



Synthesizing Simulation Environments with Generative Models

Kashyap Chitta



Simulation in Autonomous Driving

Graphics Simulators

- Great tools for research!
- Arbitrary duration & routes, E2E planning
- Hand-crafted layouts and scenarios
- Compute-heavy and high variance

Examples: CARLA, TORCS

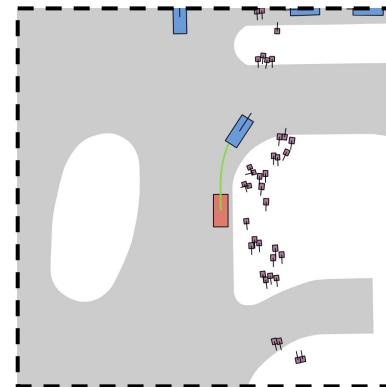


Simulation in Autonomous Driving

Log Replay Simulators

- Relatively new concept
- Replay / initialize real-world recording
- Real driving scenarios (e.g. 2 TB data)
- Fixed routes & short duration (e.g. 15s)

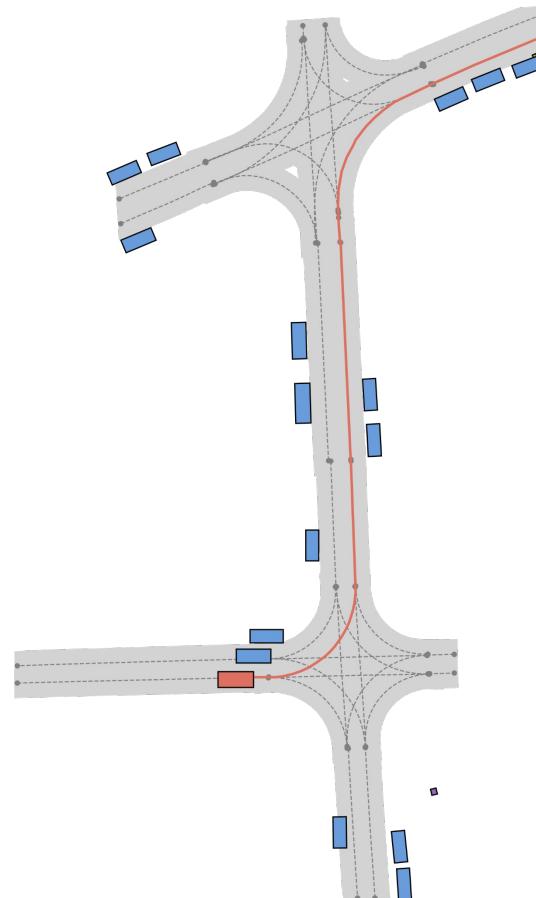
Examples: nuPlan, Waymax



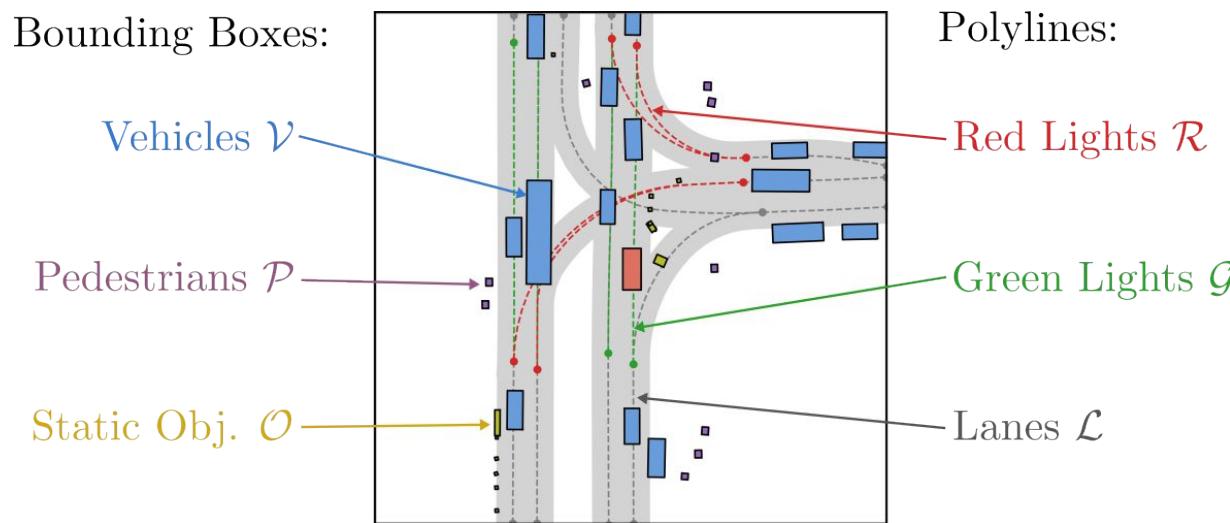
Simulation in Autonomous Driving

Generative Models

- Emerging idea
- Synthesize real-world scenes
- Arbitrary duration & routes
- Compact (only weights: 3 GB)

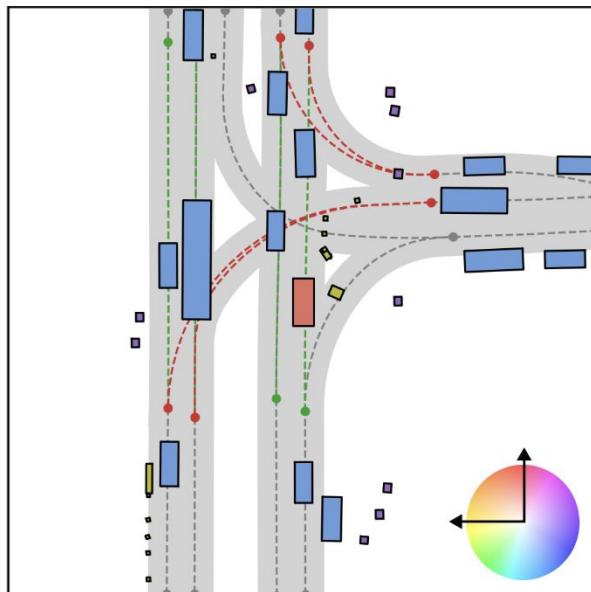
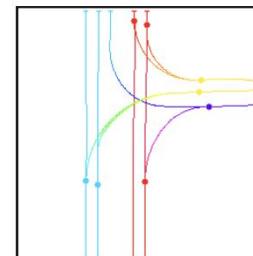
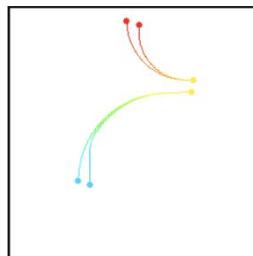
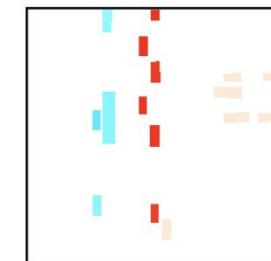
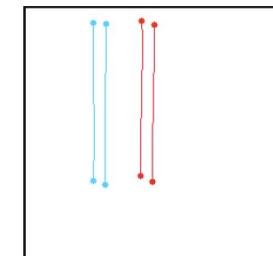


Driving Simulation: Generating Scene Elements

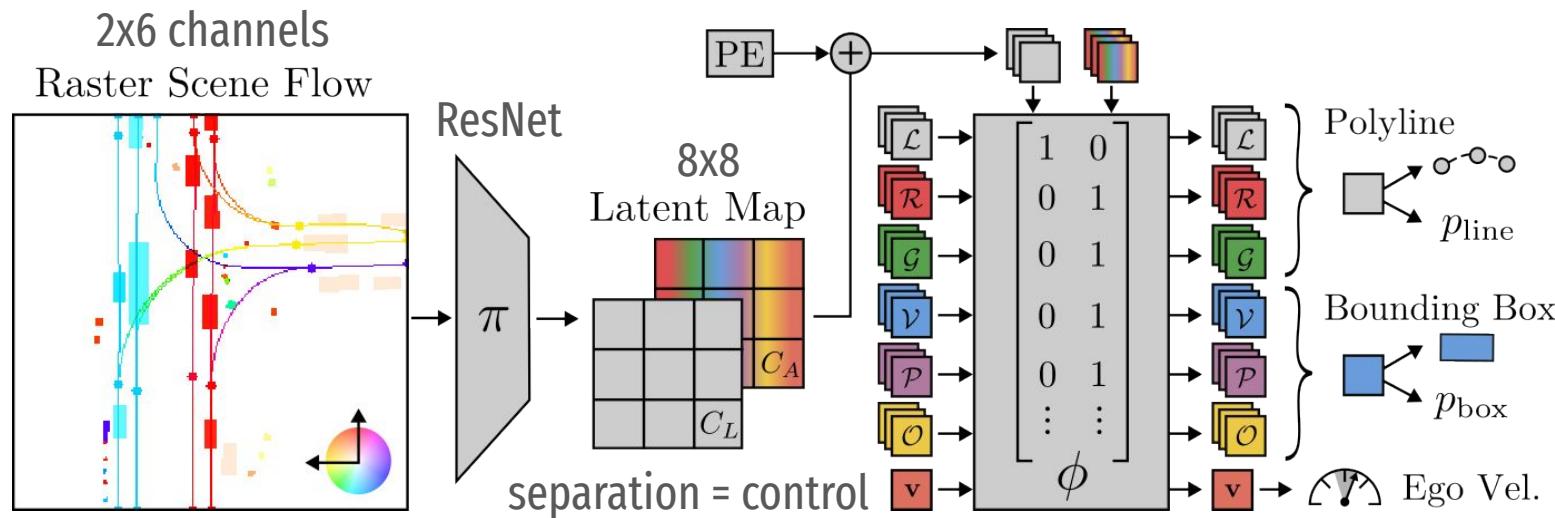


Vector representations are incompatible with most generative models
(diverse entities, variable entity count)

Vector to Raster Flow

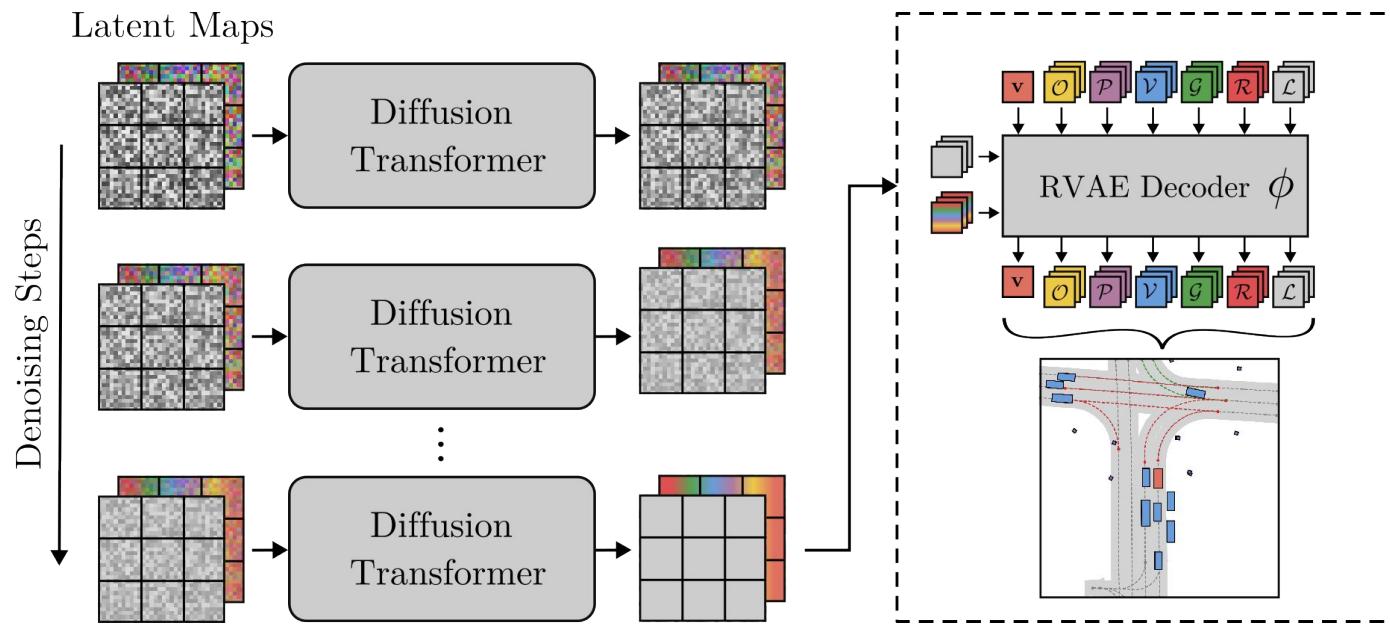
Scene State \mathcal{S} Lanes \mathcal{L} Red Lights \mathcal{R} Green Lights \mathcal{G} Pedestrians \mathcal{P} Vehicles \mathcal{V} Static Obj. \mathcal{O}

Raster to Vector Autoencoder



Bridge raster to vector representation with query-based transformer

Latent Diffusion Transformer (DiT)



Implementation Details

Dataset:

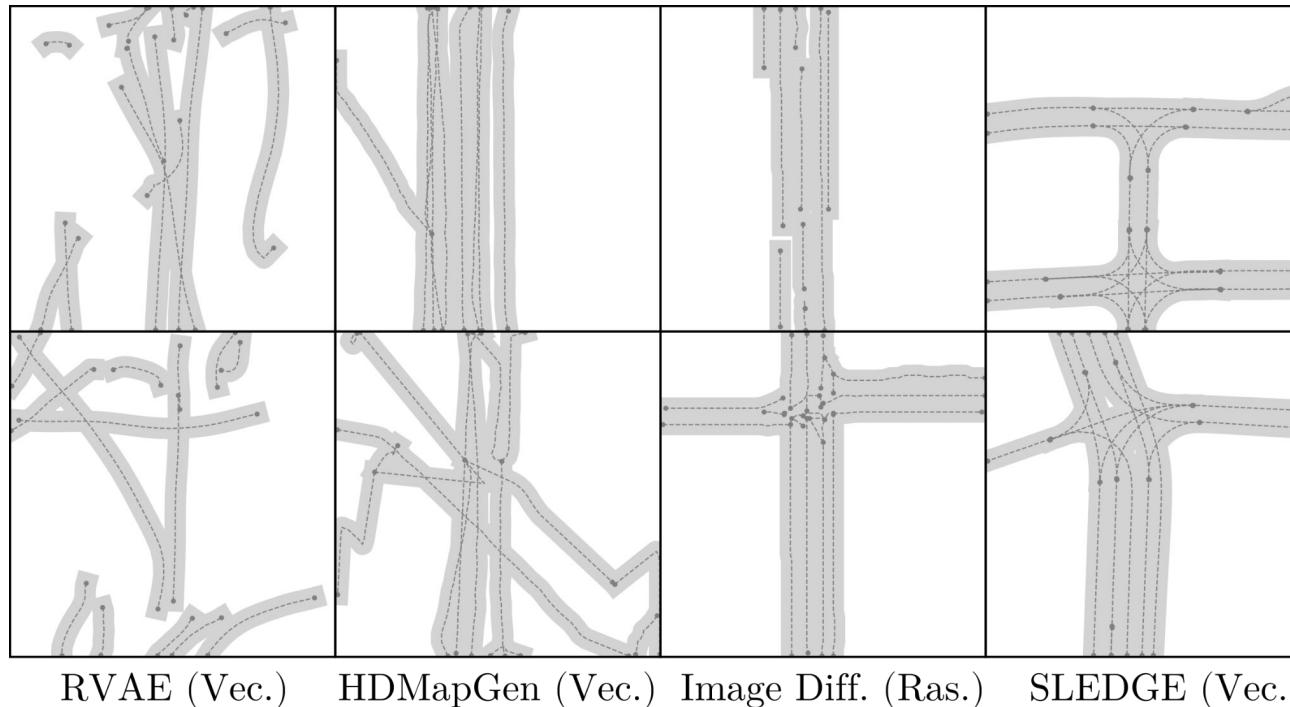
- nuPlan (450k samples)
- Cities: Las Vegas, Boston, Singapore, Pittsburgh

Training:

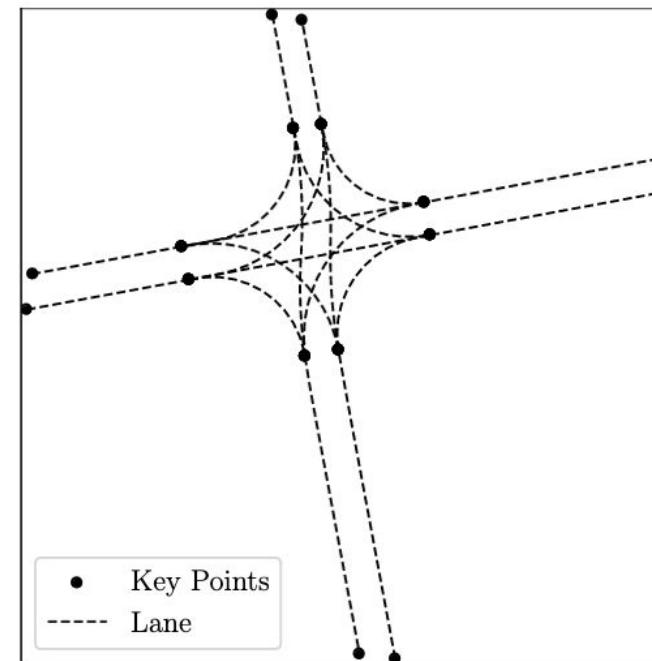
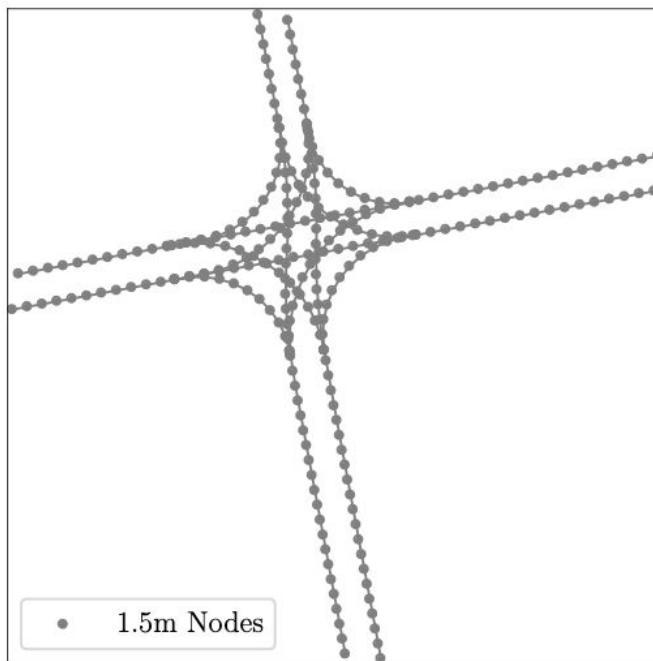
- **RVAE:** 4 x A100, 1 day
 - Hungarian Loss (BCE + L1)
 - KL-Divergence
- **DiT:** 8 x A100, 2 weeks



Task 1: HD-Map Generation



Task 1: HD-Map Generation

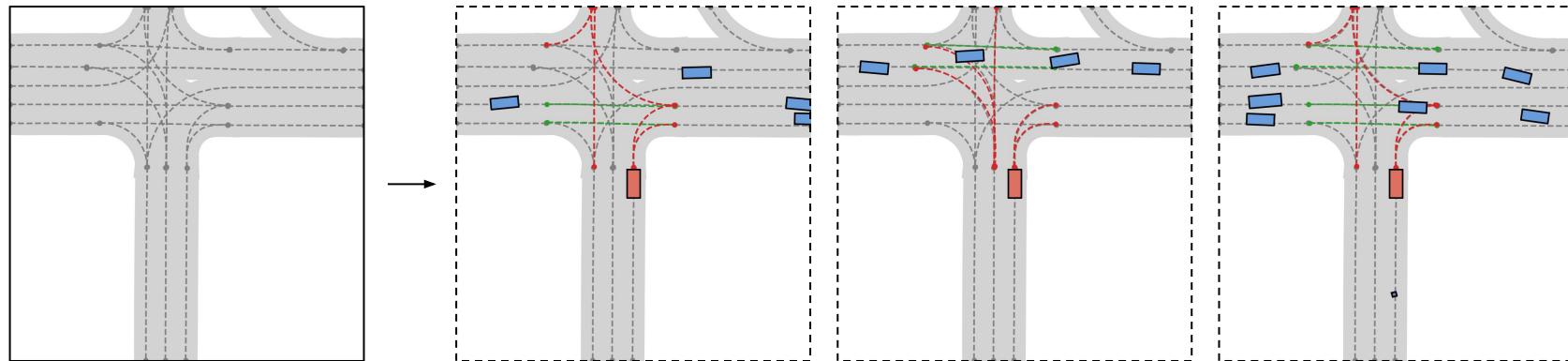


Task 1: HD-Map Generation

Architecture	Repr.	Frechet (Statistics on Keypoints) ↓			
		Connect.	Density	Reach	Conveni.
RVAE	Vector	15.63	12.57	3.08	17.72
HDMapGen	Vector	7.02	3.03	2.49	18.10
Image Diff.	Raster	6.11	15.33	1.90	3.95
SLEDGE	Latent	0.27	2.47	0.20	0.47

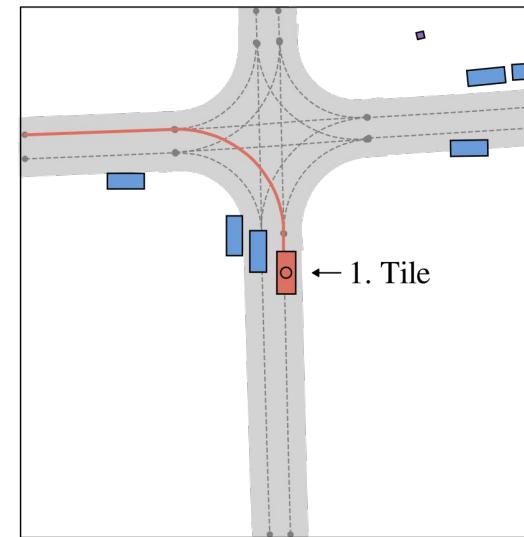
Latent diffusion enables state-of-the-art HD-Map generation

Task 2: Agent Inpainting



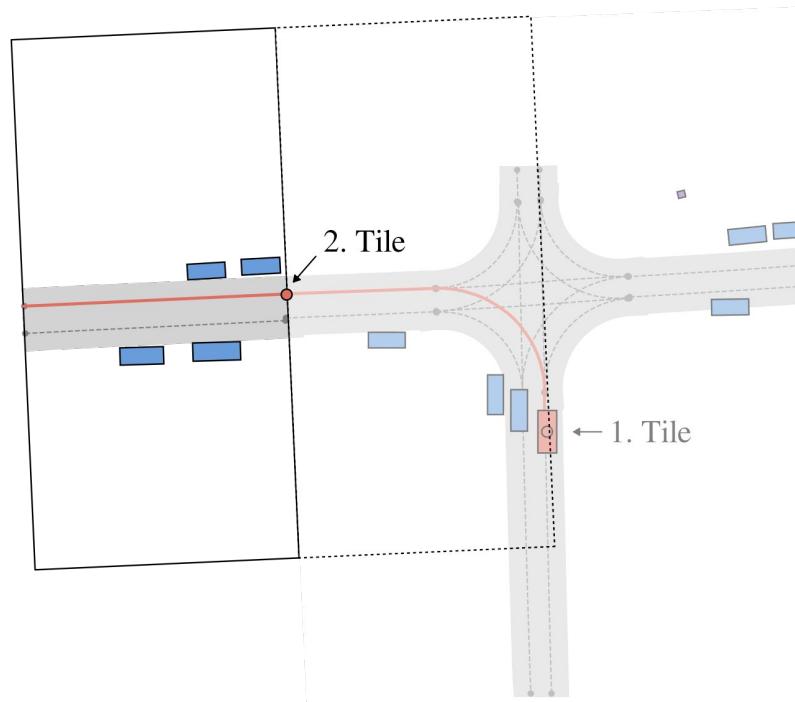
Conditioned on HD-Maps, SLEDGE inpaints actors and traffic lights

Task 3: Autoregressive Map & Agent Generation



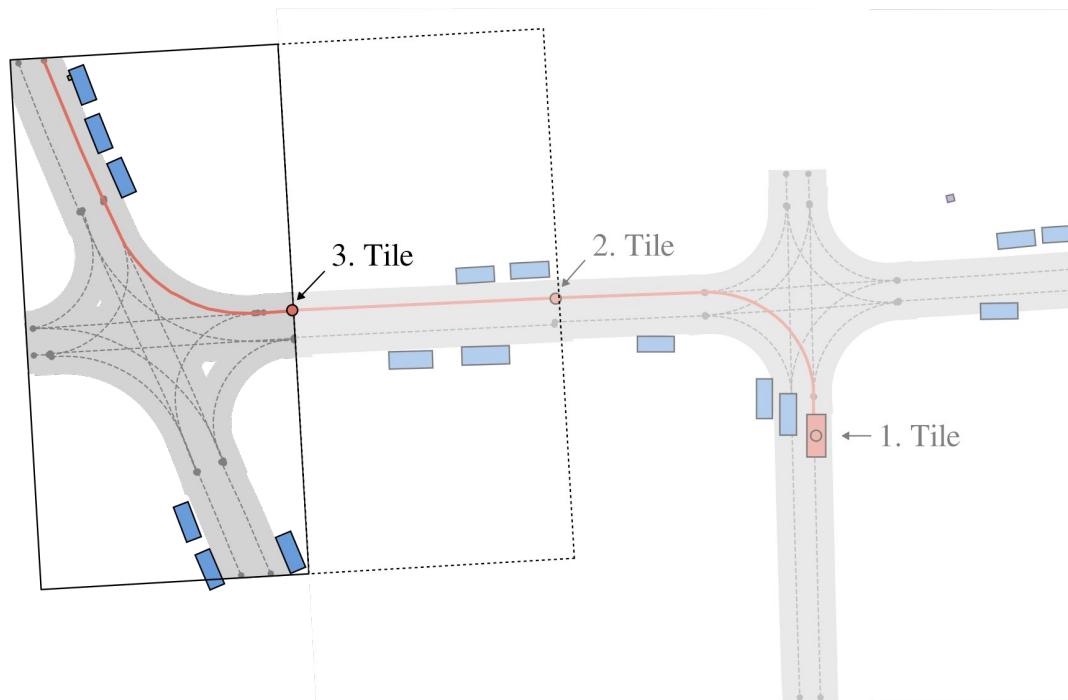
Iteratively extend map and agents via spatial outpainting

Task 3: Autoregressive Map & Agent Generation



Iteratively extend map and agents via spatial outpainting

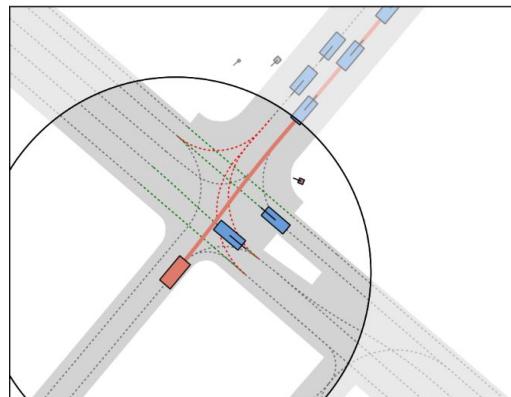
Task 3: Autoregressive Map & Agent Generation



Iteratively extend map and agents via spatial outpainting

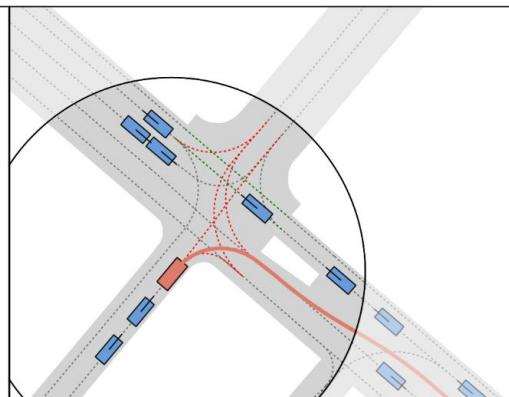
Simulation in SLEDGE

(a) Log Replay



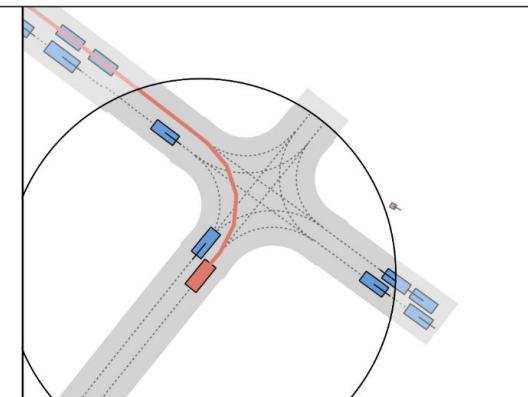
Map: from log
Agents: from log

(b) Lane → Agent



Map: from log
Agents: generated

(c) Lane & Agent



Map: generated
Agents: generated

SLEDGE simulates agents with distance $< 64m$ to the ego vehicle

SLEDGE: Synthesizing Simulation Environments for Driving Agents with Generative Models



Long-routes in SLEDGE Expose Failures in PDM-Closed

Method	Failure Rates [%]			
	Lane -> Agent		Lane & Agents	
	100m	500m	100m	500m
<i>Log-Replay</i>	6	26	-	-
SLEDGE	7	25	22	39
SLEDGE + Traffic	11	39	20	44
SLEDGE + Turns	14	28	29	45

SLEDGE: Synthesizing Simulation Environments for Driving Agents with Generative Models



- #1 Generative models enable long traffic simulations
- #2 Vectorized scene generation possible via latent diffusion
- #3 SLEDGE enhances simulation control: turns, traffic density
- #4 Simulator requires 500× less storage than logs



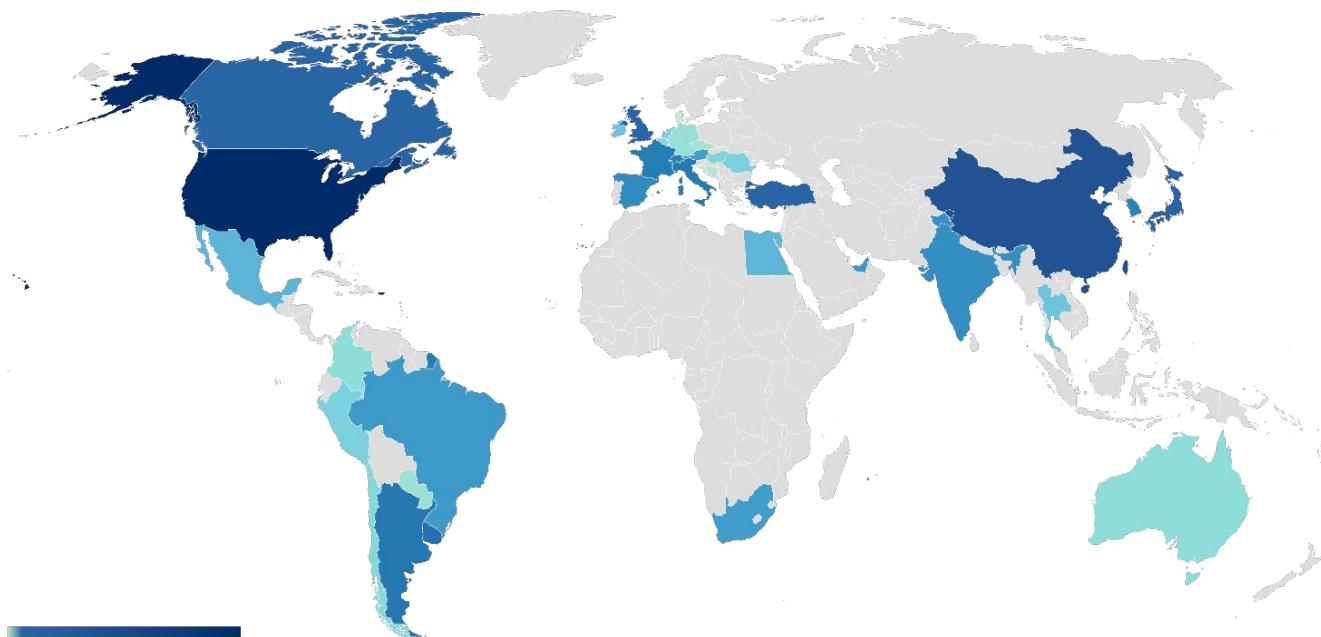
github.com/autonomousvision/sledge



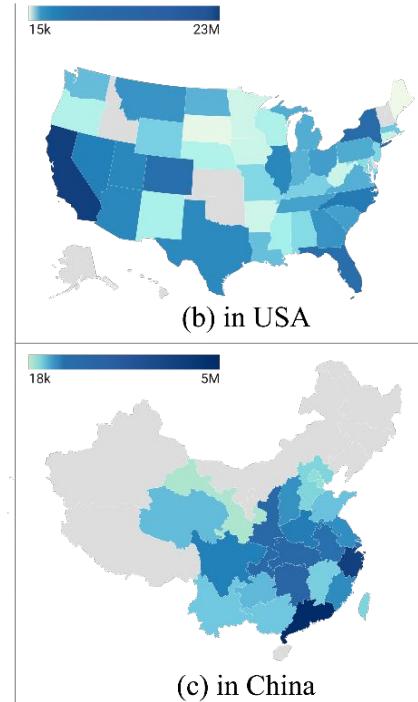
OpenDV-2k



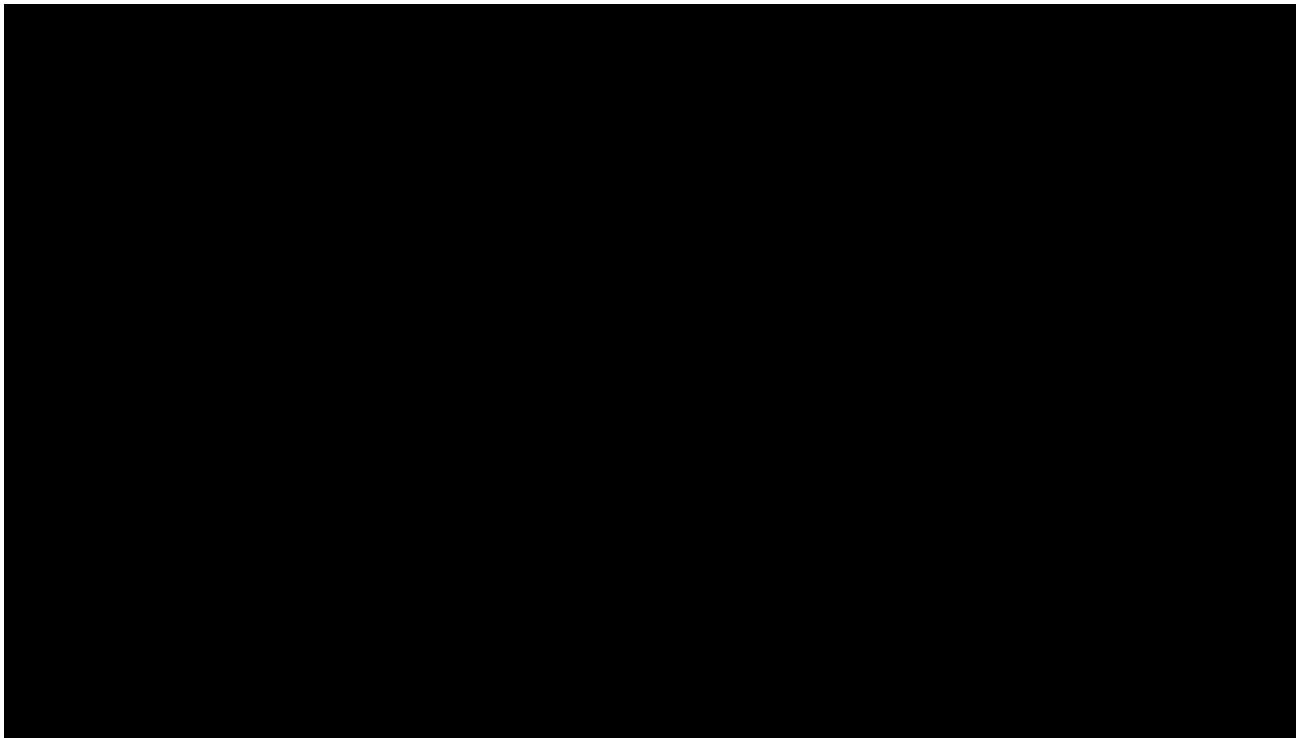
2000+ Hours, 65M+ Frames, 40+ Countries, 700+ Cities



(a) Global Distribution



2.5B Driving World Model with **Open Code and Weights!**



vista-demo.github.io