ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-1/W2-2025
13th International Conference on Mobile Mapping Technology (MMT 2025)
"Mobile Mapping for Autonomous Systems and Spatial Intelligence", 20–22 June 2025, Xiamen, China

HD Map in the Loop Framework for End-to-End Autonomous Driving

Shan He[1], Shen Ying[1]*, Lu Tao[1], Shi Chen[1], Yang Zhang[1]

Department of GI Science and Mapping, School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China

* Corresponding author: shy@whu.edu.cn

Keywords: HD map, Reinforcement learning, Imitation learning, Autonomous driving, Human in the loop.

Abstract

The generalized concept of "Human in the Loop" (HITL) enhances system performance by integrating human expertise into the decision-making process of agents. In a narrower sense, HITL specifically refers to human involvement in reinforcement learning (RL) through three key mechanisms: demonstration, intervention, and evaluation, each optimizing different stages of the training process. This approach effectively incorporates prior human knowledge, mitigates risks and sample bias in RL, and improves exploration efficiency and neural network convergence. However, existing HITL methods heavily rely on human experts for real-time annotations and guidance, leading to high implementation costs and operational complexity.
In the domain of autonomous driving, traditional hierarchical decision-making frameworks depend on high-definition (HD) maps for planning and navigation. Notably, the construction of HD maps inherently embeds expert knowledge, semantic rules, and constraint information. Inspired by this observation, this study introduces an innovative approach: "HD Map in the Loop" (HMITL), leveraging HD map features as a substitute for human expertise and establishing a corresponding application framework for autonomous driving. Specifically, this research systematically investigates three core aspects of HMITL in training end-to-end decision-control models: (1) imitation learning based on expert demonstrations from HD maps; (2) Method for constructing action interference and reward function guided by HD map spatial heterogeneity; and (3) Critic priority architecture relying on expert evaluations from HD map perception and features. These three dimensions are logically interrelated and collectively form the foundational framework of HMITL. By pioneering this methodological innovation, this study provides a novel solution to reducing reliance on real-time human intervention in autonomous driving while ensuring the reliability and safety of system decision-making.

1. Introduction

Conventional autonomous driving algorithms typically adopt a multi-level hierarchical structure consisting of planning and control modules. The planning module predicts the agent's trajectory, while the control module executes low-level actions such as steering, throttle, and braking (Le Mero et al., 2022). Although modular approaches have been widely used in early stages, they face inherent structural challenges. For instance, these methods require experts to define operational rules and environmental characteristics through hard-coded logic (Ravichandar et al., 2020). However, constructing a rule set that accounts for all possible scenarios is both complex and impractical (Zeng et al., 2019). Additionally, the explicit interfaces between modules often cause cumulative error propagation from upstream perception modules to downstream decision-making and control modules, limiting modular approaches to constrained environments.

In contrast, end-to-end autonomous driving methods treat perception, decision-making, and control as a unified learning task, inspired by the "behavioral reflex" mechanism in human driving. Instead of relying on predefined rules or explicit module interfaces, these methods directly generate waypoints or control commands from onboard sensor signals. The key advantage of this approach lies in its ability to eliminate performance bottlenecks caused by manually defined rules while simultaneously uncovering hidden information patterns.

Within the realm of end-to-end autonomous driving, Imitation Learning (IL) and Deep Reinforcement Learning (DRL) are two primary learning paradigms. IL focuses on replicating human driving behavior by mimicking expert demonstrations in given states, making it intuitive and user-friendly. However, it suffers from distributional shift issues, leading to cumulative errors over time (Codevilla et al., 2019), and its performance is inherently constrained by the capabilities of the expert policy it imitates (Wu et al., 2023). In contrast, DRL is a data-driven self-optimization algorithm that autonomously discovers control strategies through exploration and trial-and-error (Sutton and Barto, 2014). It has demonstrated remarkable potential in sequential decision-making tasks, as exemplified by its success in Go (Silver et al., 2016, 2017, 2018).

Despite their advantages, end-to-end methods—particularly those based on IL and DRL—suffer from inherent limitations due to their black-box nature. They often lack model interpretability, struggle to guarantee performance lower bounds, and raise safety concerns. To overcome these limitations, recent research has explored hybrid frameworks that integrate modular and end-to-end methods, leveraging their respective strengths. Modular approaches provide interpretability and traceability, compensating for the shortcomings of end-to-end models, while end-to-end learning enhances task complexity handling. This Modular End-to-End Learning paradigm represents a promising direction for autonomous driving (Chen et al., 2024), enabling feature modules to assist policy training within end-to-end networks.

With the rapid advancement of autonomous driving technologies, an increasing number of road networks have been mapped using high-definition (HD) maps. However, the future of autonomous driving faces a strategic divide between map-centric (HD map-dependent) and perception-centric (sensor-heavy) approaches. Under the perception-centric paradigm, existing HD map resources risk becoming significantly underutilized. Currently, HD maps primarily serve as components in traditional hierarchical frameworks for path planning and guidance. Yet, their inherent geometric precision, regulatory constraints, and semantic richness remain underexplored, and the HD map generated in modular end-to-end is only a feature expression that

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-1-W2-2025-43-2025 | © Author(s) 2025. CC BY 4.0 License.

has no connection to reality as shown in figure 1. This suggests that HD maps hold potential beyond their conventional role, particularly as structured modules within end-to-end models. However, this application should differ from existing approaches that simply use HD maps as direct input sources or generated features, instead exploring how HD maps can actively contribute within perception-centric frameworks.

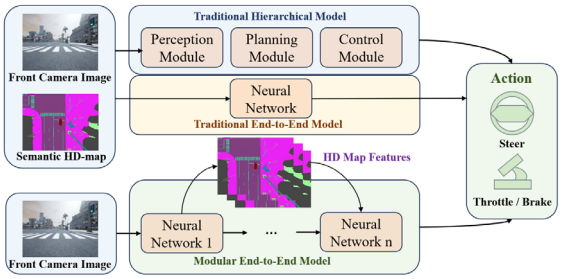

Figure 1. Application of HD Maps in Autonomous Driving Models

Therefore, this paper proposes a modular end-to-end autonomous driving framework that integrates HD maps into the loop. The core idea is to introduce the geometric, semantic, and rule-based features of HD maps into the training process of the end-to-end autonomous driving model, similar to how human expert experience is incorporated into agent training in traditional human-in-the-loop (HITL) approaches. Three specific HITL methods are proposed for this framework: (1) an improved Dataset Aggregation (DAgger) algorithms algorithm based on map navigation to mimic human demonstrations; (2) action interference and reward function construction based on HD map spatial heterogeneity, corresponding to human intervention; and (3) a RL critic-prioritized structure corresponding to human evaluation. These methods form the HD map-in-the-loop (HMITL) framework, which does not directly use HD map perceptual information as the basis for agent decision-making. Instead, it leverages HD maps as an aid to accelerate the agent's training iterations, improving training quality and reducing risks during the training process. The final policy model, however, is capable of operating independently of HD maps, representing a purely perception-centric autonomous driving model.

2. Related Works

2.1 Reinforcement Learning in Autonomous Driving and the Role of HD Maps

As a major branch of machine learning, Reinforcement Learning (RL) has demonstrated significant potential in solving complex decision-making and control tasks (Gajcin and Dusparic, 2024; Sutton and Barto, 2014). In the field of autonomous driving, Deep Reinforcement Learning (DRL) plays a crucial role in learning driving policies. Early model-free algorithms, such as Deep Q-Network (DQN) and Deep Deterministic Policy Gradient (DDPG) (Mnih et al., 2015), have been successfully applied to autonomous driving strategies (Wolf et al., 2017). Subsequently, more advanced actor-critic architectures—such as Soft Actor-Critic (SAC) (Haarnoja et al., n.d.) and Twin Delayed Deep Deterministic Policy Gradient (TD3) (Fujimoto, n.d.)—have significantly improved performance in complex driving environments, particularly in urban driving and high-speed drifting scenarios (Sallab et al., 2017).

Despite these advancements, DRL still faces two major challenges: low learning efficiency and limited scene adaptability. Additionally, DRL models often struggle with scene comprehension in complex environments, which hampers their learning performance and generalization ability. As a result, DRL methods frequently underperform compared to human drivers when handling diverse driving tasks (Huang et al., 2021; Lv et al., 2018). To address these challenges, researchers have sought to integrate human-like features into DRL models, leveraging common-sense knowledge and neuro-symbolic learning to enhance machine intelligence (Mao et al., 2019).

Beyond the lack of human prior knowledge, reward function design remains a critical challenge in DRL-based autonomous driving. The reward function not only dictates the quality of the learned policy but also affects training efficiency (Neary et al., 2021; Xu et al., 2023). In high-dimensional, complex environments, agents require extensive interaction with the environment, consuming significant time and computational resources. Moreover, sparse reward signals further degrade learning efficiency. To mitigate this issue, reward shaping (Harutyunyan et al., n.d.) and inverse reinforcement learning (IRL) (Ibarz et al., n.d.) have been explored. Additionally, combining imitation learning (IL) with DRL has emerged as a viable strategy for improving convergence speed by using expert knowledge to constrain the exploration space (Codevilla et al., 2019; Han and Yilmaz, 2022; Zhang et al., 2021). However, these methods typically impose higher requirements on model transferability, sensor fusion, and computational resources (Chitta et al., 2023).

In summary, DRL-based autonomous driving models face fundamental challenges related to the absence of human prior knowledge and inferior performance compared to IL-based approaches that leverage such knowledge (Chen et al., 2024). This results in low exploration efficiency, increased risk of unsafe behaviors during training, and difficulties in designing effective reward functions. Additionally, the high-dimensional state space of driving scenarios exacerbates convergence difficulties, placing greater demands on model representation capabilities.

The introduction of HD maps offers a novel approach to addressing these challenges. From a RL perspective, HD maps enhance perception robustness by filtering out irrelevant noise and emphasizing critical driving information, which can help mitigate state mismatch issues, improving overall stability in dynamic environments (Cultrera et al., 2024). As intelligent sensors containing rich geometric and semantic information, HD maps provide detailed road network data, speed limits, traffic signs, and other critical elements (Liu et al., 2020). Beyond traditional map usage, HD maps can generate bird's-eye-view (BEV) imagery (Chen et al., 2019; Cui et al., 2019; Maramotti et al., 2022; Zeng et al., 2019) or semantic segmentation images (Nehme and Deo, 2023; Wu et al., 2023). However, these information of HD maps are often directly used as inputs for the policy network, resulting in the final trained model HD map-dependent. It is crucial to find a method that can both utilize information of HD maps and eliminate reliance on it during practical application.

2.2 Challenges and Solutions in Imitation Learning for Autonomous Driving

Imitation Learning (IL) offers a promising approach for rapidly acquiring an initial policy by leveraging expert demonstrations (Li et al., 2022), although its performance is often constrained by the quality of the expert data. As a result, most state-of-the-art end-to-end decision-making models pretrain using imitation learning and then introduce RL to enhance overall training

efficiency and strategy abilities in complex scenarios (Hester et al., 2018; Nair et al., 2018; Saunders et al., 2017; Vecerik et al., 2018; Zhang and Ma, 2018; Zhu et al., 2018). However, imitation learning algorithms, particularly Behavior Cloning (BC) (Ibarz et al., n.d.), face several inherent challenges: state mismatch (Hua et al., 2021), data distribution shift (Reddy et al., 2019) and error accumulation (Zare et al., 2024).

To address these issues, Dataset Aggregation (DAgger) algorithms have been proposed to mitigate covariate shift and error accumulation in imitation learning (Reddy et al., 2019; Zare et al., 2024). The DAgger algorithm works by allowing the agent to interact with the environment using its current policy, collecting data, and having it annotated by experts. This annotated data is then used for the next round of policy training (Ross et al., 2011).

However, DAgger algorithms combine data augmentation and forced interventions, which creates a potential conflict in their approach. While forced intervention helps reduce error accumulation, it also leads to data imbalance, exacerbating the covariate shift problem. This results in the collected states following the mixed-policy distribution rather than the agent's current policy leading to unfamiliar states that may cause covariate shift during training (Mandlekar et al., 2020).

In summary, interactive-exploration IL algorithms like DAgger shares a similar framework with RL, where human intervention can help resolve intrinsic challenges such as distribution shift and error accumulation that negatively impact sampling efficiency. However, unlike RL, which can use importance sampling to alleviate distribution shift (Sutton and Barto, 2014), IL is still susceptible to state distribution bias caused by the mixed-policy approach. A promising solution is to incorporate HD maps into the training framework to filter harmful samples and increase the proportion of effective samples during the interaction exploration phase. By leveraging HD maps to identify and control driving scene interactions, we can improve the sampling efficiency and mitigate the impact of mixed-policy bias in imitation learning. A good pretrained model for IL will also enhance the interactivity of end-to-end RL during the fine-tuning.

2.3 Challenges and Solutions in Human in the Loop

In human-agent interaction, humans typically guide the agent's learning by providing knowledge related to the RL problem, such as Q-values, optimal actions, or real rewards. This guidance can accelerate the learning process, avoid catastrophic outcomes, and optimize exploration efficiency. However, most existing research develops interaction protocols for specific agents, without creating a universal framework (Abel et al., 2017).

For human-in-the-loop (HITL) framework, expert intelligence can participate in through several methods, including human evaluation, human demonstrations, and human intervention. In these frameworks, methods like BC (Ibarz et al., n.d.) and IRL (Ziebart et al., n.d.) are integrated into representative algorithms such as DQL (Hester et al., 2018; Saunders et al., 2017) and DDPG (Vecerik et al., 2018). Experiments indicate that these methods outperform traditional DRL approaches in robotics (Krening et al., 2017).

Current HITL framework rely on real-time annotation guidance provided by human experts through supervised learning. However, these methods still face several challenges: Long-term Supervision and Guidance Fatigue (Littman, 2015; Droździel et al., 2020; Saunders et al., 2017), Dependence on Expert Demonstrations (Hu et al., n.d.) and Non-stationarity of Human Teaching Strategies (Knox et al., 2012). A promising solution to these challenges is to integrate HD Maps planning into the DRL training framework instead of human expert experience or directly leveraging them as control gates to determine when the intervention is needed, which can reduce the uncertainty and labor consumption of human participants.

3. Method

This paper proposes a HD-map-in-the-loop (HMITL) framework for autonomous driving, which fully leverages the information resources available in HD maps such as geometric attributes and semantic properties to replace the traditional reliance on human expert, comprising the following components as shown in Figure 2, three implementation plans corresponding to human demonstration, human intervention and human evaluation through HD maps will be provided later:

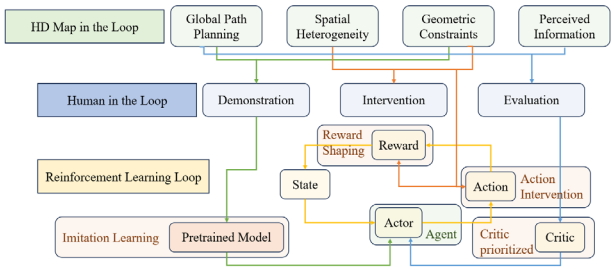

Figure 2. Illustration of HD Map in the loop Framwork

3.1 ShotgunDAgger: An HD Map-Assisted Imitation Learning Optimization Algorithm

To address the issues of covariate shift and error accumulation in imitation learning, this paper introduces the ShotgunDAgger algorithm, an improvement upon the traditional DAgger algorithm which adjusts the expert's strategy at critical moments to ensure that the collected data aligns more closely with the agent's current strategy. Additionally, a behavior constraint mechanism based on HD maps is introduced to reduce the impact of error accumulation, resulting in a more robust HITL approach based on expert demonstrations.

ShotgunDAgger provides two approaches: the first is to adjust the expert's strategy during sampling bias and continue sampling, ensuring that subsequent data follows the agent's current policy distribution; the second is to reset the simulation environment at key moments, allowing the agent to start sampling from a new state. These approaches do not forcefully intervene in the agent's decision-making but optimize at the data collection level, ensuring unbiased data distribution.

To enhance the algorithm's practical applicability in autonomous driving scenarios, this paper integrates the semantic properties of HD maps and designs an expert strategy adjustment mechanism based on lane-changing behavior. When the agent vehicle deviates from the planned path, the expert strategy is adjusted to align with the agent's strategy through two approaches above: in this process, the HD map's auxiliary decision-making mechanism, called the HD Map Gate, plays a key role. When the agent performs a lane change, this mechanism compares the current lane ID with the lane ID of the next waypoint in the planned path to determine if a deviation has occurred. When a deviation is detected, a new global path planning is performed based on the vehicle's current location and destination, or the interaction

environment is reset to begin sampling again. This operation not only ensures that the re-collected samples after deviation still follow the agent's current policy distribution, but also minimizes ineffective samples caused by data deviation.

The two implementation approaches of ShotgunDAgger are shown in Figure 3. The distinguishing feature of both methods is that the expert strategy is used only as the true label, without forcefully interfering with the agent's actions. This ensures that the collected data samples always conform to the agent's current policy distribution, avoiding the potential bias problems inherent in traditional methods. As a result, ShotgunDAgger has high applicability, particularly in the unbiased data methods commonly used in RL, ensuring the efficiency and accuracy of the data. The choice between the two methods should be based on the specific task and environment. ShotgunDAgger1 is better suited for scenarios with complex path planning, while ShotgunDAgger2 is more appropriate for tasks that can reset the environment and quickly adapt to changes.

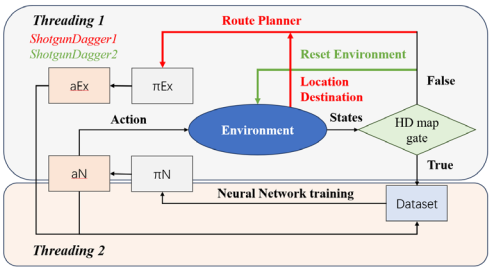

Figure 3. Illustration of HD Map Assisted Imitation Learning Algorithm ShotgunDAgger

3.2 Action Interference and Reward Function Construction Using HD Map Spatial Heterogeneity

3.2.1 Action Interference: This paper introduces HD map-based path planning information to influence the vehicle's trajectory which aim is to gradually reduce the reliance on HD maps during the RL process, allowing the agent to train more effectively in autonomous decision-making. Elements from the HD map, such as guide lines and deceleration zones, are integrated into the agent's decision-making model. Each action is assigned a probability weight, with the agent selecting the actual action to be executed based on these probabilities as shown in Figure 4:

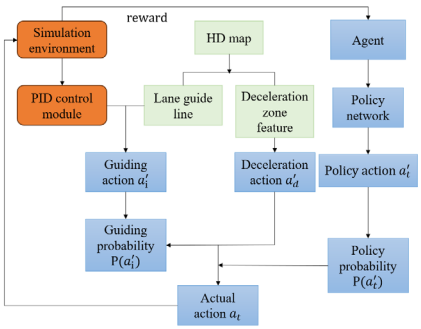

Figure 4. Illustration of Process of HD Map Intervention in action stage

A key step in the algorithm is that within the simulation platform, the vehicle's path planning information is controlled by HD map elements, such as guide lines and deceleration zones, which govern the vehicle's trajectory. During the initial stage, the agent relies more heavily on the HD map's guidance for actions, such as driving within the guide lines or automatically decelerating when entering a deceleration zone. These actions guided by HD maps or made by the RL actor policy are assigned different probability values. In the early stages of training, the probability of HD map-guided actions is higher, while the RL actor policy actions are given lower probability. As training progresses, the weight of the HD map-guided actions gradually decreases, and the influence of the RL actor policy actions increases, enabling the agent to learn standardized driving as much as possible in the early training stages, and increases exploration in the later stages to enhance models' robustness.

This method is similar to the DAgger algorithm, both based on a hybrid policy of state-action pairs as training samples. The key difference lies in the DAgger algorithm's approach of directly addressing the state distribution difference between the agent's strategy and the expert's strategy, requiring iterative correction through repeated sampling during model training. In contrast, while this study's method uses a similar hybrid policy to execute actions, the state distribution bias can be controlled through Importance Sampling (IS) in RL, making it an off-policy approach with less impact from the hybrid policy compared to IL algorithms.

Furthermore, unlike decaying action interference, the influence of the semantic functional zones in the HD map on the agent's actions does not diminish over the course of training. As shown in Figure 5, these functional areas, such as deceleration zones, immediately enforce the relevant actions (e.g., emergency braking) when the agent enters them and provide negative feedback through a comfort reward function to prevent the agent from driving at high speeds in deceleration zones. These functional areas exert continuous influence over time and are geometrically constrained.

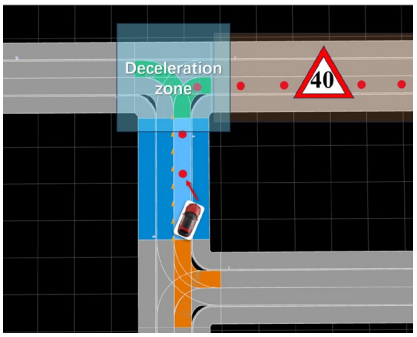

Figure 5. Illustration of Process of HD Map Intervention in semantic functional zones

3.2.2 Reward Shaping: In the reward feedback phase, the reward function is shaped by the rule-based constraints from the HD map. Using features such as lane speed limits, directions, and obstacle locations from the HD map, the reward function is updated in real-time, with different reward functions designed for various HD map elements to reflect the quality of the agent's behavior. This design makes the reward mechanism more targeted and responsive, ensuring that the feedback during the RL process aligns more closely with actual traffic rules and human cognition.

This study aims to address the sparsity of reward functions in RL cooperated with action intervention methods. Traditional reward functions often combine factors such as goal return, obstacle

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-1-W2-2025-43-2025 | © Author(s) 2025. CC BY 4.0 License.

avoidance distance, cognitive uncertainty, scene similarity, speed, and comfort into a weighted similarity calculation. The HD map provides a quantitative description of the agent's current state based on global information, which has the potential to replace human evaluation work, avoiding costs and stability issues associated with human involvement. The reward function design based on HD map spatial evaluation in this study is similar to traditional methods but is constructed from the perspectives of safety, efficiency, and stability with each aspect incorporating local and global information from the HD map as shown in figure 6:

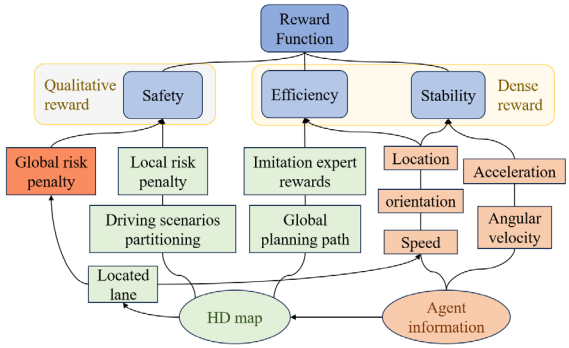

Figure 6. Schematic Diagram of Reward Function Composition based on HD Map

For safety-related rewards, traditional reward functions use discrete reward factors like collision occurrences or the dense rewards like distance to obstacles, and some neural network-based reward functions also penalize cognitive uncertainty to improve the agent's prediction accuracy. When designing reward functions from the HD map perspective, spatial heterogeneity must be considered, combining dynamic local space information with global path planning risks to provide a comprehensive evaluation of the current state. As shown in Figure 7, the agent (represented by the red vehicle) faces a blue vehicle as an obstacle in the upcoming lane. From the HD map's perspective, the risk distances to all obstacles are different: vehicles in adjacent lanes may be closer in parallel, whereas the same distance between vehicles in the same lane results in a significantly higher risk because most vehicles move in straight lines in the lane. When constructing obstacle distance penalties, higher weights should be assigned to obstacles within the same lane segment. The movement of dynamic obstacles in autonomous driving scenarios follows certain driving constraints and patterns, which can be distinguished based on the HD map's spatial heterogeneity.

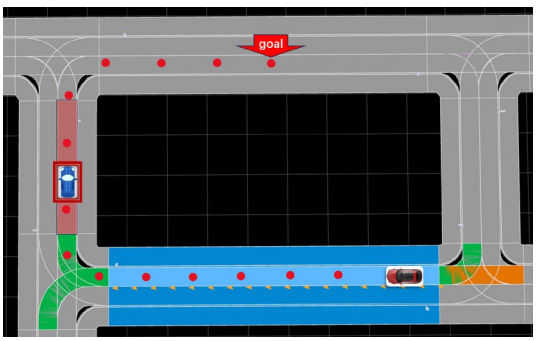

Figure 7. Illustration of Reward function shaping based on HD Map

Compared to using visual feature similarity as a target goal reward in navigation tasks, constructing goal reward using HD maps incorporates more spatial characteristics, avoiding reward sparsity issues caused by undetected effective features. Spatial similarity generally includes geometric, semantic, topological, and orientation similarities, and when the agent enters a new waypoint range, the shortest path is planned using global planning algorithms based on the HD map incorporating these four similarity factors. As shown in Figure 7, if the current distance between the red agent vehicle and the target goal is considered only geometrically using Euclidean or Manhattan distance, taking the right route would be more effective. However, based on driving behavior norms in autonomous driving scenarios, the path planned from the left is more in line with traffic regulations, particularly since the current road segment is marked with a double yellow line that prohibits U-turns. The shortest path planned by the global planning algorithms algorithm includes semantic and orientation information, where the former refers to different functional areas on the HD map, and the latter refers to a one-dimensional directional evaluation—lane segment IDs. Opposing lane segments, despite being close in space or even adjacent, represent entirely different meanings for driving decisions due to their opposite directions. The global planning algorithm itself depends on the topological representation of the road segments because ensuring the planned shortest path conforms to the topological connectivity of the HD map. By using these navigation algorithms to set the reward function, it can become more guiding. To some extent, the essence of this method belongs to the segmented reward function, but its segmentation is based on spatial heterogeneity. This encourages vehicles to make decisions based on the guidance of rewards provided by experts in different states and provides effective dense gradients combined with action interference methods beyond forced control sampling.

3.3 HD Map Perception-Based Priority Critic Network Algorithm

This paper proposes a differential actor-critic architecture, as shown in the flowchart in Figure 8. The architecture follows an actor-critic RL framework and uses an experience buffer for off-policy training. The main difference from previous RL models, which relied on HD map perceptual information, is that the HD map features are only input into the critic network, not the actor network. This means that the critic network utilizes the global view provided by the HD map for more macroscopic value evaluation, while the actor relies on the agent's local perception from visual inputs combined with the macro evaluation from the critic to iterate toward optimal actions.

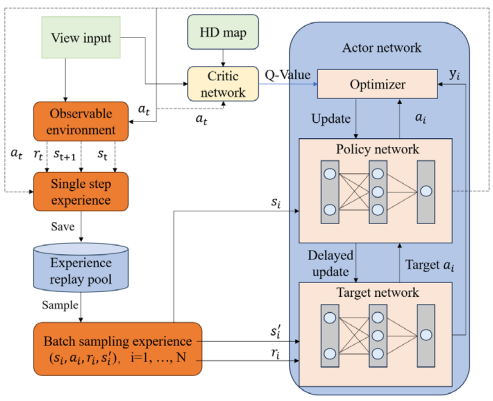

Figure 8. Framework of critic-prioritized Algorithm

In the information input phase, the actor is only provided with the forward camera's view, while the critic receives a broader range

of information, including semantic segmentation and Bird's Eye View (BEV) images from the HD map. There are three main benefits to this approach. First, compared to pure vision-based decision-making models or modular end-to-end models that start learning from random initial networks, this method incorporates many manual features in the input phase, such as obstacle distance derived from the HD map, planned paths, or semantic segmentation images. This greatly helps the critic network filter out redundant information, accelerating the understanding of the global environment. Second, in addition to enhancing the model's understanding capability, the HD map helps reduce the generation of dangerous states and extreme samples during training. For example, issues like blind-spot cornering can be mitigated by the critic's ability to give more reasonable evaluations of the current state based on the HD map, through Q-Value provided by the critic, agents can combine dangerous signs with subtle changes in forward camera images, alleviating causal confusion in RL. Third, after training, the policy network (actor) can generate action instructions solely based on the camera's visual input and auxiliary localization information, without the need for the critic or HD map assistance. However, the global information from the HD map is implicitly retained in the actor's strategy in the form of sense of direction. This is because the critic provides the global path, so the actor receives higher rewards during training when it follows the global path.

For both the actor and critic networks, in addition to differentiated perceptual inputs, there are also many shared auxiliary localization features, such as vehicle speed, acceleration, direction, and position, which can be obtained through GPS devices rather than strictly relying on the HD map. These inputs could also follow the " critic-prioritized " approach by feeding high-level features from the HD map directly to the critic network, improving its state estimation capacity. Figure 9 illustrates a specific approach based on hierarchical reinforcement learning. In existing algorithms either humans or neural networks providing semantic instructions in different scenarios, with RL only handling low-level operations to achieve rewards. In this paper, we propose replacing the human or neural network role with the HD map, as the HD map provides real-time global attributes of the agent's vehicle. Using the HD map's global information and the agent's current state, a navigation path can be planned, and high-level commands can be extracted from the perspective of human semantics (e.g., "go forward" or "turn left"). Since macro-level instructions do not require specific obstacle avoidance actions, the HD map can fully replace humans and, using global path planning algorithms like A*, can provide results even faster than a human.

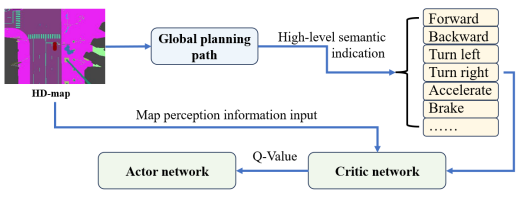

Figure 9. critic-prioritized structure based on HD Map High-level Instructions

For command input, instead of using high-level commands as input to the actor in many multi hierarchical reinforcement learning systems, this paper proposes directly inputting these commands as modular features into the critic network. This enhances the critic's ability to process path planning information. The core of this method is to maximize the utilization of all available resources to train the critic, allowing the critic to guide the actor network with greater precision. The actor, with the assistance of the vehicle's localization information and visual perception, can then achieve macro-level commands by performing low-level operations like obstacle avoidance and reaching a local destination.

4. Conclusion

Current research typically integrates HD maps as an information source into the end-to-end training process of autonomous driving models, but few studies have explored how to leverage the knowledge properties of HD maps to replace the traditional role of human experts in training agent models as HMITL. This paper expands on the application of HD maps in autonomous driving control and strategy from the perspective of HD maps. It proposes optimizing the agent's training process and sampling efficiency through the prior knowledge of HD maps and artificially constructed features within RL or partial IL based exploratory interactive policy iteration algorithms, achieving better training results with the same policy network architecture under the demands of perception-centric model. The rapid development of HD maps provides a feasible alternative for agent training, avoiding the burden and cost of expert work in HITL.
The main innovations of this paper can be summarized as follows:

a. **HD Map in the Loop Framework:** For different training stages in RL, this paper incorporates various characteristics of HD maps to build a map-based autonomous driving HMITL framework. This framework breaks the traditional limitation of using HD maps solely as an information input source in end-to-end training, broadening its role in autonomous driving training.
b. **Imitation Learning Based on HD Map Expert Demonstrations:** Combining the assistance of HD map-based judgments, this paper introduces Shotgun DAgger. This method allows the agent to fully utilize the prior knowledge provided by HD maps with limited training data, improving the efficiency of IL efficiently. The result can be used as a pretrained model for RL.
c. **Action Intervention and Reward Function Design Based on HD Map Spatial Heterogeneity:** This paper uses guidance lines and special area information from HD maps to intervene and restrict the agent's actions during training. This intervention effectively avoids risks and extreme states that may be encountered in training, reducing the distribution of extreme state samples in the collected data and improving the safety and stability of the agent's training process. By utilizing the spatial information of HD maps, this paper also provides a dense reward shaping function paired with action intervention. Quantifying the geometric features from HD maps helps mitigate the reward sparsity problem caused by qualitative descriptions in traditional reward shaping of RL.
d. **HD Map-Based Prioritized Critic Network Algorithm:** This paper feeds HD map perceptual information, global planning paths, and high-level semantic instructions—obtained via HD maps into the critic network only. The critic, equipped with global perception and HD map features, then guides the actor. While HD maps are used as an information source limited to the critic instead of actor, the final agent policy model can still make decisions independently of HD maps.

In conclusion, the innovative methods presented in this paper, by integrating HD maps at various stages of the RL process, not only

effectively improve training efficiency but also alleviate the expert-dependence problem inherent in existing HITL approaches. This provides new ideas and solutions for efficient training of end-to-end autonomous driving agents.

References

Abel, D., Salvatier, J., Stuhlmüller, A., Evans, O., 2017. Agent-Agnostic Human-in-the-Loop Reinforcement Learning. https://doi.org/10.48550/arXiv.1701.04079

Chen, J., Yuan, B., Tomizuka, M., 2019. Deep Imitation Learning for Autonomous Driving in Generic Urban Scenarios with Enhanced Safety. https://doi.org/10.48550/arXiv.1903.00640

Chen, L., Wu, P., Chitta, K., Jaeger, B., Geiger, A., Li, H., 2024. End-to-End Autonomous Driving: Challenges and Frontiers. IEEE Trans. Pattern Anal. Mach. Intell. 46, 10164–10183. https://doi.org/10.1109/TPAMI.2024.3435937

Chitta, K., Prakash, A., Jaeger, B., Yu, Z., Renz, K., Geiger, A., 2023. TransFuser: Imitation With Transformer-Based Sensor Fusion for Autonomous Driving. IEEE Trans. Pattern Anal. Mach. Intell. 45, 12878–12895. https://doi.org/10.1109/TPAMI.2022.3200245

Codevilla, F., Santana, E., Lopez, A., Gaidon, A., 2019. Exploring the Limitations of Behavior Cloning for Autonomous Driving, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Presented at the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Seoul, Korea (South), pp. 9328–9337. https://doi.org/10.1109/ICCV.2019.00942

Cui, H., Radosavljevic, V., Chou, F.-C., Lin, T.-H., Nguyen, T., Huang, T.-K., Schneider, J., Djuric, N., 2019. Multimodal Trajectory Predictions for Autonomous Driving using Deep Convolutional Networks. https://doi.org/10.48550/arXiv.1809.10732

Cultrera, L., Becattini, F., Seidenari, L., Pala, P., Bimbo, A.D., 2024. Addressing Limitations of State-Aware Imitation Learning for Autonomous Driving. IEEE Trans. Intell. Veh. 9, 2946–2955. https://doi.org/10.1109/TIV.2023.3336063

Droździel, P., Tarkowski, S., Rybicka, I., Wrona, R., 2020. Drivers 'reaction time research in the conditions in the real traffic. Open Engineering 10, 35–47. https://doi.org/10.1515/eng-2020-0004

Fujimoto, S., n.d. Addressing Function Approximation Error in Actor-Critic Methods.

Gajcin, J., Dusparic, I., 2024. Redefining Counterfactual Explanations for Reinforcement Learning: Overview, Challenges and Opportunities. ACM Comput. Surv. 56, 1–33. https://doi.org/10.1145/3648472

Haarnoja, T., Zhou, A., Abbeel, P., Levine, S., n.d. Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor.

Han, Y., Yilmaz, A., 2022. Learning to Drive Using Sparse Imitation Reinforcement Learning. https://doi.org/10.48550/arXiv.2205.12128

Harutyunyan, A., Dabney, W., Mesnard, T., Heess, N., Azar, M.G., Piot, B., van Hasselt, H., Singh, S., Wayne, G., Precup, D., Munos, R., n.d. Hindsight Credit Assignment.

Hester, T., Vecerik, M., Pietquin, O., Lanctot, M., Schaul, T., Piot, B., Horgan, D., Quan, J., Sendonaris, A., Osband, I., Dulac-Arnold, G., Agapiou, J., Leibo, J., Gruslys, A., 2018. Deep Q-learning From Demonstrations. AAAI 32. https://doi.org/10.1609/aaai.v32i1.11757

Hu, Z., Zhang, Y., Xing, Y., Zhao, Y., Cao, D., Lv, C., n.d. Towards Human-Centered Automated Driving: A Novel Spatial-Temporal Vision Transformer-Enabled Head Tracker.

Hua, J., Zeng, L., Li, G., Ju, Z., 2021. Learning for a Robot: Deep Reinforcement Learning, Imitation Learning, Transfer Learning. Sensors 21, 1278. https://doi.org/10.3390/s21041278

Huang, Z., Lv, C., Xing, Y., Wu, J., 2021. Multi-modal Sensor Fusion-Based Deep Neural Network for End-to-end Autonomous Driving with Scene Understanding. IEEE Sensors J. 21, 11781–11790. https://doi.org/10.1109/JSEN.2020.3003121

Ibarz, B., Leike, J., Pohlen, T., Irving, G., Legg, S., Amodei, D., n.d. Reward learning from human preferences and demonstrations in Atari.

Knox, W.B., Glass, B.D., Love, B.C., Maddox, W.T., Stone, P., 2012. How Humans Teach Agents: A New Experimental Perspective. Int J of Soc Robotics 4, 409–421. https://doi.org/10.1007/s12369-012-0163-x

Krening, S., Harrison, B., Feigh, K.M., Isbell, C.L., Riedl, M., Thomaz, A., 2017. Learning From Explanations Using Sentiment and Advice in RL. IEEE Trans. Cogn. Dev. Syst. 9, 44–55. https://doi.org/10.1109/TCDS.2016.2628365

Le Mero, L., Yi, D., Dianati, M., Mouzakitis, A., 2022. A Survey on Imitation Learning Techniques for End-to-End Autonomous Vehicles. IEEE Trans. Intell. Transport. Syst. 23, 14128–14147. https://doi.org/10.1109/TITS.2022.3144867

Li, Q., Peng, Z., Zhou, B., 2022. Efficient Learning of Safe Driving Policy via Human-AI Copilot Optimization. https://doi.org/10.48550/arXiv.2202.10341

Littman, M.L., 2015. Reinforcement learning improves behaviour from evaluative feedback. Nature 521, 445–451. https://doi.org/10.1038/nature14540

Liu, R., Wang, J., Zhang, B., 2020. High Definition Map for Automated Driving: Overview and Analysis. J. Navigation 73, 324–341. https://doi.org/10.1017/S0373463319000638

Lv, C., Cao, D., Zhao, Y., Auger, D.J., Sullman, M., Wang, H., Dutka, L.M., Skrypchuk, L., Mouzakitis, A., 2018. Analysis of autopilot disengagements occurring during autonomous vehicle testing. IEEE/CAA J. Autom. Sinica 5, 58–68. https://doi.org/10.1109/JAS.2017.7510745

Mandlekar, A., Xu, D., Martín-Martín, R., Zhu, Y., Fei-Fei, L., Savarese, S., 2020. Human-in-the-Loop Imitation Learning using Remote Teleoperation. https://doi.org/10.48550/arXiv.2012.06733

Mao, J., Gan, C., Kohli, P., Tenenbaum, J.B., Wu, J., 2019. The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words,

and Sentences From Natural Supervision. https://doi.org/10.48550/arXiv.1904.12584

Maramotti, P., Capasso, A.P., Bacchiani, G., Broggi, A., 2022. Tackling Real-World Autonomous Driving using Deep Reinforcement Learning. https://doi.org/10.48550/arXiv.2207.02162

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., 2015. Human-level control through deep reinforcement learning. Nature 518, 529–533. https://doi.org/10.1038/nature14236

Nair, A., McGrew, B., Andrychowicz, M., Zaremba, W., Abbeel, P., 2018. Overcoming Exploration in Reinforcement Learning with Demonstrations. https://doi.org/10.48550/arXiv.1709.10089

Neary, C., Xu, Z., Wu, B., Topcu, U., 2021. Reward Machines for Cooperative Multi-Agent Reinforcement Learning. https://doi.org/10.5555/3463952.3464063

Nehme, G., Deo, T.Y., 2023. Safe Navigation: Training Autonomous Vehicles using Deep Reinforcement Learning in CARLA. https://doi.org/10.48550/arXiv.2311.10735

Rajeswaran, A., Kumar, V., Gupta, A., Vezzani, G., Schulman, J., Todorov, E., Levine, S., 2018. Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations. https://doi.org/10.48550/arXiv.1709.10087

Ravichandar, H., Polydoros, A.S., Chernova, S., Billard, A., 2020. Recent Advances in Robot Learning from Demonstration. Annu. Rev. Control Robot. Auton. Syst. 3, 297–330. https://doi.org/10.1146/annurev-control-100819-063206

Reddy, S., Dragan, A.D., Levine, S., 2019. SQIL: Imitation Learning via Reinforcement Learning with Sparse Rewards. https://doi.org/10.48550/arXiv.1905.11108

Ross, S., Gordon, G.J., Bagnell, J.A., 2011. A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning. https://doi.org/10.48550/arXiv.1011.0686

Sallab, A.E., Abdou, M., Perot, E., Yogamani, S., 2017. Deep Reinforcement Learning framework for Autonomous Driving. ei 29, 70–76. https://doi.org/10.2352/ISSN.2470-1173.2017.19.AVM-023

Saunders, W., Sastry, G., Stuhlmueller, A., Evans, O., 2017. Trial without Error: Towards Safe Reinforcement Learning via Human Intervention. https://doi.org/10.48550/arXiv.1707.05173

Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., Hassabis, D., 2016. Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489. https://doi.org/10.1038/nature16961

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., Hassabis, D., 2018. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science 362, 1140–1144. https://doi.org/10.1126/science.aar6404

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., Van Den Driessche, G., Graepel, T., Hassabis, D., 2017. Mastering the game of Go without human knowledge. Nature 550, 354–359. https://doi.org/10.1038/nature24270

Sutton, R.S., Barto, A., 2014. Reinforcement learning: an introduction, Nachdruck. ed, Adaptive computation and machine learning. The MIT Press, Cambridge, Massachusetts.

Vecerik, M., Hester, T., Scholz, J., Wang, F., Pietquin, O., Piot, B., Heess, N., Rothörl, T., Lampe, T., Riedmiller, M., 2018. Leveraging Demonstrations for Deep Reinforcement Learning on Robotics Problems with Sparse Rewards. https://doi.org/10.48550/arXiv.1707.08817

Wolf, P., Hubschneider, C., Weber, M., Bauer, A., Hartl, J., Durr, F., Zollner, J.M., 2017. Learning how to drive in a real world simulation with deep Q-Networks, in: 2017 IEEE Intelligent Vehicles Symposium (IV). Presented at the 2017 IEEE Intelligent Vehicles Symposium (IV), IEEE, Los Angeles, CA, USA, pp. 244–250. https://doi.org/10.1109/IVS.2017.7995727

Wu, J., Huang, Z., Hu, Z., Lv, C., 2023. Toward Human-in-the-Loop AI: Enhancing Deep Reinforcement Learning via Real-Time Human Guidance for Autonomous Driving. Engineering 21, 75–91. https://doi.org/10.1016/j.eng.2022.05.017

Xu, Z., Zhang, B., Li, D., Zhang, Z., Zhou, G., Chen, H., Fan, G., 2023. Consensus Learning for Cooperative Multi-Agent Reinforcement Learning. AAAI 37, 11726–11734. https://doi.org/10.1609/aaai.v37i10.26385

Zare, M., Kebria, P.M., Khosravi, A., Nahavandi, S., 2024. A Survey of Imitation Learning: Algorithms, Recent Developments, and Challenges. IEEE Transactions on Cybernetics 1–14. https://doi.org/10.1109/TCYB.2024.3395626

Zeng, W., Luo, W., Suo, S., Sadat, A., Yang, B., Casas, S., Urtasun, R., 2019. End-To-End Interpretable Neural Motion Planner, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, pp. 8652–8661. https://doi.org/10.1109/CVPR.2019.00886

Zhang, X., Ma, H., 2018. Pretraining Deep Actor-Critic Reinforcement Learning Algorithms With Expert Demonstrations. https://doi.org/10.48550/arXiv.1801.10459

Zhang, Z., Liniger, A., Dai, D., Yu, F., Van Gool, L., 2021. End-to-End Urban Driving by Imitating a Reinforcement Learning Coach, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Presented at the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Montreal, QC, Canada, pp. 15202–15212. https://doi.org/10.1109/ICCV48922.2021.01494

Zhu, Y., Wang, Z., Merel, J., Rusu, A., Erez, T., Cabi, S., Tunyasuvunakool, S., Kramár, J., Hadsell, R., Freitas, N. de, Heess, N., 2018. Reinforcement and Imitation Learning for Diverse Visuomotor Skills. https://doi.org/10.48550/arXiv.1802.09564