A Heterogeneous Encoding for Agent Trajectories: HEAT

Eni Solomon Laughter
School of Transportation Engineering
Chang'an University, Xi'an, Shaanxi, China
2024134912@chd.edu.cn

Abstract

Closed-loop traffic behavior for heterogeneous road users (vehicles, two-wheelers, and pedestrians) requires a scene representation that captures not only spatial proximity but anticipatory risk between agent pairs. We present HEAT (Heterogeneous Edge-Aware Trajectory generation), a submission to the HetroD Challenge v1.0. HEAT encodes each scene through a two-layer graph that couples a static map of typed road regions to a per-frame agent interaction graph. Because the dataset contains heterogeneous agents ranging from vehicles and two-wheelers to pedestrians, conventional surrogate measures such as TTC and DRAC become undefined in the frames where agents are not yet closing on each other. CCAR is embedded on cross-type edges to fill this gap, carrying a continuous risk signal across every agent pair at every frame, so the graph encodes how risk is conveyed between agents rather than only where they are. That risk-aware encoding then conditions a diffusion decoder, which generates 32 stochastic 8-second futures per agent through FiLM, so each rollout reflects the interaction context of that specific agent rather than a generic trajectory. The multi-agent interaction evaluation scores four components with fixed weights: Kinematic realism (0.30), Safety (0.35), Cross-type interaction (0.25), and a Coverage bonus (0.10), where the overall score is computed as:

$$\textit{Overall} = 0.30\,K + 0.35\,S + 0.25\,X + 0.10 \cdot C \cdot K \cdot S \quad (1)$$

where K, S, X, and C denote Kinematic, Safety, Cross-type, and Coverage scores respectively. The Coverage bonus is quality-gated: it contributes only when both kinematic and safety fidelity are high. On the held-out validation split, HEAT V2 scores 0.577 overall, clearing both non-learned baselines. The primary remaining weakness is Cross-type interaction (0.518), specifically distance proximity between heterogeneous agent pairs. Because agents are decoded independently, co-simulated agents drift into shared space without awareness of each other's evolving paths. A post-hoc joint sample re-pairing procedure and a Risk-aware Self-Consistent (RaSc) auxiliary supervision objective were introduced to address this, improving the overall score from 0.507 to 0.577.

1. Introduction

Multi-agent trajectory generation has become a foundational challenge in the development of autonomous vehicles, requiring models that can simultaneously represent diverse agent types, capture their interactions, and produce physically plausible futures at scale. A substantial body of work has addressed this challenge through a range of generative architectures from generative adversarial networks and variational autoencoders (VAE) to, more recently, denoising diffusion probabilistic models (DDPM)- whereby each of these generative architectures offers distinct trade-offs in sample quality, diversity, and training stability [1, 7, 23]. While these advances have steadily improved trajectory fidelity and diversity, the challenge of generating coherent multi-agent rollouts in heterogeneous, real-world traffic scenes remains an active frontier [11].

A key design dimension in this space is the choice between open-loop and closed-loop simulation. Open-loop evaluation measures prediction accuracy against recorded ground truth, but does so under the assumption that the environment is static and the generated trajectories do not feed back into the scene, meaning compounding errors in agent behavior go undetected [4, 6]. Closed-loop simulation removes this assumption as agents respond to one another's evolving states, exposing the model to the distribution of futures it actually induces rather than the distribution it was trained on [21]. For the task of generating 32 stochastic rollouts per scenario across all agent types, as required by the HetroD Challenge, the closed-loop setting is the appropriate evaluation regime because it directly rewards behavioral realism and penalizes trajectories that place agents in physically untenable configurations.

Modeling heterogeneous agent interactions in this setting demands a representation capable of capturing structured relationships across agent types and spatial context simultaneously. Graph attention networks (GATs) have emerged as a principled approach to this problem, encoding each agent's relationship to its neighbors through learned, attention-weighted edge aggregation [5, 14, 15]. Unlike fixed-weight graph convolutions, attention-based aggregation allows the model to selectively amplify the edges most relevant to each agent's future, a property particularly valuable in heterogeneous scenes where the relevant interaction partners for a pedestrian and for a vehicle may differ sharply. Prior work has demonstrated the value of this class of model for lane-change conflict prediction, crash risk assessment, and multi-agent trajectory encoding [16].

Beyond interaction modeling, the scene representation must carry explicit risk information at every frame, not only when agents are already closing on each other. Safety is the highest-weighted component in the evaluation (0.35), and a model that cannot represent risk during the build-up phase of a conflict will only ever react to danger rather than anticipate it. Conventional surrogate safety measures such as Time-to-Collision (TTC) and Deceleration Rate to Avoid a Crash (DRAC) fail precisely here: both are undefined for the class of frames where closing motion is absent yet real risk is present [3, 17]. A vehicle following a lead vehicle that is accelerating away creates no collision course in the TTC sense, even though a sudden brake by the leader would immediately expose the follower to danger. Because safety cannot be optimised from a signal that goes silent in the frames that matter most, counterfactual risk reasoning is embedded instead: at every frame, the question asked is not whether agents are closing, but whether the follower could respond if the leader braked now and how risk grows within each pairs.

For the generative stage, diffusion models have demonstrated strong capabilities in modeling multimodal trajectory distributions, avoiding the mode collapse and training instability associated with GAN-based approaches while supporting principled stochastic sampling [12, 23]. Their application to trajectory generation has spanned pedestrian motion, vehicle motion forecasting, and controllable scene generation [18, 19, 22], with recent work establishing diffusion conditioning on scene context as a stable and expressive paradigm [2, 8, 20].

This report presents **HEAT** (**H**eterogeneous **E**dge-**A**ware **T**rajectory generation), a method developed for the HetroD Challenge v1.0 submitted under team *mxr*. HEAT couples a two-layer edge-aware GAT encoder adapted from a prior representational framework of coupled road and interaction graphs [10] with a FiLM-conditioned 1D U-Net diffusion decoder extended from a car-following generative model [9]. Cross-type edges between vehicles and pedestrian-class agents carry generalized Counterfactual Closing-Acceleration Risk (CCAR) values, providing the model with anticipatory risk information that conventional surrogate safety measures cannot supply in the same frames. The decoder operates under a v-prediction parameterization with a rescaled zero-terminal-SNR noise schedule [13], a modification adopted to resolve a generation-collapse failure diagnosed during development. The resulting system generates 32 stochastic 8-second closed-loop rollouts per scenario across all agent types, evaluated against kinematic realism, safety, and cross-type interaction criteria.

2. Method

2.1. Problem Statement

The HetroD Challenge requires generating 32 stochastic 8-second futures for every agent in a mixed urban traffic scene, evaluated on kinematic realism, spatial validity, cross-type interaction quality, and diversity of generated futures. The underlying traffic engineering problem is one of behavioural plausibility of do the generated rollouts reflect the kinds of decisions real road users make, given their type, their location, and the risk posed by nearby agents?

Three requirements follow from this framing. First, the scene representation must distinguish agents by type and location in a way that reflects how behavioural relevance is actually distributed in traffic. A pedestrian mid-crossing and a pedestrian waiting at a kerb occupy similar spatial positions but carry entirely different implications for a nearby vehicle's next action. Second, the risk state of the scene must be encoded at every frame, including frames where conventional safety measures are silent. As risk in traffic does not begin when agents are already closing on each other; it accumulates during the approach phase, and a model that cannot represent that accumulation cannot produce behaviourally realistic rollouts in the frames that precede a conflict. Third, the generative stage must produce a distribution over plausible futures rather than a single predicted trajectory. At unsignalized crossings and shared-space zones, road users face genuinely multimodal situations: a pedestrian may wait, cross, or retreat, and a model that collapses this to one outcome misrepresents the decision space that surrounding agents must plan against.

HEAT addresses these three requirements through a two-stage pipeline: a risk-aware scene encoder that produces a per-agent conditioning embedding, and a generative decoder that samples a trajectory ensemble conditioned on that embedding.

2.2. Method Overview

HEAT couples a two-layer edge-aware Graph Attention Network (GAT) encoder with a FiLM-conditioned 1D U-Net diffusion decoder. The encoder operates once per scenario frame, ingesting a macroscopic map graph, a per-frame agent interaction graph, and the spatial coupling between them, to produce a 128-dimensional conditioning vector for each agent. The decoder performs 300-step reverse diffusion independently per agent, conditioned on that embedding, to produce (x, y, yaw) rollouts across 80 future steps. The two-layer graph representation, comprising a road-level map graph coupled to a per-frame interaction graph, is adapted from a prior representational framework by the author [10]. The GAT encoder and diffusion decoder built on top of that structure are new work developed for this submission.

2.3. Model Architecture

Why GAT. In heterogeneous traffic scenes, the behavioural relevance of nearby agents is asymmetric and context-dependent. A pedestrian waiting at a kerb does not constrain a through-moving vehicle in the way a pedestrian mid-crossing does. A leading vehicle decelerating ahead is more consequential to a following vehicle than a lateral vehicle at the same distance in an adjacent lane. Fixed-weight graph convolutions treat all neighbours within a radius equally, which misrepresents how road users distribute their attention in practice. The GAT attention mechanism learns this selective weighting from data, producing an agent encoding that reflects the interaction priorities a real road user would exercise in the same scene [5, 15]. A vehicle, a two-wheeler, and a pedestrian at the same spatial distance from an ego agent carry entirely different behavioural implications, and the encoder is built to reflect that asymmetry.

Why diffusion. Traffic futures at unsignalized crossings and shared-space zones are genuinely multimodal. A pedestrian may wait, step into the road, or retreat. A two-wheeler may filter through a gap or hold position. Diffusion models recover this distribution by generating samples through iterative denoising from distinct noise initialisations, producing rollouts that span the plausible behavioural space rather than converging on a single modal prediction [11, 12]. The FiLM conditioning mechanism ensures that each rollout is anchored to the same risk-aware scene encoding, so diversity is constrained to behaviourally plausible variation rather than unconstrained scatter.

Why CCAR. Conventional surrogate safety measures such as TTC and DRAC measure risk only when agents are already on a closing trajectory. They are undefined in the frames where a lead vehicle is accelerating away but a sudden brake would immediately expose the follower to danger [3, 17]. This blind region means that standard surrogate measures cannot characterise the risk state of a scene during the build-up phase of a conflict, only during its execution. CCAR (Counterfactual Closing-Acceleration Risk) resolves this by asking a counterfactual question at every frame: if the lead agent applied a reference deceleration now, could the following agent respond in time? This produces a risk signal that is continuous, defined at all frames, and grounded in the same kinematic assumptions used in highway design standards (AASHTO, $b_L = b_F = 3.5$ m/s^2, $t_r = 1.0$ s). For HEAT, CCAR is generalised from vehicle-following pairs to arbitrary heterogeneous agent pairs by projecting each pair's relative position, velocity, and acceleration onto their line-of-sight axis and calling the validated CCAR closed-form kernel on the resulting scalars. Embedding this signal on interaction graph edges means the scene encoding carries information about latent conflict potential, not only observed closing motion.

Macroscopic map graph G. A map graph is built once per recording session from the dataset's `valid_regions` polygons. Nodes are typed by region function: vehicle, cyclist, `pedestrian_core`, `pedestrian_crosswalk`, and `pedestrian_road`. This graph encodes the static spatial structure of the scene and is shared across all scenarios from the same session, providing each agent with a stable reference for where its type is permitted to operate.

Coupling τ. At each frame, every agent is assigned to the map region(s) it physically occupies via point-in-polygon test. Multi-membership is supported: an agent inside the overlap zone of a crosswalk and a road region is flagged for both. The resulting multi-hot region-membership vector forms five of the fourteen node features, grounding each agent in the physical map without requiring lane-snapping or HD map lane connectivity. This was an important design choice given the short ($\sim$4.4 m average) lane segments in this dataset, which made lane-based connectivity fragile in early experiments.

Per-frame interaction graph. Edges are established between all agent pairs within a universal 10 m distance gate, regardless of type combination. Cross-type edges (vehicle/two-wheeler and pedestrian) carry a real-valued generalised CCAR risk score R as an edge feature, representing the latent conflict potential between behaviourally dissimilar agent pairs. Same-type edges, including vehicle-vehicle, carry only structural features (relative position, relative velocity, distance); their interaction weights are left for the GAT to learn from data without hand-engineered risk priors.

GAT encoder. The encoder is a two-layer edge-aware graph attention network. Layer 1 uses 4 attention heads with hidden dimension 64; Layer 2 uses 1 head with output dimension 128, producing the final per-agent conditioning embedding. Node features are 14-dimensional: position (2), velocity (2), heading encoded as $(\cos, \sin)$ (2), agent-type one-hot (3), and τ region-membership flags (5, multi-hot). Edge features are 11-dimensional: relative position (2), relative velocity (2), distance (1), CCAR R (1, nonzero on cross-type edges only), and a cross-type binary flag (1), with remaining structural terms. Edge features are concatenated to source node features before the attention computation at each layer.

Diffusion decoder. The decoder is a 1D U-Net conditioned via FiLM, extending the single-agent car-following architecture from the author's prior work [9] to multi-agent (x, y, yaw) output. The conditioning signal fed to every FiLM block is the concatenation of the GAT embedding (128-dim), the agent-type one-hot (3-dim), and the sinusoidal diffusion timestep embedding. The decoder generates x, y, and yaw jointly over 80 future steps; the z-coordinate is held constant at its last observed value. The model uses v-prediction parameterization with a rescaled zero-terminal-SNR noise schedule following Lin *et al*. [13] across 300 diffusion steps.

2.4. Training Procedure

Training uses the full public train and validation splits of the HetroD dataset (5,087 + 955 scenarios in ScenarioNet format) across 48 recording sessions. After filtering to scenarios with at least one agent whose full 80-step future is valid, 6,041 of 6,042 total scenarios are retained and split 85/15 into 5,135 training and 906 held-out validation scenarios.

Each scenario is conditioned on a single frame (`current_time_index`) with the 80-step real future as the generation target. Target (x, y, yaw) are normalized to zero mean and unit variance using train-split statistics only; denormalization is applied after sampling at inference time to prevent leakage. The model is optimized with Adam at a learning rate of 1×10^{-3}, with gradient clipping at `max_norm` = 1.0 to stabilize diffusion training. The batch unit is a full scenario: all valid agents in a given frame are processed together in a single forward pass, allowing the GAT to encode inter-agent structure over the complete scene graph. Training runs for 30 epochs with checkpointing every epoch; the checkpoint selected for submission achieves the highest overall score on the held-out validation set under the official `hetrod_eval.py` evaluator.

Generation-collapse diagnosis and fix. An earlier model configuration produced near-identical outputs for all agents regardless of identity, with a 99.7–100% collision rate. Diagnosis confirmed the GAT encoder was producing differentiated per-agent embeddings, isolating the fault to the diffusion schedule. The cosine noise schedule did not reach true zero signal at its terminal training step (non-zero terminal SNR), leaving residual real-position signal at the final training step that was absent at actual sampling time. The fix applied three changes simultaneously: target normalization, a rescaled noise schedule following Lin *et al*. [13], and the switch to v-prediction. This raised the overall validation score from 0.397 to 0.507.

2.5. Data and Pretrained Models Used

The primary training data is the HetroD Challenge dataset in ScenarioNet format. Two components from prior work by the author are reused: the validated CCAR closed-form kernel (`ccar.py`, v1.1.0), called directly to compute generalised risk values on arbitrary agent pairs, and the FiLM-conditioned 1D U-Net diffusion architecture from the car-following paper [9], extended from single-agent speed-only to multi-agent (x, y, yaw) output and retrained from scratch on the HetroD dataset. No third-party pretrained weights are used. The one external published technique adopted is the noise schedule rescaling and v-prediction fix from Lin *et al*. [13], implemented from that paper's specification.

2.6. Inference and Rollout-Generation Settings

At inference time, 32 independent rollouts are generated per scenario by sampling distinct initial Gaussian noise vectors and running the full 300-step reverse diffusion process for each. No step-skipping or DDIM-style acceleration is applied. Each agent's 80-step (x, y, yaw) future is produced by its own conditioned diffusion chain anchored to its GAT embedding, with no cross-agent coordination during generation. Rollouts are denormalized using the train-split (x, y, yaw) statistics after sampling. The z-coordinate is filled by copying the agent's last observed value across all 80 future steps. Inference runs on CPU; measured throughput on the held-out validation set is approximately 2 minutes per scenario, with full test-set timing to be confirmed once the 955-scenario generation run completes.

3. Results

All results are reported against the official HetroD evaluation suite on the held-out validation split. The proposed model is compared against two non-learned baselines and a GT Oracle reference.

3.1. Overall Performance

Two honest baselines bound the evaluation from below. A frozen rollout holds every agent at its last observed position for all 80 future steps. A constant-velocity extrapolation projects each agent forward linearly from its observed heading and speed. Both are parameter-free and require no learning. The GT Oracle, which replays actual ground-truth trajectories as simulated rollouts, scores 0.897 overall and represents a ceiling that reflects perfect behavioural fidelity to one observed future. The more meaningful threshold for a learned generative model is clearing both non-learned baselines, because a model that fails to do so has not demonstrated that its scene representation contributes anything. The proposed model scores 0.577 overall, clearing the constant-velocity baseline (0.503) and the frozen baseline (0.488).

3.2. Component-by-Component Breakdown

Table 1 summarises the score breakdown across all four evaluation components. Kinematic realism is 0.612, confirming that the generated motion profiles are not kinematically implausible across vehicles, two-wheelers, and pedestrians.

Safety scores 0.682 overall, the strongest relative improvement over V1. Collision improved from 0.422 to 0.613, driven primarily by the joint sample re-pairing procedure. The RaSc supervision objective contributed by improving the model's attention to high-risk agent pairs: attention-R correlation increased from $+0.146$ to $+0.181$, so the encoder was directing more weight toward edges where conflict potential was highest before decoding began. Valid region conformance improved from 0.702 to 0.751. These two sub-components are addressed separately in Section 3.5.

Cross-type interaction scores 0.518. This component is addressed in the V2 improvements section below.

Coverage scores 0.593. The reduction from V1's 0.687 reflects the CCAR-off configuration adopted in V2, which trades some sample diversity for improved kinematic and safety fidelity.

Table 1. Validation results on the held-out 160-scenario split under the official `hetrod_eval.py` evaluator.

Method	Overall	Kinem.	Safety	X-type	Cov.
Frozen baseline	0.488	–	–	–	–
Constant-velocity	0.503	–	–	–	–
GT Oracle	0.897	1.000	0.992	1.000	0.000
HEAT V1	0.507	0.579	0.562	0.458	0.687
HEAT V2 (ours)	**0.577**	**0.612**	**0.682**	**0.518**	**0.593**

Kinem. = Kinematic realism; X-type = Cross-type interaction; Cov. = Coverage. Dashes indicate metrics not individually reported for non-learned baselines.

3.3. V2 Improvements: CCAR-Off, Re-Pairing, and RaSc

Three targeted interventions improved the overall score from 0.507 to 0.577, each motivated by a specific failure identified in the preceding configuration.

CCAR-off configuration. Although CCAR R was embedded as a passive edge feature in V1, probing the trained model revealed the signal was not being used. The GAT's attention-R correlation was measured at only $+0.146$, while distance dominated attention at a correlation of -0.436. Because the model was routing around the risk signal rather than through it, the feature was adding noise to the embeddings without contributing to the encoding. Removing it improved the overall score from 0.507 to 0.534, with gains in both kinematic and cross-type components.

Joint sample re-pairing. Removing the passive CCAR feature improved embedding quality but left same-sample collision rates elevated, since agents were still decoded independently without awareness of each other's evolving paths. A post-hoc collision resolution procedure was applied, using iterative Hungarian-algorithm-based permutation solving across entangled multi-agent components within each scenario's 32-sample rollout set. Because the procedure operates on the full set of 32 rollouts rather than individual trajectories, it resolves collisions by reassigning which rollout each agent belongs to rather than altering any generated trajectory. This improved the overall score from 0.534 to 0.560.

Risk-aware Self-Consistent (RaSc) auxiliary supervision [?]. Re-pairing addressed collisions post-hoc but did not change how the encoder represented risk. Because the passive CCAR feature had been removed, the model had no mechanism to learn risk structure from the interaction graph during training. RaSc was introduced to close this gap: a separate RiskHead module (MLP: $256 \to 64 \to 1$, Sigmoid) was added alongside the GAT encoder, predicting pairwise CCAR R values from agent embeddings and supervised against ground-truth R computed from the dataset. This shifted the model from a configuration with no active risk signal to one that learns to predict where conflict potential is highest directly from the interaction graph.

Post-training probing confirmed the mechanism worked. Attention-R correlation improved from $+0.146$ to $+0.181$, and mean attention on high-risk edges ($R > 0.5$) rose from 0.244 to 0.271, showing the encoder was now directing weight toward the edges that mattered. Embedding cosine similarity dropped from 0.306 to 0.159, indicating the representations were structurally differentiated rather than collapsed. Combined with re-pairing, this produced the final score of 0.577.

3.4. Coverage Score Analysis

The GT Oracle scores 0.000 on Coverage. Due to a structural consequence of what ground truth is: one observed future, replicated 32 times. When 32 identical copies of a single trajectory are submitted as rollouts, the occupancy distribution across those 32 samples is a point mass with zero diversity, and the Coverage score is zero by construction. The Oracle reports: here is the one thing that actually happened. Coverage asks: across your 32 imagined futures, how much of the plausible behavioural space did you explore while remaining realistic and safe?

A generative model and a deterministic reference are incommensurable on this metric. HEAT samples 32 distinct noise vectors per scenario and decodes 32 genuinely different futures, producing a Coverage score of 0.593. The quality gate in the evaluation, where Coverage is weighted by Kinematic and Safety scores before contributing to the overall figure, means diversity is only rewarded when the generated futures are also kinematically plausible and spatially valid. The fact that the Coverage term contributes positively confirms that the generated rollouts are not reckless scatter across the map.

3.5. Safety Score Analysis

Safety decomposes as $0.5 \times$ Collision $+ 0.5 \times$ Valid Region. The two sub-components reflect different dimensions of behavioural plausibility and should not be read as a unified safety verdict.

Valid Region scores 0.751, reflecting that agents largely remain within their type-designated operating zones: vehicles on carriageway, two-wheelers in the shared corridor, pedestrians in core and crosswalk regions. The macroscopic graph structure, which assigns agents to typed region nodes at each frame, is successfully grounding each agent's generated trajectory in a type-appropriate part of the scene.

Collision scores 0.613. The agent-type breakdown is the more informative figure. In V1, vehicle collision rate reached 70.2% against a GT rate of 4.0%, and two-wheeler collision rate reached 60.9% against a GT rate of 11.8%. The GT rates reflect a structural property of the reference: the Oracle replays one observed future identically across all 32 rollouts, so co-simulated agents never diverge from their recorded paths and collisions are rare by construction. The generated rollouts, by contrast, sample distinct futures for each agent independently, which produces path combinations that never occurred in the recording and creates collision opportunities the GT submission cannot. The joint repairing procedure substantially reduced these rates in V2.

Pedestrian collision rate at 37.8% remains below the GT rate of 42.4%. Real pedestrians in dense crossing environments make physical contact frequently, and the GT Oracle faithfully replays that contact across all 32 copies. The generator, learning from those same scenes, produces futures in which pedestrians spread across the plausible behavioural space rather than converging on the recorded path, resulting in a lower contact rate than the ground truth itself.

The vehicle and two-wheeler rates point to a specific gap in the behavioural representation. In car-following theory, a safe following response requires that the following agent anticipate the leader's future deceleration, not merely react to current speed. The generated rollouts fail on exactly this dimension. Each agent's 80-step future is produced from the scene encoding at the observation frame, but without any representation of where co-simulated agents will be at intermediate timesteps during generation. A lead vehicle that decelerates in its generated trajectory produces no corresponding signal in the following agent's generation process. The result is structurally identical to a following vehicle with zero anticipation horizon: the most dangerous configuration in car-following models, applied across all agent pairs wherever paths converge in the rollout.

3.6. Ablation: Small-Scale Results Misled, Full-Scale Corrected

This ablation was conducted on V1, where CCAR R was still embedded as a passive edge feature. An ablation comparing CCAR-conditioned rollouts against unconditioned rollouts was first evaluated on 18 held-out validation scenarios. At that scale, the unconditioned model outperformed CCAR conditioning on every metric, suggesting the risk signal was harmful to behavioural quality. The full-scale evaluation at 906 scenarios reversed this finding, as shown in Table 2. CCAR conditioning produced a higher overall score (0.397 vs. 0.368), a higher kinematic score (0.479 vs. 0.384), and a higher cross-type score (0.589 vs. 0.582). The 18-scenario result was noise from an unrepresentative sample.

The V2 probing result contextualises what was happening in V1. Even when CCAR conditioning produced a measurable score benefit at full scale, the attention-R correlation of $+0.146$ confirmed the model was not genuinely routing through the risk signal. The feature was contributing indirectly, likely by perturbing the embedding space in ways that helped generalisation, rather than by teaching the encoder to attend to high-risk pairs. This is what motivated the shift in V2 from passive feature inclusion to active RaSc supervision: the full-scale ablation showed CCAR information was useful, and the correlation measurement showed the passive form was not the right way to deliver it.

The methodological point remains direct: ablations on simulation benchmarks with high scenario-level variance require sufficient sample size before any conclusion is drawn. A negative result on 18 scenarios would have been a plausible reason to remove CCAR conditioning entirely. Doing so would have been wrong, and the V2 trajectory confirms it.

Table 2. CCAR ablation results at full scale (906 validation scenarios).

Configuration	Overall	Kinematic	Cross-type
CCAR = Off	0.368	0.384	0.582
CCAR = On	0.397	0.479	0.589

3.7. Collision Breakdown: What the Failure Reveals

Examining the collision geometry rather than the aggregate collision rate produced a finding that contradicted the initial hypothesis. The prior expectation from the scenario geometry, predominantly 4-way unsignalized crossings, was that crossing and turning conflicts would dominate: left-turning vehicles cutting across oncoming paths, two-wheelers executing hook turns across pedestrian desire lines. The breakdown showed the opposite. Rear-end and parallel-path collisions between same-type agents are the majority failure mode, with crossing conflicts secondary.

Rear-end and parallel collisions at this rate are a signature of agents that do not share a representation of each other's future path. Crossing conflicts arise from failures to model gap acceptance and right-of-way. Rear-end conflicts at this scale arise from something more fundamental: agent B has no information about where agent A is going after the observation frame. This is not a gap-acceptance failure. It is the anticipation failure that car-following theory has characterised for decades: the follower cannot respond to a deceleration it has no representation of, regardless of how well the scene was encoded at the start of the rollout.

The valid-region result and the collision result together describe the representation accurately. The model knows where the road is. It knows which agents pose a cross-type risk at the observation frame. It does not carry forward a representation of where co-simulated agents will be during the rollout itself. The information is present at encoding time but is not propagated through the generation process. Closing that gap requires joint decoding, not improved encoding.

3.8. Generation Failure Found Mid-Project, Diagnosed, and Fixed

During development, rollout inspection revealed that generated agents were effectively stationary, producing trajectories in which (x, y) positions changed negligibly across all 80 future steps despite the model receiving valid kinematic history. The cause was a normalization failure. Scene coordinates were not normalized to an agent-centric frame before being passed to the model, leaving absolute (x, y) values in a range the network could not meaningfully differentiate. The model learned to predict near-zero displacements because that minimized loss on unnormalized coordinates, producing rollouts in which all agents remain at their last observed position regardless of their actual kinematic state.

Normalizing agent positions relative to each agent's own reference frame at the observation endpoint resolved the failure. Safety improved from 0.298 to 0.562 as a direct consequence. This is reported as part of the experimental record rather than set aside, because the failure mode will recur in any pipeline that ingests raw global coordinates without agent-centric normalization, and the before-and-after numbers confirm the fix was the operative change.

4. Limitations and Future Work

Several directions were investigated during development that did not improve beyond the reported result. They are documented here for reproducibility.

Extending RaSc supervision to vehicle-vehicle pairs (same-type VV CCAR) increased Coverage from 0.593 to 0.705 but reduced both Safety and Cross-type scores, producing a net loss of 0.037 on the overall figure. The gain in diversity was real but came at the cost of precision in the components that carry the higher evaluation weights.

A broader attention supervision approach was also tested, in which all attention weights were directly supervised proportional to R across every edge in the graph. This boosted the absolute magnitude of attention scores but produced no improvement in discrimination between high-risk and low-risk edges, and the overall score degraded. The model learned to attend more strongly in general rather than more selectively to the edges that mattered.

A sparse variant of attention supervision was then evaluated, restricting supervision to only the highest-R neighbour per agent while explicitly suppressing low-R edges. This produced partial distance decoupling, with $\mathrm{corr}(\mathrm{attn}, \mathrm{dist})$ shifting from -0.489 to -0.204, but the correlation gain between attention and risk remained marginal.

The most direct path to closing the remaining Cross-type gap is multimodal behavioural branching conditioned on predicted R, where yield and assert modes are sampled from $P(M \mid R)$ and calibrated against real conditional rates observed in behavioural clustering of the dataset (crosswalk yield: 27% real opportunity rate; car-following response: 52% aggregate). This would replace the current single-mode generation per agent with a mechanism that commits to a behavioural intent before decoding the trajectory, which is the structural change the cross-type proximity scores are sensitive to. Predecessor tracing for vehicle agents in high-risk situations, following the approach of Liu *et al*. [**?**], is a second candidate that could improve anticipatory response in following scenarios specifically. The full 955-scenario validation evaluation was running at time of submission; all results reported here are from the 160-scenario held-out development subset.

AI Assistance Disclosure

Generative AI was used in the preparation of this manuscript, limited to drafting, editing, and refining the written presentation of ideas. All technical content, methodological choices, experimental results, and arguments are the product of the author's own work and reasoning, developed through an iterative process in which AI-assisted drafts were critically reviewed, corrected, and shaped by the author at every stage.

References

[1] Sam Bond-Taylor, Adam Leach, Yang Long, and Chris G. Willcocks. Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models. *IEEE TRANSACTIONS ON*

PATTERN ANALYSIS AND MACHINE INLIGENCE, 44(11): 7327–7347, 2022. Type: Review. 1

[2] Entao Chang, Jiawei Fu, Wenjie Gao, Shitao Chen, Jingmin Xin, and Nanning Zheng. CNDM: Customized noise diffusion model for trajectory prediction. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, 2026. Type: Article; Early Access tex.affiliation: Chen, ST (Corresponding Author), Xi An Jiao Tong Univ, Natl Engn Res Ctr Visual Informat & Applicat, Natl Key Lab Human Machine Hybrid Augmented Intell, Xian 710049, Shaanxi, Peoples R China. Chen, ST (Corresponding Author), Xi An Jiao Tong Univ, Inst Artificial Intelligence & Robot, Xian 710049, Shaanxi, Peoples R China. Chang, Entao; Gao, Wenjie; Chen, Shitao; Xin, Jingmin; Zheng, Nanning, Xi An Jiao Tong Univ, Natl Engn Res Ctr Visual Informat & Applicat, Natl Key Lab Human Machine Hybrid Augmented Intell, Xian 710049, Shaanxi, Peoples R China. Chang, Entao; Gao, Wenjie; Chen, Shitao; Xin, Jingmin; Zheng, Nanning, Xi An Jiao Tong Univ, Inst Artificial Intelligence & Robot, Xian 710049, Shaanxi, Peoples R China. Fu, Jiawei, Chinese Univ Hong Kong, Hong Kong, Peoples R China. tex.affiliations: Xi'an Jiaotong University; Xi'an Jiaotong University; Chinese University of Hong Kong tex.author-email: qq969827455@stu.xjtu.edu.cn jwfu@cse.cuhk.edu.hk gaowenjie999@gmail.com chenshitao@xjtu.edu.cn jxin@mail.xjtu.edu.cn nnzheng@mail.xjtu.edu.cn tex.cited-references: BLOM HAP, 1988, IEEE T AUTOMAT CONTR, V33, P780, DOI 10.1109/9.1299. Casas S, 2021, PROC CVPR IEEE, P14398, DOI 10.1109/CVPR46437.2021.01417. Chai Y., 2019, ARXIV. Chen GY, 2021, 2021 IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV 2021), P15560, DOI 10.1109/ICCV48922.2021.01529. Chen J., 2023, P ADV NEUR INF PROC, P1863. Cui A, 2023, IEEE INT CONF ROBOT, P7801, DOI 10.1109/ICRA48891.2023.10160984. Deo N, 2018, IEEE COMPUT SOC CONF, P1549, DOI 10.1109/CVPRW.2018.00196. Ettinger S, 2021, 2021 IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV 2021), P9690, DOI 10.1109/ICCV48922.2021.00957. Gao X, 2023, IEEE ROBOT AUTOM LET, V8, P2946, DOI 10.1109/LRA.2023.3262150. Gilles T., 2022, PROC INT C LEARN REP. Gilles T, 2022, IEEE INT CONF ROBOT, P9107, DOI 10.1109/ICRA46639.2022.9812253. Gilles T, 2021, IEEE INT C INTELL TR, P500, DOI 10.1109/ITSC48978.2021.9564944. Gu JR, 2021, 2021 IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV 2021), P15283, DOI 10.1109/ICCV48922.2021.01502. Gu TP, 2022, PROC CVPR IEEE, P17092, DOI 10.1109/CVPR52688.2022.01660. Gupta A, 2018, PROC CVPR IEEE, P2255, DOI 10.1109/CVPR.2018.00240. Ho Jonathan, 2020, ADV NEURAL INFORM PR, V33, P6840. Huang Z., 2022, ARXIV. Huang ZY, 2023, IEEE I CONF COMP VIS, P3880, DOI 10.1109/ICCV51070.2023.00361. Huang ZY, 2024, IEEE T NEUR NET LEAR, V35, P15222, DOI 10.1109/TNNLS.2023.3283542. Jiang C, 2023, PROC CVPR IEEE, P9644, DOI 10.1109/CVPR52729.2023.00930. Kang M, 2024, IEEE T INTELL TRANSP, V25, P12285, DOI 10.1109/TITS.2024.3381631. Karras T, 2022, ADV NEUR IN. Konev S., 2022, ARXIV. Liu YJ, 2025, SENSORS-BASEL, V25, DOI 10.3390/s25154603. Loshchilov I., 2017, ARXIV. Luo ST, 2021, PROC CVPR IEEE, P2836, DOI 10.1109/CVPR46437.2021.00286. Mangalam K, 2020, Img Proc Comp Vis Re, V12347, P759, DOI 10.1007/978-3-030-58536-5_45. Mo X., 2021, PROC WORKSHOP AUTODR, V6, P1. Ngiam J., 2022, P INT C LEARN REPR. Park D., 2023, ARXIV. Poole B., 2022, ARXIV. Prevost CG, 2007, P AMER CONTR CONF, P4123. Sadeghian A, 2019, PROC CVPR IEEE, P1349, DOI 10.1109/CVPR.2019.00144. Shi Shaoshuai, 2022, ADV NEURAL INFORM PR, V35, P6531. Sun Q, 2022, PROC CVPR IEEE, P6533, DOI 10.1109/CVPR52688.2022.00643. Treiber M, 2000, PHYS REV E, V62, P1805, DOI 10.1103/PhysRevE.62.1805. Wagner R., 2023, ARXIV. Wang B., 2025, ARXIV. Wang MK, 2023, IEEE INT CONF ROBOT, P1609, DOI 10.1109/ICRA48891.2023.10160468. Wang RZ, 2023, AAAI CONF ARTIF INTE, P4685. Wilson B., 2023, ARXIV. Wu D., 2021, ARXIV. Zeng WY, 2019, PROC CVPR IEEE, P8652, DOI 10.1109/CVPR.2019.00886. Zhu YY, 2023, IEEE I CONF COMP VIS, P8226, DOI 10.1109/ICCV51070.2023.00759. tex.da: 2026-05-22 tex.doc-delivery-number: GZ2GI tex.earlyaccessdate: APR 2026 tex.eissn: 1558-0016 tex.funding-acknowledgement: National Natural Science Foundation of China [62088102, 62106192] tex.funding-text: This work was supported by the National Natural Science Foundation of China under Grant 62088102 and Grant 62106192. tex.journal-iso: IEEE Trans. Intell. Transp. Syst. tex.number-of-cited-references: 44 tex.research-areas: Engineering; Transportation tex.researcherid-numbers: Chen, Shitao/KDM-6589-2024 tex.times-cited: 0 tex.unique-id: WOS:001748557500001 tex.usage-count-last-180-days: 0 tex.usage-count-since-2013: 0 tex.web-of-science-categories: Engineering, Civil; Engineering, Electrical & Electronic; Transportation Science & Technology tex.web-of-science-index: Science Citation Index Expanded (SCI-EXPANDED). 2

[3] Tanmay Das, M. Shoaib Samandar, Meagan Kittle Autry, and Nagui M. Rouphail. Surrogate Safety Measures: Review and Assessment in Real-World Mixed Traditional and Autonomous Vehicle Platoons. *IEEE ACCESS*, 11:32682–32696, 2023. Inst Transportat Res & Educ. 2, 3

[4] Boris Ivanovic, Yifeng Lin, Shubham Shrivastava, Punarjay Chakravarty, and Marco Pavone. Propagating State Uncertainty Through Trajectory Forecasting, 2022. arXiv:2110.03267 [cs.RO]. 1

[5] Xiaosong Jia, Penghao Wu, Li Chen, Yu Liu, Hongyang Li, and Junchi Yan. HDGT: Heterogeneous Driving Graph Transformer for Multi-Agent Trajectory Prediction via Scene Encoding. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INLIGENCE*, 45(11):13860–13875, 2023. Type: Article. 2, 3

[6] Peter Karkus, Boris Ivanovic, Shie Mannor, and Marco Pavone. DiffStack: A Differentiable and Modular Control

Stack for Autonomous Vehicles, 2022. arXiv:2212.06437 [cs.RO]. 1

[7] Jimin Kim and Jongwoo Song. A review on the generative models and its performance metrics. *COMMUNICATIONS FOR STATISTICAL APPLICATIONS AND METHODS*, 32 (2):235–248, 2025. Type: Review. 1

[8] Zhengxing Lan, Lingshan Liu, Yilong Ren, Zhiyong Cui, and Haiyang Yu. Diffutory: a future feature and mode association augmented diffusion model for trajectory prediction in autonomous vehicles. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, 2025. Type: Article; Early Access tex.affiliation: Ren, YL; Yu, HY (Corresponding Author), Beihang Univ, Sch Transportat Sci & Engn, Beijing 100191, Peoples R China. Lan, Zhengxing; Liu, Lingshan; Ren, Yilong; Cui, Zhiyong; Yu, Haiyang, Beihang Univ, Sch Transportat Sci & Engn, Beijing 100191, Peoples R China. Lan, Zhengxing; Liu, Lingshan; Ren, Yilong; Cui, Zhiyong; Yu, Haiyang, State Key Lab Intelligent Transportat Syst, Beijing 100191, Peoples R China. Ren, Yilong; Yu, Haiyang, Zhongguancun Lab, Beijing 100194, Peoples R China. Ren, Yilong; Yu, Haiyang, Beihang Univ, Hangzhou Innovat Inst, Hangzhou 310023, Peoples R China. tex.affiliations: Beihang University; Zhongguancun Laboratory; Beihang University tex.author-email: zhengxlan@buaa.edu.cn llsbuaa666@buaa.edu.cn yilongren@buaa.edu.cn zhiyongc@buaa.edu.cn hyyu@buaa.edu.cn tex.da: 2026-05-22 tex.doc-delivery-number: 4IN5D tex.earlyaccessdate: JUN 2025 tex.eissn: 1558-0016 tex.funding-acknowledgement: National Key Research and Development Project of China [2022YFB4300400] tex.funding-text: This work was supported by the National Key Research and Development Project of China under Grant 2022YFB4300400. The Associate Editor for this article was R. Jhaveri. tex.journal-iso: IEEE Trans. Intell. Transp. Syst. tex.number-of-cited-references: 77 tex.research-areas: Engineering; Transportation tex.researcherid-numbers: Yu, Haiyang/JFK-3502-2023 Ren, Yilong/MXL-4162-2025 Cui, Zhiyong/AGV-6207-2022 tex.times-cited: 2 tex.unique-id: WOS:001518840500001 tex.usage-count-last-180-days: 10 tex.usage-count-since-2013: 26 tex.web-of-science-categories: Engineering, Civil; Engineering, Electrical & Electronic; Transportation Science & Technology tex.web-of-science-index: Science Citation Index Expanded (SCI-EXPANDED). 2

[9] Eni Solomon Laughter. *Generative Scenario Expansion for ADAS Safety Evaluation: A Conditional Diffusion Model for Uncertainty-Aware Car-Following Risk*. 2026. 2, 4

[10] Eni Solomon Laughter. *Multi-Layer Graph Representation of Road and Traffic for Autonomous Driving*. 2026. 2, 3

[11] Chaoneng Li, Xiaolong Wang, Shuxu Zhao, Xiaohu Wang, and Ze Ye. DiffMATP: Interaction-Aware Multi-Agent Trajectory Prediction via Denoising Diffusion Models. *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, 75(5): 7222–7235, 2026. Type: Article. 1, 3

[12] Lequan Lin, Zhengkun Li, Ruikun Li, Xuliang Li, and Junbin Gao. Diffusion models for time-series applications: a survey. *FRONTIERS OF INFORMATION TECHNOLOGY & ELECTRONIC ENGINEERING*, 25(1):19–41, 2024. Type: Review. 2, 3

[13] Shanchuan Lin, Bingchen Liu, Jiashi Li, and Xiao Yang. Common Diffusion Noise Schedules and Sample Steps are Flawed. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5392–5399, 2024. ISSN: 2642-9381. 2, 4

[14] Xiaojia Liu, Jun Yang, Dexin Yu, Bowei Liu, and Yu Yang. A GAT-Transformer-Based Framework for Lane-Change Conflict Prediction in Highway Merging and Diverging Zones. *IEEE ACCESS*, 13:144143–144158, 2025. Type: Article. 2

[15] Shihan Ma and Jidong J. Yang. Modeling Road User Interactions with Dynamic Graph Attention Networks for Traffic Crash Prediction. *APPLIED SCIENCES-BASEL*, 16(3), 2026. Type: Article. 2, 3

[16] Arnav Vaibhav Malawade, Shih-Yuan Yu, Brandon Hsu, Deepan Muthirayan, Pramod P. Khargonekar, and Mohammad Abdullah Al Faruque. Spatiotemporal Scene-Graph Embedding for Autonomous Vehicle Collision Prediction. *IEEE INTERNET OF THINGS JOURNAL*, 9(12):9379–9388, 2022. Type: Article. 2

[17] Dimitrios Nikolaou, Apostolos Ziakopoulos, and George Yannis. A review of surrogate safety measures uses in historical crash investigations. *Sustainability*, 15(9), 2023. 2, 3

[18] Luke Rowe, Roger Girgis, Anthony Gosselin, Liam Paull, Christopher Pal, and Felix Heide. Scenario dreamer: Vectorized latent diffusion for generating driving simulation environments. In *2025 ieee/cvf conference on computer vision and pattern recognition (cvpr)*, pages 17207–17218, 10662 LOS VAQUEROS CIRCLE, PO BOX 3014, LOS ALAMITOS, CA 90720-1264 USA, 2025. IEEE COMPUTER SOC / Institute of Electrical and Electronics Engineers Inc; IEEE Computer Society. Type: Proceedings Paper tex.affiliation: Rowe, L (Corresponding Author), Mila, Montreal, PQ, Canada. Rowe, L (Corresponding Author), Univ Montreal, Montreal, PQ, Canada. Rowe, L (Corresponding Author), Torc Robot, Blacksburg, VA 24060 USA. Rowe, Luke; Girgis, Roger; Gosselin, Anthony; Paull, Liam; Pal, Christopher, Mila, Montreal, PQ, Canada. Rowe, Luke; Paull, Liam; Pal, Christopher, Univ Montreal, Montreal, PQ, Canada. Girgis, Roger; Gosselin, Anthony; Pal, Christopher, Polytech Montreal, Montreal, PQ, Canada. Heide, Felix, Princeton Univ, Princeton, NJ USA. Paull, Liam; Pal, Christopher, CIFAR AI Chair, Toronto, ON, Canada. Rowe, Luke; Girgis, Roger; Heide, Felix, Torc Robot, Blacksburg, VA 24060 USA. tex.affiliations: Mila Quebec Artificial Intelligence Institute; Universite de Montreal; Universite de Montreal; Polytechnique Montreal; Princeton University tex.bookgroup-author: IEEE COMPUTER SOC tex.cited-references: Abramson J, 2024, NATURE, V630, DOI [10.1038/s41586-024-07487-w, 10.1038/s41586-024-08416-7]. Alonso Eloi, 2024, NeurIPS, V2. Bergamini L, 2021, IEEE INT CONF ROBOT, P5119, DOI 10.1109/ICRA48506.2021.9561666. Bruce Jake, 2024, ICML. Cai Panpan, 2020, ICRA. Chen D, 2024, IEEE T INTELL VEHICL, V9, P4730, DOI 10.1109/TIV.2024.3367919. Chen J., 2023, NeurIPS, V2, P5. Chen Shaoyu, 2024, arXiv. Chen Yurui, 2025,

IEEE TPAMI, P2. Chitta Kashyap, 2024, ECCV. Chiyu Max, 2023, CVPR. Cornelisse Daphne, 2024, Reinforcement Learning Journal, V1, P2. Dosovitskiy A, 2017, PR MACH LEARN RES, V78. Ettinger Scott, 2021, ICCV. Feng L, 2023, IEEE INT CONF ROBOT, P3567, DOI 10.1109/ICRA48891.2023.10160296. FHWA, 2022, Average annual miles per driver by age group. Gao JY, 2020, PROC CVPR IEEE, P11522, DOI 10.1109/CVPR42600.2020.01154. Gao Shenyuan, 2024, NeurIPS. Girgis Roger, 2022, ICLR. Gulino Cole, 2023, NeurIPS. Guo Zhiming, 2023, ArXiv, abs/2309.00615, P2. Higgins Irina, 2017, ICLR. Hoogeboom Emiel, 2022, ICML. Hu Anthony, 2023, arXiv. Hu Yihan, 2024, ECCV. Huang Zhiyu, 2024, arXiv. Igl Maximilian, 2022, ICRA. Jiang Bo, 2023, ICCV. Jiang Chiyu Max, 2024, NeurIPS. Karnchanachari Napat, 2024, ICRA. Kazemkhani Saman, 2025, ICLR. Kesting A, 2007, TRANSPORT RES REC, P86, DOI 10.3141/1999-10. Kingma Diederik P., 2014, ICLR. Lavington Jonathan Wilder, 2024, arXiv, P2. Lee Kuang-Huei, 2022, ADV NEURAL INFORM PR. Leurent Edouard, 2018, An environment for autonomous driving decision-making. Li QY, 2023, IEEE T PATTERN ANAL, V45, P3461, DOI 10.1109/TPAMI.2022.3190471. Li Yueyuan, 2024, IEEE Trans. Intell. Veh., V1. Liang M, 2020, Img Proc Comp Vis Re, V12347, P541, DOI 10.1007/978-3-030-58536-5_32. Liao BC, 2025, INT J COMPUT VISION, V133, P1352, DOI 10.1007/s11263-024-02235-z. Liao Bencheng, 2023, ICLR. Ljungbergh William, 2024, ECCV. Lu Jack, 2024, ICRA, P6. Mahjourian Reza, 2024, ICRA. Mi Lu, 2021, CVPR. Ngiam Jiquan, 2022, ICLR. Ost J, 2021, PROC CVPR IEEE, P2855, DOI 10.1109/CVPR46437.2021.00288. Peebles W, 2023, IEEE I CONF COMP VIS, P4172, DOI 10.1109/ICCV51070.2023.00387. Philion Jonah, 2024, ICLR. Pronovost E, 2023, ADV NEUR IN. Ramesh Aditya, 2022, arXiv. Rowe Luke, 2024, PR MACH LEARN RES, V270. Rowe Luke, 2023, CVPR. Scanlon JM, 2021, ACCIDENT ANAL PREV, V163, DOI 10.1016/j.aap.2021.106454. Schulman John, 2017, arXiv. Sun S, 2024, IEEE ROBOT AUTOM LET, V9, P7007, DOI 10.1109/LRA.2024.3416792. Suo S, 2021, PROC CVPR IEEE, P10395, DOI 10.1109/CVPR46437.2021.01026. Suo Simon, 2023, CVPR. Tan Shuhan, 2021, CVPR. Tan Shuhan, 2023, CORL. Tonderski Adam, 2024, CVPR. Treiber M, 2000, PHYS REV E, V62, P1805, DOI 10.1103/PhysRevE.62.1805. Valevski Dani, 2025, ICLR. Vaswani Ashish, 2017, NeurIPS. Vinitsky Eugene, 2022, ADV NEURAL INFORM PR. Wu Jialong, 2024, NEURIPS. Wu Wei, 2024, NEURIPS. Xie Ziyang, 2023, ICLR. Xu Danfei, 2023, ICRA. Xu Minkai, 2023, ICML. Yan Yunzhi, 2024, ECCV. Yang JM, 2026, INT J MATH, V37, DOI 10.1142/S0129167X24400020. Yang Sherry, 2024, ICLR. Yang Xuemeng, 2024, arXiv, V1, P2. Yang Ze, 2023, CVPR. Yu Youwei, 2024, CORL. Zhang Zhejun, 2023, ICRA. Zhong ZY, 2023, IEEE INT CONF ROBOT, P3560, DOI 10.1109/ICRA48891.2023.10161463. Zhong Ziyuan, 2023, CORL. Zhou Hongyu, 2024, CVPR. Zhou Hongyu, 2024, arXiv, P2. Zhou Ming, 2020, Smarts: Scalable multi-agent reinforcement learning training school for autonomous driving. Zhou XY, 2024, PROC CVPR IEEE, P21634, DOI 10.1109/CVPR52733.2024.02044. Zhou ZK, 2022, PROC CVPR IEEE, P8813, DOI 10.1109/CVPR52688.2022.00862. Zhou Zikang, 2024, NEURIPS. tex.da: 2026-05-22 tex.doc-delivery-number: BZ5ZI tex.funding-acknowledgement: NSF CAREER Award [2047359]; Packard Foundation Fellowship; Sloan Research Fellowship; Sony Young Faculty Award; Project X Innovation Award; Amazon Science Research Award; IVADO; NSERC; CIFAR AI Chairs programs tex.funding-text: Felix Heide was supported by an NSF CAREER Award (2047359), a Packard Foundation Fellowship, a Sloan Research Fellowship, a Sony Young Faculty Award, a Project X Innovation Award, and an Amazon Science Research Award. C. Pal thanks the IVADO, the NSERC Discovery Grants and the CIFAR AI Chairs programs for their support. tex.keywords-plus: DIVERSE tex.number-of-cited-references: 85 tex.research-areas: Computer Science tex.times-cited: 1 tex.unique-id: WOS:001601141700484 tex.usage-count-last-180-days: 4 tex.usage-count-since-2013: 4 tex.web-of-science-categories: Computer Science, Artificial Intelligence; Computer Science, Interdisciplinary Applications; Computer Science, Theory & Methods tex.web-of-science-index: Conference Proceedings Citation Index - Science (CPCI-S). 2

[19] Yiming Xu, Hao Cheng, and Monika Sester. Controllable diverse sampling for diffusion based motion behavior forecasting. In *2024 35th ieee intelligent vehicles symposium, ieee iv 2024*, pages 2397–2404, 345 E 47TH ST, NEW YORK, NY 10017 USA, 2024. IEEE / IEEE; IEEE Intelligent Transportat Soc. Type: Proceedings Paper tex.affiliation: Xu, YM (Corresponding Author), Leibniz Univ Hannover, Inst Cartog & Geoinformat, Appelstr 9a, D-30167 Hannover, Germany. Xu, Yiming; Sester, Monika, Leibniz Univ Hannover, Inst Cartog & Geoinformat, Appelstr 9a, D-30167 Hannover, Germany. Cheng, Hao, Univ Twente, Fac Geoinformat Sci & Earth Observat, NL-7500 AE Enschede, Netherlands. tex.affiliations: Leibniz University Hannover; University of Twente tex.author-email: Yiming.Xu@ikg.uni-hannover.de h.cheng-2@utwente.nl Monika.Sester@ikg.uni-hannover.de tex.book-group-author: IEEE tex.cited-references: Bansal M., 2018, ARXIV. Barquero G, 2023, IEEE I CONF COMP VIS, P2317, DOI 10.1109/ICCV51070.2023.00220. Chai Y., 2019, ARXIV. Chang MF, 2019, PROC CVPR IEEE, P8740, DOI 10.1109/CVPR.2019.00895. Choi J, 2021, 2021 IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV 2021), P14347, DOI 10.1109/ICCV48922.2021.01410. Cui HG, 2019, IEEE INT CONF ROBOT, P2090, DOI [10.1109/ICRA.2019.8793868, 10.1109/icra.2019.8793868]. Deo Nachiket., 2022, CORL PMLR, P203. Dhariwal P, 2021, ADV NEUR IN, V34. Feng C, 2023, IEEE ROBOT AUTOM LET, V8, P6795, DOI [10.1109/lra.2023.3311351, 10.1109/LRA.2023.3311351]. Gao JY, 2020, PROC CVPR IEEE, P11522, DOI 10.1109/CVPR42600.2020.01154. Gao X, 2023, IEEE ROBOT AUTOM LET, V8, P2946,

DOI 10.1109/LRA.2023.3262150. Gilles T., 2022, INT C LEARN REPR. Gilles T, 2021, IEEE INT C INTELL TR, P500, DOI 10.1109/ITSC48978.2021.9564944. Gu TP, 2022, PROC CVPR IEEE, P17092, DOI 10.1109/CVPR52688.2022.01660. Gupta A, 2018, PROC CVPR IEEE, P2255, DOI 10.1109/CVPR.2018.00240. Ho J., ARXIV. Ho Jonathan, 2022, arXiv. Ho Jonathan, 2020, ADV NEURAL INFORM PR, V33, P6840. Janner M, 2022, PR MACH LEARN RES. Jiang C, 2023, PROC CVPR IEEE, P9644, DOI 10.1109/CVPR52729.2023.00930. Liang RQ, 2024, IEEE T NEUR NET LEAR, V35, P16504, DOI 10.1109/TNNLS.2023.3294998. Lin CH, 2023, PROC CVPR IEEE, P300, DOI 10.1109/CVPR52729.2023.00037. Liu M, 2023, ARXIV. Liu MM, 2023, IEEE INT CONF COMP V, P3245, DOI 10.1109/ICCVW60793.2023.00349. Liu YC, 2021, PROC CVPR IEEE, P7573, DOI 10.1109/CVPR46437.2021.00749. Loshchilov I., 2017, INT C LEARNING REPRE, DOI [10.48550/arXiv.1608.03983, DOI 10.48550/ARXIV.1608.03983, 10.48550/arxiv.1608.03983]. Loshchilov I., 2019, INT C LEARN REPR, DOI DOI 10.48550/ARXIV.1711.05101. Makansi O, 2019, PROC CVPR IEEE, P7137, DOI 10.1109/CVPR.2019.00731. Mangalam K, 2020, Img Proc Comp Vis Re, V12347, P759, DOI 10.1007/978-3-030-58536-5_45. Mao WB, 2023, PROC CVPR IEEE, P5517, DOI 10.1109/CVPR52729.2023.00534. Pellegrini S, 2009, IEEE I CONF COMP VIS, P261, DOI 10.1109/ICCV.2009.5459260. Poole B., 2022, arXiv. Ramesh Aditya, 2022, arXiv preprint arXiv:2204.06125, P3, DOI DOI 10.48550/ARXIV.2204.06125. Reynolds DouglasA., 2009, Encyclopedia of biometrics, V741, P659, DOI [10.1007/978-1-4899-7488-4_196, DOI 10.1007/978-0-387-73003-5_196]. Richardson E, 2018, ADV NEUR IN, V31. Saadatnejad S, 2023, IEEE INT CONF ROBOT, P8246, DOI 10.1109/ICRA48891.2023.10160399. Sadeghian A, 2019, PROC CVPR IEEE, P1349, DOI 10.1109/CVPR.2019.00144. Salzmann Tim, 2020, Computer Vision - ECCV 2020. 16th European Conference. Proceedings. Lecture Notes in Computer Science (LNCS 12363), P683, DOI 10.1007/978-3-030-58523-5_40. Sohn Samuel S., 2022, Computers & Graphics, V106, P130, DOI [10.1016/j.cag.2022.05.010, 10.1016/j.cag.2022.05.010]. Varadarajan B, 2022, IEEE INT CONF ROBOT, P7814, DOI 10.1109/ICRA46639.2022.9812107. Vaswani A, 2017, ADV NEUR IN, V30, DOI 10.1007/s11356-023-31437-0. Wang XS, 2023, PROC CVPR IEEE, P21995, DOI 10.1109/CVPR52729.2023.02106. Wilson B., 2021, P NEUR INF PROC SYST. Xu P, 2022, LECT NOTES COMPUT SC, V13664, P511, DOI 10.1007/978-3-031-19772-7_30. Yang R., Entropy, V25, P1469. Yuan Y., 2019, arXiv. Zhang M., 2022, ARXIV. Zhao H, 2020, PR MACH LEARN RES, V155, P895. Zhou ZK, 2023, PROC CVPR IEEE, P17863, DOI 10.1109/CVPR52729.2023.01713. Zhou ZK, 2022, PROC CVPR IEEE, P8813, DOI 10.1109/CVPR52688.2022.00862. tex.da: 2026-05-22 tex.doc-delivery-number: BX3DC tex.number-of-cited-references: 50 tex.orcid-numbers: sester, monika/0000-0002-6656-8809 Xu, Yiming/0009-0004-6303-7801 tex.research-areas: Automation & Control Systems; Computer Science; Robotics; Transportation tex.researcherid-numbers: Cheng, Hao/JCO-2150-2023 sester, monika/A-2285-2013 tex.times-cited: 0 tex.unique-id: WOS:001275100902067 tex.usage-count-last-180-days: 0 tex.usage-count-since-2013: 0 tex.web-of-science-categories: Automation & Control Systems; Computer Science, Artificial Intelligence; Robotics; Transportation Science & Technology tex.web-of-science-index: Conference Proceedings Citation Index - Science (CPCI-S). 2

[20] Junwei You, Rui Gan, Weizhe Tang, Zilin Huang, Jiaxi Liu, Zhuoyu Jiang, Haotian Shi, Keshu Wu, Keke Long, Sicheng Fu, Sikai Chen, and Bin Ran. FollowGen: A scaled noise conditional diffusion model for car-following trajectory prediction. *COMMUNICATIONS IN TRANSPORTATION RESEARCH*, 5, 2025. Number: 100215 Type: Article tex.affiliation: Shi, HT (Corresponding Author), Tongji Univ, Coll Transportat, Shanghai 201804, Peoples R China. Shi, HT; Chen, SK (Corresponding Author), Univ Wisconsin Madison, Dept Civil & Environm Engn, Madison, WI 53706 USA. Shi, Haotian, Tongji Univ, Coll Transportat, Shanghai 201804, Peoples R China. You, Junwei; Gan, Rui; Tang, Weizhe; Huang, Zilin; Liu, Jiaxi; Shi, Haotian; Long, Keke; Fu, Sicheng; Chen, Sikai; Ran, Bin, Univ Wisconsin Madison, Dept Civil & Environm Engn, Madison, WI 53706 USA. Jiang, Zhuoyu, Nanyang Technol Univ, Coll Comp & Data Sci, Singapore 639798, Singapore. Wu, Keshu, Texas A&M Univ, Zachry Dept Civil & Environm Engn, College Stn, TX 77843 USA. tex.affiliations: Tongji University; University of Wisconsin System; University of Wisconsin Madison; Nanyang Technological University; Texas A&M University System; Texas A&M University College Station tex.author-email: shihaotian95@tongji.edu.cn sikai.chen@wisc.edu tex.cited-references: Altché F, 2017, IEEE INT C INTELL TR. De Miguel MA, 2022, IEEE ACCESS, V10, P32742, DOI 10.1109/ACCESS.2022.3161661. Chen Jianlv, 2024, CORR. Chen KH, 2023, IEEE INT C INTELL TR, P746, DOI 10.1109/ITSC57777.2023.10421892. Chen L, 2022, IET INTELL TRANSP SY, V16, P1682, DOI 10.1049/itr2.12174. Chen WH, 2021, PR MACH LEARN RES, V157, P454. Chen XB, 2022, IEEE T INSTRUM MEAS, V71, DOI 10.1109/TIM.2022.3192056. Chen XB, 2022, IEEE T INTELL TRANSP, V23, P19471, DOI 10.1109/TITS.2022.3170551. Cheng H, 2023, ISPRS J PHOTOGRAMM, V205, P163, DOI 10.1016/j.isprsjprs.2023.10.001. Choi S, 2021, IEEE INT C INTELL TR, P1515, DOI 10.1109/ITSC48978.2021.9564674. Cong Peichao, 2024, IEEE Transactions on Intelligent Vehicles, V9, P5339, DOI [10.1109/tiv.2023.3321656, 10.1109/TIV.2023.3321656]. Dey R, 2017, MIDWEST SYMP CIRCUIT, P1597, DOI 10.1109/MWSCAS.2017.8053243. DUHAMEL P, 1990, SIGNAL PROCESS, V19, P259, DOI 10.1016/0165-1684(90)90158-U. Federal Highway Administration, 2017, ROSA P. Feng XD, 2019, IEEE INT C INTELL TR, P3514, DOI

[10.1109/itsc.2019.8917482, 10.1109/ITSC.2019.8917482]. Geng MS, 2023, TRANSPORT RES C-EMER, V154, DOI 10.1016/j.trc.2023.104272. Gheini M, 2021, 2021 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP 2021), P1754. Gu TP, 2022, PROC CVPR IEEE, P17092, DOI 10.1109/CVPR52688.2022.01660. Guo K, 2022, PROC CVPR IEEE, P2232, DOI 10.1109/CVPR52688.2022.00228. Huang ZL, 2024, COMMUN TRANSP RES, V4, DOI 10.1016/j.commtr.2024.100127. Jiang C, 2023, PROC CVPR IEEE, P9644, DOI 10.1109/CVPR52729.2023.00930. Kim D, 2021, IEEE ACCESS, V9, P169348, DOI 10.1109/ACCESS.2021.3138502. Lee M, 2022, PROC CVPR IEEE, P2211, DOI 10.1109/CVPR52688.2022.00226. Li GP, 2023, IEEE INT C INTELL TR, P5818, DOI 10.1109/ITSC57777.2023.10422574. Li ZY, 2023, KNOWL-BASED SYST, V280, DOI 10.1016/j.knosys.2023.110990. Liao HC, 2024, PROCEEDINGS OF THE THIRTY-THIRD INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, IJCAI 2024, P7331. Liao HC, 2024, AAAI CONF ARTIF INTE, P10332. Long KK, 2025, COMMUN TRANSP RES, V5, DOI [10.1016/j.commtr.2025.100166, 10.1016/j.commtr.2025.100166]. Ma Yecheng Jason, 2021, [No title captured], P13279. Mo XY, 2021, IEEE INT C INTELL TR, P1934, DOI 10.1109/ITSC48978.2021.9564929. Neumeier M, 2021, IEEE INT C INTELL TR, P820, DOI 10.1109/ITSC48978.2021.9565120. Qin Y, 2023, IEEE T VEH TECHNOL, V72, P9746, DOI 10.1109/TVT.2023.3253695. Ronneberger O, 2015, LECT NOTES COMPUT SC, V9351, P234, DOI 10.1007/978-3-319-24574-4_28. Schmidt J, 2022, IEEE INT CONF ROBOT, P7799, DOI 10.1109/ICRA46639.2022.9811637. Sheng ZH, 2024, COMMUN TRANSP RES, V4, DOI 10.1016/j.commtr.2024.100142. Sheng ZH, 2022, IEEE T INTELL TRANSP, V23, P17654, DOI 10.1109/TITS.2022.3155749. Shi KS, 2022, PHYSICA A, V599, DOI 10.1016/j.physa.2022.127303. Shi LS, 2023, IEEE I CONF COMP VIS, P9641, DOI 10.1109/ICCV51070.2023.00887. Tang YJ, 2024, IEEE SENS J, V24, P17208, DOI 10.1109/JSEN.2024.3382406. Tang YJ, 2024, NEUROCOMPUTING, V580, DOI 10.1016/j.neucom.2024.127526. Wang RZ, 2023, AAAI CONF ARTIF INTE, P4685. Wang Y, 2022, IEEE T INTELL TRANSP, V23, P236, DOI 10.1109/TITS.2020.3009762. Wei TL, 2024, IEEE T KNOWL DATA EN, V36, P7940, DOI 10.1109/TKDE.2024.3460051. Westny T, 2024, Arxiv, DOI arXiv:2403.11643. Wu KS, 2024, IEEE T INTELL VEHICL, V9, P3630, DOI 10.1109/TIV.2023.3341071. Xie G, 2020, SCI CHINA INFORM SCI, V63, DOI 10.1007/s11432-019-2761-y. Xu M, 2023, COMMUN TRANSP RES, V3, DOI 10.1016/j.commtr.2023.100099. Zhang KP, 2023, IEEE T INTELL VEHICL, V8, P73, DOI 10.1109/TIV.2022.3155236. Zhang ZB, 2024, IEEE T INTELL VEHICL, V9, P4749, DOI 10.1109/TIV.2023.3327501. Zhao TY, 2019, PROC CVPR IEEE, P12118, DOI 10.1109/CVPR.2019.01240. Zheng H, 2015, IEEE IJCNN. Zhu M., 2022, arXiv. Zuo ZQ, 2024, IEEE T INTELL VEHICL, V9, P2171, DOI 10.1109/TIV.2023.3319024. tex.da: 2026-05-22 tex.doc-delivery-number: 8XU4J tex.earlyaccessdate: OCT 2025 tex.journal-iso: Commun. Transp. Res. tex.number-of-cited-references: 53 tex.oa: Green Submitted, gold tex.orcid-numbers: Huang, Zilin/0000-0003-3754-4821 Wu, Keshu/0000-0002-8085-0990 You, Junwei/0009-0002-6447-8276 tex.research-areas: Transportation tex.researcherid-numbers: SHI, HAOTIAN/HGV-0826-2022 gan, rui/HMO-7146-2023 Huang, Zilin/JAO-2256-2023 Long, Keke/KCX-9761-2024 tex.times-cited: 2 tex.unique-id: WOS:001599881600001 tex.usage-count-last-180-days: 21 tex.usage-count-since-2013: 23 tex.web-of-science-categories: Transportation; Transportation Science & Technology tex.web-of-science-index: Science Citation Index Expanded (SCI-EXPANDED); Social Science Citation Index (SSCI). 2

[21] Ruixuan Zhang, Shichun Yang, Bingtao Ren, Zehua Wu, Zhenxiong Xu, and Jingxiang Huang. Diffusion-based corner scenario generation method for autonomous driving with a dynamics-based decoder. In *2025 9th caa international conference on vehicular control and intelligence, cvci*, pages 511–516, 345 E 47TH ST, NEW YORK, NY 10017 USA, 2025. IEEE. Type: Proceedings Paper tex.affiliation: Ren, BT (Corresponding Author), Beihang Univ, Sch Transportat Sci & Engn, Beijing, Peoples R China. Ren, BT (Corresponding Author), Beihang Univ, Hangzhou Int Innovat Inst, Hangzhou, Peoples R China. Zhang, Ruixuan; Yang, Shichun; Ren, Bingtao; Wu, Zehua; Xu, Zhenxiong; Huang, Jingxiang, Beihang Univ, Sch Transportat Sci & Engn, Beijing, Peoples R China. Ren, Bingtao, Beihang Univ, Hangzhou Int Innovat Inst, Hangzhou, Peoples R China. tex.affiliations: Beihang University; Beihang University tex.author-email: zhangruixuan@buaa.edu.cn renbt1706@buaa.edu.cn tex.book-group-author: IEEE tex.cited-references: Baumler M., 2024, IEEE ACCESS. Demetriou A, 2023, SN Computer Science, V4, DOI [10.1007/s42979-023-01714-3, 10.1007/s42979-023-01714-3, DOI 10.1007/S42979-023-01714-3]. Ding WH, 2023, IEEE T INTELL TRANSP, V24, P6971, DOI 10.1109/TITS.2023.3259322. Garikapati D, 2024, BIG DATA COGN COMPUT, V8, DOI 10.3390/bdcc8040042. Guo ZP, 2023, PROC INT C TOOLS ART, P717, DOI 10.1109/ICTAI59109.2023.00111. Hao K., 2023, IEEE T INTELL VEHICL. Koenig A, 2017, SAE INT J TRANSP SAF, V5, P81, DOI 10.4271/2017-01-1403. Li Q., 2025, ARXIV. Li Y, 2024, ACCIDENT ANAL PREV, V205, DOI 10.1016/j.aap.2024.107667. Lu J, 2024, IEEE INT CONF ROBOT, P16908, DOI 10.1109/ICRA57147.2024.10610324. Parekh D, 2022, ELECTRONICS-SWITZ, V11, DOI 10.3390/electronics11142162. Ponn T, 2020, 2020 FIFTEENTH INTERNATIONAL CONFERENCE ON ECOLOGICAL VEHICLES AND RENEWABLE ENERGIES (EVER). Pronovost E., 2023, ADV NEURAL INFORM

PR, P68873. Rajamani R., 2012, VEHICLE DYNAMICS CON. Saifuzzaman M, 2014, TRANSPORT RES C-EMER, V48, P379, DOI 10.1016/j.trc.2014.09.008. Wang S, 2024, IEEE INT C INT ROBOT, P14241, DOI 10.1109/IROS58592.2024.10801623. Wang XS, 2019, ACCIDENT ANAL PREV, V124, P127, DOI 10.1016/j.aap.2019.01.006. Winner H, 2019, LECT N MOBIL, P185, DOI 10.1007/978-3-319-94896-6_16. tex.da: 2026-05-22 tex.doc-delivery-number: BA2YB tex.funding-acknowledgement: National Key Research and Development Program of China [2022YFB2503400]; National Natural Science Foundation of China [52302486, U22A202101]; Research Start-up Funds of Hangzhou International Innovation Institute of Beihang University [2024KQ103] tex.funding-text: This work was supported by the National Key Research and Development Program of China (2022YFB2503400), the National Natural Science Foundation of China (52302486, U22A202101), and the Research Start-up Funds of Hangzhou International Innovation Institute of Beihang University (2024KQ103). tex.number-of-cited-references: 18 tex.research-areas: Automation & Control Systems; Computer Science; Transportation tex.researcherid-numbers: Ren, Bingtao/AEU-6929-2022 tex.times-cited: 0 tex.unique-id: WOS:001706365700087 tex.usage-count-last-180-days: 5 tex.usage-count-since-2013: 5 tex.web-of-science-categories: Automation & Control Systems; Computer Science, Interdisciplinary Applications; Transportation Science & Technology tex.web-of-science-index: Conference Proceedings Citation Index - Science (CPCI-S). 1

[22] 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, 345 E 47TH ST, NEW YORK, NY 10017 USA, 2023. IEEE / IEEE; IEEE Robot & Automat Soc. Type: Proceedings Paper tex.affiliation: Zhong, ZY (Corresponding Author), Columbia Univ, New York, NY 10027 USA. Zhong, ZY (Corresponding Author), NVIDIA, Santa Clara, CA 95051 USA. Zhong, Ziyuan; Ray, Baishakhi, Columbia Univ, New York, NY 10027 USA. Zhong, Ziyuan; Rempe, Davis, NVIDIA, Santa Clara, CA 95051 USA. Rempe, Davis; Pavone, Marco, Stanford Univ, Stanford, CA 94305 USA. Xu, Danfei, Georgia Tech, Atlanta, GA USA. Xu, Danfei; Chen, Yuxiao; Veer, Sushant; Che, Tong; Pavone, Marco, NVIDIA Res, Santa Clara, CA USA. tex.affiliations: Columbia University; Nvidia Corporation; Stanford University; University System of Georgia; Georgia Institute of Technology tex.author-email: ziyuan.zhong@columbia.com drempe@stanford.edu danfei@gatech.edu yuxiaoc@nvidia.com sveer@nvidia.com tongc@nvidia.com rayb@cs.columbia.edu pavone@stanford.edu tex.book-group-author: IEEE tex.cited-references: Abeysirigoonawardena Y, 2019, IEEE INT CONF ROBOT, P8271, DOI [10.1109/ICRA.2019.8793740, 10.1109/icra.2019.8793740]. Bergamini L, 2021, IEEE INT CONF ROBOT, P5119, DOI 10.1109/ICRA48506.2021.9561666. Brockfeld E, 2003, TRANSPORT RES REC, P124, DOI 10.3141/1852-16. Caesar H, 2020, PROC CVPR IEEE, P11618, DOI 10.1109/CVPR42600.2020.01164. Chai Yuning., 2020, CORL, P86. Chen B., 2021, IEEE T INTELL TRANSP, V23, P1. Chen N, 2021, P INT C LEARNING REP. Chen YX, 2022, PROC CVPR IEEE, P17082, DOI 10.1109/CVPR52688.2022.01659. Dhariwal P, 2021, ADV NEUR IN, V34. Dosovitskiy A, 2017, PR MACH LEARN RES, V78. He KM, 2016, PROC CVPR IEEE, P770, DOI 10.1109/CVPR.2016.90. Ho J, 2022, arXiv preprint arXiv:2207.12598. Ho Jonathan, 2022, arXiv:2204.03458. Ho Jonathan, 2020, ADV NEURAL INFORM PR, V33, P6840. Janner M, 2022, PR MACH LEARN RES. Kalra N., 2016, Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?, DOI DOI 10.7249/J.CTT1BTC0XW. Kong Z., 2021, P INT C LEARNING REP. Leung K., 2022, Semi-supervised trajectory-feedback controller synthesis for signal temporal logic specifications. LG Electronics, 2021, SVL SIM AUT VEH SIM. Li X L., 2022, DIFFUSION LM IMPROVE. Lopez PA, 2018, IEEE INT C INTELL TR, P2575, DOI 10.1109/ITSC.2018.8569938. Maler O, 2004, LECT NOTES COMPUT SC, V3253, P152, DOI 10.1007/978-3-540-30206-3_12. Nichol A, 2021, PR MACH LEARN RES, V139. Paszke A, 2019, ADV NEUR IN, V32. Popov V, 2021, PR MACH LEARN RES, V139. Rempe D, 2022, PROC CVPR IEEE, P17284, DOI 10.1109/CVPR52688.2022.01679. Salzmann Tim, 2020, Computer Vision - ECCV 2020. 16th European Conference. Proceedings. Lecture Notes in Computer Science (LNCS 12363), P683, DOI 10.1007/978-3-030-58523-5_40. Sohl-Dickstein J, 2015, PR MACH LEARN RES, V37, P2256. Suo S, 2021, PROC CVPR IEEE, P10395, DOI 10.1109/CVPR46437.2021.01026. Wang JK, 2021, PROC CVPR IEEE, P9904, DOI 10.1109/CVPR46437.2021.00978. Wang JK, 2022, PROC CVPR IEEE, P17113, DOI 10.1109/CVPR52688.2022.01662. Waymo, 2021, Waymo safety report. Xu D., 2022, Bits: Bi-level imitation for traffic simulation. tex.da: 2026-05-22 tex.doc-delivery-number: BV4NP tex.eissn: 2577-087X tex.number-of-cited-references: 34 tex.orcid-numbers: Pavone, Marco/0000-0002-0206-4337 tex.research-areas: Automation & Control Systems; Computer Science; Engineering; Robotics tex.researcherid-numbers: Veer, Sushant/AAG-4773-2019 xu, danfei/OBO-2343-2025 tex.times-cited: 99 tex.unique-id: WOS:001036713002134 tex.usage-count-last-180-days: 1 tex.usage-count-since-2013: 8 tex.web-of-science-categories: Automation & Control Systems; Computer Science, Artificial Intelligence; Engineering, Electrical & Electronic; Robotics tex.web-of-science-index: Conference Proceedings Citation Index - Science (CPCI-S). 2

[23] Weiping Zhu, Renhe Jiang, Dongyuan Li, Xiangjie Kong, and Xuan Song. Trajectory generative models: a survey from unconditional and conditional perspectives. *GEOINFORMATICA*, 29(4):1033–1065, 2025. 1, 2