Safedrive Dreamer: Navigating Safety-Critical Scenarios in the Real-world with World Models

Anonymous CVPR submission

Paper ID *****

Abstract

Ensuring safety in dynamic and unpredictable environments is a crucial challenge in the rapidly evolving field of autonomous driving. In this work, we propose the Safedrive Dreamer, a novel vision-based navigation framework that integrates world models with safety-critical decision ability, enabling autonomous vehicles to navigate complex situations safely in the real world. Our approach proactively learns potential dangers and plans safer routes, leveraging the predictive capabilities of world models and significantly reducing the reliance on extensive trial-and-error learning in the real world. The effectiveness of Safedrive Dreamer is validated through a series of experiments in real-world sim-to-real driving conditions, covering a diverse range of safety-critical scenarios, such as abrupt obstacle avoidance. Our results show that Safedrive Dreamer achieves superior performance in safety metrics, such as collision avoidance and risk minimization, compared to other end-to-end solutions. This framework advances autonomous driving safety and offers insights into integrating world models for enhancing decision-making in safety-critical applications. Safedrive Dreamer paves the way for developing more resilient and trustworthy autonomous driving systems that are adept at handling the dynamics and uncertainties of the real world.

1. Introduction

The advancement of machine learning (ML) in autonomous driving (AD) represents a paradigm shift, offering a nuanced approach to navigating complex, dynamic environments [15] [26]. As a safety-critical application [16] [25], the autonomous driving system faces challenges regarding robustness and safety in the real-world deployment process [28] [2]. Unreliable autonomous driving systems may threaten human life and the surrounding environment [20].

Direct learning in real-world environments is costly and potentially dangerous [9]. Most of the time, agents are trained within designed simulated environments before being deployed into reality, referred to as *"sim-to-real"* [14]. The real world is characterized by uncertainties including stochastic interactions with other road users and the possibility of encountering rare weather or lighting conditions [9]. Thus, creating a perfect high-fidelity training environment is computationally costly and impractical [22]. The inevitable discrepancy between simulation and reality leads to the potential degradation of an agent's performance upon real deployment [12] [5], known as the *"reality gap"* (RG). One solution for bridging the RG is domain randomization [12] [20] [13], which involves exposing extensive training environments with randomized parameters to the agent during the learning stage, enhancing its adaptability to variable real-world conditions after deployment. Although this method usually works well, it lacks a guarantee of reliability.

While ensuring the transferability of the agent, another challenge is to guarantee the safety of the agent's real-world behaviors. In the absence of safety constraints, the intermediate policies during the training may lead to severe physical damage, as data-driven approaches such as the Reinforcement learning (RL) method explore all possible actions to derive the optimal policy through trial and error [4]. Real-world behavior safety also suffers from the inevitability of the reality gap. Some rare but safety-critical real-world scenarios such as abrupt obstacles or actors that are hard to identify due to obstructions [25], may not be commonly featured in the simulation but still play a crucial role in forming the safety metrics [1].

To tackle these challenges, we introduce *"Safedrive Dreamer"*, a framework that integrates advanced world models with safety-aware learning algorithms to bridge the sim-to-real transition *(reality gap)*. Furthermore, this framework is validated using a test vehicle in the real world. *"Safedrive Dreamer"* aims to make predictions and navigate through safety-critical scenarios with unprecedented reliability and safety, marking a significant step forward in the quest for autonomous driving. Our main contributions are:

- We integrate world models with safety-critical decision-making to enhance autonomous driving safety and efficiency.
- We close the reality gap between sim-to-real in safety-critical scenarios through our safe sim-to-real framework.
- We demonstrate superior performance in safety metrics like collision avoidance and risk minimization through real-world testing.

2. Related work

Generating and testing safety-critical scenarios is crucial in autonomous driving testing. Wang et al. [25] proposed an adversarial framework designed to generate safety-critical scenarios for LiDAR-based autonomous driving systems. Hanselmann et al. [11] introduce KING, a method for generating safety-critical driving scenarios using the CARLA simulator. They employ a kinematic bicycle model to optimally perturb background traffic trajectories, enhancing the generation of challenging scenarios for self-driving systems. However, they didn't evaluate the performance in a real-world setting.

Our framework is closely related to Model-based Reinforcement Learning (MBRL), which involves learning a system dynamics model from the environment. The accuracy of MBRL heavily depends on the model's fidelity [18]. While constructing an accurate model presents challenges, compared to model-free RL approaches, MBRL generally has a higher sample efficiency and requires less real data [7], [3], [8], [17], [6]. For example, MBRL [4] offers high-probability safety assurances of stability by leveraging Lyapunov functions, with regularity assumptions in terms of a Gaussian process prior. However, constructing a Lyapunov function is often challenging and involves hand-crafted elements without a universal principle [8]. Zanon et al. [30] combine RL's adaptability with MPC's ability to enforce safety and stability constraints. However, linear MPC might fail to provide satisfactory performance and safety in systems with strong nonlinearities.

Numerous previous studies investigated how to bridge the sim2real gap while providing a way to ensure generalizability. Wang et al. [24] introduced a novel reinforcement learning framework for autonomous driving that combines traditional modular pipelines with end-to-end approaches. They addressed key challenges such as effective representation learning, sim-to-real generalization to complex real-world scenarios, and training cost balance, followed by validation on a real-world vehicle. Akhauri et al. [1] employ a CNN-LSTM network that undergoes a two-phase training process to improve robustness, capitalize on the invariance of spatio-temporal features across domains and utilizes salient data augmentation to aid target domain training. A bi-directional domain adaptation (BDA) method with high sample efficiency proposed by Truong et al. [23], comprises a real-to-sim observation adaptation module (OA) and a sim-to-real dynamic adaptation module (DA), bridges the vision domain the dynamic domain gaps. Yuan et al. [29] introduce a learning-efficient DRQfD framework for modeling lane-changing decisions within a hierarchical decision-making architecture for learning-based autonomous driving (HAD). They employ a twin high-fidelity simulator based on ROS-Gazebo and use a domain randomization method to bridge the sim-to-real gap. Mozifian et al. [19] present an Intervention-Based Invariant Transfer Learning (IBIT) approach, merging domain randomization with data augmentation, which allows the agent to focus on essential visual features for task completion, therefore enhances the agent's generalization across real-world scenarios. Although these past studies have improved the generalization performance and quantified generalizability to some extent, they still lack an index to quantify the guaranteed degree of generalization performance. Moreover, in these studies, although the testing unseen scenarios differ from the training scenarios, they are still relatively similar, which means there needs to be more investigation on the safety performance and generalizability for some rare, uncommon scenarios that are even hard to generate in the simulator.

Further contributing to this field, Ren et al. [21] employed a two-stage approach where it first constructs a policy distribution through a conditional variational autoencoder (cVAE) with expert demonstrations. It then refines a posterior distribution over latent variables in fresh environments, focusing on optimizing a generalization performance bound derived from PAC-Bayes theory. However, to ensure a high guarantee of the generalization performance, it relies on the assumption of the same underlying distributions between training and novel environments, which is challenging to satisfy in a sim-to-real process. Moreover, it also needs proof of robustness in safety-critical scenarios.

In summary, although novel frameworks proposed in past research have bridged the gap between simulation and reality (sim2real) and achieved excellent generalization performance compared to their baselines, these studies did not delve deeply into sim-to-real transfer in uncommon, safety-critical scenarios. Additionally, several studies among the related work still needed to be validated in real-world environments. The insights gained from these previous studies have been organized into a table, which intuitively compares their sample efficiency (measured by training sample size), the deployment process of experimental validation (sim2sim: trained in a simulated environment and then deployed to another unseen simulated environment; sim2real: trained in a simulated environment and then deployed to an unseen real environment), the specific training task, and whether there is a way to quantify the guaranteed of generalization performance.

3. Method

We propose the Safedrive Dreamer framework which integrates world models with safety-aware learning to address the challenges of autonomous driving in safety-critical scenarios. At its core, the framework adapts the concept of Safe Reinforcement Learning (SafeRL) through a Constrained Markov Decision Process (CMDP) setup, enabling the autonomous system to learn policies that maximize safety and performance simultaneously.

Safedrive Dreamer leverages a world model to simulate future states and actions, allowing the autonomous agent to anticipate and navigate through complex driving scenarios safely. The world model is trained on data collected from both real-world driving and high-fidelity simulations, ensuring a comprehensive understanding of diverse driving conditions. This model facilitates the agent's ability to predict outcomes of actions before execution, crucial for making informed decisions in dynamic environments.

$$z_{t+1}, r_{t+1}, c_{t+1} = \text{WorldModel}(s_t, a_t) \quad (1)$$

where z_{t+1} is the predicted next state, r_{t+1} the anticipated reward, and c_{t+1} the potential cost or risk associated with action a_t from state s_t.

The world model on which Safedrive Dreamer is based is depicted in Fig. 1. In this world model, the input is defined as the content stored in the Replay Buffer. The use of the Replay Buffer facilitates the removal of correlations among data, thereby enhancing the diversity of the samples. The input data

Table 1. Wang et al. [24] considered some rare scenarios, such as dense pedestrian flow and high beam lighting conditions during testing but did not investigate the model's performance in more safety-critical environments. Ren et al. [21], during the testing, added an object with a relatively unique geometry. However, the results showed that the trained model could not perfectly complete the task, which means still lacks consideration for such rare scenarios.

Article	Framework	Training sample size	Deployment Type		Training Task	Consideration of Safety-critical/rare scenarios during testing	Quantified guarantee for generalization performance
			sim2sim	sim2real			
Wang et al.	Versatile and efficient autonomous driving framework	●	√	√	Autonomous Driving (lane-following, turning dynamic obstacle avoidance)	◐	-
Akhauri et al.	Spatio-temporal features transfer with salient data augmentation	●	√	-	Autonomous Driving (Collision Classification, turning)	○	-
Truong et al.	Bi-directional Domain Adaptation (BDA)	●	√	-	Autonomous Navigation (turning, obstacle avoidance)	○	-
Yuan et al.	Deep Recurrent Q-learning from demonstration (DRQfD)	○	-	√	Autonomous Driving (Car-following, Lane changing)	○	-
Mozifian et al.	Intervention-based Invariant Transfer learning (IBIT)	-	√	√	Robotic Manipulation (grasping objects)	○	-
Ren et al.	Two-tier traininig pipeline with PAC-Bayes Control	○	-	√	Robotic Manipulation (grasping objects, pushing objects, navigation)	◐	√

○ - total training samples smaller than 3000 samples, or 2 hours

● - total training samples greater than 5000 samples, or 7 hours

◐ - in between

○ - doesn't consider effects of special/rare/safety-critical scenarios

● - consider effects of special/rare/safety-critical scenarios to some extent

◐ -fully consider effects of special/rare/safety-critical scenarios

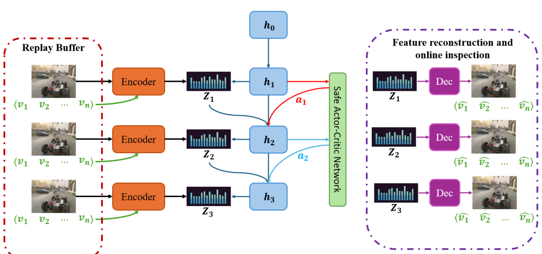


Figure 1. The architecture of Safedrive Dreamer's World Model consists of two main components: On the left-hand side, there is the Replay Buffer, responsible for processing input data and facilitating learning through the policy network and value network trained within Dreamer. On the right-hand side, there is the feature reconstruction and online observation area. Although this section does not directly participate in the decision-making process, the feedback it provides is crucial for model evaluation and performance calibration.

includes not only RGB images but also additional modal information, Such as vehicle speed information, radar, and simulated imagery. These high-dimensional sensory inputs are processed by an encoder, which then transforms them into discrete, low-dimensional state variables. These discrete, low-dimensional state variables are combined with information h_t from the hidden layer to obtain the latent state z_t. The hidden layer's h_t encompasses all prior observations and actions up to the current timestep (next state, reward, cost, etc.), enabling Safedrive Dreamer to make decisions based on the entire sequence of observations.

Model updating for the prediction of the current action constitutes a key emphasis within the Safedrive Dreamer algorithm. The hidden layer state h_t captures all antecedent observations and actions, which, combined with the latent state z_{t-1}, facilitate the forecasting of future actions and states within the latent space. Prognostications are conducted via the Safe Actor-Critic Network, which extrapolates not just the ensuing latent state but also prospective rewards, costs, and additional salient information. These "imagined" results are internally construed without direct engagement with the tangible environment, thus empowering the model to internally evaluate potential outcomes of disparate behaviors prior to actual implementation. This modality mitigates the exigency for empirical exploration in volatile environments, thereby ameliorating the security and efficacy of the learning trajectory.

We define ψ as parameters of Safedrive Dreamer that are continuously adjusted during the optimization process to better predict state transitions and rewards. Within the Safedrive Dreamer framework, the update of world model parameters is governed by a loss function as defined below(equation2), which synthesizes regularization loss, future prediction loss, observation loss, reward loss, and cost loss. In addition, to bolster the model's exploratory capabilities, an entropy loss is introduced. These collective loss components guide the adjustment of model parameters towards minimizing predictive errors and enhancing behavioral diversity([10]). The $sg(*)$ represents the gradient stopping operation, employed to regulate or stabilize the learning process.

$$\begin{aligned}\mathcal{L}(\psi) = \sum_{t=1}^{T} \alpha_1 KL(z_t \| sg(\hat{z}_t)) + \ \alpha_2 KL(sg(z_t) \| \hat{z}_t) \\ -\beta_1 \ln O_\psi(o_t|s_t) - \ \beta_2 \ln R_\psi(r_t|s_t) \\ -\beta_3 \ln C_\psi(c_t|s_t) + \xi H(\pi_\psi(\cdot|s_t))\end{aligned} \tag{2}$$

Additionally, the latent state z_t can be reconstructed into RGB images via the decoder, allowing the model to evaluate the quality of its state representation and predictions. By comparing the output of the decoder with actual observations, an error signal can be generated to guide the learning process of the model, and the accuracy of the model's predictions regarding obstacles or traffic conditions on the road can be visually inspected through online observation.

4. World Model-based Safe RL and Sim-to-Real Transition

In the framework of Constrained Markov Decision Processes (CMDP), we seek an optimal policy π' that maximizes expected return and satisfies predefined constraints. This is expressed as:

$$\pi' = \arg \max_{\pi_\theta \in \Pi_C} J^r(\pi_\theta), \tag{3}$$

where $J^r(\pi_\theta)$ is the return function under policy π_θ, and Π_C represents the policy space meeting all constraints.

We extend the model-based transition probability P to $P_{WorldModel}$, enabling simulation of actions' outcomes through the world model to optimize policy while managing risks.

5. Experimental Setup and Results

In our study, a comprehensive series of experiments were conducted within simulation environments crafted to replicate the driving conditions of the real world, encompassing urban traffic flows, highway travel, and scenarios involving pedestrians and cyclists. We evaluated the performance of Safedrive Dreamer against benchmark methods in terms of safety metrics, such as the number of safety incidents, and performance metrics, like average travel time. Likewise, Safedrive Dreamer was deployed on the Pix-Hooke platform and subjected to a variety of challenges in the real world through a series of meticulously designed experiments.

5.1. Experiment setup

Hardware Setup: The hardware utilized in this study is built on the PIX-Hooke open-source autonomous driving development platform, which integrates perception, decision-making, and control into a single system. The test vehicle is powered by a 72-volt lead-acid battery and equipped with high-precision steering, braking, and propulsion systems. Moreover, the PIX-Hooke platform operates on the Ubuntu 18.04 operating system and is equipped with a Core I7-8700 processor and an NVIDIA RTX2080 GPU, providing substantial computational power for autonomous driving tasks. The platform is also equipped with various perception hardware, including LiDAR and RGB cameras, as shown in Fig. 2.

Evaluation Metrics: To thoroughly evaluate the performance of the Safedrive Dreamer algorithm across various scenarios, the defined evaluation metrics are as follows:

- Meters Per Intervention(MPI, m): This metric measures the distance traveled between interventions. For example, if the vehicle travels 200 meters before an intervention is needed, the MPI is

200. A higher MPI value indicates better performance, as it signifies fewer interventions.

- Travel Time (TT, s): The total time taken to travel from the start point to the endpoint. This metric helps evaluate the efficiency of the autonomous vehicle; shorter travel times indicate higher efficiency.
- Success Rate (SR, %): The percentage of the journey completed successfully without any interventions before the first one occurs. A higher SR indicates that the vehicle can navigate longer distances independently, which is a sign of better performance.
- Standard Deviation of Speed (Std[v], m/s): This represents the consistency in speed variation and is related to the longitudinal smoothness of the travel trajectory. A lower standard deviation indicates a smoother driving experience.

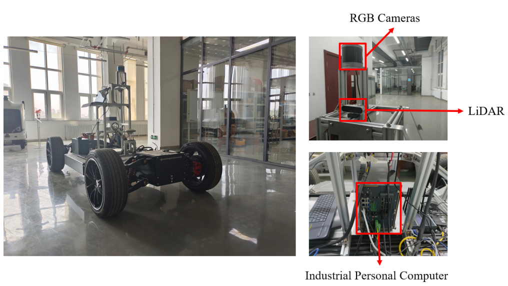
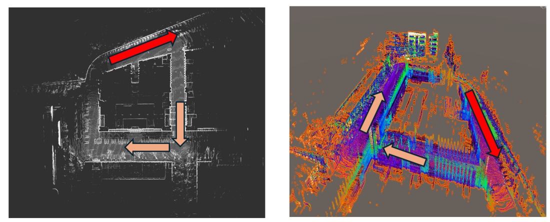


Figure 2. Pix-Hooke hardware description.

5.2. Real-world physical scenarios test

Experiment Description: To evaluate the performance of Safedrive Dreamer in real-world physical environments, we established a series of test environments based on actual vehicular scenarios. as shown in Fig. 3, we constructed planar and three-dimensional representations of the entire scene using LiDAR scanning. Based on the transition from simulation to reality, real-vehicle experiments were conducted as depicted in the figure, with the scene segmented into simple straight roads and complex environments.

The complexities of these environments are diverse, encompassing interactions with external agents of varying scales. Specifically, we designed a variety of agent quantities and condition combinations within these environments and progressively demonstrated how the Safedrive Dreamer's capability to understand the environment evolves with increasing training durations.



(a) (b)

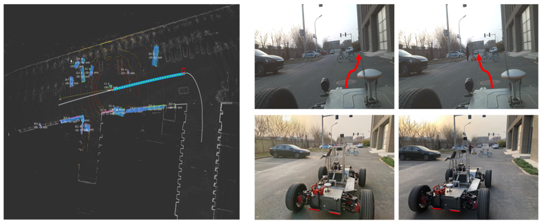
Figure 3. LiDAR scanning is utilized for the visualization of real physical scenarios, with red arrows indicating the driving trajectories in simple vehicle scenes, and pale orange arrows depicting the trajectories in more complex scenarios.

Progressive Scenario Analysis: In setting up the environment for scenarios and scaling interactions with external agents, we selected five typical scenarios to analyze and validate the evolution of Safedrive Dreamer's interactive capabilities with the environment and agents at different stages. As depicted in Fig. 4, the design of the scenarios and interactions was progressively developed.

During the scenario construction process, we implemented a progressively increasing difficulty design strategy, akin to the "level-by-level challenge" mode found in games. Within the Bridge environment, we initially collected a set of data based on the Carla platform, covering:

- basic simple straight-line driving scenarios.
- more complex scenarios combining straight roads and curves.

This collected data was used for preliminary training in Safedrive Dreamer to ensure that the world model-based agent could initially adapt to and understand the traffic environment. Subsequently, the training results obtained in Bridge were transferred to real-vehicle environments for validation and application. Given the relatively limited training data from Carla, we had to continue more in-depth train-



(a) Simple scenarios, from left to right, the sequence displays the LiDAR path, a straight driving scenario with only a single simple static obstacle, and a straight driving scenario that includes a manually operated small remote-controlled vehicle in addition to the static obstacle.

(b) Complex scenarios, from left to right, the sequence displays the LiDAR path, a curve and straight path scene with fewer pedestrians, a curve and straight path scene with more pedestrians, and a curve and straight path scene incorporating both dynamic obstacles and pedestrians.

Figure 4. Experiment scenarios for Safedrive Dreamer performance evaluation

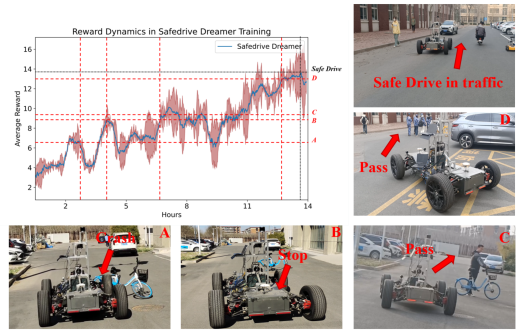


Figure 5. We present the curve showing the variation of the average reward of Safedreamer over time during training. On this curve, the actual reward at several specific time points is recorded. Concurrently, the vehicle states corresponding to these time points are displayed and illustrated through images A to D. For instance, during the process of generalizing the vehicle to real-world scenarios for learning, there was an increase in the reward curve, indicating an action to avoid obstacles. However, a collision still occurred, leading to a subsequent decrease in reward. This collision is represented on the reward curve by dashed line A, with the corresponding state time point documented.

ing in the real-vehicle environment.

As shown in Fig. 5, in the real-vehicle training phase, following the design strategy previously described, we placed the vehicle in a straight-line driving scenario with a simple static obstacle. Manual interventions were made to address unsafe behaviors as the vehicle learned the forward progression strategy, with the scenario being reset multiple times for enhanced learning. Once the vehicle mastered the simple static obstacle scenario, we increased the number of interactive agents within the scene, introducing additional challenges to the training process.

After the vehicle had mastered specific strategies within the simple static obstacle environment and demonstrated robustness in interactions with agents, we generalized its capabilities to the more complex scenarios of turns and straight lines that had been defined in both the Bridge and Real environments. This process mirrored the learning approach in simpler scenarios, where the number of interactive agents was incrementally increased.

5.3. Comparison with Baseline Model

In the comparison with the baseline model, we conducted analyses against advanced safety models and the World Model to demonstrate the performance advantages of our model. Specifically, in Table 1, we present the results of our performance comparison between our model and the Daydreamer model, as well as the Efficient Reinforcement Learning Framework for Autonomous Driving. This comparison serves to illustrate the superior performance of our model and underscores its potential in the realm of autonomous driving.

The Dreamer algorithm[27]: by planning within a learned world model, effectively reduces trial and error and has demonstrated superior performance to pure reinforcement learning in video games. Experiments have shown that Dreamer can rapidly adapt to environmental changes and accomplish

CVPR #*****

Table 2. Performance Comparison with Baseline Model

Model	MPI(m)	TT(s)	SR(%)	Std[V]	MPI(m)	TT(s)	SR(%)	Std[V]
		Simple scenario				**Complex scenario**		
DayDreamer	86.1	**21**	82.3	0.25	55.2	25	59.5	0.44
Efficient-IL	94.2	25	83.7	0.31	71.4	28	66.3	0.38
Efficient-RL	91.6	27	77.5	0.27	62.8	26	64.7	0.36
Our	**97.8**	21	**91.3**	**0.22**	**80.7**	**23**	**71.8**	**0.33**

complex tasks when applied to autonomous vehicles.

Efficient Reinforcement Learning Framework[24]: A fully functional autonomous vehicle was constructed for real-world validation, exhibiting exceptional generalizability and training efficiency through the integration of end-to-end and modular approaches.

In Table 1, we present the performance of the Safedrive Dreamer model across four key metrics and compare it with other models to demonstrate its performance under different evaluation criteria. The analysis indicates that, in both simple and complex scenarios, our model achieves the best performance in terms of Meters Per Intervention (MPI), demonstrating its capability to generalize from simple to complex scenarios and exhibiting strong robustness. In terms of travel time, Safedrive Dreamer performs on par with DayDreamer in simple scenarios, reaching the lowest level, surpassing all other algorithms in complex environments, and maintaining high efficiency. This underscores the model's strong adaptability in complex environments.

Additionally, regarding the success rate and standard deviation of speed, Although the success rate in complex scenarios is slightly lower than in simple ones, the model still demonstrates stability and maintains optimal performance, further reflecting the enhancement in safety brought about by employing safe reinforcement learning in the Safedrive Dreamer.

6. Conclusion

In this work, we introduced Safedrive Dreamer, a novel framework integrating world models with safety-critical decision-making for autonomous driving in dynamic and uncertain real-world conditions. Our approach enhances autonomous vehicles' ability to navigate safely by proactively learning potential dangers and planning safer routes. Through a comprehensive series of experiments based on sim-to-real scenarios, Safedrive Dreamer demonstrated superior performance in safety metrics, including collision avoidance and risk minimization, outperforming existing end-to-end solutions.

Our findings demonstrate the effectiveness of leveraging predictive world models for decision-making in safety-critical applications. Furthermore, the transition from simulation-based training to real-world deployment highlighted the importance of bridging the sim-to-real gap, ensuring the reliability and robustness of autonomous driving systems in handling diverse and unpredictable traffic conditions. However, due to safety concerns, we didn't evaluate the model in some other more extreme scenarios such as high speed, congested intersections, and multi-vehicle collaboration scenarios. We will add more baselines and compare them in more extreme scenarios.

In conclusion, Safedrive Dreamer shows insights of developing more resilient and trustworthy autonomous driving systems that can navigate the complexities and uncertainties of the real world. Future work will focus on extending the framework to incorporate more diverse scenarios and further improving the sim-to-real transferability to ensure even higher levels of safety and efficiency in autonomous driving.

References

[1] Shivam Akhauri, Laura Zheng, Tom Goldstein, and Ming Lin. Improving generalization of transfer learning across domains using spatio-temporal features in autonomous driving, 2021. 2

[2] Jean Pierre Allamaa, Panagiotis Patrinos, Herman Van der Auweraer, and Tong Duy Son. Sim2real for autonomous vehicle control using executable digital twin. *IFAC-PapersOnLine*, 55(24):385–391, 2022. 10th IFAC Symposium on Advances in Automotive Control AAC 2022. 1

[3] C.G. Atkeson and J.C. Santamaria. A comparison of direct and model-based reinforcement learning. In *Proceedings of International Conference on Robotics and Automation*, pages 3557–3564 vol.4, 1997. 2

[4] Felix Berkenkamp, Matteo Turchetta, Angela P. Schoellig, and Andreas Krause. Safe model-based reinforcement learning with stability guarantees. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 908–919, Red Hook, NY, USA, 2017. Curran Associates Inc. 2

[5] Iván García Daza, Rubén Izquierdo, Luis Miguel Martínez, Ola Benderius, and David Fernández Llorca. Sim-to-real transfer and reality gap modeling in model predictive control for autonomous driving. *Applied Intelligence*, 53(10): 12719–12735, 2022. 1

[6] Fei Deng, Ingook Jang, and Sungjin Ahn. Dreamerpro: Reconstruction-free model-based reinforcement learning with prototypical representations, 2021. 2

[7] Frederik Ebert, Chelsea Finn, Sudeep Dasari, Annie Xie, Alex Lee, and Sergey Levine. Visual foresight: Model-based deep reinforcement learning for vision-based robotic control, 2018. 2

[8] Shangding Gu, Long Yang, Yali Du, Guang Chen, Florian Walter, Jun Wang, Yaodong Yang, and Alois Knoll. A review of safe reinforcement learning: Methods, theory and applications, 2023. 2

[9] Cole Gulino, Justin Fu, Wenjie Luo, George Tucker, Eli Bronstein, Yiren Lu, Jean Harb, Xinlei Pan, Yan Wang, Xiangyu Chen, John D. Co-Reyes, Rishabh Agarwal, Rebecca Roelofs, Yao Lu, Nico Montali, Paul Mougin, Zoey Yang, Brandyn White, Aleksandra Faust, Rowan McAllister, Dragomir Anguelov, and Benjamin Sapp. Waymax: An accelerated, data-driven simulator for large-scale autonomous driving research. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, 2023. 1

[10] Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023. 5

[11] Niklas Hanselmann, Katrin Renz, Kashyap Chitta, Apratim Bhattacharyya, and Andreas Geiger. King: Generating safety-critical driving scenarios for robust imitation via kinematics gradients. In *European Conference on Computer Vision*, pages 335–352. Springer, 2022. 2

[12] Kai-Chieh Hsu, Allen Z Ren, Duy P Nguyen, Anirudha Majumdar, and Jaime F Fisac. Sim-to-lab-to-real: Safe reinforcement learning with shielding and generalization guarantees. *Artificial Intelligence*, 314:103811, 2023. 1

[13] Chuqing Hu, Sinclair Hudson, Martin Ethier, Mohammad Al-Sharman, Derek Rayside, and William Melek. Sim-to-real domain adaptation for lane detection and classification in autonomous driving. In *2022 IEEE Intelligent Vehicles Symposium (IV)*, pages 457–463, 2022. 1

[14] Xuemin Hu, Shen Li, Tingyu Huang, Bo Tang, Rouxing Huai, and Long Chen. How simulation helps autonomous driving:a survey of sim2real, digital twins, and parallel intelligence, 2023. 1

[15] Parth Kothari, Christian Perone, Luca Bergamini, Alexandre Alahi, and Peter Ondruska. Drivergym: Democratising reinforcement learning for autonomous driving, 2021. 1

[16] Quanyi Li, Zhenghao Peng, Lan Feng, Zhizheng Liu, Chenda Duan, Wenjie Mo, and Bolei Zhou. Scenarionet: Open-source platform for large-scale traffic scenario simulation and modeling, 2023. 1

[17] Fan-Ming Luo, Tian Xu, Hang Lai, Xiong-Hui Chen, Weinan Zhang, and Yang Yu. A survey on model-based reinforcement learning, 2022. 2

[18] S M Nahid Mahmud, Scott A Nivison, Zachary I. Bell, and Rushikesh Kamalapurkar. Safe model-based reinforcement learning for systems with parametric uncertainties, 2021. 2

[19] Melissa Mozifian, Amy Zhang, Joelle Pineau, and David Meger. Intervention design for effective sim2real transfer, 2020. 3

[20] James Queeney and Mouhacine Benosman. Risk-averse model uncertainty for distributionally robust safe reinforcement learning, 2023. 1

[21] Allen Z. Ren, Sushant Veer, and Anirudha Majum-

dar. Generalization guarantees for imitation learning, 2020. 3, 4

[22] Erica Salvato, Gianfranco Fenu, Eric Medvet, and Felice Andrea Pellegrino. Crossing the reality gap: A survey on sim-to-real transferability of robot controllers in reinforcement learning. *IEEE Access*, 9: 153171–153187, 2021. 1

[23] Joanne Truong, Sonia Chernova, and Dhruv Batra. Bi-directional domain adaptation for sim2real transfer of embodied navigation agents. *IEEE Robotics and Automation Letters*, 6(2):2634–2641, 2021. 2

[24] Guan Wang, Haoyi Niu, Desheng Zhu, Jianming Hu, Xianyuan Zhan, and Guyue Zhou. A versatile and efficient reinforcement learning framework for autonomous driving. *arXiv preprint arXiv:2110.11573*, 2021. 2, 4, 8

[25] Jingkang Wang, Ava Pun, James Tu, Sivabalan Manivasagam, Abbas Sadat, Sergio Casas, Mengye Ren, and Raquel Urtasun. Advsim: Generating safety-critical scenarios for self-driving vehicles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9909–9918, 2021. 1, 2

[26] Jingda Wu, Yanxin Zhou, Haohan Yang, Zhiyu Huang, and Chen Lv. Human-guided reinforcement learning with sim-to-real transfer for autonomous navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45:14745–14759, 2023. 1

[27] Philipp Wu, Alejandro Escontrela, Danijar Hafner, Pieter Abbeel, and Ken Goldberg. Daydreamer: World models for physical robot learning. In *Conference on Robot Learning*, pages 2226–2240. PMLR, 2023. 7

[28] Muharrem Ugur Yavas, Tufan Kumbasar, and Nazim Kemal Ure. A real-world reinforcement learning framework for safe and human-like tactical decision-making. *IEEE Transactions on Intelligent Transportation Systems*, 24(11):11773–11784, 2023. 1

[29] Mingfeng Yuan, Jinjun Shan, and Kevin Mi. From naturalistic traffic data to learning-based driving policy: A sim-to-real study. *IEEE Transactions on Vehicular Technology*, 73(1):245–257, 2024. 3

[30] Mario Zanon and Sebastien Gros. Safe reinforcement learning using robust mpc. *IEEE Transactions on Automatic Control*, 66(8):3638–3652, 2021. 2