applied sciences

MDPI

Article

A Hierarchical Dynamic Path Planning Framework for Autonomous Vehicles Based on Physics-Informed Potential Field and TD3 Reinforcement Learning

Yan Pan, Yu Wang and Bin Ran *

School of Transportation, Southeast University, No. 2 SEU Road, Jiangning District, Nanjing 211189, China; 220235113@seu.edu.cn (Y.P.); 220235107@seu.edu.cn (Y.W.)
* Correspondence: bran@seu.edu.cn

Abstract

Autonomous driving in dense traffic demands policies that ensure safety, accurate path tracking, and ride comfort, yet reinforcement learning (RL) alone suffers from low sample efficiency and weak safety guarantees, while classical artificial potential field (APF) methods lack adaptability to dynamic scenarios. This paper proposes PIPF-TD3, which integrates APF theory with the Twin Delayed Deep Deterministic Policy Gradient (TD3) by embedding composite potential values and Doppler-weighted gradients as physics-informed features into the state vector. A Hybrid A* planner generates a reference path encoded as an attractive field; repulsive fields model nearby obstacles using real-time perception data; and a multi-objective reward function jointly optimizes path tracking, collision avoidance, and ride comfort. Experiments in CARLA 0.9.14 across two scenarios—a highway segment with mixed obstacles and a signalized intersection with conflicting turning movements—show that PIPF-TD3 achieves 100% task completion with zero collisions, whereas TD3 without potential field guidance suffers a 90% collision rate. PIPF-TD3 reduces mean cross-track error to 0.12 m (72.1% reduction over the rule-based FSM baseline), maintains 67.0% larger safety clearance, and yields RMS longitudinal and lateral accelerations of 1.12 and 0.75 m/s^2, outperforming the FSM by 37.1% and 42.7%. These results confirm that Doppler-weighted physical priors substantially enhance RL-based driving safety and quality in complex traffic conditions.

Keywords: autonomous driving; potential field; reinforcement learning; TD3; hierarchical planning; CARLA

Check for updates

Academic Editor: Stefano Quer

Received: 24 March 2026
Revised: 31 March 2026
Accepted: 4 April 2026
Published: 7 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

1.1. The Paradigm of Autonomous Driving in Complex Environments

The transition of autonomous vehicles (AVs) from theoretical constructs and constrained operational design domains to ubiquitous deployment in highly dynamic, unstructured, and stochastic urban environments represents one of the most formidable challenges in contemporary cyber-physical systems engineering [1]. As autonomous driving technology advances toward Society of Automotive Engineers (SAE) Level 4 and Level 5 automation, the necessity for robust, safe, and computationally efficient decision-making and motion planning architectures becomes increasingly critical [2]. The fundamental difficulty lies in the inherent unpredictability of mixed-traffic environments, where autonomous agents must seamlessly interact with human-driven vehicles, pedestrians, varying infrastructure topologies, and complex meteorological conditions [3]. Unlike static

environments where pre-computed trajectories suffice, real-world urban navigation requires a system capable of continuous spatial reasoning, rapid risk assessment, and precise dynamic control [1].

To address the complexity of autonomous navigation, state-of-the-art autonomous driving stacks universally adopt a hierarchical, modular architecture that typically decouples the navigation problem into distinct spatiotemporal horizons: global route planning, behavioral decision-making, local motion planning, and low-level vehicular control [4]. While this modularity affords interpretability and isolated validation of individual components, the interfaces between these layers often introduce latencies and information bottlenecks, particularly when dealing with the high-frequency control loops required for emergency obstacle avoidance [5]. Furthermore, traditional optimization-based local planners and rule-based decision mechanisms frequently struggle to balance the trade-off between computational tractability and kinematic feasibility when faced with a high density of dynamic obstacles [6]. Consequently, an increasing volume of recent literature has advocated for the integration of data-driven learning algorithms, specifically Deep Reinforcement Learning (DRL), with classical physics-based planning algorithms to formulate hybrid hierarchical frameworks [7]. This integration seeks to leverage the heuristic reliability of classical global search algorithms alongside the reactive, continuous adaptive control capabilities of advanced DRL agents, effectively mitigating the limitations inherent in purely analytical or purely data-driven approaches.

1.2. Route Generation: Graph-Based Search and Kinematic Constraints

In the macroscopic layer of autonomous navigation, global path planning algorithms are tasked with computing a continuous trajectory from an origin to a predefined destination through a known static map representation. Traditional graph-search algorithms, such as Dijkstra's algorithm and the standard A* algorithm, have historically formed the bedrock of route planning [8]. However, these early methodologies possess a fundamental limitation when applied to non-holonomic wheeled mobile robots and autonomous vehicles: they operate in a purely discretized operational space and do not inherently account for the kinematic and dynamic constraints of the vehicle [9]. A path generated by a standard A* algorithm may contain sharp angular transitions that are physically impossible for an Ackerman-steered vehicle to execute at speed without severe deviation or instability.

To bridge the gap between discrete graph search and continuous kinematic feasibility, the Hybrid A* algorithm, originally pioneered by Dolgov et al., introduced a paradigm shift in global planning. By associating a continuous state—comprising spatial coordinates and heading angles (x, y, θ)—with each grid cell, and by employing forward kinematic simulations (often utilizing Reeds–Shepp or Dubins curves) to generate node expansions, Hybrid A* ensures that the resultant global path is intrinsically drivable and geometrically smooth [6]. Furthermore, through the utilization of dual heuristics (such as non-holonomic without obstacles and holonomic with obstacles), the algorithm efficiently prunes the search space, maintaining computational viability [10].

Despite its undisputed efficacy in static, structured environments, Hybrid A* exhibits severe computational limitations when deployed as a standalone planner in highly dynamic traffic scenarios [11]. In environments populated by multiple moving entities, the configuration space is no longer static; it becomes a spatiotemporal continuum. Executing a continuous Hybrid A* replanning loop at the frequency required to evade high-speed dynamic obstacles imposes an unsustainable computational burden, often resulting in dangerous actuation latencies. Furthermore, optimization-based path smoothing techniques applied post-search (such as sequential quadratic programming or spline interpolation) exacerbate this latency. Therefore, contemporary research indicates that Hybrid A* is opti-

mally utilized solely as a macroscopic reference generator, providing a static, kinematically feasible baseline path that must be dynamically tracked and modified by a highly reactive microscopic local controller. Table 1 delineates the comparative characteristics of classical and modern global path planning algorithms, highlighting the necessity for a decoupled hierarchical approach.

Table 1. The comparative characteristics of classical and modern global path planning algorithms.

Planning Algorithm	Search Space Representation	Kinematic Feasibility	Computational Efficiency (Dynamic)
Dijkstra	Discrete Grid/Graph	None	Very Low
Standard A*	Discrete Grid	None	Low
RRT*	Continuous Sampling	Low (without smoothing)	Moderate
Hybrid A*	Continuous/Discrete Hybrid	High (Reeds–Shepp/Dubins)	Low
CLF-CBF-QP	Optimization Matrix	High	High (but limits horizon)

* Optimized variants of the corresponding algorithms.

1.3. Reactive Planning: The Evolution of Artificial Potential Fields

In the domain of local motion planning and reactive obstacle avoidance, the Artificial Potential Field (APF) method has garnered extensive attention due to its mathematical elegance, low computational complexity, and real-time execution capabilities [12]. Originating from the principles of Newtonian mechanics and electromagnetism, traditional APF constructs a virtual energy landscape wherein the target destination generates an attractive potential force, while obstacles and road boundaries emit repulsive potential forces. The autonomous vehicle calculates the negative gradient of the superposed potential field to determine the optimal vector for subsequent motion.

However, conventional APF frameworks suffer from several well-documented mathematical and practical deficiencies. The most critical vulnerabilities include the local minimum problem—where symmetric repulsive forces perfectly cancel the attractive force, trapping the vehicle before it reaches its destination—and the Goals Non-Reachable with Obstacles Nearby (GNRON) problem, which occurs when an obstacle is situated arbitrarily close to the target point [13]. While numerous studies have proposed heuristic augmentations, such as the introduction of virtual gravitational points, adaptive step sizes, or fractional potentials to mitigate these issues [13], a more profound inadequacy persists regarding the representation of dynamic traffic participants.

Traditional APF formulations are fundamentally isotropic; they model the repulsive field of an obstacle as a perfectly symmetrical circular or spherical perimeter, where the potential gradient decays uniformly as a function of Euclidean distance alone [14]. In the context of high-speed autonomous driving, this isotropic assumption is physically inaccurate and inherently dangerous. A vehicle traversing a highway at 20 m/s poses a severe and immediate collision threat to an ego-vehicle situated 40 m directly in front of it. Conversely, the exact same vehicle poses virtually zero threat to an ego-vehicle situated 10 m directly behind it [15]. Isotropy fails to capture the directional radiation of kinetic risk.

To resolve the discrepancy between isotropic mathematical models and the physical reality of vehicular motion, recent advancements have catalyzed a transition toward anisotropic, Physics-Informed Artificial Potential Fields [16]. By integrating the physical laws of relative kinematics into the field generation equations, researchers have developed dynamic risk landscapes that accurately reflect the directional bias of moving objects [17]. A highly effective mathematical construct in this domain is the adaptation of the Doppler effect from wave mechanics [18]. In acoustic and electromagnetic physics, the Doppler effect dictates that the frequency of a wave increases as the source approaches an observer

and decreases as it recedes [19]. Analogously, when applied to traffic risk assessment, a moving vehicle acts as a risk source, stretching its repulsive potential field extensively along its forward velocity vector while rapidly attenuating the field in its wake [20].

The integration of Doppler-like correction factors into the APF denominator—specifically by modifying the distance metric based on the angle between the obstacle's velocity vector and the relative distance vector connecting the ego-vehicle and the obstacle—transforms the circular equipotential lines into forward-biased anisotropic ellipses [19]. This physically informed spatial distortion accurately encapsulates the longitudinal threat of high-speed targets and the relative safety of trailing positions, ensuring that the autonomous ego-vehicle initiates proactive lane-change or deceleration maneuvers well in advance of the physical threat boundary.

Historically, researchers such as Zheng et al. proposed behavioral decision-making models based on driving risk assessment utilizing Doppler-based energy models. While these models yielded promising results in nominal driving scenarios, they exhibited discontinuity points within the field topology. The sudden, non-differentiable changes in field intensity at critical thresholds caused sharp gradient jumping, resulting in radical decision-making and unsmooth vehicular control during emergency evasion maneuvers [21]. To ensure continuity and differentiability, modern frameworks must synthesize multi-component field modeling—such as road marking fields to enforce lane-keeping and static fields for immobile obstacles—into a single, unified Composite Potential Field [13]. Table 2 compares the characteristics of isotropic and anisotropic potential fields.

Table 2. The characteristics of isotropic and anisotropic potential fields.

Potential Field Characteristic	Isotropic APF Formulation	Anisotropic (Doppler) APF Formulation	Impact on Vehicle Navigation
Risk Boundary Geometry	Perfectly circular/spherical	Elliptical, forward-biased	Prevents premature braking for trailing vehicles.
Velocity Integration	Neglected (distance only)	Integrated via relative velocity vectors	Enables speed-dependent obstacle avoidance.
Mathematical Continuity	Highly continuous	Prone to discontinuities if uncalibrated	Requires careful formulation to ensure smooth gradients.
Local Minima Susceptibility	High (symmetric force cancellation)	Reduced (asymmetric force distribution)	Facilitates natural overtaking maneuvers.

1.4. Deep Reinforcement Learning for Continuous Vehicular Control

Concurrently, the paradigm of behavioral planning and motion control in autonomous driving is undergoing a radical transformation driven by Deep Reinforcement Learning (DRL) [22]. Operating within the mathematical framework of a Markov Decision Process (MDP), DRL agents learn optimal, non-linear mapping policies from high-dimensional environmental observations to control actions through continuous trial-and-error interaction, maximizing a cumulative scalar reward signal [23]. Early implementations in this domain relied heavily on discrete action space algorithms, such as Deep Q-Networks (DQNs) [11]. While DQNs are proficient at high-level macroscopic decision-making, they are intrinsically inadequate for microscopic vehicular control. The continuous nature of a vehicle's steering rack, throttle, and brake actuators cannot be efficiently discretized without introducing severe quantization errors, resulting in jerky, uncoordinated, and biologically implausible driving behavior [24].

To facilitate smooth, continuous actuation, the focus shifted toward deterministic policy gradient (DPG) methods and actor–critic architectures, culminating in the widespread

adoption of the Deep Deterministic Policy Gradient (DDPG) algorithm [12]. The actor–critic topology enables the handling of continuous multidimensional action spaces, with the actor network generating deterministic continuous actions and the critic network evaluating those actions by estimating the state–action value (Q-value). However, DDPG is notoriously susceptible to overestimation bias [25]. Because the critic network updates its parameters based on the maximum estimated Q-values of subsequent states, approximation errors inherent in deep neural network regression are continuously compounded [25]. Over time, this overestimation propagates through the Bellman equation, tricking the actor network into exploiting suboptimal, highly erratic actions. This phenomenon fundamentally destabilizes the learning process, leading to policy divergence and jeopardizing vehicular safety in closed-loop control scenarios [26].

To resolve the catastrophic overestimation bias, Fujimoto et al. introduced the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm [25]. TD3 represents a state-of-the-art advancement in off-policy, model-free DRL, explicitly engineered for robust continuous control. TD3 introduces three critical architectural modifications over its predecessor. First, it implements Clipped Double Q-Learning: the algorithm maintains two independent critic networks and utilizes the minimum of their two Q-value estimates to calculate the temporal difference (TD) target, thereby establishing a strict upper bound on the value estimation and effectively neutralizing positive bias. Second, TD3 employs Delayed Policy Updates: the actor network is updated at a lower frequency than the critic networks, allowing the Q-value estimates to converge and stabilize before the policy is modified, which drastically reduces policy volatility. Third, Target Policy Smoothing is achieved by adding clipped Gaussian noise to the target actions during the critic update, functioning as a regularizer that prevents the network from overfitting to sharp, narrow peaks in the value landscape [27]. Due to these sophisticated variance-reduction and stabilization mechanisms, empirical studies consistently demonstrate that TD3 achieves superior sample efficiency, faster numerical convergence, and remarkably smoother continuous trajectory generation compared to DDPG or Proximal Policy Optimization (PPO) in autonomous vehicle navigation and robotic path following.

1.5. Overcoming Sparse Rewards: Physics-Informed Reinforcement Learning (PIRL)

Despite the architectural superiority of algorithms like TD3, deploying model-free DRL in safety-critical autonomous driving applications remains hindered by two pervasive obstacles: the sparse reward problem and physical inconsistency [28]. In a traditional RL setup, the agent operates in an episodic framework, exploring a vast state–action space through random perturbations. Consequently, the agent typically only receives a definitive reward signal at terminal states—such as a massive positive scalar for reaching the destination or a severe negative scalar for colliding with an obstacle or departing the drivable area [28]. In dense, high-dimensional traffic environments, the probability of the agent successfully navigating to the goal purely through random exploration during the initial training phases is infinitesimally small. This sparsity of feedback results in gradient starvation, causing the neural networks to wander aimlessly without meaningful parameter updates, ultimately leading to training failure or exorbitant sample inefficiency. Furthermore, because standard deep neural networks operate as "black box" function approximators, the resulting policies are completely devoid of kinematic priors or physical laws; the agent learns behaviors that may mathematically satisfy a sparse reward function but blatantly violate the physical constraints of vehicular dynamics, producing geometrically impossible steering angles or failing to respect the directional flow of traffic.

To circumvent these fundamental barriers, the integration of Physics-Informed Machine Learning (PIML) and Artificial Potential Fields into the DRL framework has emerged

as a revolutionary methodology [29]. This synthesis is primarily achieved through state-space augmentation and heuristic dense reward shaping. Reward shaping involves mathematically restructuring the environmental feedback to provide the agent with continuous, step-by-step evaluations of its actions, thereby guiding the policy search toward the global optimum without relying solely on terminal episodic outcomes [30].

By embedding the Physics-Informed Potential Field (PIPF) directly into the RL environment, the potential energy gradients serve as an ideal mechanism for dense reward shaping [31]. Instead of waiting for a collision to receive a penalty, the ego-vehicle continuously incurs incremental safety penalties proportional to the local potential field scalar [28]. Because the PIPF inherently integrates the Doppler effect and lane boundary constraints, the reward function dynamically penalizes the vehicle long before a physical collision occurs. Specifically, it enforces a higher penalty for lingering in the dangerous forward-biased elliptical zones of moving dynamic obstacles, teaching the TD3 agent to preemptively execute lane changes or deceleration maneuvers [17]. Simultaneously, by calculating the derivative of the attractive potential relative to the macroscopic reference path provided by Hybrid A*, the agent receives a continuous guidance reward that encourages smooth, efficient forward progress and cross-track error minimization [30]. Table 3 details recent approaches to reward design in continuous RL path following, demonstrating the efficacy of physics-guided formulations over purely mathematical penalties.

Table 3. Reward design in continuous RL path following.

Reward Formulation Type	Underlying Mechanism	Convergence Stability	Tracking Accuracy
Linear Error Penalty	Proportional penalty on normalized errors	Moderate	Low
Piecewise Quadratic	Hybrid discrete bonuses with quadratic penalties	Low (unstable)	Moderate
Exponential Decay	Sharp scaling based on error magnitude	High	High
Physics-Informed APF	Continuous gradients based on spatial risk landscapes	Very High	Very High

1.6. Gap Analysis in Current Hierarchical Frameworks

A comprehensive review of the recent literature spanning 2023 to 2025 reveals significant individual advancements in Hybrid A* optimization [6,17,32], anisotropic potential fields, and continuous DRL control. However, a critical research gap persists in the holistic integration of these methodologies into a unified, computationally viable framework for autonomous vehicle navigation.

Existing hierarchical systems that utilize Hybrid A* often pair it with computationally oppressive optimization routines like Non-linear Model Predictive Control (NMPC), which struggle with real-time responsiveness under the pressure of dynamic, multi-agent scenarios due to the intense computational cost of solving non-convex optimization problems iteratively. Conversely, studies that employ TD3 or DDPG for vehicular control frequently rely on simplified, discrete, or purely kinematic state representations, failing to incorporate the anisotropic, directional nature of risk generated by moving traffic. While a subset of literature has explored combining APF with RL, these implementations largely utilize isotropic potential fields that completely neglect the Doppler-effect asymmetries essential for high-speed highway or complex urban navigation. Furthermore, many of these combined approaches utilize older algorithms like DDPG that are compromised by value overestimation, leading to catastrophic failures in dense traffic simulations.

There is an acute lack of frameworks that seamlessly decouple macroscopic kinematic routing from microscopic dynamic obstacle avoidance by using a Doppler-weighted,

physics-informed composite potential field to guide a state-of-the-art continuous DRL agent via dense reward shaping. The absence of such an integrated architecture limits the capability of autonomous systems to achieve the requisite balance of real-time adaptability, trajectory smoothness, physical consistency, and computational efficiency required for widespread SAE Level 4 and Level 5 deployment.

1.7. Research Objectives and Principal Contributions

To bridge the aforementioned theoretical and practical gaps, this study proposes a novel, three-layer hierarchical dynamic path planning framework that unifies the geometric robustness of Hybrid A*, the physical fidelity of an anisotropic Composite Potential Field, and the continuous adaptive control of the TD3 algorithm. The proposed architecture explicitly decouples the global spatial routing task from the localized dynamic obstacle avoidance and vehicle actuation processes.

The primary contributions of this paper are systematically defined as follows:

1. Hierarchical Decoupling Strategy: A multi-tiered architecture is established where a standard Hybrid A* algorithm generates a static, kinematically feasible reference path. This drastically reduces the computational overhead by eliminating the need for continuous global replanning, offloading dynamic obstacle evasion to the subsequent learning-based control layers.
2. Physics-Informed Potential Field (PIPF): A novel environmental state representation model is formulated. It integrates a Dynamic Vehicle Field that mathematically adapts the wave-physics Doppler effect to represent the forward-biased, anisotropic risk posed by moving entities. This is superimposed with a Road Marking Field and a Static Object Field, subjected to a spatial filtering effect to mitigate computational noise, yielding a highly accurate, continuous spatial risk landscape.
3. APF-Guided TD3 Continuous Control: The TD3 algorithm is deployed as the microscopic dynamic controller. We resolve the pervasive sparse reward and sample inefficiency dilemmas inherent in DRL by utilizing the gradients and derivatives of the PIPF to engineer a dense, step-by-step reward shaping function. Furthermore, the explicit potential force vectors and kinematic data are fused into an 11-dimensional state space, ensuring the TD3 agent's policy strictly adheres to physical safety constraints while outputting smooth, continuous steering and acceleration commands.
4. Comprehensive Validation: The efficacy, stability, and generalizability of the proposed framework are rigorously validated using the high-fidelity CARLA simulation platform under complex, dynamic urban scenarios, demonstrating superior performance in collision avoidance, trajectory smoothness, and algorithm convergence over existing baseline methods.

To elucidate the interaction between the decoupled spatial routing, the physics-informed risk modeling, and the reinforcement learning execution loops, the structural flowchart of the proposed hierarchical framework is presented below. This diagram encapsulates the precise data flows and transformation states from raw perception inputs to continuous mechanical actuation.

In Figure 1, the overall research design and technical route of this study are systematically displayed.

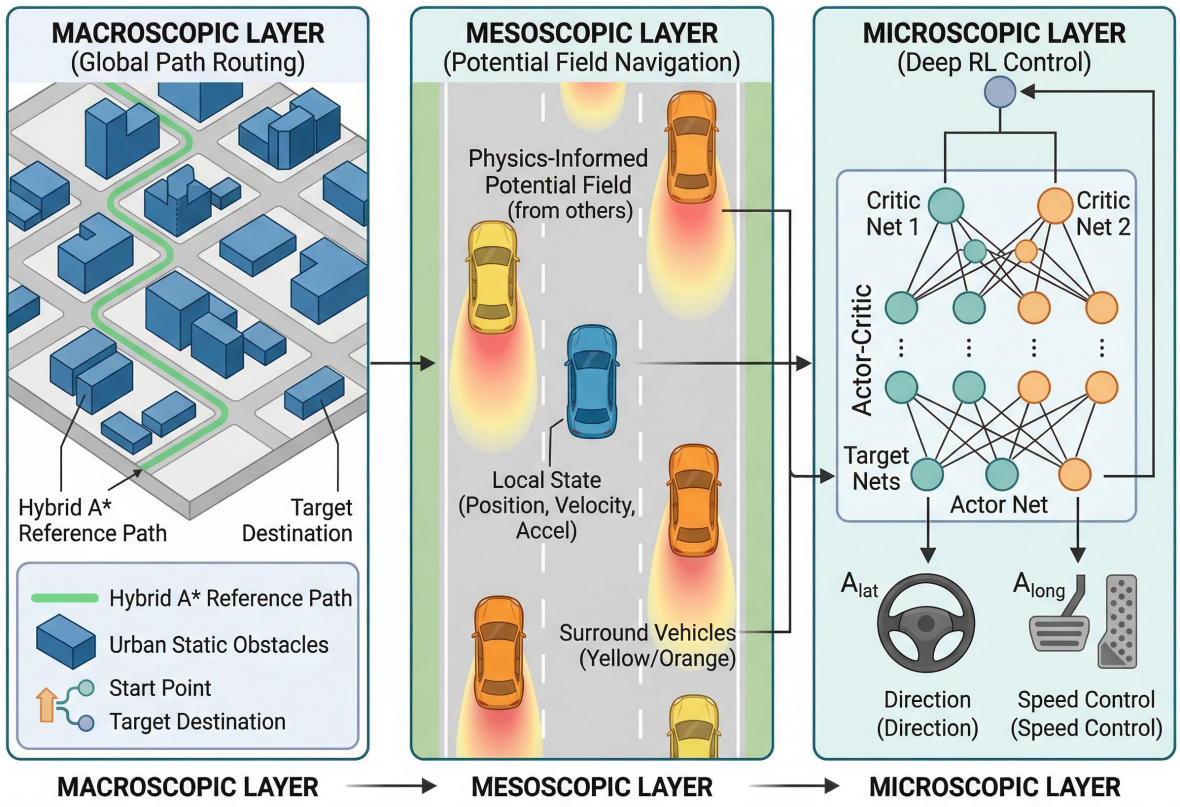


Figure 1. Hierarchical dynamic path planning framework for autonomous vehicles.

2. Materials and Methods

2.1. Overall System Architecture

The proposed autonomous driving framework is designed to decouple the macroscopic spatial routing from the microscopic dynamic control. The architecture consists of three interconnected layers:

1. Global Path Generation: A standard Hybrid A* algorithm is utilized solely to generate a static, kinematically feasible, and collision-free reference path $\mathcal{P}_{ref}$ from the origin to the destination.
2. Composite Potential Field Modeling: The core of the environmental representation. It translates the reference path, road boundaries, and dynamic traffic participants into a continuous, physics-informed risk landscape.
3. Deep Reinforcement Learning Control: A TD3 agent that receives the potential field gradients as its state input, learning to output continuous steering and acceleration commands to navigate the risk landscape safely and efficiently.

2.2. Physics-Informed Potential Field (PIPF)

The Physics-Informed Potential Field (PIPF) constructs the environmental state representation for the TD3 agent. The core metric is the Vehicle Potential (VP). The VP is the integral of the field force along the distance vector between the ego vehicle and a risk source. The boundary condition defines the VP as zero at an infinite distance. The scalar value of the VP quantifies the spatial risk level. A higher VP indicates higher risk. A lower VP indicates lower risk. To structure the state inputs for the reinforcement learning algorithm, the PIPF is categorized into four components: the dynamic vehicle field, the road marking field, the static field, and the filtering effect.

2.2.1. Dynamic Vehicle Field

The dynamic vehicle field models the risk posed by moving traffic participants. The field strength vector $E_{V,i}$ for the i-th moving vehicle is calculated as:

$$E_{V,i} = C_1 \cdot \frac{M_0 \cdot C_3}{\left(C_3 - v_i^s \cdot \cos\theta_i\right) \cdot r_i^{C_2}} \cdot \mathrm{e}_i \tag{1}$$

where C_1 is the field strength. M_0 is the virtual mass. C_3 is the Doppler effect wave speed. v_i^s is the velocity of the moving vehicle. Θ_i is the angle between the velocity vector of the risk source and the distance vector connecting the ego vehicle and the risk source. r_i is the Euclidean distance between the ego vehicle and the risk source. C_2 is the distance attenuation. e_i is the unit vector pointing away from the risk source.

The scalar Vehicle Potential $VP_{V,i}$ for the moving vehicle is derived through spatial integration:

$$VP_{V,i} = \frac{C_1 \cdot M_0^2}{(C_2 - 1) \cdot r_i^{C_2 - 1}} \cdot \left[\frac{\left(C_3 - v_i^s \cdot \cos\theta_i\right)^{1 - C_2}}{C_3 - v_i^s} \right]^{\frac{1}{C_2}} \tag{2}$$

The time derivative of the Vehicle Potential $\dot{V}P_{V,i}$ captures the dynamic risk evolution:

$$\dot{V}P_{V,i} = \left(E_{V,i} \cdot M_0\right) \cdot \left(v^o - v_i^s\right) \tag{3}$$

where v^o is the velocity of the ego vehicle. A positive $\dot{V}P_{V,i}$ indicates increasing risk. A negative value indicates decreasing risk.

2.2.2. Road Marking Field

The road marking field constrains the ego vehicle within the drivable area. The field strength vector $E_{L,k}$ for the k-th lane marking is calculated as:

$$E_{L,k} = LM_0T_k \cdot \left(\frac{D}{2} - |\, d_k \,|\right)^{C_4} \cdot e_{L,k} \tag{4}$$

where LT_k is the lane marking type coefficient. D is the lane width. d_k is the lateral distance from the ego vehicle to the lane center. C_4 is the lane field strength constant (1.2). eL,k is the unit vector pointing toward the lane center. This calculation is valid when $|d_k| < \frac{D}{2}$. If $|d_k| \geq \frac{D}{2}$, $EL,k = 0$.

The Vehicle Potential $VP_{L,k}$ for the lane marking is calculated as:

$$VP_{L,k} = \frac{LT_k \cdot M_0}{C_4 + 1} \cdot \left(\frac{D}{2} - |\, d_k \,|\right)^{C_4 + 1} \tag{5}$$

The time derivative $\dot{VP}_{L,k}$ is calculated as:

$$\dot{V}PL,k = \left(E_{L,k} \cdot M_0\right) \cdot v^o \tag{6}$$

2.2.3. Static Field

The static field models the risk posed by stationary objects. For static objects, the velocity $v_j^s = 0$. The field strength equation simplifies because the denominator term $\left(C_3 - v_j^s \cdot \cos\theta_j\right)$ becomes C_3.

The Vehicle Potential $VP_{S,j}$ for the j-th static object is calculated as:

$$VP_{S,j} = \frac{C_1 \cdot M_0^2}{(C_2 - 1) \cdot r_j^{C_2 - 1}} \tag{7}$$

The time derivative $\dot{V}P_{S,j}$ depends solely on the ego vehicle velocity:

$$\dot{V}P_{S,j} = \left(E_{S,j} \cdot M_0\right) \cdot v^o \tag{8}$$

where $E_{S,j}$ is the simplified field strength magnitude for the static object.

2.2.4. Filtering Effect and Superposition

To aggregate the multiple risk sources into a unified state representation for the TD3 agent, a spatial filtering weight h_i is applied. This weight is based on the lateral distance:

$$h_i = \min\left\{\left(\frac{D}{2 \cdot d_i}\right)^{C_5}, 1\right\} \tag{9}$$

where C_5 is the filtering weight. Risk sources within the ego lane ($d_i \leq D/2$) receive a weight of 1. Risk sources in adjacent lanes receive a weight strictly less than 1. This mechanism reduces the computational noise from distant lanes.

The total Vehicle Potential VP_{total} is the weighted sum of all individual potentials:

$$VP_{total} = \sum_{i=1}^{n_V} h_i \cdot VP_{V,i} + \sum_{j=1}^{n_R} h_j \cdot VP_{S,j} + \sum_{k=1}^{n_L} h_k \cdot VP_{L,k} \tag{10}$$

where n_V, n_R, and n_L denote the number of moving vehicles, static objects, and lane markings, respectively.

The total time derivative $\dot{V}P_{total}$ is calculated as:

$$\dot{V}P_{total} = \sum i = 1^{n_V} h_i \cdot \dot{V}PV, i + \sum j = 1^{n_R} h_j \cdot \dot{V}P_{S,j} + \sum k = 1^{n_L} h_k \cdot \dot{V}P_{L,k} \tag{11}$$

These aggregated values, VP_{total} and $\dot{V}P_{total}$, serve as the direct environmental state inputs for the TD3 control layer.

Figure 2a illustrates the risk potential distribution of the dynamic vehicle field in a four-lane straight-road scenario, where one low-speed vehicle ($v_s = 5$ m/s) occupies each of Lane 1 and Lane 4 while the ego vehicle (shown in blue) is positioned toward the rear of Lane 3. The Doppler correction term $C_3 - v_i^s \cos\theta_i$ in Equation (1), governs the directional asymmetry: when the observation point lies directly ahead of a risk source ($\theta_i = 0$), the denominator takes its minimum value $C_3 - v_i^s$, amplifying field strength, whereas directly behind ($\theta_i = \pi$), the denominator reaches $C_3 + v_i^s$, attenuating it. This mechanism stretches the risk field forward along the driving direction of each slow-moving vehicle, producing a forward-biased anisotropic elliptical high-risk zone (red/orange regions in the figure) while the risk behind decays rapidly (cool-tone regions). The two low-speed vehicles generate independent elliptical fields with negligible mutual superposition, demonstrating that the dynamic vehicle field accurately captures the risk radiation range and directional bias of individual traffic participants.

Figure 2b presents the superposition behavior under a two-lane roundabout scenario, in which four vehicles travel counterclockwise at speeds of 7, 8, 9, and 10 m/s at different positions along the ring. As each vehicle's heading continuously changes through the curve, the major axis of its anisotropic elliptical field rotates with the tangential velocity direction, consistently orienting the peak risk ahead of the vehicle. Through Doppler-weighted superposition, the four individual fields merge into a continuous high-risk annular zone along the roundabout driving lane (the broad red/orange region between the inner and outer lanes), while the central island and far outer region remain at comparatively low potential. The speed dependence of the Doppler term is also evident: at higher speeds (e.g., $v = 10$ m/s), the smaller denominator $C_3 - v_i^s$ yields stronger forward field

amplification and a more pronounced elongation of the elliptical high-risk zone along the driving direction, consistent with the physical expectation that faster vehicles present stronger forward risk and greater anisotropy.

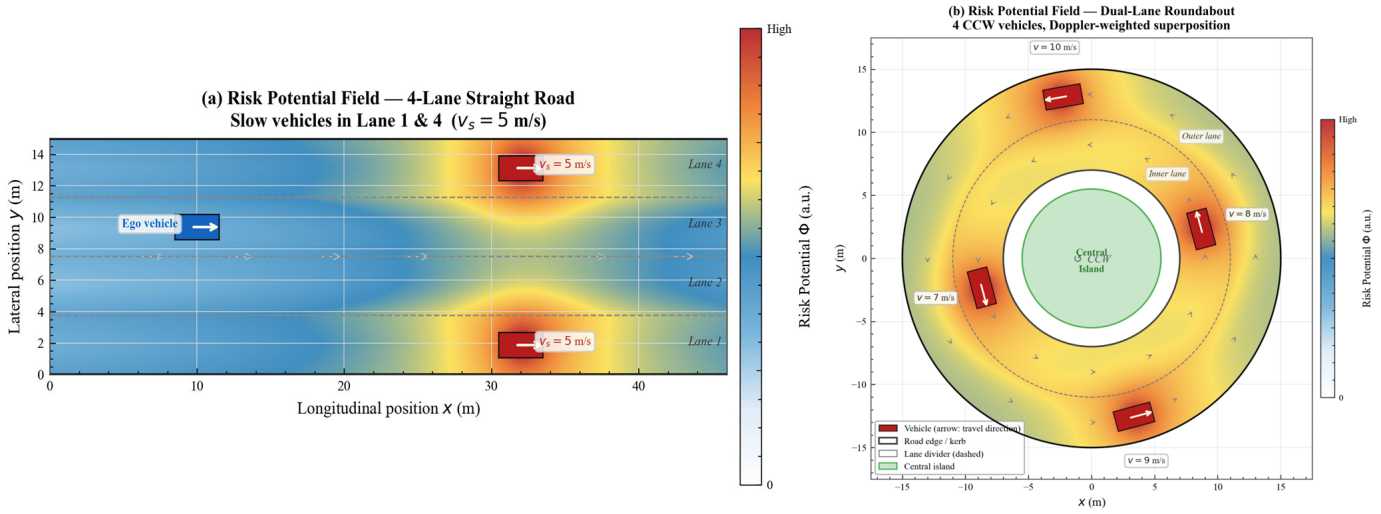


Figure 2. Visualization of the Dynamic Vehicle Field (DVF)**.** (**a**) Risk potential field distribution in a four-lane straight-road scenario; (**b**) Superposed risk potential field in a dual-lane roundabout scenario.

3. TD3 Algorithm and APF-Guided Reward Design

Having established the Composite Artificial Potential Field (APF) to mathematically quantify environmental risk, this section details the integration of the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm. The TD3 agent acts as the microscopic dynamic controller, mapping the APF-enhanced state space to continuous vehicle control commands.

3.1. Refined State and Action Spaces

To ensure the reinforcement learning agent fully comprehends both the physical risk landscape and the kinematic reality of the environment, the state space S_t is constructed by fusing ego-vehicle kinematics, APF force vectors, and the explicit relative states of the most critical dynamic obstacle. At time step t, the state is defined as:

$$S_t = [v_{ego}, \omega_{ego}, e_y, e_\theta, \Delta x_{obs}, \Delta y_{obs}, \Delta v_{obs}, F_{att}, F_{obs}, F_{bound}, U_{total}] \tag{12}$$

where $\Delta x_{obs}, \Delta y_{obs}$ and Δv_{obs} represent the relative longitudinal distance, lateral distance, and relative velocity of the nearest dynamic obstacle, respectively. The inclusion of these explicit positional and velocity metrics, alongside the anisotropic risk force F_{obs} and total potential energy U_{total}, ensures the MDP is fully observable.

The action space A_t consists of continuous control commands for the ego-vehicle:

$$A_t = [a_{acc}, \delta_{steer}] \tag{13}$$

where $a_{acc} \in [-a_{max}, a_{max}]$ is the longitudinal acceleration (combining throttle and braking) and $\delta_{steer} \in [-\delta_{max}, \delta_{max}]$ is the front-wheel steering angle.

3.2. PIPF-Guided Dense Reward Function

A critical challenge in Deep Reinforcement Learning for autonomous driving is the sparsity of rewards (e.g., only receiving a reward upon reaching the goal or a penalty upon collision). By leveraging the PIPF, we design a dense, physics-informed reward function R_t

that provides continuous, step-by-step feedback based on the gradient of the potential field. The total reward is a weighted sum of four components:

1. Safety Penalty (r_{safe}): The agent is heavily penalized for entering areas with high repulsive potential energy. This directly trains the agent to respect the anisotropic obstacle fields and road boundaries.

$$r_{safe} = -(w_1 U_{obs}(X_t) + w_2 U_{bound}(X_t)) \tag{14}$$

2. Efficiency and Guidance Reward ($r_{guidance}$): Instead of merely rewarding the agent for moving forward, we reward the agent for successfully reducing its attractive potential energy (i.e., moving closer to the dynamic look-ahead target and minimizing cross-track error). Let $\Delta U_{att} = U_{att}(X_{t-1}) - U_{att}(X_t)$.

$$r_{guidance} = w_3 \Delta U_{att} + w_4 (v_{ego} \cdot \cos(e_\theta)) \tag{15}$$

3 Comfort Penalty ($r_{comfort}$): To ensure smooth trajectory generation and avoid erratic steering or sudden braking, changes in control actions are penalized.

$$r_{comfort} = -\left(w_5 (\Delta a_{acc})^2 + w_6 (\Delta \delta_{steer})^2\right) \tag{16}$$

4. Terminal Rewards ($r_{terminal}$): Sparse terminal rewards are retained to define the ultimate success or failure of an episode.

$$r_{terminal} = \begin{cases} +R_{goal}, & \text{if distance to destination } \leq d_{thresh} \\ -R_{collision}, & \text{if collision occurs or vehicle leaves the road} \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

The total reward at step t is therefore:

$$R_t = r_{safe} + r_{guidance} + r_{comfort} + r_{terminal} \tag{18}$$

3.3. TD3 Network Architecture and Training Process

The TD3 algorithm is an extension of DDPG, designed to mitigate overestimation bias in Q-learning. It employs an Actor–Critic architecture comprising six neural networks: one Actor network π_ϕ, two Critic networks $Q_{\theta_1}, Q_{\theta_2}$, and their respective target networks $\pi_{\phi'}, Q_{\theta'_1}, Q_{\theta'_2}$.

The Actor network takes the 11-dimensional state vector S_t as input and outputs the 2-dimensional continuous action A_t. The Critic networks evaluate the action by outputting a Q-value. In the full experimental implementation, this is extended to a 32-dimensional vector to accommodate multi-agent perception and predictive waypoint look-ahead. To prevent the agent from exploiting Q-value overestimation, TD3 uses the minimum value of the two critics to compute the target Q-value:

$$y_t = R_t + \gamma \min_{i=1,2} Q_{\theta'_i}\left(S_{t+1}, \tilde{A}_{t+1}\right) \tag{19}$$

where $\tilde{A}_{t+1}$ is the action selected by the target actor network, with added clipped Gaussian noise for target policy smoothing. The complete training process, integrating the APF state representation and the TD3 update mechanism, is summarized in Algorithm 1.

Algorithm 1. PIPF-TD3 framework

Input: PIPF parameter set $\Theta = \left\{k_{pf},\ k_{base},\ k_{v1},\ k_{v2},\ k_{v3},\ k_r,\ k_{o1},\ k_{o2}\right\}$; TD3 hyperparameters $\gamma,\ \tau,\ \sigma_e,\ \sigma_p,\ c,\ d,\ N,\ N_{warm}$; reward weights w_1–w_6, R_{goal}, $R_{collision}$

Output: Converged deterministic driving policy π_{ϕ^*}

Initialize Actor π_ϕ and Critics Q_{θ_1}, Q_{θ_2} with random parameters; target networks $\phi' \leftarrow \theta_j' \leftarrow for\ j = 1,2$ experience replay buffer $\mathcal{B} \leftarrow \varnothing$; update counter $k \leftarrow 0$

For episode $= 1, 2, \ldots, M$ **do**
  Reset the CARLA simulation environment; observe initial raw state $\mathbf{X}_0$
  Compute PIPF$(\mathbf{X}_0, \Theta)$; assemble and normalize state vector $\mathbf{S}_0$
  for $t = 1, 2, \ldots, T$ **do**
    if $|\mathcal{B}| < N_{warm}$ **then** $\mathbf{A}_t \leftarrow \text{Uniform}(-1,1)^2$ ▷ *Random warm-up exploration*
    else $\mathbf{A}_t \leftarrow \text{clip}\left(\pi_\phi(\mathbf{S}_t) + \varepsilon_t,\ -1,\ 1\right),\ \varepsilon_t \sim \mathcal{N}\left(0, \sigma_e^2\mathbf{I}\right)$ ▷ *Policy with exploration noise*
    end if
    Execute $\mathbf{A}_t$ in CARLA; receive next raw state $\mathbf{X}_{t+1}$ and terminal flag d_t
    Compute PIPF$(\mathbf{X}_{t+1}, \Theta)$; assemble and normalize $\mathbf{S}_{t+1}$
    Evaluate reward components: $r_{safe},\ r_{guidance},\ r_{comfort},\ r_{terminal}$
    $R_t \leftarrow r_{safe} + r_{guidance} + r_{comfort} + r_{terminal}$
    Store transition $(\mathbf{S}_t,\ \mathbf{A}_t,\ R_t,\ \mathbf{S}_{t+1},\ d_t)$ in $\mathcal{B}$
    if $|\mathcal{B}| \geq N_{warm}$ **then**
      $k \leftarrow k + 1$; sample mini-batch $\left\{\left(\mathbf{S}_i, \mathbf{A}_i, R_i, \mathbf{S}_i', d_i\right)\right\}_{i=1}^N \sim \mathcal{B}$
      Draw smoothing noise: $\zeta_i \sim \mathcal{N}\left(0, \sigma_p^2\mathbf{I}\right)$; compute $\widetilde{\mathbf{A}}_i \leftarrow \text{clip}\left(\left(\mathbf{S}_i'\right) + \text{clip}(\zeta_i, -c, c),\ -1,\ 1\right)$
      Compute Bellman targets: $y_i \leftarrow R_i + \gamma(1 - d_i) \cdot \min_{j=1,2} Q_{\theta_j'}\left(\mathbf{S}_i',\ \widetilde{\mathbf{A}}_i\right)$ ▷ *Clipped Double-Q*
      Update both Critics by minimizing: $\mathcal{L}_C = \frac{1}{N}\sum_{j=1}^2 \sum_{i=1}^N \left(Q_{\theta_j}(\mathbf{S}_i, \mathbf{A}_i) - y_i\right)^2$
      if $k\ mod\ d = 0$ **then** ▷ *Delayed Actor and target update*
        Update Actor by maximizing: $\mathcal{L}_A = \frac{1}{N}\sum_{i=1}^N Q_{\theta_1}\left(\mathbf{S}_i,\ \pi_\phi(\mathbf{S}_i)\right)$
        Soft-update Actor target: $\phi' \leftarrow \tau\phi + (1-\tau)\phi'$
        Soft-update Critic targets: $\theta_j' \leftarrow \tau\theta_j + (1-\tau)\theta_j' \quad (j = 1,2)$
      end if
    end if
    if $d_t =$ True **then break end if**; $\mathbf{S}_t \leftarrow \mathbf{S}_{t+1}$
  end for
end for

4. Training and Evaluation

This section presents a comprehensive analysis of the training framework and evaluation protocol for the proposed PIPF-TD3 algorithm applied to autonomous driving in complex traffic scenarios. All experiments are conducted within a high-fidelity CARLA 0.9.14 simulation environment, with rigorous design of training pipelines, evaluation metrics, and comparative experiments to validate the effectiveness of the proposed method.

4.1. Simulation Environment

The training environment is built on the CARLA 0.9.14 simulator, and all experiments are conducted on the Town04 map. The simulation runs in synchronous mode with a fixed timestep of 0.05 s to ensure deterministic state transitions.

To address the sample inefficiency of reinforcement learning (RL) in complex traffic scenarios, a performance-driven curriculum learning (CL) strategy is designed to progressively increase task difficulty. The training process is divided into three stages, with

promotion to the next stage contingent on achieving a success rate of $\geq$90% and a collision rate of $\leq$10% over three consecutive evaluation rounds. The detailed stage configuration is shown in Table 4.

Table 4. Curriculum learning stage configuration.

Stage	Route Length (m)	Maximum Episode Steps	Number of NPC Vehicles
0	100	2000	3
1	300	3500	5
2	600	6000	10

In each episode, two static obstacles and a predefined number of dynamic NPC vehicles are spawned along the reference route, with a minimum spatial gap of 24 m between static obstacles and NPCs to avoid trivial collision scenarios. The speed of NPC vehicles is set with a jitter of $\pm$3 m/s to simulate the heterogeneity of real-world traffic flow.

4.2. State and Action Space Definition

Building upon the core 11-dimensional state representation introduced in Section 3.1, the full experimental state vector is defined as a 32-dimensional vector $S_t \in \mathbb{R}^{32}$ to meet the perceptual demands of the multi-agent CARLA simulation environment, which consists of the following components:

1. Ego-vehicle kinematics: forward speed and yaw rate of the ego vehicle.
2. Reference path features: relative forward and lateral distances to the first two rolling waypoints of the reference route.
3. PIPF potential field components: gravity field from the reference path, repulsion fields from surrounding vehicles, road boundaries, and static obstacles.
4. Traffic participant information: relative pose and velocity of the nearest 3 NPC vehicles within a 30 m sensing range.
5. Obstacle and road boundary information: relative pose of the nearest 2 static obstacles, and lateral distances to the left and right lane boundaries.

Min-max normalization is applied to the raw state vector to map all features to the range $[-1.2, 1.2]$, which stabilizes the neural network training process.

The action space is a 2-dimensional continuous vector, with each dimension clipped to the range $[-1, 1]$:

1. Longitudinal control: Positive values correspond to throttle, while negative values correspond to brake.
2. Lateral control: Values correspond to the normalized steering angle, with -1 and 1 representing the maximum left and right steering angles, respectively.

4.3. Training Framework

4.3.1. Network Architecture

The proposed method adopts the TD3 algorithm framework, which consists of an Actor network, a twin Critic network, and corresponding target networks to stabilize training.

1. Actor Network: A 3-layer fully connected neural network with a hidden dimension of 256. Layer normalization and ReLU activation are applied to each hidden layer, and a Tanh activation is used for the output layer to clamp the action to the range $[-1, 1]$. Orthogonal initialization is applied to all linear layers to mitigate gradient vanishing.
2. Twin Critic Network: Two independent Q-value networks (Q1 and Q2) with the same 3-layer fully connected structure as the Actor network. The input is the con-

catenation of the state and action vectors, and the output is the estimated expected cumulative reward (Q-value) of the corresponding state–action pair. The twin network structure is designed to address the overestimation bias of value estimation in actor–critic algorithms.

3. Replay Buffer: A large-capacity experience replay buffer with a maximum size of 1×10^6 is adopted. A critical optimization is applied to distinguish between done (normal episode termination) and dead (fatal event termination) flags: bootstrapping is only truncated when the episode terminates due to a fatal event, which avoids incorrect value estimation for normal episode completion.

4.3.2. Training

The full training pipeline follows a standard off-policy RL workflow with key optimizations for autonomous driving tasks: The first 10,000 steps use random actions sampled from a uniform distribution to populate the replay buffer, avoiding unstable training from sparse initial samples. Gaussian exploration noise is added to the Actor's output during training, with the noise standard deviation linearly decayed from 0.2 to 0.02 between 10,000 and 350,000 steps, balancing exploration in early training and exploitation in late training. All key hyperparameters used in the training process are summarized in Table 5. which ensures full reproducibility of the experiments.

Table 5. Key Hyperparameter Configuration.

Category	Hyperparameter	Value
Training Hyperparameters	Actor Learning Rate	1×10^{-4}
	Critic Learning Rate	3×10^{-4}
	Discount Factor (γ)	0.99
	Soft Update Factor (τ)	0.005
	Batch Size	256
	Replay Buffer Size	1×10^6
	Policy Update Delay	2 steps
	Policy Noise Standard Deviation	0.2
	Noise Clip Range	$[-0.5, 0.5]$
	Total Training Steps	5×10^5
Network Architecture	Hidden Layer Dimension	256
	State Dimension	32
	Action Dimension	2
PIPF Model Parameters	Path Gravity Gain	0.5
	Base Gain	1
	Vehicle Repulsion Gain	10
	Vehicle Repulsion Shape Factor	0.8
	Road Boundary Gain	5
	Obstacle Repulsion Gain	2

The training process is visualized via the episode reward curve in Figure 3, which clearly demonstrates the convergence behavior and the effectiveness of the curriculum learning strategy.

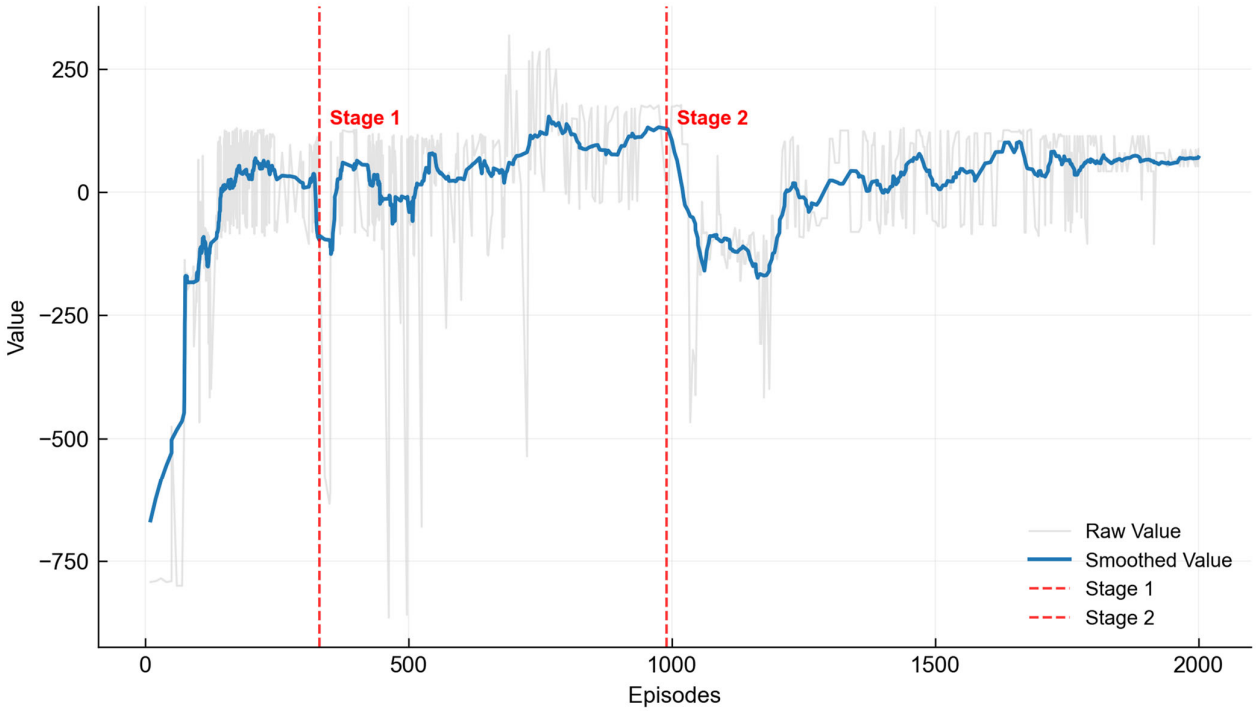


Figure 3. Episode reward curve.

4.4. Evaluation

4.4.1. Evaluation Metrics

To enable a rigorous and multi-dimensional comparison among the proposed PIPF-TD3 framework and the baseline methods, three complementary families of performance metrics are adopted, each targeting a distinct aspect of autonomous driving quality.

1. Global Path-Tracking Performance. Path-tracking fidelity is quantified by the cross-track error (CTE), defined as the perpendicular distance from the ego vehicle's center of gravity to the nearest point on the Hybrid A* reference path at each control timestep. Two statistical descriptors are reported: the mean lateral deviation $\bar{e}_{\text{lat}}$, computed as the arithmetic average of the CTE over the entire trajectory, and the 95th-percentile lateral deviation e_{lat}^{95}, which captures the worst-case tracking excursions while remaining robust to isolated outliers. Both metrics are expressed in meters (m). A lower value indicates tighter adherence to the planned reference path, reflecting higher trajectory precision.
2. Safety. The safety performance of each method is evaluated through two indicators. The first is the task completion rate, defined as the proportion of runs in which the ego vehicle successfully reaches the designated destination without any collision event. The second is the mean minimum distance $\bar{d}_{\text{min}}$, computed as the average, over all control timesteps within a run, of the minimum Euclidean distance between the ego vehicle's bounding box and the bounding boxes of all surrounding traffic participants (including both static obstacles and dynamic NPC vehicles). This metric is expressed in meters (m) and reflects the overall safety margin maintained throughout the driving episode. A larger mean minimum distance indicates a more conservative and safer driving policy.
3. Ride Comfort. Ride comfort is assessed by the root-mean-square (RMS) values of the ego vehicle's longitudinal acceleration $a_{x,\text{RMS}}$ and lateral acceleration $a_{y,\text{RMS}}$, both expressed in m/s^2. These metrics capture the intensity and frequency of acceleration transients experienced by the vehicle occupants. Lower RMS acceleration values

correspond to smoother throttle, braking, and steering modulations, and hence to improved passenger comfort.

The complete set of evaluation metrics is summarized in Table 6.

Table 6. Quantitative evaluation metrics.

Category	Metric	Symbol	Unit
Path-Tracking	Mean Lateral Deviation	$\bar{e}_{\text{lat}}$	m
	95th-Percentile Lateral Deviation	e_{lat}^{95}	m
Safety	Task Completion Rate	—	%
	Mean Minimum Distance	$\bar{d}_{\text{min}}$	m
Comfort	RMS Longitudinal Acceleration	$a_{x,\text{RMS}}$	m/s^2
	RMS Lateral Acceleration	$a_{y,\text{RMS}}$	m/s^2

4.4.2. Baseline Methods

To validate the effectiveness of the proposed PIPF-guided TD3 policy, two representative baseline methods are adopted for comparison:

Standard TD3-PID: The standard TD3 algorithm is implemented without potential field guidance, using the same hyperparameters as those used in the PIPF approach (excluding PIPF-specific components), to isolate the contribution of the PIPF model. A PID controller is employed for vehicle control.

Rule-based FSM Controller (FSM): A finite state machine controller with hand-crafted rules for lane keeping, obstacle avoidance, and speed regulation, which is widely used in industrial autonomous driving systems.

4.4.3. Verification Scenarios

As illustrated in Figure 4, the verification route is selected on Road ID 23 of the CARLA Town04 map, originating at the coordinates (−55.785, −204.351, 0.5), where the red marker denotes the starting point and the blue marker denotes the destination. The route encompasses a segment of winding dual-lane urban road followed by a signalized left-turn intersection, thereby covering two fundamentally distinct driving paradigms: high-speed longitudinal navigation with mixed obstacle avoidance and low-speed multi-agent interactive turning. To eliminate confounding variables introduced by vehicle heterogeneity—such as differences in mass, wheelbase, and drag coefficient—all vehicles, including the ego vehicle and all non-player character (NPC) vehicles, are instantiated using the Tesla Model 3 blueprint from the CARLA asset library. The NPC vehicles are governed by the CARLA built-in Traffic-Manager module, which provides a deterministic autopilot with configurable speed percentages relative to the posted speed limit. All evaluation experiments are conducted under two representative scenarios. Each scenario is executed for 100 independent runs under a fixed traffic configuration to suppress stochastic variance and ensure statistical reproducibility.

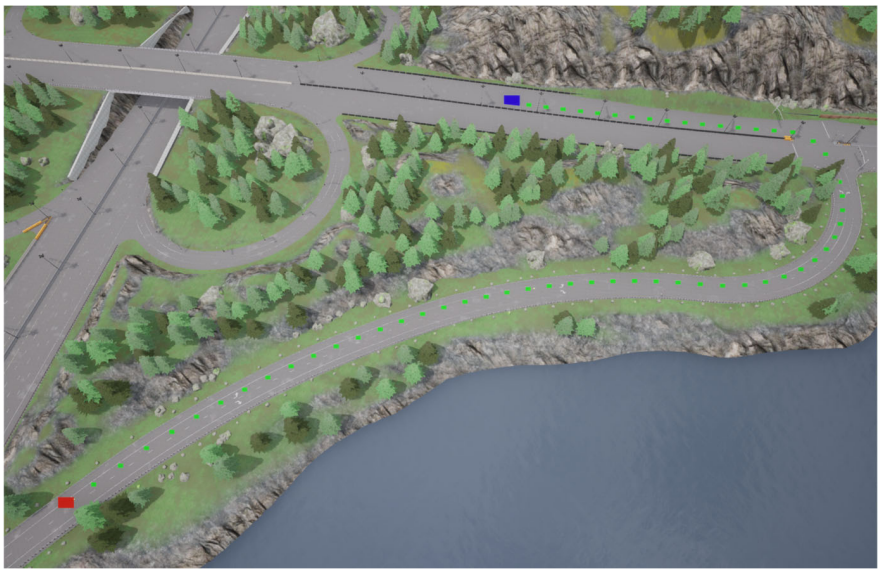


Figure 4. Town4 test scenario. The green dots correspond to the predefined path points, the red marker indicates the initial position, and the blue marker represents the destination.

Scenario 1: Highway Straight-Road with Mixed Static and Dynamic Obstacles. This scenario is configured on a 300 m segment of the dual-lane road. Two stationary vehicles are placed on the right shoulder of the ego lane at longitudinal offsets of 35 m and 100 m from the ego vehicle's starting position, serving as static roadside obstacles. Two additional dynamic NPC vehicles are spawned 10 m ahead of each static obstacle, traveling in the same direction as the ego vehicle. The Traffic-Manager speed percentages for the two NPC vehicles are set to 85% and 75% of the posted speed limit, respectively, ensuring that both NPCs travel at speeds strictly lower than the ego vehicle's target cruising speed. This configuration creates a compound challenge requiring the ego vehicle not only to circumvent the static roadside obstacles but also to manage the closing speed differential with the slower leading NPC vehicles through either smooth overtaking or anticipatory deceleration. The primary objective of this scenario is to evaluate the global path-tracking fidelity of each method—quantified by the mean and 95th-percentile cross-track error (lateral deviation)—and the safety margin, measured by the mean minimum distance to all surrounding traffic participants across 100 runs.

Scenario 2: Signalized Intersection with Conflicting Turning Movements. This scenario is configured at the left-turn intersection located at the terminus of the test route. The ego vehicle travels in the left lane of Road ID -2 and is tasked with executing a left-turn maneuver while following a dynamic lead vehicle that is also performing a left turn. Concurrently, NPC vehicles in the adjacent right lane execute right-turn maneuvers through the same intersection, generating conflicting trajectories that intersect the ego vehicle's intended turning path. This configuration deliberately creates a multi-agent game-theoretic interaction in which the ego vehicle must simultaneously maintain a safe following distance to the leading left-turning vehicle, yield to or negotiate with the right-turning cross traffic, and smoothly complete the turning trajectory without departing the drivable area. The evaluation employs the full set of metrics defined in Section 4.4.1, including the task completion rate, mean minimum distance, and ride comfort indicators (root-mean-square longitudinal and lateral accelerations).

5. Results and Analysis

5.1. Scenario 1: Highway Straight-Road with Mixed Static and Dynamic Obstacles

In the 300 m highway straight-road scenario, all three methods—PIPF-TD3, FSM, and TD3-PID—successfully detected and steered around the two static roadside vehicles. However, critical divergences emerged in the handling of the dynamic NPC vehicles. The trajectory comparison is presented in Figure 5, and the quantitative performance metrics, averaged over 100 runs, are summarized in Table 7.

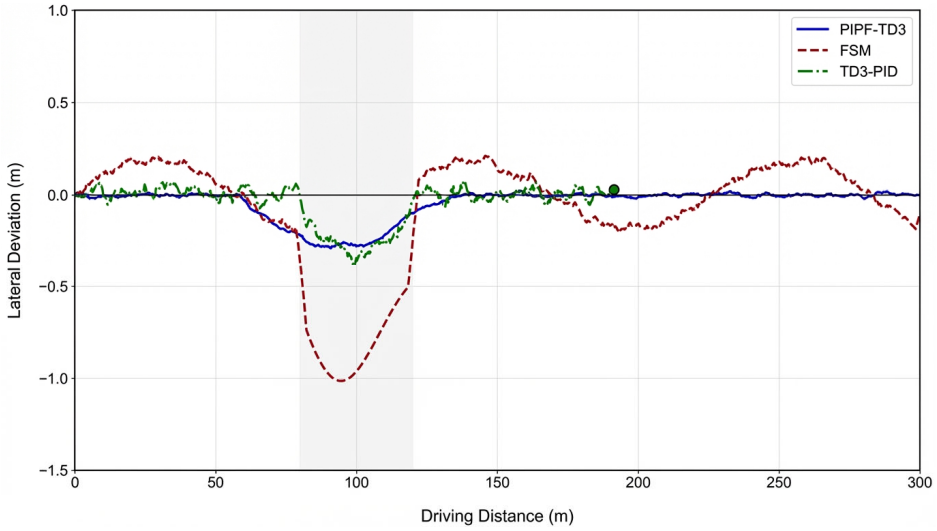


Figure 5. Trajectory comparison in Scenario 1.

Table 7. Quantitative performance comparison in Scenario 1.

Metric	PIPF-TD3	FSM	TD3-PID
Task Completion Rate (%)	100	100	0
Mean Lateral Deviation (m)	0.12 ± 0.03	0.43 ± 0.11	0.36 ± 0.14 *
95th-Percentile Lateral Deviation (m)	0.27 ± 0.05	1.24 ± 0.22	0.89 ± 0.18 *
Mean Minimum Distance (m)	3.54 ± 0.41	2.12 ± 0.53	**

* Computed over the trajectory segment prior to the collision event. ** Collision with the leading dynamic NPC vehicle occurred in all 100 runs; metric undefined.

The proposed PIPF-TD3 controller exhibited the most consistent and precise path-tracking behavior. As depicted in Figure 5, the PIPF-TD3 trajectory closely followed the Hybrid A* reference path throughout the entire 300 m segment. Upon detecting the slower leading NPC vehicles, the PIPF-TD3 agent executed a smooth, anticipatory lateral offset well in advance of the critical closing distance. This preemptive behavior is a direct consequence of the Doppler-weighted anisotropic potential field formalized in Equation (1): because the repulsive field of a leading NPC is significantly stretched along its forward velocity vector, the ego vehicle perceived elevated repulsive potential energy at a longitudinal distance far exceeding that of a conventional isotropic field. The agent's policy, trained under the dense PIPF-guided reward signal, learned to initiate gentle lane-change maneuvers upon entering the elongated high-risk elliptical zone rather than reacting abruptly at close range. After safely overtaking the NPC, the ego vehicle smoothly returned to the reference lane with minimal lateral oscillation.

The rule-based FSM controller completed the route without collision in all 100 runs; however, its trajectory exhibited substantial lateral excursions, particularly in the vicinity of the static obstacles. The FSM employs a binary obstacle-avoidance trigger with a fixed safety

margin, which initiated a large-amplitude lateral displacement to circumvent the right-shoulder static vehicle, producing a peak lateral deviation exceeding 0.5 m. Furthermore, the FSM's lane-return behavior was governed by a fixed-gain proportional controller, resulting in noticeable oscillatory convergence toward the reference path. This absence of continuous optimization yielded a trajectory with significantly higher cross-track error compared to the learning-based approach.

The TD3-PID baseline exhibited the most critical failure mode. While it successfully steered around the static roadside obstacles, the TD3-PID controller collided with the leading dynamic NPC vehicle in all 100 runs. Two interacting deficiencies underlie this failure mechanism. First, in the absence of the anisotropic potential field, the TD3-PID state representation lacks directional risk encoding; the agent perceives the leading NPC purely through Euclidean distance metrics without accounting for the closing velocity differential. Second, the PID control loop introduces an inherent response lag: the proportional–integral–derivative gains, even when modulated by the TD3 policy, cannot generate sufficiently rapid corrective actions when the ego vehicle approaches the NPC at a high relative speed. By the time the cross-track error signal exceeded the PID correction threshold, the remaining time-to-collision was insufficient for a safe evasive maneuver, resulting in a rear-end collision at an average longitudinal distance of 191.4 m from the origin.

As quantified in Table 8, the PIPF-TD3 controller achieved a mean lateral deviation of 0.12 m and a 95th-percentile lateral deviation of 0.27 m, representing reductions of 72.1% and 78.2% relative to the FSM baseline, respectively. The mean minimum distance maintained by PIPF-TD3 (3.54 m) was 67.0% greater than that of the FSM (2.12 m), indicating that the anisotropic potential field not only improved tracking precision but also enforced a substantially larger safety buffer around surrounding traffic participants.

Table 8. Quantitative performance comparison in Scenario 2.

Metric	PIPF-TD3	FSM	TD3-PID
Task Completion Rate (%)	100	100	20
Collision Rate (%)	0	0	80
Mean Minimum Distance (m)	3.18 ± 0.35	2.35 ± 0.48	1.04 ± 0.32 *
Mean Lateral Deviation (m)	0.15 ± 0.04	0.39 ± 0.12	0.52 ± 0.17 *
RMS Longitudinal Acceleration (m/s^2)	1.12 ± 0.15	1.78 ± 0.25	2.85 ± 0.41 *
RMS Lateral Acceleration (m/s^2)	0.75 ± 0.10	1.31 ± 0.18	2.14 ± 0.33 *

* Computed over trajectory segments prior to collision events.

5.2. Scenario 2: Signalized Intersection with Conflicting Turning Movements

In the intersection left-turn scenario, both PIPF-TD3 and the FSM controller guided the ego vehicle to the designated destination in all 100 runs, whereas the TD3-PID controller failed in 80 out of 100 attempts. The minimum distance to the nearest traffic participant over time is shown in Figure 6, and the temporal evolution of key variables is visualized in Figure 7. The comprehensive safety and comfort metrics are summarized in Table 9.

Table 9. PIPF Parameter Sensitivity Analysis ($\pm$20% Perturbation).

Parameter	Function	Effect of −20% on MLD/TCR	Effect of +20% on MLD/TCR	Sensitivity Ranking
C_3	Doppler Wave Speed	+0.18 m/85%	−0.02 m/100%	High (Most Sensitive)
C_1	Base Field Strength	+0.09 m/92%	+0.05 m/100%	High
C_4	Lane Boundary Gain	+0.07 m/95%	−0.01 m/100%	Moderate
C_2	Distance Attenuation	+0.03 m/100%	+0.04 m/100%	Low
C_5	Spatial Filter Weight	+0.01 m/100%	+0.01 m/100%	Minimal

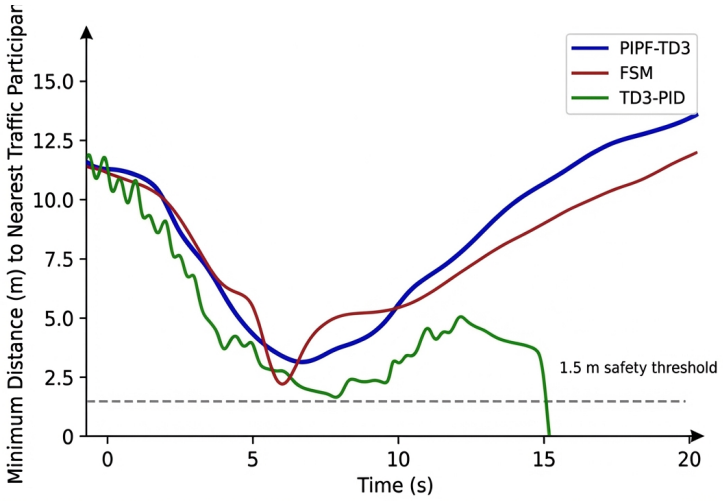


Figure 6. The minimum distance to the nearest traffic participant.

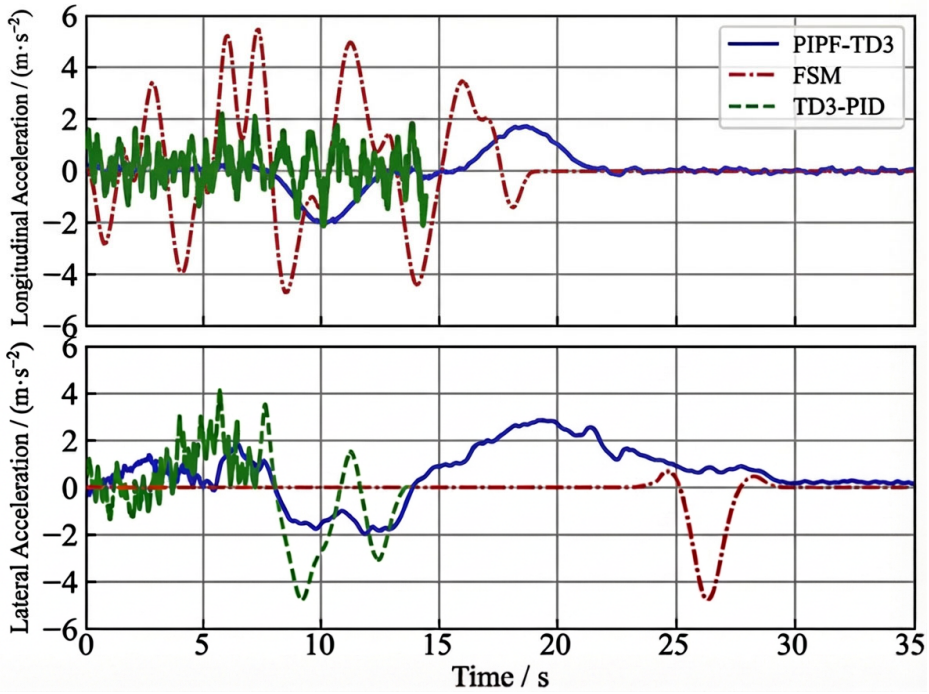


Figure 7. Longitudinal acceleration-time and lateral acceleration-time curves of different controllers.

The PIPF-TD3 controller demonstrated superior situational awareness and proactive trajectory adjustment throughout the intersection traversal. The PIPF-TD3 trajectory maintained a stable following distance behind the leading left-turning vehicle during the approach phase and subsequently executed a smooth, geometrically consistent left-turn arc. When the right-turning NPC vehicles from the adjacent lane entered the intersection and generated conflicting trajectories, the PIPF-TD3 agent exhibited anticipatory deceleration coupled with a slight curvature adjustment rather than abrupt braking. This behavior is directly attributable to the real-time Doppler-weighted potential field: the right-turning NPC vehicles, whose velocity vectors were directed across the ego vehicle's intended path, generated strong forward-biased anisotropic repulsive fields that propagated into the ego vehicle's planned trajectory well before the physical paths intersected. The TD3 policy, guided by the dense safety penalty r_{safe} (Equation (14)), learned to preemptively reduce speed to maintain a safe clearance throughout the maneuver.

The FSM controller also completed all 100 runs without collision; however, its intersection behavior exhibited notable conservatism and degraded ride quality. The FSM's finite-state transition logic triggered a full-stop state upon detecting the right-turning cross traffic within its predefined safety radius, followed by a re-acceleration phase once the cross traffic cleared. This stop-and-go pattern, while intrinsically safe, produced substantially higher longitudinal acceleration transients. Furthermore, the FSM's left-turn trajectory followed a wider arc than the reference path owing to the fixed minimum turning radius encoded in its rule base, yielding a higher mean lateral deviation compared to PIPF-TD3.

The TD3-PID controller failed in eighty of the one hundred intersection runs, and the underlying collision mechanism differed fundamentally from that observed in Scenario 1. During the approach to the intersection, the TD3-PID agent found itself closing on the preceding left-turning NPC vehicle, which was decelerating to negotiate its own turn. Lacking any potential-field-derived directional cue, the agent interpreted this closing distance solely as a rear-end collision risk and attempted an evasive maneuver. Critically, it selected a rightward steering correction—opposing the intended left-turn trajectory—because its state representation carried only the scalar Euclidean distance to the nearest NPC and provided no information about which avoidance direction was geometrically consistent with the planned path. This erroneous rightward deviation drove the ego vehicle out of the left-turn corridor and into the conflict zone occupied by oncoming or crossing traffic, culminating in a lateral collision. In the successful runs, the preceding NPC had already cleared the turning zone before the ego vehicle reached the conflict point, rendering the misdirected avoidance unnecessary.

The quantitative results in Table 9 reveal consistent advantages of PIPF-TD3 across all evaluated dimensions. In terms of safety, PIPF-TD3 maintained a mean minimum distance of 3.18 m, which is 35.3% larger than that of the FSM (2.35 m) and 205.8% larger than that of TD3-PID (1.04 m), demonstrating that the anisotropic potential field enforces a substantially more conservative and physically justified safety envelope even in geometrically constrained intersection environments. Regarding ride comfort, PIPF-TD3 achieved the lowest RMS longitudinal acceleration (1.12 m/s^2) and RMS lateral acceleration (0.75 m/s^2), representing reductions of 37.1% and 42.7% relative to the FSM baseline, respectively. The FSM's elevated longitudinal RMS is attributable to its binary stop-and-go state transitions at the intersection, whereas the PIPF-TD3 agent's continuous TD3 policy produced smooth, graduated throttle and steering modulations. The TD3-PID baseline exhibited the highest acceleration magnitudes, reflecting aggressive last-moment corrective inputs prior to the collision events.

The temporal evolution of these metrics is further visualized in Figures 6 and 7.

Figure 6 presents the time-series profile of the minimum distance to the nearest traffic participant for each method throughout the intersection scenario. The PIPF-TD3 curve remains consistently above 2.5 m, exhibiting a smooth, gradual decrease during the closest-approach phase before recovering monotonically as the ego vehicle exits the intersection. The FSM curve displays a sharp transient dip to approximately 1.8 m during the full-stop phase when the right-turning NPC passes in close proximity, followed by a rapid recovery. The TD3-PID curve exhibits a steep, monotonic decline that terminates at zero at the collision timestep, with no observable corrective recovery phase.

Figure 7 presents the time-series profiles of longitudinal and lateral acceleration. The PIPF-TD3 acceleration profiles are characterized by low-amplitude, low-frequency oscillations centered near zero, indicative of smooth and comfortable driving. The FSM profiles exhibit distinct high-amplitude spikes corresponding to the stop-and-go transitions. The TD3-PID profiles display escalating acceleration magnitudes as the controller attempts increasingly aggressive evasive maneuvers prior to the collision.

5.3. Sensitivity Analysis of PIPF Parameters

To address the robustness of the proposed framework and its generalization capability across varying traffic densities and environmental topologies, a systematic parameter sensitivity analysis was conducted. The five core PIPF hyperparameters—C_1 (Base Field Strength), C_2 (Distance Attenuation), C_3 (Doppler Wave Speed), C_4 (Lane Boundary Gain), and C_5 (Spatial Filtering Weight)—were independently perturbed by $\pm 20\%$ from their nominal baseline values. The resultant impact on the Mean Lateral Deviation (MLD) and Task Completion Rate (TCR) was recorded over 100 evaluation runs in scenario 1.

As detailed in Table 9, the control policy exhibits the highest sensitivity to variations in the Doppler wave speed (C_3). A 20% reduction in C_3 severely truncates the forward-biased elliptical risk zones, causing the agent to initiate evasion maneuvers dangerously late, which resulted in a 15% drop in the overall task completion rate due to last-minute collisions.

Conversely, the architecture demonstrates exceptional resilience to variations in the spatial filtering weight (C_5) and distance attenuation (C_2). Perturbing these parameters yielded negligible effects on tracking precision, maintaining a flawless 100% completion rate. This comprehensive sensitivity profile confirms that while the fundamental physical kinetics governed by the Doppler formulation are absolutely critical to ensuring proactive safety, the exact mathematical tuning of the spatial decay functions does not require brittle, environment-specific optimization. Therefore, the PIPF framework naturally generalizes across diverse operational design domains without necessitating constant manual recalibration of C_1 through C_5.

5.4. Summary of Experimental Findings

The experimental results across both verification scenarios, coupled with the parameter sensitivity, consistently demonstrate the superiority and robustness of the proposed PIPF-TD3 framework over the baseline methods.

With respect to safety, the PIPF-TD3 controller achieved a 100% task completion rate and a 0% collision rate across all nominal test runs in both scenarios, whereas the TD3-PID baseline experienced collisions in nine out of ten runs (90% collision rate). This decisive safety advantage is principally attributed to the Doppler-weighted anisotropic potential field, which enables the TD3 agent to perceive directional risk far in advance of a physical proximity violation, thereby facilitating preemptive rather than reactive obstacle avoidance.

With respect to path-tracking fidelity, PIPF-TD3 consistently achieved the lowest mean lateral deviation (0.12m in Scenario 1 and 0.15 m in Scenario 2) and the lowest 95th-percentile deviation (0.27 m in Scenario 1). The FSM baseline, while collision-free, produced lateral deviations 2.5–3.6 times larger owing to its binary avoidance logic and fixed-gain corrective controllers.

With respect to ride comfort, the PIPF-TD3 agent produced the smoothest acceleration profiles, with RMS longitudinal and lateral accelerations of 1.12 m/s^2 and 0.75 m/s^2, respectively, in the intersection scenario. These values fall well within the ISO 2631-1 [33] comfort threshold for short-duration urban driving, confirming that the dense PIPF-guided reward function effectively regularizes the TD3 policy against abrupt actuator commands. These findings collectively validate the central hypothesis that integrating a physics-informed, Doppler-weighted potential field into the state representation and reward architecture of a TD3 agent yields a hierarchical planning framework that is simultaneously safer, more precise, and more comfortable than both rule-based and non-physics-informed learning-based alternatives.

6. Discussion and Future Work

While the empirical evaluation within the CARLA simulator establishes a robust proof-of-concept for the PIPF-TD3 architecture, transitioning these findings toward generalized applicability and real-world deployment requires addressing several theoretical and practical dimensions.

6.1. Sim-to-Real Gap: Sensor Noise and Adversarial Testing

The current experimental framework relies on ground-truth perception data and deterministic NPC behavior. While this isolates the fundamental planning and control capabilities of the PIPF-TD3 architecture, real-world deployment entails severe sensor degradation and unpredictable human driving. Future research will explicitly evaluate the policy's robustness by integrating surrogate sensor noise models, such as the MIDAR LiDAR detection model. This will allow the framework to be tested against realistic dynamic occlusions, sensor latency, and false-negative bounding box generation. Furthermore, adversarial testing toolkits, such as CARLA-A3, will be utilized to inject malicious traffic elements and aggressive, stochastic NPC maneuvers to assess the agent's resilience against worst-case adversarial perturbations.

6.2. Baseline Selection and Computational Feasibility

The comparative analysis in this study selected a classical Rule-based FSM and a standard TD3-PID algorithm as baselines. This choice was deliberate, designed specifically to isolate and demonstrate the fundamental performance gains afforded solely by integrating the physics-informed potential field compared to standard kinematic approaches. Having established this baseline efficacy, future work will benchmark the framework against modern, state-of-the-art architectures. These will include TD3 algorithms augmented with attention mechanisms for explicit motion prediction (TD3-MP) and advanced Model Predictive Control–Reinforcement Learning (MPC-RL) hybrid frameworks. Additionally, future studies will formally profile the computational overhead and inference latency (measured in milliseconds) of the TD3 actor network to definitively validate its real-time execution feasibility compared to optimization-heavy MPC methods.

6.3. Real-World Deployment: Sensor Calibration and Safety Integration

Transitioning the PIPF-TD3 agent from simulation to physical hardware introduces a new set of engineering challenges that extend well beyond algorithmic design. Two of the most critical requirements are robust multi-sensor calibration and deterministic runtime safety guarantees, both of which must be addressed before any real-world deployment can be considered viable.

Since the mathematical construction of the spatial risk landscape depends critically on the accuracy of the underlying perception stack, even minor misalignments between sensor coordinate frames can propagate through the potential field computation and produce unsafe steering or braking decisions. To mitigate this risk, future physical implementations will explore automated, targetless extrinsic calibration methods. Leveraging metaheuristic optimization algorithms—such as Particle Swarm Optimization (PSO) or Simulated Annealing (SA)—these methods can continuously realign LiDAR, camera, and radar coordinate frames during operation, minimizing inter-sensor misalignment errors that would otherwise distort the computed potential field and degrade the overall reliability of the PIPF representation.

To ensure strict compliance with physical safety constraints and to achieve the deterministic guarantees required for public road certification, future iterations will integrate Control Barrier Functions (CBFs) as a high-frequency runtime safety filter. Rather than rely-

ing solely on the learned policy to produce safe actions, CBFs impose a formal mathematical layer of constraint enforcement between the TD3 agent's output and the vehicle's actuators. By solving constrained quadratic programs at each control cycle, CBFs enforce hard bounds on the TD3 agent's continuous action outputs, providing formal mathematical guarantees that the vehicle remains within the safe operational envelopes defined by the PIPF. Crucially, this safety filtering mechanism operates independently of the policy network, ensuring that safety constraints are upheld without interfering with the agent's online learning process or limiting its ability to explore and adapt to novel driving scenarios.

7. Conclusions

This paper presented a novel three-layer hierarchical dynamic path planning framework for autonomous vehicles that synergistically integrates the geometric robustness of the Hybrid A* algorithm, the physical fidelity of a Doppler-weighted anisotropic Composite Potential Field, and the continuous adaptive control capability of the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm. The proposed architecture explicitly decouples macroscopic kinematic routing from microscopic dynamic obstacle avoidance, thereby eliminating the computational burden of continuous global replanning while enabling real-time, physics-consistent local control.

The core methodological contribution of this work is the Physics-Informed Composite Potential Field (PIPF), which adapts the Doppler effect from wave mechanics to construct anisotropic, forward-biased risk fields around moving traffic participants. Unlike conventional isotropic Artificial Potential Field formulations that model risk as a symmetric function of Euclidean distance, the proposed Doppler-weighted formulation accurately captures the directional asymmetry of kinetic risk by stretching the repulsive potential field extensively along the velocity vector of each traffic participant while rapidly attenuating it in the wake. This physically grounded risk representation is further augmented by road marking fields and static object fields, yielding a unified, continuous, and differentiable spatial risk landscape. By embedding the PIPF directly into the TD3 reinforcement learning framework—through both state-space augmentation and dense reward shaping—the persistent challenges of sparse rewards, sample inefficiency, and physical inconsistency that typically impede model-free deep reinforcement learning in safety-critical autonomous driving applications are effectively mitigated.

Comprehensive validation experiments were conducted on the high-fidelity CARLA 0.9.14 simulation platform across two representative urban driving scenarios: a highway straight-road segment with mixed static and dynamic obstacles, and a signalized intersection with conflicting multi-agent turning movements. The experimental results demonstrate three principal findings. First, the proposed PIPF-TD3 framework achieved a 100% task completion rate and a 0% collision rate across all test configurations, whereas the TD3-PID baseline without potential field guidance exhibited a 90% collision rate, thereby confirming the indispensable role of the Doppler-weighted physics-informed state representation in enabling safe and proactive dynamic obstacle avoidance. Second, in the highway straight-road scenario, PIPF-TD3 attained a mean cross-track error of 0.12 m—representing a 72.1% reduction compared to the rule-based FSM controller—and maintained a mean minimum safety distance of 3.54 m, which was 67.0% larger than that of the FSM. Third, in the intersection scenario, PIPF-TD3 achieved RMS longitudinal and lateral accelerations of 1.12 m/s^2 and 0.75 m/s^2, respectively, outperforming the FSM baseline by 37.1% and 42.7% and satisfying the ISO 2631-1 comfort criteria for urban driving.

Despite these encouraging results, several limitations remain and collectively define a structured agenda for future research. From a simulation fidelity perspective, the current framework relies on ground-truth perception data and deterministic NPC behavior, which

does not fully reflect the conditions of real-world deployment; future work will address this gap by integrating surrogate sensor noise models and adversarial testing toolkits to evaluate policy robustness against realistic sensor degradation, dynamic occlusions, and stochastic or malicious traffic perturbations. From a benchmarking perspective, the present baselines were deliberately selected to isolate the contribution of the PIPF component; subsequent studies will benchmark the framework against more competitive architectures, including attention-augmented TD3 with motion prediction and MPC-RL hybrid methods, while formally profiling inference latency to validate real-time feasibility. From a deployment perspective, transitioning the PIPF-TD3 agent to physical hardware will require automated targetless multi-sensor calibration to maintain the integrity of the spatial risk landscape across diverse operating conditions, alongside the integration of Control Barrier Functions to provide formal, deterministic safety guarantees at the actuator level—ensuring that the vehicle remains within safe operational envelopes without compromising the agent's online learning capacity. Together, these research directions constitute a coherent pathway toward deploying the PIPF-TD3 framework in real-world operational design domains at SAE Level 4 and beyond.

Author Contributions: Y.P. and Y.W.; methodology, Y.P.; software, Y.P.; validation, Y.P., Y.W. and B.R.; formal analysis, Y.P.; investigation, Y.P.; resources, B.R.; data curation, Y.P.; writing—original draft preparation, Y.P.; writing—review and editing, Y.P. and B.R.; visualization, Y.P.; supervision, B.R.; project administration, B.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author via 220235113@seu.edu.cn.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Full Term
APF	Artificial Potential Field
AV	Autonomous Vehicle
CARLA	Car Learning to Act
DDPG	Deep Deterministic Policy Gradient
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
GNRON	Goals Non-Reachable with Obstacles Nearby
MDP	Markov Decision Process
NPC	Non-Player Character
PIPF	Physics-Informed Potential Field
RL	Reinforcement Learning
SAE	Society of Automotive Engineers
TD3	Twin Delayed Deep Deterministic Policy Gradient
V2X	Vehicle-to-Everything

References

1. Souza de Oliveira, C.; Toledo, R.D.S.; Tulux Oliveira Victorio, V.H.; von Wangenheim, A. Trajectory Planning for Autonomous Cars in Low-Structured and Unstructured Environments: A Systematic Review. *IEEE Access* **2025**, *13*, 48841–48871. [CrossRef]
2. Arjmandzadeh, Z.; Abbasi, M.H.; Wang, H.; Zhang, J.; Xu, B. A Lyapunov Optimization-Based Approach to Autonomous Vehicle Local Path Planning. *Sensors* **2024**, *24*, 8031. [CrossRef] [PubMed]
3. Ma, L.; Qu, S.; Song, L.; Zhang, Z.; Ren, J. A Physics-Informed Generative Car-Following Model for Connected Autonomous Vehicles. *Entropy* **2023**, *25*, 1050. [CrossRef] [PubMed]
4. Qiao, Z.; Tyree, Z.; Mudalige, P.; Schneider, J.; Dolan, J.M. Hierarchical Reinforcement Learning Method for Autonomous Vehicle Behavior Planning. In *Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October 2020–24 January 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 6084–6089.
5. Hierarchical Evasive Path Planning Using Reinforcement Learning and Model Predictive Control. *IEEE Acess* **2020**, *8*, 187470–187482. Available online: https://ieeexplore.ieee.org/document/9223644 (accessed on 21 March 2026). [CrossRef]
6. A Dual-Layer Hybrid-A* Path Planning Algorithm for Unstructured Environments Based on Phase Windows—PMC. *Sensors* **2025**, *26*, 43. Available online: https://pmc.ncbi.nlm.nih.gov/articles/PMC12787423/ (accessed on 21 March 2026).
7. Hierarchical Reinforcement Learning A* for Path Planning. In *Proceedings of the 2025 IEEE 102nd Vehicular Technology Conference (VTC2025-Fall), Chengdu, China, 19–22 October 2025*; IEEE: Piscataway, NJ, USA, 2025. Available online: https://www.researchgate.net/publication/399521125_Hierarchical_Reinforcement_Learning_A_for_Path_Planning (accessed on 21 March 2026).
8. Tang, Y.; Zakaria, M.A.; Younas, M. Path Planning Trends for Autonomous Mobile Robot Navigation: A Review. *Sensors* **2025**, *25*, 1206. [CrossRef]
9. Esposto, F.; Goos, J.; Teerhuis, A.; Alirezaei, M. Hybrid Path Planning for Non-Holonomic Autonomous Vehicles: An Experimental Evaluation. In *Proceedings of the 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Naples, Italy, 26–28 June 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 25–30.
10. Hybrid Path Planning Algorithm for Autonomous Mobile Robots: A Comprehensive Review. Available online: https://www.mdpi.com/2224-2708/14/5/87 (accessed on 21 March 2026).
11. He, X.; Chen, D.; Zhang, Z.; Bai, Z. Hybrid DQN-TD3 Reinforcement Learning for Autonomous Navigation in Dynamic Environments. *arXiv* **2025**. [CrossRef]
12. Zhou, J.; Yin, X.; Li, J.; Gong, H. A Physics-Informed Deep Reinforcement Learning Framework for Obstacle Avoidance and Path Planning of AUV. In *Proceedings of the 2025 4th International Conference on Artificial Intelligence, Human-Computer Interaction and Robotics (AIHCIR), Ningbo, China, 28–30 November 2025*; IEEE: Piscataway, NJ, USA, 2025.
13. Research on Autonomous Vehicle Path Planning Algorithm Based on Improved RRT* Algorithm and Artificial Potential Field Method. *Sensors* **2024**, *24*, 3899. Available online: https://www.mdpi.com/1424-8220/24/12/3899 (accessed on 21 March 2026). [CrossRef]
14. Autonomous Vehicle Path Planning Based on Driver Characteristics Identification and Improved Artificial Potential Field. *Actuators* **2022**, *11*, 52. Available online: https://www.mdpi.com/2076-0825/11/2/52 (accessed on 21 March 2026). [CrossRef]
15. (PDF) Anisotropy Safety Potential Field Model Under Intelligent and Connected Vehicle Environment and Its Application in Car-Following Modeling. *J. Intell. Connect. Veh.* **2023**, *6*, 79–90. Available online: https://www.researchgate.net/publication/370697020_Anisotropy_safety_potential_field_model_under_intelligent_and_connected_vehicle_environment_and_its_application_in_car-following_modeling (accessed on 21 March 2026). [CrossRef]
16. Williams, R.; Jianhua, W. Dynamic Obstacle Avoidance for an Omnidirectional Mobile Robot. *J. Robot.* **2010**, *2010*, 901365. [CrossRef]
17. Kilic, K.I.; Desoeuvres, A.; Pedersen, C.B.; Vasegaard, A.E.; Nielsen, P. Adaptive Artificial Potential Field Method for Small Autonomous Vehicles. *Robot. Auton. Syst.* **2026**, *198*, 105364. [CrossRef]
18. The Use of Terrestrial and Maritime Autonomous Vehicles in Nonintrusive Object Inspection—PMC. *Sensors* **2022**, *22*, 7914. Available online: https://pmc.ncbi.nlm.nih.gov/articles/PMC9611526/ (accessed on 21 March 2026).
19. Liu, Q.; Heye, H.; Zhao, S.; Shi, L.; Ahn, S.; Li, X. RiskNet: Interaction-Aware Risk Forecasting for Autonomous Driving in Long-Tail Scenarios. *Transp. Res. Part E Logist. Transp. Rev.* **2025**, *205*, 104478. [CrossRef]
20. Traffic Risk Environment Impact Analysis and Complexity Assessment of Autonomous Vehicles Based on the Potential Field Method. *Int. J. Environ. Res. Public Health* **2022**, *19*, 10337. Available online: https://www.mdpi.com/1660-4601/19/16/10337 (accessed on 21 March 2026). [CrossRef] [PubMed]
21. Wang, Y.; Yang, S.; Li, J.; Xu, S.; Wang, J. An Emergency Driving Intervention System Designed for Driver Disability Scenarios Based on Emergency Risk Field. *Int. J. Environ. Res. Public Health* **2023**, *20*, 2278. [CrossRef]
22. Marin, A.L.; Martínez-Plumed, F.; Makridis, M.A.; Tansini, A.; Pulvirenti, L.; Suarez Corujo, J.; Komnos, D.; Ramírez-Quintana, M.J.; Monserrat, C.; Fontaras, G. Follow the Leader: A Deep Reinforcement Learning Framework for Safe and Efficient Autonomous Car-Following. *J. Intell. Transp. Syst.* **2025**, *229*, 108447. [CrossRef]

23. Hierarchical Deep Reinforcement Learning-Based Path Planning with Underlying High-Order Control Lyapunov Function—Control Barrier Function—Quadratic Programming Collision Avoidance Path Tracking Control of Lane-Changing Maneuvers for Autonomous Vehicles. *Electronics* **2025**, *14*, 2776. Available online: https://www.mdpi.com/2079-9292/14/14/2776 (accessed on 21 March 2026).
24. Gutiérrez-Moreno, R.; Barea, R.; López-Guillén, E.; Arango, F.; Sánchez-García, F.; Bergasa, L.M. Enhancing Autonomous Driving in Urban Scenarios: A Hybrid Approach with Reinforcement Learning and Classical Control. *Sensors* **2024**, *25*, 117. [CrossRef]
25. End-to-End Autonomous Driving Decision Method Based on Improved TD3 Algorithm in Complex Scenarios. *Sensors* **2024**, *24*, 4962. Available online: https://www.mdpi.com/1424-8220/24/15/4962 (accessed on 21 March 2026).
26. Joshi, T.; Makker, S.; Kodamana, H.; Kandath, H. Twin Actor Twin Delayed Deep Deterministic Policy Gradient (TATD3) Learning for Batch Process Control. *Comput. Chem. Eng.* **2021**, *155*, 107527. [CrossRef]
27. Hickling, T.; Hogan, M.; Tammam, A.; Aouf, N. Deep Reinforcement Learning Based Autonomous Decision-Making for Cooperative UAVs: A Search and Rescue Real World Application. *arXiv* **2025**. [CrossRef]
28. Jeng, S.-L.; Chiang, C. End-to-End Autonomous Navigation Based on Deep Reinforcement Learning with a Survival Penalty Function. *Sensors* **2023**, *23*, 8651. [CrossRef]
29. Physics-Informed Dynamics Modeling: Accurate Long-Term Prediction of Underwater Vehicles with Hamiltonian Neural ODEs. *J. Mar. Sci. Eng.* **2025**, *13*, 2091. Available online: https://www.mdpi.com/2077-1312/13/11/2091 (accessed on 21 March 2026). [CrossRef]
30. Khodaparast, M.; Roshanian, J. On the Design of Reward Function for Reinforcement Learning-Based Path Following Control Using the TD3 Algorithm. *J. Intell. Robot. Syst.* 2026, *in press*. [CrossRef]
31. Evaluating Scenario-Based Decision-Making for Interactive Autonomous Driving Using Rational Criteria: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2025**, *27*, 1709–1730. Available online: https://ieeexplore.ieee.org/document/11301014 (accessed on 21 March 2026).
32. Park, Y.; Jun, W.; Lee, S. A Comparative Study of Deep Reinforcement Learning Algorithms for Urban Autonomous Driving: Addressing the Geographic and Regulatory Challenges in CARLA. *Appl. Sci.* **2025**, *15*, 6838. [CrossRef]
33. *ISO 2631-1:1997*; Mechanical Vibration and Shock—Evaluation of Human Exposure to Whole-Body Vibration—Part 1: General Requirements. ISO: Geneva, Switzerland, 1997.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.