KnowMoformer: Knowledge-Conditioned Motion Transformer for Controllable Traffic Scenario Simulation

Honglin He Shu Li Jingxuan Yang Linxuan He
Yi Zhang Qiujing Lu Shuo Feng ✉
Department of Automation, Tsinghua University
{hehl21,li-s23,yangjx20,helx20}@mails.tsinghua.edu.cn
{zhyi,qiujinglu,fshuo}@mail.tsinghua.edu.cn

Abstract

*Simulation is of crucial importance for development and testing of autonomous vehicles. To minimize the sim-to-real gap, the simulator should generate realistic scenarios. And to meet diverse needs, the simulator should be controllable to make vehicles follow specific trajectories or rules. Heuristic-based simulators offer strong controllability but lack of realism. Existing data-driven approaches can generate scenarios with human-like behaviors. However, generated scenarios are usually not controllable since the models are black-box. In this work, we introduce **Know**ledge-Conditioned **Mo**tion Trans**former** (KnowMoformer), integrating knowledge of traffic as control to the neural network, to make model offer both of realism and controllability. KnowMoformer incorporates long-term routes and model-based actions to the model by spatial attention. The results demonstrate that KnowMoformer can generate realistic and controllable traffic scenarios.*

1. Introduction

Since deploying autonomous vehicles (AVs) in real-world incurs significant cost and risk, simulation becomes the crucial method for research and development of AVs. The core challenges of simulation are **1)** realism and **2)** controllability. For realism, we want to minimize the sim-to-real gap, so that behaviors of AVs tested in the simulator are very close to its performance in the real-world [15, 32, 38]. For the controllability, we want the simulator can be controlled to generate scenarios that meets diverse needs [41].

There are many simulators such as CARLA [9], AirSim [25], DriveSim [22] that focus on realism of rendering and physics. However, these approaches lack realism since vehicles cannot react to AVs. The challenge still remains in the generation of realistic behaviors of background vehicles (BVs). To address this issue, recent works used conditional generative model to generate trajectories given the scene context by learning the conditional distribution from data [11, 23, 30, 31, 37, 38, 41]. Some works [23, 41] used diffusion model [16, 28] and integrated rules or trajectories as guidance to control the states. However, due to the instability of denoising and manipulation in the latent space, the controllability is still with lots of problems.

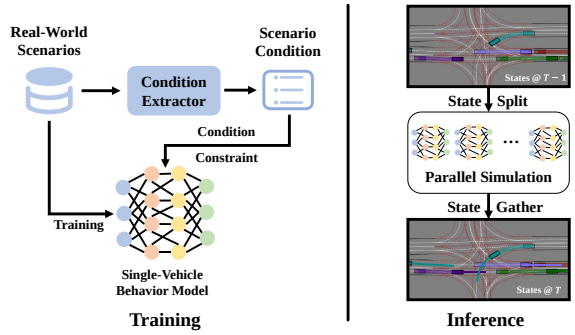

Figure 1. **Overview of training and inference pipeline of our method.** We extract information of route and model-based actions for each scenario as condition to train our single-vehicle behavior model. At each step during inference, the scenario is split to several agent-centric scenarios and simulated in parallel.

To address these challenges, we propose **Know**ledge-Conditioned **Mo**tion Trans**former** (KnowMoformer), a unified framework that takes advantages of both of model-based and data-driven approaches. In KnowMoformer, we adopt two spatial attention mechanism to integrate prior knowledge of agents' movement in traffic scenarios.

Our contributions can be summarized as follows: **1)** We propose a realistic and controllable traffic simulator **KnowMoformer** by integrating long-term route and model-based action to the neural network, which improves the controllability of the simulator when generating scenarios with distribution-level accuracy. **2)** We present a closed-loop training framework to enhance the robustness of the model and a parallel simulation framework for traffic simulation.

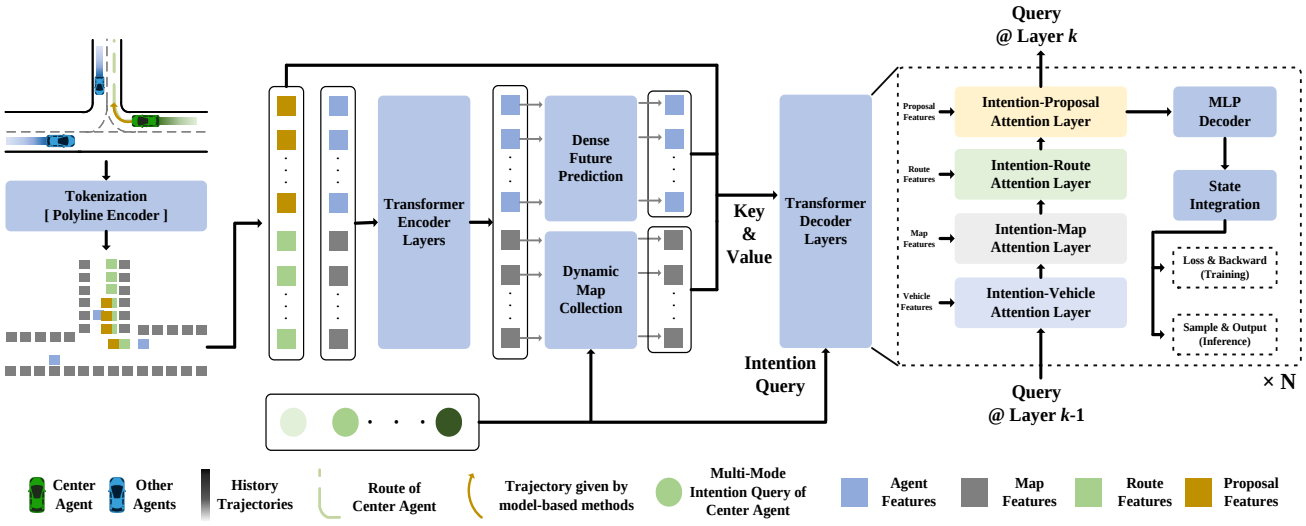

Figure 2. **The architecture of our KnowMoformer.** The input of the model is time-series of states of all agents, spatial information of map, route and IDM output of interested agent. And the output is the distribution of predicted trajectories of interested agent with GMM.

2. Related Work

Motion Prediction. Motion prediction has been a widely studied topic thanks to the development of autonomous driving. There are many similarities between prediction and simulation such as data structure, model design, etc. The prediction task usually takes map information and agent history states in the scenario as input. For different subtasks, there are agent-centric models for single-agent motion forecasting such as Wayformer [21] and MTR [26], as well as scene-level models for multi-agent joint prediction [5, 13, 18, 20, 27, 36, 39, 42]. For scene representation, early works rasterize the scene to image and use CNNs to extract the global feature [2, 4, 6, 8], recent works adopt graph representation [5, 18] or vectorized representation [12] due to their simplicity, efficiency and scalability [14, 26, 27, 29, 39]. For the model architecture, Transformer [34] is widely used in recent works [26, 27, 39, 42].

Recent works based on Waymo Open Motion Dataset (WOMD) [10] focus on long-term forecasting. Although some models achieve promising results, they cannot be used for simulation directly since prediction is fundamentally different from simulation. Prediction is open-loop, *i.e.* producing results within a pre-defined horizon in once but simulation is closed-loop, calling for more robust model that can generate realistic results given synthetic trajectories.

Data-Driven Traffic Simulation. The fidelity of rule-based traffic simulators like SUMO [19] and VISSIM [1] is constrained due to the manually set rules and inherent error of parametric behavior models. Recent approaches focus on data-driven strategies, using neural networks to simulate complex and highly interactive scenarios. Trafficsim [30] adopts implicit latent model [5] to generate socially-consistent actions for all agents jointly. TrafficGen [11] designs a novel vector-based representation for vehicle placement and trajectory generation. Trafficbots [40] integrates destination to the simulator. CTG [41] uses diffusion model to learn the distribution and integrates control to the model by denoising guidance. Mixsim [31] introduced a hierarchical model for mixed reality simulation. MVTA [35] adopted a variant model of MTR [26] that produces trajectories for all agents jointly. RealGen [7] introduced a retriever to query dataset and generate similar scenarios based on retrieved information. Waymax [15] and ScenarioNet [17] introduced efficient and data-uniformed platforms to train traffic models on large-scale data. NeuralNDE [38] proposed a model that can generate complex driving environments with statistical realism, especially for long-tail safety-critical scenarios.

3. Method

3.1. Problem Formulation

The learning objective of the simulator is to learn a model that generates states in next few steps given the scene context (*i.e.*, past agent states and road map states), the objective function can be formulated as

$$\min_{\theta} D_{KL}(P(X_T|X_{0:T-1};m)||P_\theta(X_T|X_{0:T-1};m)) \quad (1)$$

where $X_{0:T-1}$ denotes the past states of all agents in the scene, X_T the next states, m the road map, P and P_θ are real distribution and learned distribution, respectively.

If we assume that back ground vehicles share the same behavior model π and the learned model π_θ, for a scenario with N agents $X_t = (x_{1,t}, ..., x_{N,t})$, then we can obtain

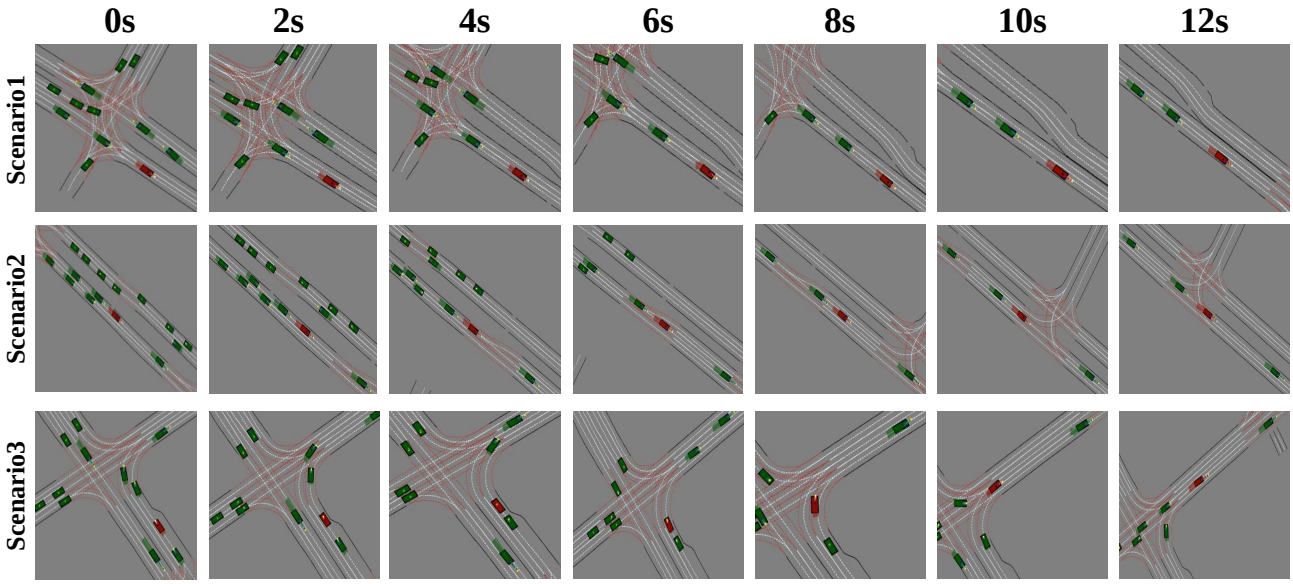

Figure 3. **Simulated traffic scenarios sampled from KnowMoformer.** Agent in red is the randomly selected center for visualization.

$$P(X_T|X^i_{0:T-1};m) = \prod_{i=1}^{N} \pi(x_{i,T}|X^i_{0:T-1};m) \quad (2)$$

where $X^i_{0:T-1}$ denotes the transformed scenario that ith agent in the center position.

So the Eq. 1 can be equivalent to

$$\min_{\theta} D_{KL}(\pi(x_{i,T}|X^i_{0:T-1};m)||\pi_\theta(x_{i,T}|X^i_{0:T-1};m)) \quad (3)$$

Additionally, we assume that each agent has its independent long-term planning that only related to the map, *i.e.*, route $r_i(m)$. And we can obtain a IDM [33]-based trajectory of each agent as the coarse estimation $I_i(X_{0:T-1})$, so the problem can be translated to

$$\min_{\theta} D_{KL}(\pi^c(x_{i,T}|X^i,m,r_i,I_i)||\pi^c_\theta(x_{i,T}|X^i,m,r_i,I_i)) \quad (4)$$

which is our final learning objective function.

3.2. Model Architecture

Main Architecture. The main architecture of KnowMoformer is illustrated in Fig. 2. Our model is agent-centeric, scene context is initialized by processing agents, map, route and IDM-proposal data through polyline encoder as in MTR [26]. Features of agents and map are updated through the transformer encoder. After that, features will be updated by the transformer decoder, each layer contains four cross-attention components, as shown in Fig. 2.

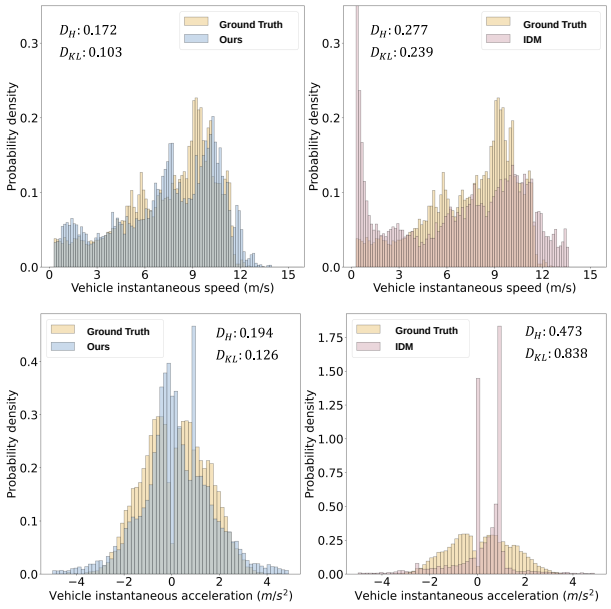

Figure 4. **Statistical realism of normal behaviors.**

Transformer Encoder. Given the past states of agents, road map, route and IDM-proposed-trajectories, we use the vectorized representation [12] and process variables through the polyline encoder which is constructed by MLP with maxpooling [26]. After that, we can obtain the agent feature $A_p \in \mathbf{R}^{N\times D}$, map feature $M_m \in \mathbf{R}^{P_m\times D}$, route feature $M_r \in \mathbf{R}^{P_r\times D}$ and proposal feature $I_p \in \mathbf{R}^{T_p\times D}$, where N is the number of agents, P_m is the number of map polylines, P_r is the number of route polylines, T_p is the time horizon of IDM output and D is feature dimension. The transformer encoder leverages local self-attention [26] on the con-

	minADE ↓	KL_{vel} ↓	KL_{acc} ↓	COLLISION ↓	OFFROAD ↓
IDM [33]	1.059	0.239	0.838	**0.0098**	0.389
MTR [26]	0.753	1.277	0.184	0.0134	0.442
Ours-NR	0.761	0.475	0.134	0.0337	0.414
Ours-NP	**0.549**	0.973	0.421	0.0218	0.397
Ours	0.682	**0.103**	**0.126**	0.0117	**0.382**

Table 1. **Quantitative evaluation on nuPlan [3] Boston.**

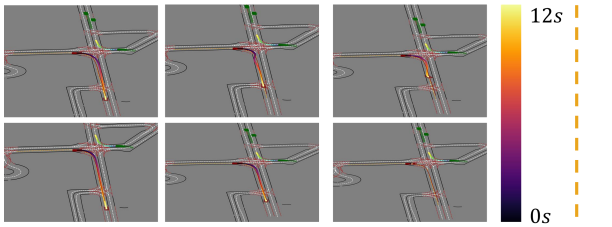

Figure 5. **Random samples given the same route conditions.**

catenated features $[A_p, M_m]$ to generate scene context of decoder layers. We also apply the dense future prediction and dynamic map collection branches proposed by MTR [26].

Transformer Decoder. To generate multi-modal results based on the given scene context, we use intention query[26] as prior, which is a scene-free and pre-computed set of waypoints, indicating the motion intention of the agent. Given the intention query $Q \in \mathbf{R}^{K\times D}$ with K modals, the scene context and knowledge of traffic rules (*i.e.*, route and IDM proposals) will be used as key and values in the cross-attention mechanisms. Specifically, each decoder layer consists of four components, the first and the second ones are as same as the ones in MTR [26]. The third and the final modules are query-route and query-proposal mechanisms, respectively. These two modules are adopted to fuse the features of agents and knowledge priors of traffic scenarios, allowing the model to capture agent-map, agent-agent interactions in a more effective way. In each decoder layer, a prediction head (MLP) is used to produce GMM results $\{p_k; a_{x,k}, a_{y,k}, \sigma_{x,k}, \sigma_{y,k}, \rho_k, dh_k\}_{k=1}^{K}$. It is notable that we use acceleration rather than position as the model output, followed by a action-to-state module. This mechanism is used to ensure that all trajectories are dynamically-feasible [24].

3.3. Training

Training Losses. As proved in [5], Eq. 4 is equivalent to

$$\min_{\theta} -E_{\{X^i, x_{i,T}^{GT}\}\sim P_{real}}[log(\pi_\theta^c(x_{i,T}^{GT}|X^i, m, r_i, I_i))] \quad (5)$$

so the Gaussian regression loss implemented based on the negative log-likelihood loss L_{NLL} is used for distribution of positions. $L1$ regression loss is used for both of heading and velocity of agents. Additionally, $L1$ regression loss is also utilized for the dense future branch [26]. The total training loss of our model is

$$L_{NLL} + \lambda_h L_h + \lambda_v L_v + \lambda_{dense} L_{dense} \quad (6)$$

where $\lambda_h = 20, \lambda_v = 0.5, \lambda_{dense} = 1$. Each decoder layer is trained to predict 2s trajectory with 0.1s interval.

3.4. Simulation

Parallel Simulation. During simulation, we transform the scenario with N agents to N parallel scenarios, and the ith scenario is centered at the lasted position of ith agent. After that, we use the model to infer each agent's action, which will be transformed to the scenario coordinate system through inverse transformation. It's notable that routes and proposals are as the interaction inforamtion given by the simulator, ensuring reasonable interactive behaviors.

Step-by-Step Simulation. During inference, we only use the first 5 steps (0.5s) of the output to ensure that the simulator is interactive. At next unroll step, the input trajectory of each agent is updated by concatenating with last inference results temporally.

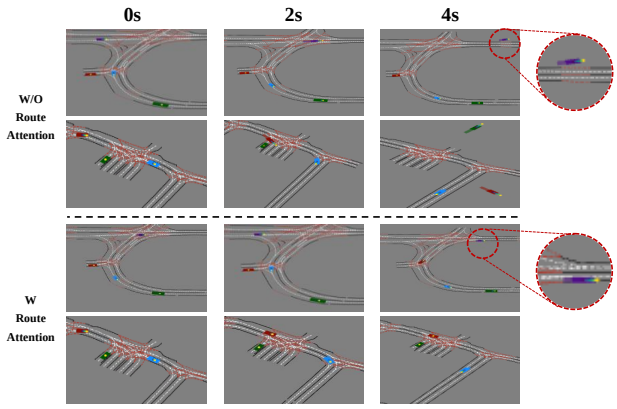

Figure 6. **Ablation on intention-route attention mechanism.**

4. Experiments

4.1. Implementation Details

We use Boston split in nuPlan [3] dataset to train our model. There are 12000, 2157, 855 processed scenarios in the training, validation, and test set, respectively. We extract the route and IDM proposal based on the ground truth data by $L1$ distance between processed results and gt trajectories. For the intention query, we use k-means clustering [26] on the endpoints of gt trajectories.

4.2. Experimental Results and Ablation Study

Qualitative Results. Fig. 3 gives qualitative visualization of complex and diverse scenarios generated by our method. Fig. 6 demonstrates the effectiveness of of our route-integration modules. Fig. 5 shows the controllability of our model.

Quantitative Results. Tab. 1 and Fig. 4 show quantitative results (Ours-NR and Ours-NP are models trained without route- or proposal-attention, respectively). minADE is evaluated within 2s, others are 12s. Our approach achieve the best overall performance. As shown in Fig. 4 and Tab. 1, our model has better distribution consistency with real-world than both of model-based and condition-free learning-based approaches.

5. Conclusion

In this work, we present a traffic simulation model **KnowMoformer** for realistic and controllable traffic scenario generation. Our model employs two well-designed attention-based modules to integrate prior knowledge to the model to address the challenges of insufficient controllability of data-driven approaches. Through qualitative and quantitative results, our method proves its applicability on high-fidelity traffic scenario generation.

References

[1] Vissim, 2012. 2

[2] Yuriy Biktairov, Maxim Stebelev, Irina Rudenko, Oleh Shliazhko, and Boris Yangel. Prank: motion prediction based on ranking. *Advances in neural information processing systems*, 33:2553–2563, 2020. 2

[3] Holger Caesar, Juraj Kabzan, Kok Seang Tan, Whye Kit Fong, Eric Wolff, Alex Lang, Luke Fletcher, Oscar Beijbom, and Sammy Omari. nuplan: A closed-loop ml-based planning benchmark for autonomous vehicles. *arXiv preprint arXiv:2106.11810*, 2021. 4

[4] Sergio Casas, Cole Gulino, Renjie Liao, and Raquel Urtasun. Spagnn: Spatially-aware graph neural networks for relational behavior forecasting from sensor data. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9491–9497. IEEE, 2020. 2

[5] Sergio Casas, Cole Gulino, Simon Suo, Katie Luo, Renjie Liao, and Raquel Urtasun. Implicit latent variable model for scene-consistent motion forecasting. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16*, pages 624–641. Springer, 2020. 2, 4

[6] Sergio Casas, Abbas Sadat, and Raquel Urtasun. Mp3: A unified model to map, perceive, predict and plan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14403–14412, 2021. 2

[7] Wenhao Ding, Yulong Cao, Ding Zhao, Chaowei Xiao, and Marco Pavone. Realgen: Retrieval augmented generation for controllable traffic scenarios. *arXiv preprint arXiv:2312.13303*, 2023. 2

[8] Nemanja Djuric, Vladan Radosavljevic, Henggang Cui, Thi Nguyen, Fang-Chieh Chou, Tsung-Han Lin, Nitin Singh, and Jeff Schneider. Uncertainty-aware short-term motion prediction of traffic actors for autonomous driving. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2095–2104, 2020. 2

[9] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017. 1

[10] Scott Ettinger, Shuyang Cheng, Benjamin Caine, Chenxi Liu, Hang Zhao, Sabeek Pradhan, Yuning Chai, Ben Sapp, Charles R Qi, Yin Zhou, et al. Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9710–9719, 2021. 2

[11] Lan Feng, Quanyi Li, Zhenghao Peng, Shuhan Tan, and Bolei Zhou. Trafficgen: Learning to generate diverse and realistic traffic scenarios. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3567–3575. IEEE, 2023. 1, 2

[12] Jiyang Gao, Chen Sun, Hang Zhao, Yi Shen, Dragomir Anguelov, Congcong Li, and Cordelia Schmid. Vectornet: Encoding hd maps and agent dynamics from vectorized representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11525–11533, 2020. 2, 3

[13] Roger Girgis, Florian Golemo, Felipe Codevilla, Martin Weiss, Jim Aldon D'Souza, Samira Ebrahimi Kahou, Felix Heide, and Christopher Pal. Latent variable sequential set transformers for joint multi-agent motion prediction. *arXiv preprint arXiv:2104.00563*, 2021. 2

[14] Junru Gu, Chen Sun, and Hang Zhao. Densetnt: End-to-end trajectory prediction from dense goal sets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15303–15312, 2021. 2

[15] Cole Gulino, Justin Fu, Wenjie Luo, George Tucker, Eli Bronstein, Yiren Lu, Jean Harb, Xinlei Pan, Yan Wang, Xiangyu Chen, et al. Waymax: An accelerated, data-driven simulator for large-scale autonomous driving research. *Advances in Neural Information Processing Systems*, 36, 2024. 1, 2

[16] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 1

[17] Quanyi Li, Zhenghao Mark Peng, Lan Feng, Zhizheng Liu, Chenda Duan, Wenjie Mo, and Bolei Zhou. Scenarionet: Open-source platform for large-scale traffic scenario simulation and modeling. *Advances in neural information processing systems*, 36, 2024. 2

[18] Ming Liang, Bin Yang, Rui Hu, Yun Chen, Renjie Liao, Song Feng, and Raquel Urtasun. Learning lane graph representations for motion forecasting. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 541–556. Springer, 2020. 2

[19] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. Microscopic traffic simulation using sumo. In *2018 21st international conference on intelligent transportation systems (ITSC)*, pages 2575–2582. IEEE, 2018. 2

[20] Wenjie Luo, Cheol Park, Andre Cornman, Benjamin Sapp, and Dragomir Anguelov. Jfp: Joint future prediction with interactive multi-agent modeling for autonomous driving. In *Conference on Robot Learning*, pages 1457–1467. PMLR, 2023. 2

[21] Nigamaa Nayakanti, Rami Al-Rfou, Aurick Zhou, Kratarth Goel, Khaled S Refaat, and Benjamin Sapp. Wayformer: Motion forecasting via simple & efficient attention networks. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2980–2987. IEEE, 2023. 2

[22] NVIDIA. Nvidia drive sim, 2021. 1

[23] Ethan Pronovost, Meghana Reddy Ganesina, Noureldin Hendy, Zeyu Wang, Andres Morales, Kai Wang, and Nick Roy. Scenario diffusion: Controllable driving scenario generation with diffusion. *Advances in Neural Information Processing Systems*, 36:68873–68894, 2023. 1

[24] Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco Pavone. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*, pages 683–700. Springer, 2020. 4

[25] Shital Shah, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and Service Robotics: Results of the 11th International Conference*, pages 621–635. Springer, 2018. 1

[26] Shaoshuai Shi, Li Jiang, Dengxin Dai, and Bernt Schiele. Motion transformer with global intention localization and local movement refinement. *Advances in Neural Information Processing Systems*, 35:6531–6543, 2022. 2, 3, 4

[27] Shaoshuai Shi, Li Jiang, Dengxin Dai, and Bernt Schiele. Mtr++: Multi-agent motion prediction with symmetric scene modeling and guided intention querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2

[28] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 1

[29] Qiao Sun, Xin Huang, Junru Gu, Brian C Williams, and Hang Zhao. M2i: From factored marginal trajectory prediction to interactive prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6543–6552, 2022. 2

[30] Simon Suo, Sebastian Regalado, Sergio Casas, and Raquel Urtasun. Trafficsim: Learning to simulate realistic multi-agent behaviors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10400–10409, 2021. 1, 2

[31] Simon Suo, Kelvin Wong, Justin Xu, James Tu, Alexander Cui, Sergio Casas, and Raquel Urtasun. Mixsim: A hierarchical framework for mixed reality traffic simulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9622–9631, 2023. 1, 2

[32] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 23–30. IEEE, 2017. 1

[33] Martin Treiber and Arne Kesting. Traffic flow dynamics. *Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg*, pages 983–1000, 2013. 3, 4

[34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 2

[35] Yu Wang, Tiebiao Zhao, and Fan Yi. Multiverse transformer: 1st place solution for waymo open sim agents challenge 2023. *arXiv preprint arXiv:2306.11868*, 2023. 2

[36] Zhongning Wang, Jianwei Zhang, Jicheng Chen, and Hui Zhang. Spatio-temporal context graph transformer design for map-free multi-agent trajectory prediction. *IEEE Transactions on Intelligent Vehicles*, 2023. 2

[37] Danfei Xu, Yuxiao Chen, Boris Ivanovic, and Marco Pavone. Bits: Bi-level imitation for traffic simulation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2929–2936. IEEE, 2023. 1

[38] Xintao Yan, Zhengxia Zou, Shuo Feng, Haojie Zhu, Haowei Sun, and Henry X Liu. Learning naturalistic driving environment with statistical realism. *Nature communications*, 14 (1):2037, 2023. 1, 2

[39] Ye Yuan, Xinshuo Weng, Yanglan Ou, and Kris M Kitani. Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9813–9823, 2021. 2

[40] Zhejun Zhang, Alexander Liniger, Dengxin Dai, Fisher Yu, and Luc Van Gool. Trafficbots: Towards world models for autonomous driving simulation and motion prediction. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1522–1529. IEEE, 2023. 2

[41] Ziyuan Zhong, Davis Rempe, Danfei Xu, Yuxiao Chen, Sushant Veer, Tong Che, Baishakhi Ray, and Marco Pavone. Guided conditional diffusion for controllable traffic simulation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3560–3566. IEEE, 2023. 1, 2

[42] Zikang Zhou, Zihao Wen, Jianping Wang, Yung-Hui Li, and Yu-Kai Huang. Qcnext: A next-generation framework for joint multi-agent trajectory prediction. *arXiv preprint arXiv:2306.10508*, 2023. 2