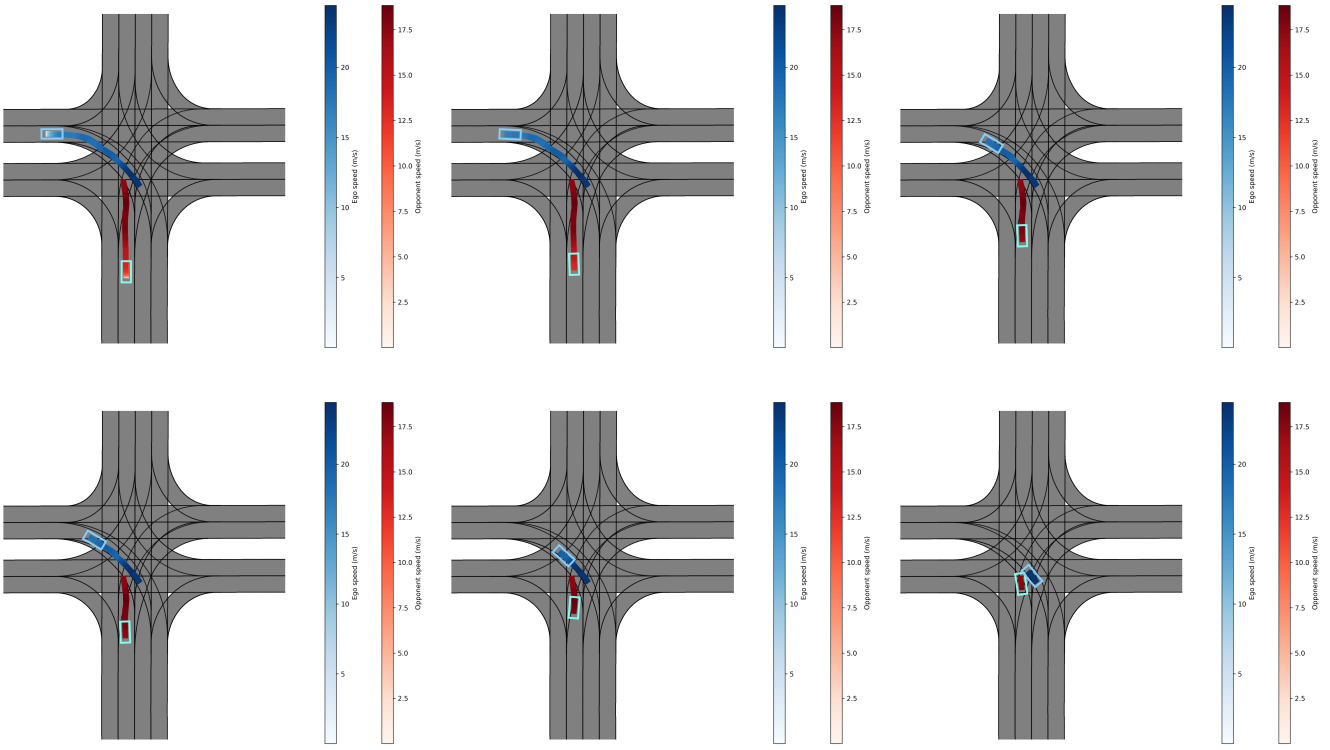

Figure 9. Visualization of Potential Collision Scenarios at Unprotected Four-Way Intersections

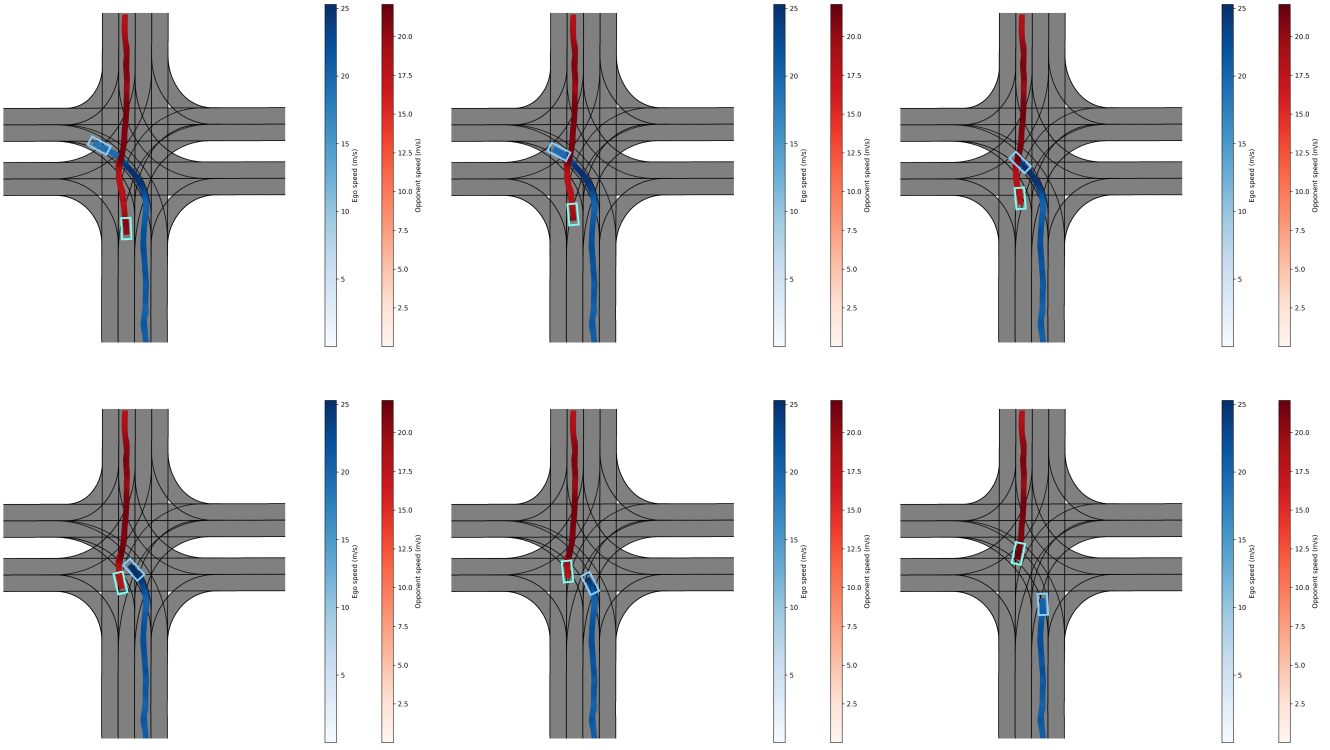

Figure 10. Visualization of the Yielding Strategy at Unprotected Four-Way Intersections

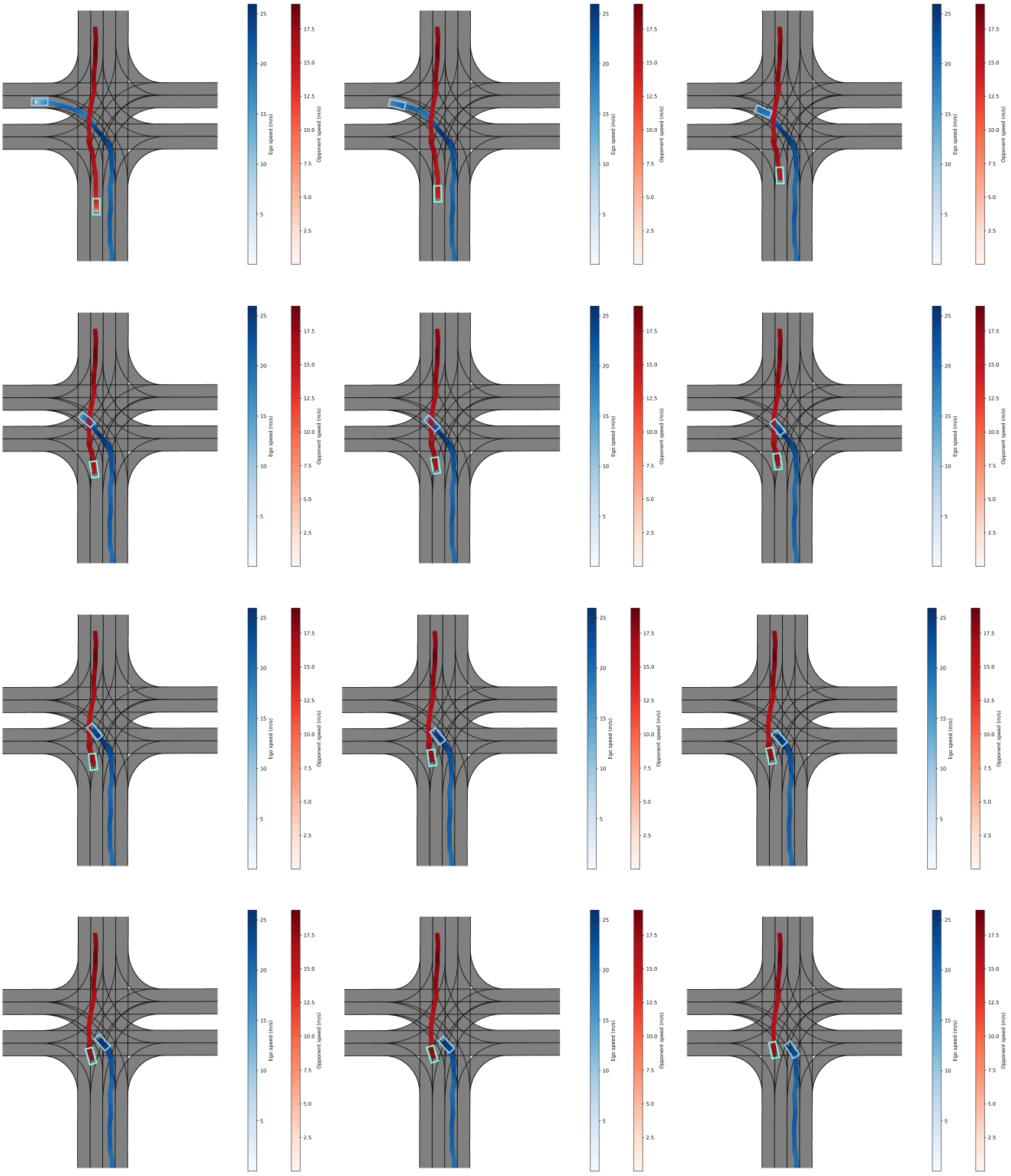

Figure 11. Visualization of the Conservative Driving Strategy at Unprotected Four-Way Intersections

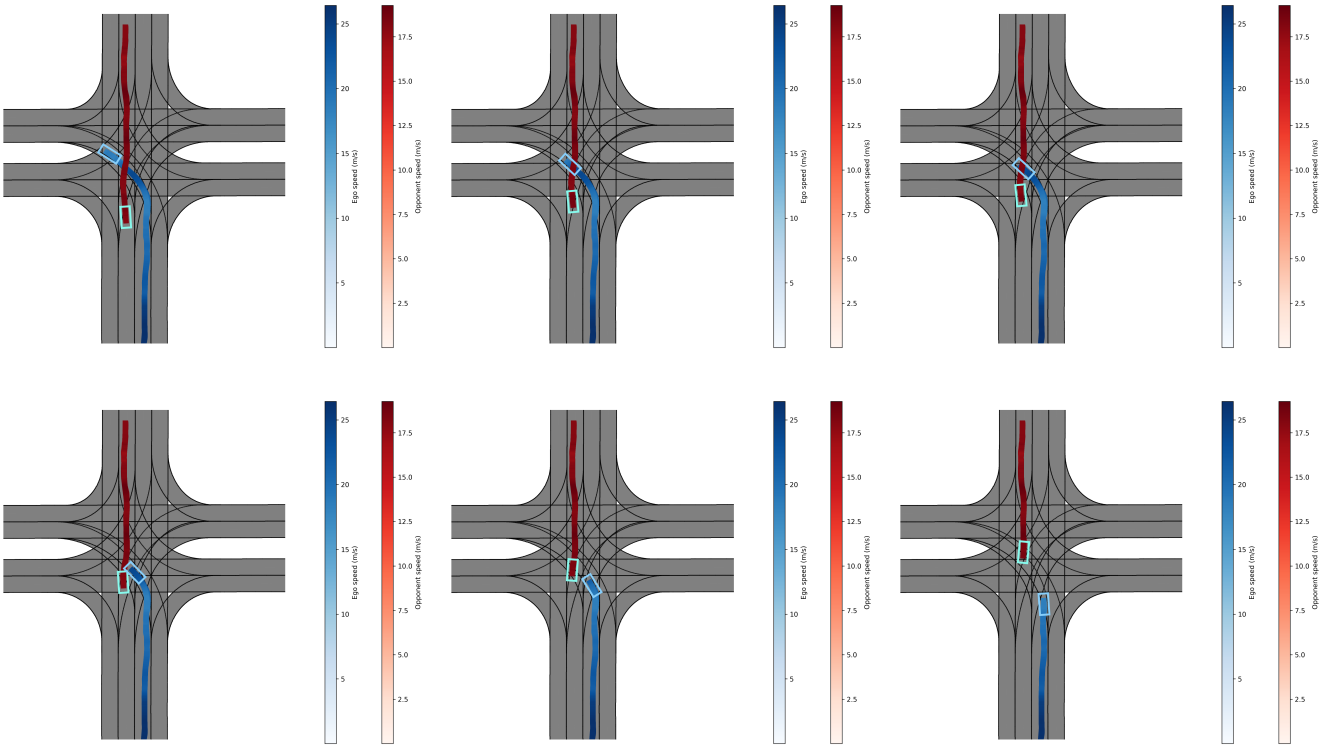

Figure 12. Visualization of the Aggressive Driving Strategy at Unprotected Four-Way Intersections

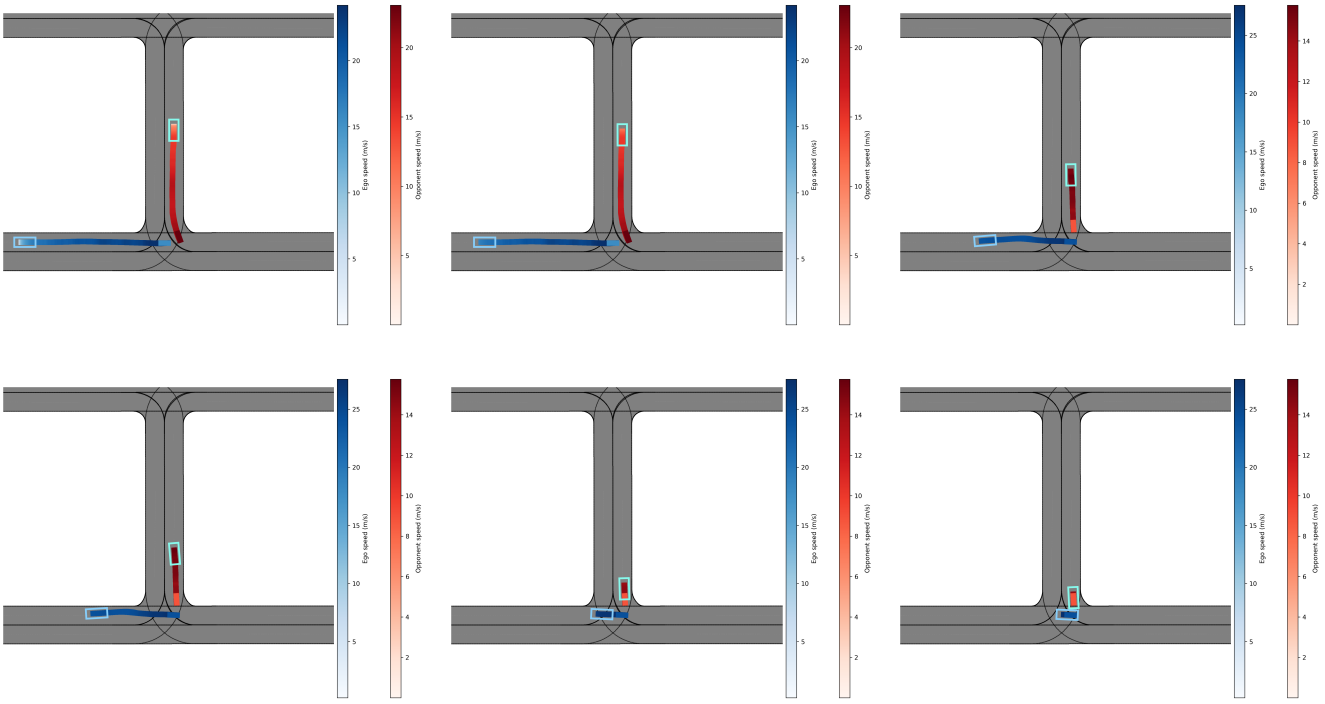

Figure 13. Visualization of Potential Collision Scenarios at Unprotected T-Shaped Intersections

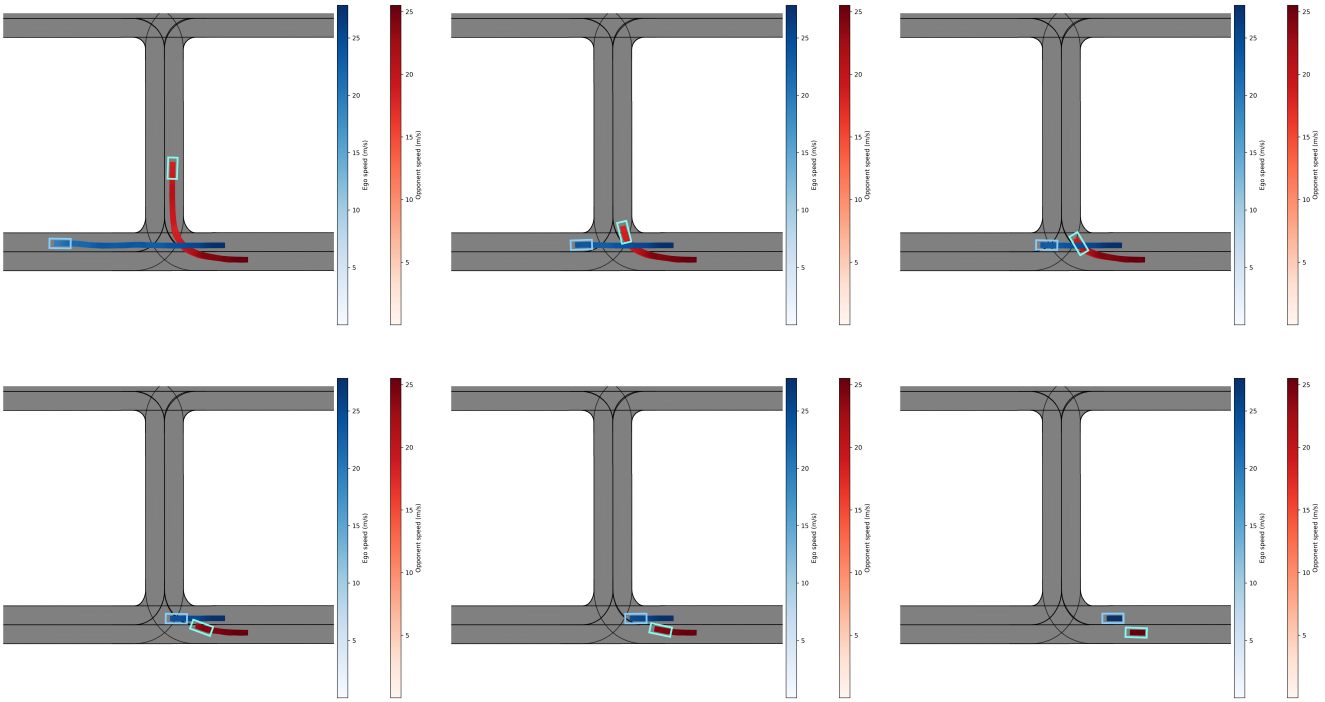

Figure 14. Visualization of the Ego-Vehicle Yielding Strategy at Unprotected T-Shaped Intersections

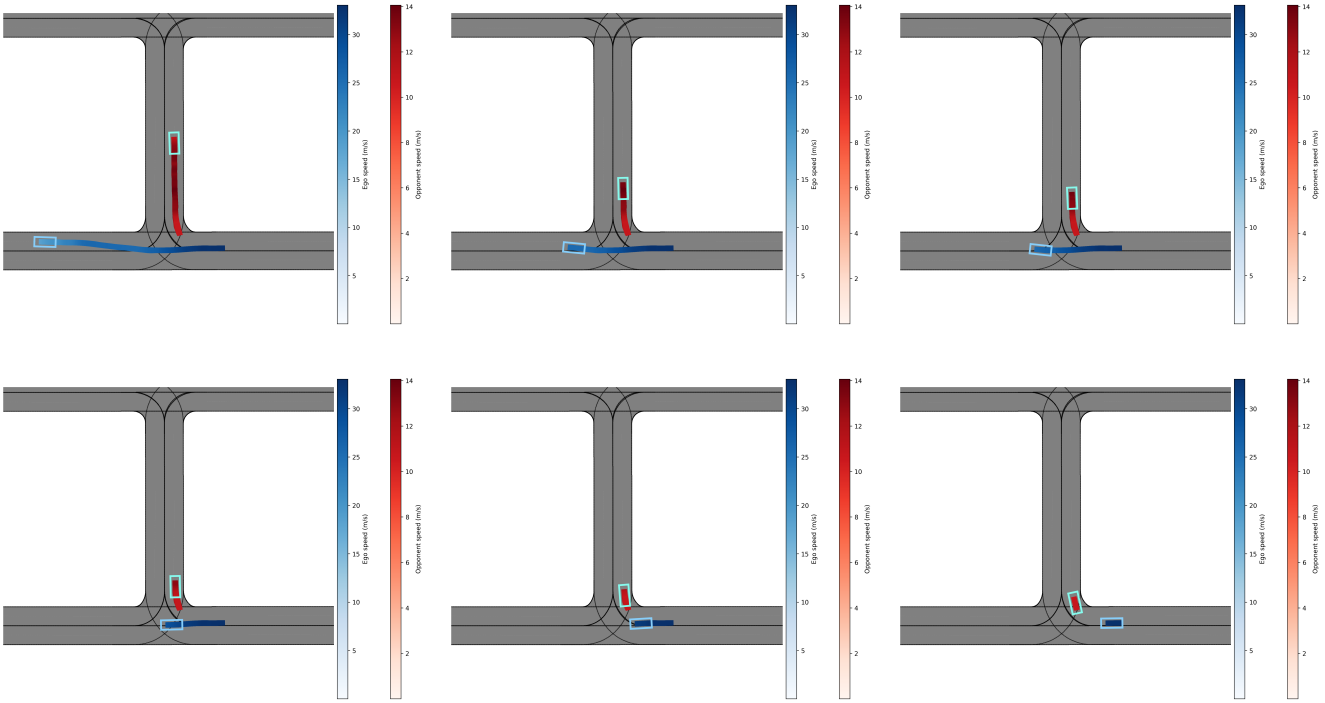

Figure 15. Visualization of the Opposing-Vehicle Yielding Strategy at Unprotected T-Shaped Intersections

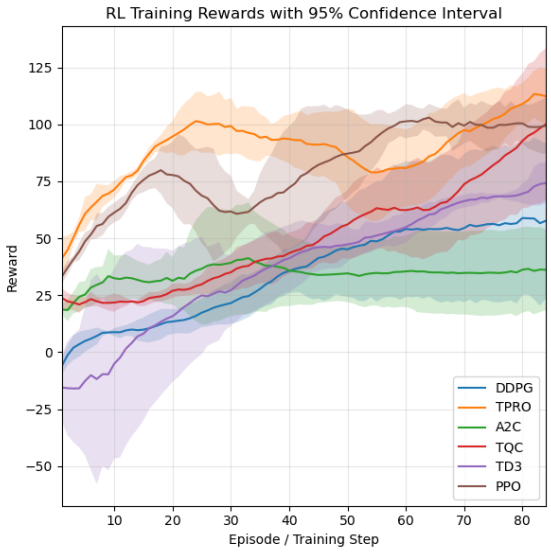

Figure 16. using a single sparse reward signal

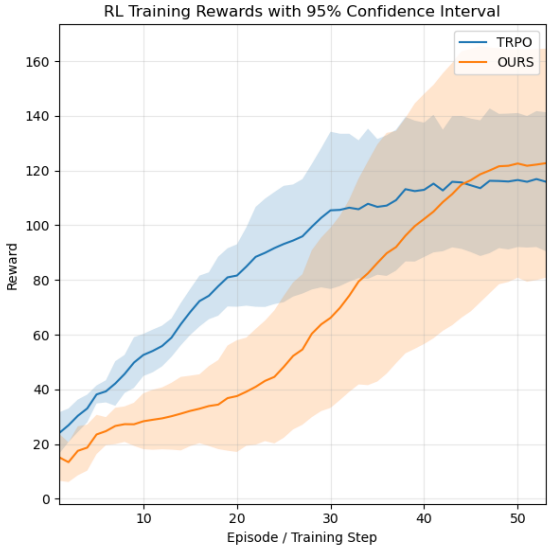

Figure 17. using shaped rewards for reward shaping