2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Trailer-aware End-to-end Autonomous Driving for Tractor-Trailers with Deep Reinforcement Learning

Congfei Li, Yang Li, Peigen Liu, Rongqi Gu, Zuolei Sun, and Yuxiang Sun

***Abstract*— End-to-end autonomous driving has been greatly advanced in recent years. However, most of existing work focuses on small vehicles (e.g., cars). Driving articulated trucks, such as tractor-trailers, still remains less being explored. The underactuated nature and extended wheelbase of tractor-trailers pose considerable driving challenges, especially when navigating narrow roads. For example, when a left-hand-drive tractor-trailer makes a right turn on a two-way two-lane narrow road, the tractor usually needs to encroach some spaces in the opposing lane. Otherwise, the trailer may have insufficient spaces to turn right and strike curbside objects. To provide a solution to this problem, we employ deep reinforcement learning to train an end-to-end autonomous driving policy with a trailer-aware reward function. Through planar rigid-body kinematics analysis, we locate the reference points on the tractor and the trailer. We also build a tractor-trailer model for CARLA. Experimental results demonstrate the effectiveness and superiority of our method in CARLA.**

I. Introduction

In recent years, end-to-end autonomous driving has attracted great interest from both academia and industry [1]–[11]. Many effective networks have been proposed. However, most of them are designed for driving small vehicles (e.g, cars). Driving articulated trucks, such as tractor-trailers, still remains a challenge, especially at road intersections with small turning radii. Different from driving small vehicles, where the focus is mainly on aligning the vehicle with road centerline, driving articulated trucks at interactions requires careful consideration of the trailer. If the trailer's behaviors are neglected, the trailer may encroach into sidewalks or collide with roadside infrastructure.

Fig. 1 shows the swept area depicted by the bicycle model [12] for a tractor-trailer turning at an interaction. When the tractor follows the blue centerline, the trailer encroaches into the inner side of the interaction. Smaller turning radii could further increase the risk of the trailer striking curbside objects. To ensure safety, dedicated road markings are usually printed on roads to indicate dangerous areas. Fig. 2 shows a typical scenario. We can see that a yellow area marked with "DANGER ZONE" warns pedestrians not to stand in this area. In addition, a no-stopping area with cross markings (highlighted in red) is used to warn vehicles not to stop at this area, because tractor-trailers may need larger turning radii when turning right, and hence encroach into this area.

This work was supported by Shanghai Westwell Technology Co., Ltd. *(Corresponding author: Yuxiang Sun.)*

Congfei Li, Yang Li, and Yuxiang Sun are with Department of Mechanical Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong (email: yx.sun@cityu.edu.hk, sun.yuxiang@outlook.com)

Peigen Liu, Rongqi Gu, and Zuolei Sun are with Shanghai Westwell Technology Co., Ltd.

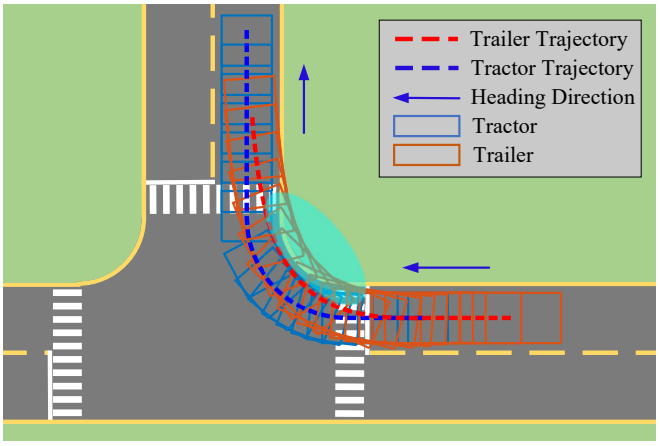


Fig. 1. The envelopes of a tractor-trailer when the tractor makes a right turn. The tractor starts from the right side and turn right following the road center line. The blue and red dashed lines represent the trajectory of the tractor's front axle center and the trajectory of the trailer's rear axle, respectively. The yellow solid lines represent road boundaries. The yellow dashed lines represent guidance lines. The encroached area by the trailer is highlighted by the green ellipse.

Fig. 3(a) and Fig. 3(b) show video snapshots capturing two tractor-trailers with human drivers turning right in real-world road intersections. Unlike small vehicles, the tractor refrains from making a sharp turn upon entering the intersection to avoid encroaching into the danger zone. Instead, the tractor postpones the turning until the trailer's rear wheels have fully entered the interaction, thereby minimizing the encroachment. During the final phase of the turning, the tractor encroaches into the no-stopping area and straighten its trajectory. The two examples illustrate that navigating tractor-trailers requires not only avoiding collisions between the trailer and the roadside, but also managing the often unavoidable encroachment into the opposing lane during right turns.

To provide a solution to this problem, in this paper, we propose a trailer-aware reward function to train a autonomous driving policy with deep reinforcement learning. The contributions of this work are summarized as follows:

- We design a novel trailer-aware reward function that enables a policy to safely navigate interactions.
- We investigate the influence of the reference points placement on both the tractor and the trailer.
- We build a tractor-trailer vehicle model for CARLA [13], and conduct closed-loop comparative experiments. Our code and model are publicly available[1].

[1]https://github.com/lab-sun/Trailer-aware-AD

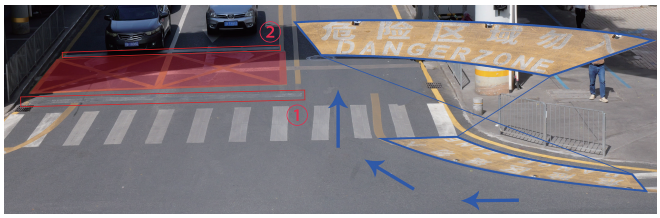


Fig. 2. Typical road markings to ensure safe right turns for tractor-trailers. The blue highlighted area indicates a dangerous zone, warning pedestrians to avoid standing in this area. The red highlighted area indicates a no-stopping zone, prohibiting vehicles from stopping in this area. It can be observed that the stop line ① has been redrawn to ②. This likely reflects a response to increased traffic involving tractor-trailers, that is, their presence may have been few initially, prompting a later revision of the stop line to accommodate more frequent turnings of tractor-trailers. The photo was captured in Shenzhen, China. Please zoom in for details.

(a) A perspective view of an articulated truck making a right turn.

(b) A bird-eye view of an articulated truck making a right turn.

Fig. 3. Video snapshots capturing two different tractor-trailers turning right in two different real-world road intersections. The numbered snapshots depict the turning sequence as follows: (1) the tractor enters the interaction; (2) the trailer follows into the interaction; (3) the tractor-trailer reaches the midpoint of the interaction; (4) the tractor begins to exit the interaction; (5) the trailer initiates its exit; and (6) the whole tractor-trailer has fully cleared the interaction.

II. Related Work

A. End-to-end Autonomous Driving

There are mainly two training frameworks for end-to-end autonomous driving: reinforcement learning (RL) [14]–[18] and imitation learning (IL) [19]–[24]. Since this paper employs reinforcement learning, we focus on methods using reinforcement learning. Most of the existing end-to-end methods are proposed for small vehicles. For example, Coelho *et al.* [25]. integrated RL with IL through online expert demonstrations, thereby improving data sample efficiency. Peng *et al.* [26]. further refined an IL-pretrained policy using RL, effectively leveraging the data efficiency of IL and the generalization capability of RL to enhance policy reliability. Huang *et al.* [14]. incorporated a pre-trained Vision-Language Model (VLM) as a reward function model, circumventing the need for complex reward shaping. Liu *et al.* [27]. designed a novel Transformer-based scene representation learning framework that significantly enhanced the performance of RL-based decision-making systems.

B. Autonomous Driving for Articulated Trucks

To achieve autonomous driving for articulated trucks, existing methods could be generally divided into traditional rule-based methods and deep learning-based methods. Many research efforts has been devoted to the former, primarily focusing on motion planning [28], [29] and trajectory optimization [30]–[32]. These methods often rely on static maps and incorporate collision avoidance constraints to ensure safe navigation.

For deep learning-based methods, Zhang *et al.* [33]. proposed a semi-supervised learning-based path planner that minimizes the swept area while ensuring collision avoidance. Wang *et al.* [34]. developed a decision making module trained with Double Deep Q-Network (DDQN), which determines whether to keep on the current lane or perform a lane change based on the surrounding environment. Attard *et al.* [35]. equipped an articulated truck with 29 range sensors to perceive the surrounding environment. The trained policy successfully enables the vehicle to safely navigate roundabouts with varying radii. Yan *et al.* [36]. designed a reward function based on an optimized reference trajectory to guide the agent in mimicking the behavior of a trajectory optimizer, enabling automatic reversing for articulated trucks. Attard *et al.* [35]. formulated a reward function using the distance from the truck to the road centerline, while incorporating distances to surrounding obstacles as part of the environmental state, allowing the agent to navigate roundabouts of various sizes without collisions. Kang *et al.* [37]. employed the Proximal Policy Optimization algorithm to train a trajectory-tracking policy for articulated trucks, which improves computational efficiency compared to traditional model-based controllers.

III. The Proposed Method

A. Preliminaries

1) Markov Decision Process: RL learns to make optimal decisions by interacting with an environment using reward signals as feedback. The theoretical foundation of RL is mainly formalized by Markov Decision Process (MDP). A canonical MDP is defined by a tuple (S, A, P, R), where S is state space, A is action space, P is state transition probability, and R is reward function. State $s_t \in S$ is a complete description of the environment at the time step t. The future state and reward depend only on the present state and action, independent of history: $P(s_{t+1}, r_t|s_t, a_t, ..., s_0, a_0) = P(s_{t+1}, r_t|s_t, a_t)$. Action $a \in A$ is an operation that the agent can execute in a given state. The state transition probability P is the probability that the environment transitions to state s_{t+1} given the current state s_t and action a_t, denoted

Fig. 4. Since there is no tractor-trailer model available in CARLA, we build a tractor-trailer model, named `QTruck`, by following the method outlined in [38].

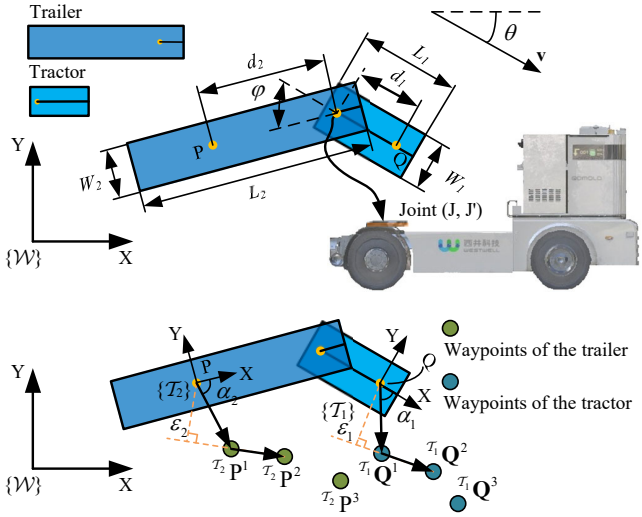


Fig. 5. The parameters used in the kinematic analysis of a tractor-trailer system. Point J denotes the hinge joint connecting the tractor and the trailer. Points R_1 and R_2 respectively represent the reference points on tractor and trailer for tracking their corresponding waypoints. The distances from R_1 and R_2 to point J are respectively denoted as d_1 and d_2. L_1 and L_2 represent the lengths of the tractor and the trailer, while W_1 and W_2 denote their widths. The vector $\mathbf{v}$ represents the tractor velocity. The coordinate frames $\{\mathcal{T}_1\}$ and $\{\mathcal{T}_2\}$ are right-handed coordinates, fixed on R_1 and R_2 respectively. Their X-axises are aligned with the heading directions of the tractor and the trailer. $^{\mathcal{T}_1}\mathbf{Q}^i$ denotes the i-th waypoint of the tractor in frame $\{\mathcal{T}_1\}$. $^{\mathcal{T}_2}\mathbf{P}^i$ denotes the i-th waypoint of the trailer in frame $\{\mathcal{T}_2\}$. The angle θ denotes the orientation of the tractor with respect to the world frame $\{\mathcal{W}\}$, and φ represents the articulation angle between the tractor and the trailer. ε_1 and α_1 are respectively the lateral distance and angle deviation between the tractor and its waypoints. ε_2 and α_2 are respectively the lateral distance and angle deviation between the trailer and its waypoints.

as $P(s_{t+1}|s_t, a_t)$. The reward function R is the immediate numerical feedback signal received from the environment after taking an action, expressed as $r_t = R(s_t, a_t, s_{t+1})$. In a partially observable setting, the agent cannot access the full state s, but instead receives an observation o_t correlated with the state. Such problems are generally modeled as Partially Observation MDP (POMDP). The overall goal of RL is to learn a policy $\pi_\phi(a|o)$ that maximizes the cumulative future reward, known as the expected return G_t. The return is defined as the summation of discounted future rewards: $G_t = \Sigma_{k=0}^{\infty}\gamma^k R_{t+k+1}$.

2) Soft Actor-Critic: The Soft Actor-Critic (SAC) algorithm is an off-policy actor-critic method based on the maximum entropy reinforcement learning framework [39]. The core idea is to simultaneously maximize expected cumulative returns and the entropy of the policy. There are five networks in a typical SAC framework, an actor network (policy network π_ϕ), two critic networks C_{θ_1}, C_{θ_2} and two target critic networks $C_{\bar{\theta}_1}$, $C_{\bar{\theta}_2}$. The double critic networks are learned by optimizing the soft Bellman residual $\mathcal{L}_{\theta_k} = \mathbb{E}_{o_t,a_t,o_{t+1}\sim\mathcal{B}}[(C_{\theta_k}(o_t, a_t) - y_t)^2], k = 1, 2$, $y_t = r_t + \gamma[\min_{k=1,2} C_{\bar{\theta}_k}(s_{t+1}, a_{t+1}) - \beta \log \pi_\phi(a_{t+1}|s_{t+1})]$, where k is the index of two critic networks, $k = 1, 2$, γ is the discount factor, $\mathcal{B}$ is the replay buffer, β is the temperature parameter, "$\sim$" means sampling $\{o_t, a_t, o_{t+1}\}$ from the replay buffer $\mathcal{B}$ and $\mathbb{E}$ represents expectation. The policy network is updated to maximize the expected future return plus entropy $\mathcal{L}_\phi = \mathbb{E}_{o_t\sim\mathcal{B}}[\beta \log \pi(a_t|s_t) - \min_{k=1,2} C_{\theta_k}(s_t, a_t)]$.

B. SAC with Imitation Learning

During training, we adopt the SAC framework with imitation learning [25]. An online expert π^* produces expert action a_t^* at the time step t. Therefore, the replay buffer stores the tuple $o_t, a_t, a_t^*, r_t, o_{t+1}$. The same batch of transitions used by RL is used to create a Gaussian distribution p_ϕ. The log-likelihood of the expert action a^* is maximized under the policy distribution $\mathcal{L}_\phi^*$:

$$\mathcal{L}_\phi^* = -\mathbb{E}_{o_t,a_t^*\sim\mathcal{D}}[\log p_\phi(a_t^*|s_t)]. \quad (1)$$

The expert policy π^* is allowed to access privilege information, such as road information and surrounding vehicles states. Readers could refer to [25] for more details about the expert, which is implemented as a set of simple heuristics. Beyond their roles in training, expert actions also aid in exploration. This is achieved by having the expert action override the policy's action whenever the policy's uncertainty is too high, ensuring safe and effective exploration.

C. Observation and Action

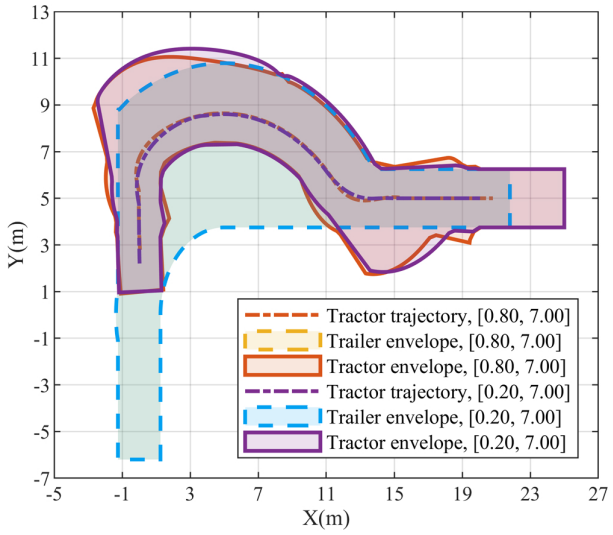
The observation o_t at time t is formulated as a concatenation of information from two most recent time steps: $o_t = \{(I_{t-1}, P_{t-1}, M_{t-1}), (I_t, P_t, M_t)\}$, where:

- I is a $3 \times 256 \times 256$ RGB image from a front-facing camera.
- P is a set of 10 upcoming 2D waypoints (in the ego truck's coordinate frame) from the global planner.
- M is the whole articulated truck state vector, which includes the truck's velocity, steering angle, and the relative angle between the truck and its trailer.

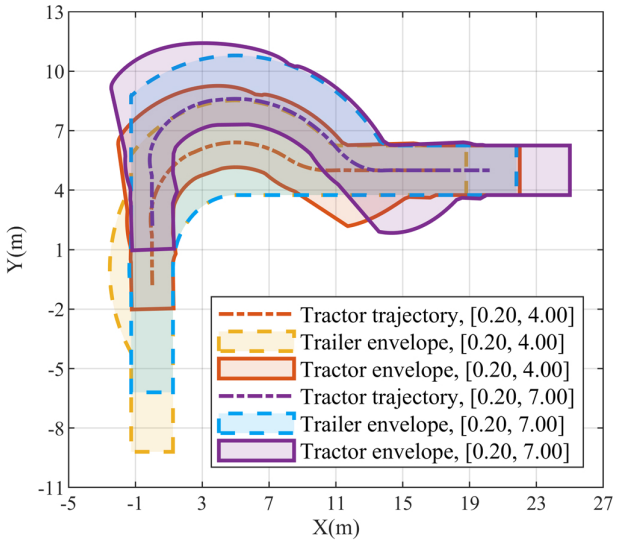
The action space is a continuous 3×1 vector: $a = [a_{\text{throttle}}, a_{\text{brake}}, a_{\text{steer}}]^T$, where $a_{\text{throttle}} \in [0, 1]$, $a_{\text{brake}} \in [0, 1]$, and $a_{\text{steer}} \in [-1, 1]$.

D. Reward Shaping

In this section, we analyze the motion of articulated truck during right-turning maneuvers with planar rigid-body



(a) d_1 changed, d_2 fixed.



(b) d_1 fixed, d_2 changed.

Fig. 6. The sub-figures (a) and (b) respectively show the tractor and trailer trajectories, as well as their envelopes with changed d_1 and d_2. The values in $[\cdot, \cdot]$ in the legends are d_1 and d_2. For clarity, the trailer trajectories are not displayed.

kinematics. The primary motivation of the reward design is to encourage the tractor to follow a trajectory that avoids the trailer from colliding with the right roadside. One potential approach is to apply trajectory optimization methods [35] that incorporate obstacles as constraints to find an optimal path for the tractor to track. However, such optimization-based techniques are computationally intensive [40], leading to prohibitively long training times. A solution is to use inverse kinematics for articulated trucks, though this involves complex algebraic computations [41].

Note that the reward function is not to enforce strict path tracking of a precomputed trajectory, but to shape the policy's action distribution towards desirable behaviors. So, in this work, we propose a simple yet effective reward function based on the planar rigid-body kinematics, which is designed to guide the articulated truck to navigate the interaction safely.

1) Motion Analysis: Based on our analysis in Fig. 3, the motion control of a tractor-trailer should be trailer-aware, that is, the trailer should follow the centerline of the road to ensure sufficient clearance on both sides. Therefore, using the trailer as the reference frame, we derive the target waypoints for the tractor with the planar rigid-body kinematics.

As shown in Fig. 5, we respectively choose a reference point, namely R_1 and R_2, from the tractor and trailer to track their waypoints. At the connection between the tractor and trailer, pivot points JJ and J' are attached to the trailer and tractor, respectively. The other geometric parameters are described in Fig. 5.

Given N waypoints of the trailer in the world coordinate frame, denoted as ${}^{\mathcal{W}}\mathbf{P} \in \mathbb{R}^{N\times 2}$, they are transformed into the trailer's body coordinate frame to obtain ${}^{\mathcal{T}_2}\mathbf{P}$: ${}^{\mathcal{T}_2}\mathbf{P} = [{}^{\mathcal{T}_2}\mathbf{R}_{\mathcal{W}}({}^{\mathcal{W}}\mathbf{P})^T]^T$, where ${}^{\mathcal{T}_2}\mathbf{R}_{\mathcal{W}} \in \mathbb{R}^{2\times 2}$ is the rotation matrix representing the orientation of the trailer relative to the world coordinate frame. The requirement that the trailer follows the road centerline implies that the reference point R_2 on the trailer moves along the spline curve formed by ${}^{\mathcal{T}_2}\mathbf{P}$. Specifically, R_2 coincides with a point on this spline, and its direction of motion aligns with the tangent direction at that point. Using the cubic spline interpolation, the heading direction ${}^{\mathcal{T}_2}\boldsymbol{\theta}_2$ corresponding to each waypoint in the trailer's body frame is derived. The waypoints of the pivot point J in the trailer's body frame can be computed as:

$$ {}^{\mathcal{T}_2}\mathbf{P}_{\mathrm{J}} = [{}^{\mathcal{T}_2}\mathbf{R}_{\mathcal{W}}({}^{\mathcal{W}}\mathbf{P})^T]^T + \mathbf{D}_2, \tag{2}$$

where $\mathbf{D}_2 = [d_2, 0; ...; d_2, 0]_{N\times 2}$.

Consistent with earlier computations, the heading direction ${}^{\mathcal{T}_2}\boldsymbol{\theta}_{\mathrm{J}}$ is derived using the cubic spline interpolation. Since points J and J' coincide, ${}^{\mathcal{T}_2}\mathbf{P}_{\mathrm{J}'} = {}^{\mathcal{T}_2}\mathbf{P}_{\mathrm{J}}$ and ${}^{\mathcal{T}_2}\boldsymbol{\theta}_{\mathrm{J}'} = {}^{\mathcal{T}_2}\boldsymbol{\theta}_{\mathrm{J}}$. ${}^{\mathcal{T}_2}\boldsymbol{\theta}_{\mathrm{J}'}$ is the articulated angle between the tractor and the trailer. Through another coordinate transformation, the waypoints of the reference point R_1 on the tractor in the tractor's body coordinate frame are derived as:

$$ {}^{\mathcal{T}_1}\mathbf{Q} = [{}^{\mathcal{T}_1}\mathbf{R}_{\mathcal{T}_2}({}^{\mathcal{T}_2}\mathbf{W}_{\mathrm{J}'})^T]^T + \mathbf{D}_1, \tag{3}$$

where ${}^{\mathcal{T}_1}\mathbf{R}_{\mathcal{T}_2} \in \mathbb{R}^{2\times 2}$ is the rotation matrix representing the orientation of the tractor relative to the trailer coordinate frame, and $\mathbf{D}_1 = [d_1, 0; ...; d_1, 0]_{N\times 2}$. Once more, the cubic spline interpolation is applied to obtain the heading direction of the tractor in the tractor's body frame, ${}^{\mathcal{T}_1}\boldsymbol{\theta}_{\mathrm{P}_1}$. At this stage, the reference waypoints and heading angles for both the trailer and the tractor in their respective local coordinate frames have been derived.

The selection of reference points is crucial for driving the behaviors of articulated trucks during turning at an interaction. We choose different combinations of $[d_1, d_2]$ for the reference points and plot the enveloping regions of both the tractor and trailer, as shown in Fig. 1, to analyze the

influence of R_1 and R_2 positions on the truck's driving behaviors during right-turning maneuvers.

As displayed in Fig. 6(a), the swept envelope of the trailer remains consistent when d_2 is fixed, while changes in d_1 have only a marginal effect on the tractor's envelope. As displayed in Fig. 6(b), a reduction in d_2 from 7.0 to 4.0 causes the trailer tail to sweep across the left lane boundary when initiating a turning maneuver, which increases the risk of collision with vehicles from the opposing lane. Meanwhile, the tractor envelope shifts toward the inner side of the interaction, resulting in reduced encroachment into the opposing lane.

2) Reward Function: The reward function is designed based on the lateral distance and angle deviation between tractor and trailer and their corresponding waypoints. The reward function consists of four components:

$$r = cr_{p,\text{trailer}} + cr_{\theta,\text{trailer}} + \lambda cr_{p,\text{tractor}} + \lambda cr_{\theta,\text{tractor}}, \quad (4)$$

where:

$$r_{p,\text{trailer}} = \{[^{\mathcal{T}_2}\mathbf{P}^1 - \mathbf{0}] \times [^{\mathcal{T}_2}\mathbf{P}^2 - ^{\mathcal{T}_2}\mathbf{P}^1]\}^2 \quad (5)$$

$$r_{p,\text{tractor}} = \{[^{\mathcal{T}_1}\mathbf{Q}^1 - \mathbf{0}] \times [^{\mathcal{T}_1}\mathbf{Q}^2 - ^{\mathcal{T}_1}\mathbf{Q}^1]\}^2, \quad (6)$$

$$r_{\theta,\text{trailer}} = \arctan[^{\mathcal{T}_2}\mathbf{P}^1_y / ^{\mathcal{T}_2}\mathbf{P}^1_x], \quad (7)$$

$$r_{\theta,\text{tractor}} = \arctan[^{\mathcal{T}_1}\mathbf{Q}^1_y / ^{\mathcal{T}_1}\mathbf{Q}^1_x], \quad (8)$$

c is a penalty weight that discourages deviation from the reference waypoints, and λ represents the ratio of the tractor's penalty weight to that of the trailer. The subscript x and y denote the x-component and y-component of the waypoint, respectively.

IV. Experimental Results and Discussions

A. Experimental Setup

This work addresses the challenge in enabling articulated trucks to safely navigate narrow interactions. Therefore, our primary focus is on avoiding collisions between the truck (both the tractor and trailer) and road sides. So, we adopt a simplified version of the no-crash benchmark, in which both training and testing are conducted in the absence of pedestrians and other dynamic objects. We use CARLA 0.9.15, and train and evaluate our policy in Town01. The roads are all two-lane two-way in both train and test processes, and the width of single road is 3.5m. The evaluation metrics include success rate (SR), success rate with no crash (SRNC), route completion rate (RCR), number of tractor collisions with layouts (e.g., traffic lights, street lights, pavements, green belts and etc.) per kilometer (CL-tractor), number of trailer collisions with layouts per kilometer (CL-trailer).

B. Comparative Analysis

To verify whether autonomous driving policies designed for small vehicles could also safely drive the tractor-trailer, we first test Transfuser++ (TF++) [42] with its pre-trained weights to control the tractor with trailer. Then, we test RLfOLD [25]. Since the original RLfOLD policy does not provide pretrained weights, we re-implement RLfOLD and train a policy using only the tractor model. This reproduced policy is then evaluated on the whole tractor-trailer. Additionally, we propose an enhanced policy, RLfOLD with Trailer (RLfOLDwT), which is trained on the full tractor-trailer system and incorporates an additional collision penalty specifically for the trailer. Finally, we develop another policy, named Trailer on Road (ToR), using the whole tractor-trailer with a reward function designed specifically for the tractor-trailer, as described in [35]. This reward encourages both the trailer and the tractor to closely follow the road centerline. Our proposed policy is named Trailer Guides Tractor (TGT). The results are displayed in Tab. I.

TABLE I

Quantitative results of the comparative experiment. The best results are highlighted in bold.

Policy	SR %	SRNC %	RCR %	CL-trailer #/Km	CL-tractor #/Km
RLfOLD	98.00	28.00	98.92	4.03	0.05
RLfOLDwT	100.00	46.00	100.00	3.07	0.09
TF++	24.00	0.00	54.21	10.73	7.74
ToR	90.00	42.00	96.31	4.63	0.21
TGT	**100.00**	**84.00**	**100.00**	**1.90**	**0.00**

Since TF++ was trained with small vehicles, its policy tends to replicate the driving behaviors typical of small vehicle drivers, including control patterns in speed, turning and braking. With TF++ controlling the tractor-trailer, the truck exhibits high speed (about 60 Km/h) during straight-line driving, and brakes too late before turning to reduce the speed to a safe value. This results in a very low SR for TF++, failing to complete routes without collisions. In contrast, for RLfOLD, RLfOLDwT, and ToR, a desired speed of 4 Km/h was specified during training, significantly improving stability during turning and increasing the SR.

Adding the trailer collision penalty, RLfOLDwT improves SRNC by 64.3% compared to RLfOLD as shown in Tab. I. The number of collisions per kilometer between the trailer and the layout is also decreased by 23.8%. For ToR, adding a reward function about trailer following road centerline does not improve SRNC. This indicates that for wide turns (e.g., left-hand-drive tractor-trailers turning left), trailer collision penalties allow the policy to learn how to pass through without collisions. However, tight turns (e.g., left-hand-drive tractor-trailers turning right on two-way two-lane roads), even with both collision penalties and the "Trailer on Road" reward, the policy fails to learn how to traverse the current intersection without collisions.

C. Ablation Study

To further verify the effectiveness of the proposed reward, we conduct ablation studies on different combinations of the reference points from the trailer and the tractor $[d_1, d_2]$, as well as on the penalty weight ratio λ.

1) Ablation on Reference Points: Tab. II displays the results of ablation study on the reference points. As we can see, large values of d_1 and d_2 (e.g., $d_1 = 1.0$, $d_2 = 8.0$) cause the trailer to enter the interaction too late, making the tractor prone to colliding with the layout on the opposing lane.

TABLE II

THE RESULTS OF THE ABLATION STUDY ON DIFFERENT COMBINATIONS OF REFERENCE POINTS. THE NUMBERS IN $[\cdot, \cdot]$ ARE $[d_1, d_2]$. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

$[d_1, d_2]$	SR %	SRNC %	RCR %	CL-trailer #/Km	CL-tractor #/Km
[1.0, 8.0]	0.00	0.00	23.15	4.58	105.72
[0.6, 8.0]	28.00	18.00	46.82	99.76	27.35
[0.2, 8.0]	36.00	24.00	52.07	182.61	30.79
[1.0, 6.0]	88.00	28.00	93.63	4.20	0.59
[0.6, 6.0]	98.00	70.00	98.11	2.23	4.48
[0.2, 6.0]	60.00	46.00	72.63	3.40	15.85
[1.0, 4.0]	**100.00**	**84.00**	**100.00**	1.90	**0.00**
[0.6, 4.0]	**100.00**	72.00	**100.00**	**1.88**	0.12
[0.2, 4.0]	70.00	54.00	83.6	9.07	1.18

TABLE III

THE RESULTS OF THE ABLATION STUDY ON DIFFERENT PENALTIES WEIGHTS ON TRACTOR. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Penalty Weight λ	SR %	SRNC %	RCR %	CL-trailer #/Km	CL-tractor #/Km
0.0	**100.00**	80.00	**100.00**	1.91	**0.00**
0.3	98.00	**84.00**	98.67	**1.52**	0.48
0.6	98.00	82.00	98.43	1.77	0.10
0.9	98.00	64.00	97.74	1.93	0.29
1.2	78.00	40.00	88.64	2.83	2.48

This issue is significantly alleviated with reduced d_1 or d_2. However, when d_2 remains at 8.0 and only d_1 is decreased, the tractor successfully navigates out of the interaction, but the trailer retains a deviation angle from the lane direction, resulting in a collision with the roadside layouts.

As d_2 decreases, all metrics are improved. Our policy achieves the best overall performance with $d_2 = 4.0$ and $d_1 = 1.0$, though the number of collisions between the trailer and the layout is not the best. This is because, during left turning, the tractor-trailer cannot perceive the surrounding environment, leading to collisions with the layout on that side.

2) Ablation on Penalty on Tractor: Tab. III displays the ablation study results on penalty weight. We can see that the variations in the penalty weight have little effect on SR. However, SRNC decreases as the penalty weight increases, while the route completion metric remains largely consistent with SR. Moreover. Higher penalty weight imposed on deviations of the tractor from the waypoints increases CL-trailer.

D. Qualitative Demonstrations

Fig. 7 demonstrates two typical scenarios of the tractor-trailer performing left-turning and right-turning. During left-turning, the tractor initiates the interaction earlier than the calculated waypoints to prevent the trailer's rear from colliding with the obstacles on the right. However, this causes the left side of the trailer to encroach into the opposing lane. While this turning can be safely performed in the absence of other traffic participants, it poses a collision risk when other vehicles are present. During right-turning, the tractor intentionally encroaches into the opposing lane to increase its turning radius, thereby keeping the trailer aligned with the lane center and avoiding collisions with surrounding layouts.

(a) Bird-eye view of our tractor-trailer making a left turn on a narrow road.

(b) Bird-eye view of our tractor-trailer making a right turn on a narrow road.

Fig. 7. Qualitative demonstrations of our policy during left and right turnings for the tractor-trailer. The images are captured from a camera mounted on the trailer in CARLA.

V. CONCLUSIONS AND FUTURE WORK

We proposed here a method toward trailer-aware autonomous driving for tractor-trailers, demonstrating the effectiveness of combining deep reinforcement learning with kinematic insights to achieve safe and efficient driving behaviors. Through ablation studies, we verified the effectiveness of the reference points combinations and the penalty weight. The comparative experiments among various policies further showcase our superiority. However, this work still exhibits limitations. Our study focuses solely on scenarios without other road users. Since driving in dynamic traffic environments is a required capability in real-world applications, we will focus on designing policies with interactions with other road users in the future.

REFERENCES

[1] Z. Xu, Y. Bai, Y. Zhang, Z. Li, F. Xia, K.-Y. K. Wong, J. Wang, and H. Zhao, "Drivegpt4-v2: Harnessing large language model capabilities for enhanced closed-loop autonomous driving," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 17 261–17 270.

[2] Y. Feng, Z. Feng, W. Hua, and Y. Sun, "Multimodal-xad: Explainable autonomous driving based on multimodal environment descriptions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 12, pp. 19 469–19 481, 2024.

[3] B. Liao, S. Chen, H. Yin, B. Jiang, C. Wang, S. Yan, X. Zhang, X. Li, Y. Zhang, Q. Zhang *et al.*, "Diffusiondrive: Truncated diffusion model for end-to-end autonomous driving," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 12 037–12 047.

[4] Y. Feng and Y. Sun, "Polarpoint-bev: Bird-eye-view perception in polar points for explainable end-to-end autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 11, pp. 6753–6763, 2024.

[5] C. Min, D. Zhao, L. Xiao, J. Zhao, X. Xu, Z. Zhu, L. Jin, J. Li, Y. Guo, J. Xing *et al.*, "Driveworld: 4d pre-trained scene understanding via world models for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 15 522–15 533.

[6] Y. Feng, W. Hua, and Y. Sun, "Nle-dm: Natural-language explanations for decision making of autonomous driving based on semantic scene understanding," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 9, pp. 9780–9791, 2023.

[7] Z. Xu, Y. Zhang, E. Xie, Z. Zhao, Y. Guo, K.-Y. K. Wong, Z. Li, and H. Zhao, "Drivegpt4: Interpretable end-to-end autonomous driving via large language model," *IEEE Robotics and Automation Letters*, 2024.

[8] P. Cai, H. Wang, Y. Sun, and M. Liu, "Dq-gat: Towards safe and efficient autonomous driving with deep q-learning and graph attention networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 21 102–21 112, 2022.

[9] J. Araluce, L. M. Bergasa, M. Ocana, A. Llamazares, and E. López-Guillén, "Leveraging driver attention for an end-to-end explainable decision-making from frontal images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 8, pp. 10 091–10 102, 2024.

[10] H. Wang, P. Cai, Y. Sun, L. Wang, and M. Liu, "Learning interpretable end-to-end vision-based motion planning for autonomous driving with optical flow distillation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 13 731–13 737.

[11] P. Cai, H. Wang, Y. Sun, and M. Liu, "Dignet: Learning scalable self-driving policies for generic traffic scenarios with graph neural networks," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 8979–8984.

[12] M. M. Michałek, B. Patkowski, and T. Gawron, "Modular kinematic modelling of articulated buses," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8381–8394, 2020.

[13] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.

[14] Z. Huang, Z. Sheng, Y. Qu, J. You, and S. Chen, "Vlm-rl: A unified vision language models and reinforcement learning framework for safe autonomous driving," *Transportation Research Part C: Emerging Technologies*, vol. 180, p. 105321, 2025.

[15] H. Taghavifar, C. Hu, C. Wei, A. Mohammadzadeh, and C. Zhang, "Behaviorally-aware multi-agent rl with dynamic optimization for autonomous driving," *IEEE Transactions on Automation Science and Engineering*, 2025.

[16] W. Huang, H. Liu, Z. Huang, and C. Lv, "Safety-aware human-in-the-loop reinforcement learning with shared control for autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[17] Q. Li, X. Jia, S. Wang, and J. Yan, "Think2drive: Efficient reinforcement learning by thinking with latent world model for autonomous driving (in carla-v2)," in *European Conference on Computer Vision*. Springer, 2024, pp. 142–158.

[18] P. Cai, X. Mei, L. Tai, Y. Sun, and M. Liu, "High-speed autonomous drifting with deep reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1247–1254, 2020.

[19] J. Cheng, Y. Chen, and Q. Chen, "Pluto: Pushing the limit of imitation learning-based planning for autonomous driving," *arXiv preprint arXiv:2404.14327*, 2024.

[20] Y. Duan, Q. Zhang, and R. Xu, "Prompting multi-modal tokens to enhance end-to-end autonomous driving imitation learning with llms," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 6798–6805.

[21] G. C. K. Couto and E. A. Antonelo, "Hierarchical generative adversarial imitation learning with mid-level input generation for autonomous driving on urban environments," *IEEE Transactions on Intelligent Vehicles*, 2024.

[22] C. Gong, C. Lu, Z. Li, Z. Liu, J. Gong, and X. Chen, "Beyond imitation: A life-long policy learning framework for path tracking control of autonomous driving," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 7, pp. 9786–9799, 2024.

[23] P. Cai, Y. Sun, H. Wang, and M. Liu, "Vtgnet: A vision-based trajectory generation network for autonomous vehicles in urban environments," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 3, pp. 419–429, 2021.

[24] H. Wang, P. Cai, R. Fan, Y. Sun, and M. Liu, "End-to-end interactive prediction and planning with optical flow distillation for autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2021, pp. 2229–2238.

[25] D. Coelho, M. Oliveira, and V. Santos, "Rlfold: Reinforcement learning from online demonstrations in urban autonomous driving," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 10, 2024, pp. 11 660–11 668.

[26] Z. Peng, W. Luo, Y. Lu, T. Shen, C. Gulino, A. Seff, and J. Fu, "Improving agent behaviors with rl fine-tuning for autonomous driving," in *European Conference on Computer Vision*. Springer, 2024, pp. 165–181.

[27] H. Liu, Z. Huang, X. Mo, and C. Lv, "Augmenting reinforcement learning with transformer-based scene representation learning for decision-making of autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 3, pp. 4405–4421, 2024.

[28] A. Widyotriatmo, P. I. Siregar, and Y. Y. Nazaruddin, "Line following control of an autonomous truck-trailer," in *2017 International Conference on Robotics, Biomimetics, and Intelligent Computational Systems (Robionetics)*. IEEE, 2017, pp. 24–28.

[29] J. Diestra and S. Skouras, "Trajectory planning for automated driving of articulated heavy vehicles," 2019.

[30] Z. Wang, H. Zhang, J. Wang, and W. Chen, "Optimization-based trajectory planning for tractor-trailer vehicles on curvy roads: A progressively increasing sampling number method," *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[31] S. Han, K. Yoon, G. Park, and K. Huh, "Robust lane keeping control for tractor with multi-unit trailer under parametric uncertainty," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 2333–2347, 2023.

[32] A. Rahimi, W. Huang, T. Sharma, and Y. He, "An autonomous driving control strategy for multi-trailer articulated heavy vehicles with enhanced active trailer safety," in *The IAVSD International Symposium on Dynamics of Vehicles on Roads and Tracks*. Springer, 2021, pp. 769–782.

[33] X. Zhang, J. Eck, and F. Lotz, "A path planning approach for tractor-trailer system based on semi-supervised learning," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 3549–3555.

[34] D. Wang, L. Gao, Z. Lan, W. Li, J. Ren, J. Zhang, P. Zhang, P. Zhou, S. Wang, J. Pan *et al.*, "An intelligent self-driving truck system for highway transportation," *Frontiers in neurorobotics*, vol. 16, p. 843026, 2022.

[35] D. Attard and J. Bajada, "Reinforcement learning for autonomous control of articulated vehicles in roundabout intersections," in *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE, 2024, pp. 450–454.

[36] H. Yan, M. A. Zohdy, E. Alhawsawi, and A. Mahmoud, "Trajectory state model-based reinforcement learning for truck-trailer reverse driving," in *2024 8th International Conference on Robotics, Control and Automation (ICRCA)*. IEEE, 2024, pp. 197–205.

[37] Q. Kang, A. Hartmannsgruber, S.-H. Tan, X. Zhang, and C.-M. Chew, "Deep reinforcement learning based tractor-trailer tracking control," in *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2024, pp. 3147–3153.

[38] A. Behera, S. Kharrazi, and E. Frisk, "Collision avoidance analysis of an articulated heavy vehicle in carla," in *The IAVSD International Symposium on Dynamics of Vehicles on Roads and Tracks*, 2025.

[39] J. Gao, Y. Li, Y. Chen, Y. He, and J. Guo, "An improved sac-based deep reinforcement learning framework for collaborative pushing and grasping in underwater environments," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–14, 2024.

[40] Y. Yang, L. Gao, D. Li, B. Jia, Z. Hu, S. Xie, and M. Fu, "A review of trajectory planning and tracking methods for the tractor-trailer system," *IEEE Intelligent Transportation Systems Magazine*, pp. 2–27, 2025.

[41] D. Kreimer, P. Fleck, T. Kernbauer, and C. Arth, "Assisted trailer parking using a reverse camera system and inverse kinematics," in *2025 IEEE Intelligent Vehicles Symposium (IV)*, 2025, pp. 1618–1625.

[42] J. Zimmerlin, J. Beißwenger, B. Jaeger, A. Geiger, and K. Chitta, "Hidden biases of end-to-end driving datasets," *arXiv preprint arXiv:2412.09602*, 2024.