SceneControl: Diffusion for Controllable Traffic Scene Generation
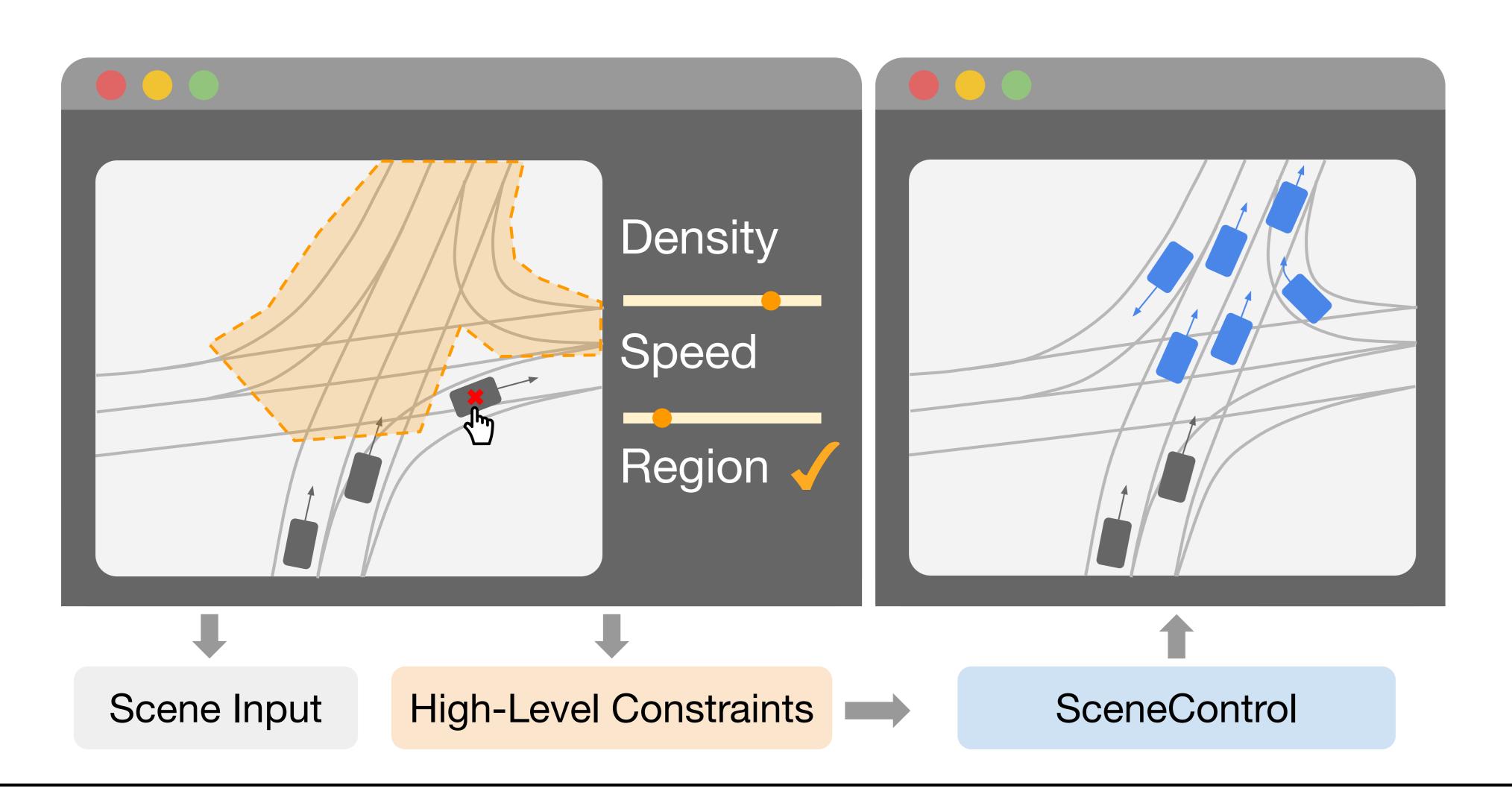
Jack Lu*, Kelvin Wong*, Chris Zhang, Simon Suo, Raquel Urtasun
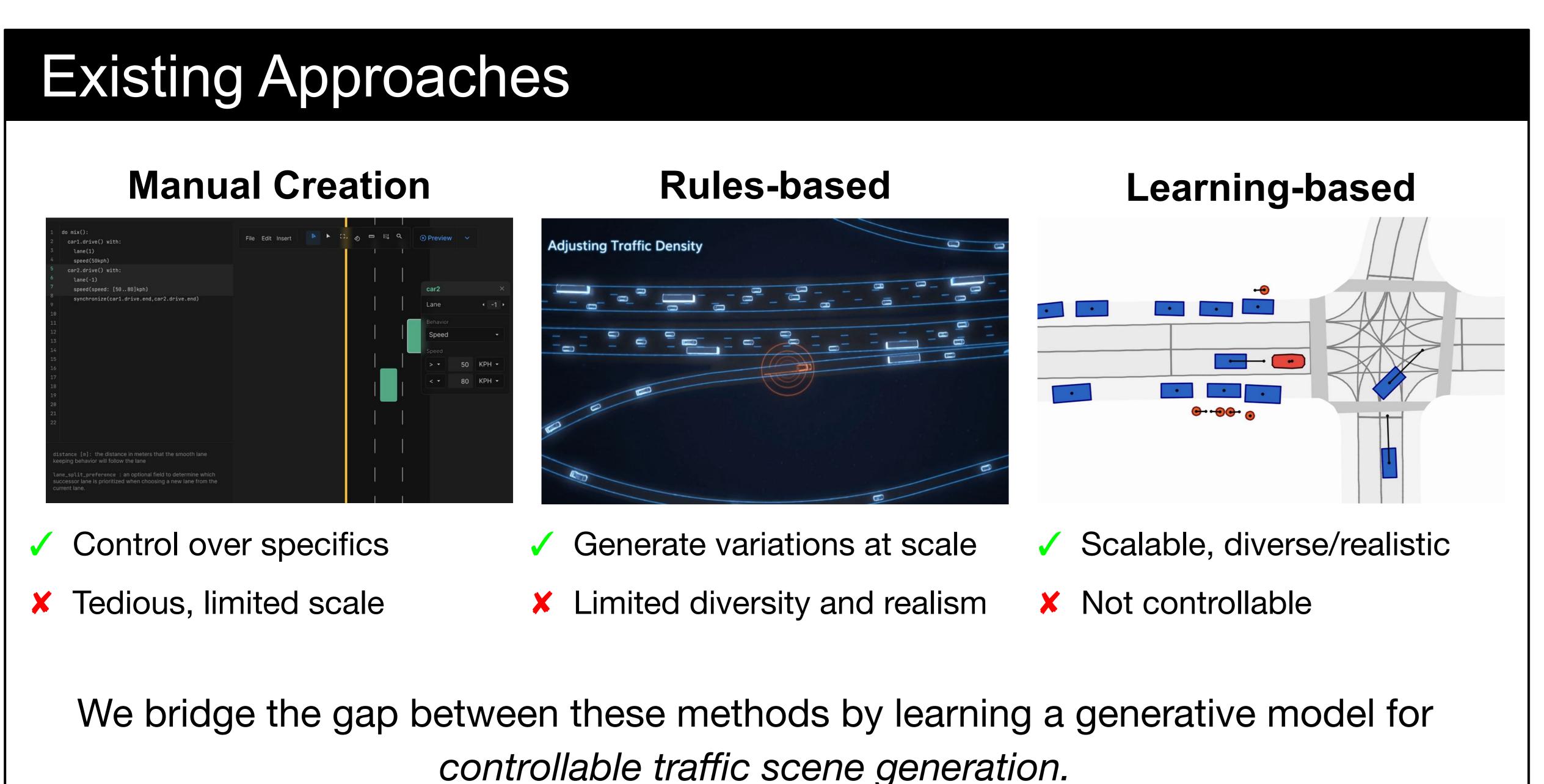
waabi

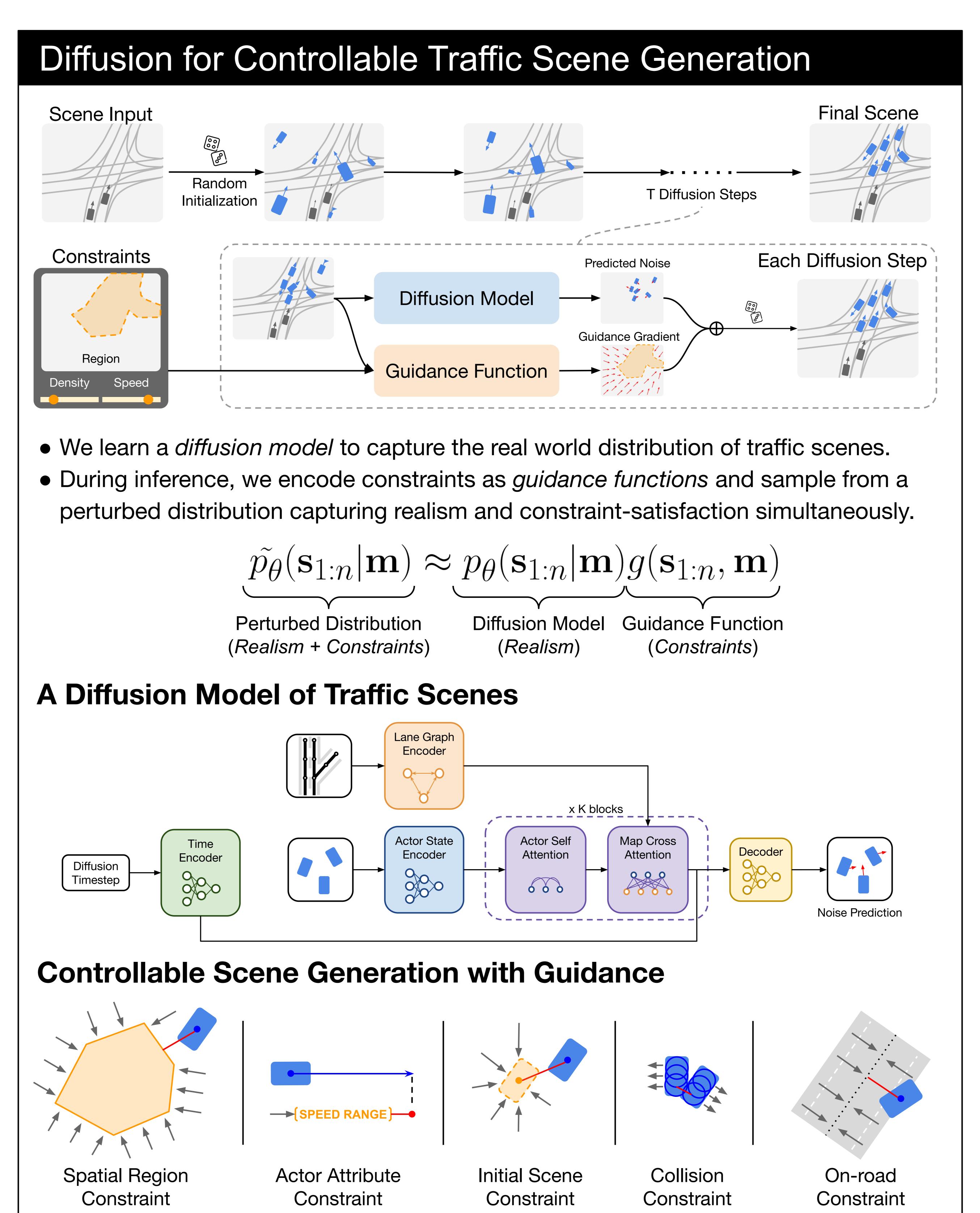**ICRA2024** YOKOHAMA | JAPAN

Traffic Scene Generation

Motivation: Simulation is an important tool to safely and scalably develop self-driving vehicles. The ability to simulate realistic traffic scenarios is essential.

Goal: Learn to generate realistic traffic scenes to initialize simulation.

Task: Given an HD map of the traffic scene, generate the placement and attributes (*e.g.*, size, speed, *etc.*) of the actors in the scene.

Density · Speed · Region ✓

Scene Input → High-Level Constraints → SceneControl

Existing Approaches

Manual Creation
- ✓ Control over specifics
- ✗ Tedious, limited scale

Rules-based
- ✓ Generate variations at scale
- ✗ Limited diversity and realism

Learning-based
- ✓ Scalable, diverse/realistic
- ✗ Not controllable

We bridge the gap between these methods by learning a generative model for *controllable traffic scene generation.*

Diffusion for Controllable Traffic Scene Generation

Scene Input — Random Initialization — T Diffusion Steps — Final Scene

Constraints — Region, Density, Speed — Diffusion Model → Predicted Noise; Guidance Function → Guidance Gradient — Each Diffusion Step

- We learn a *diffusion model* to capture the real world distribution of traffic scenes.
- During inference, we encode constraints as *guidance functions* and sample from a perturbed distribution capturing realism and constraint-satisfaction simultaneously.

$$\tilde{p}_\theta(\mathbf{s}_{1:n}|\mathbf{m}) \approx p_\theta(\mathbf{s}_{1:n}|\mathbf{m})\, g(\mathbf{s}_{1:n}, \mathbf{m})$$

Perturbed Distribution (*Realism + Constraints*) · Diffusion Model (*Realism*) · Guidance Function (*Constraints*)

A Diffusion Model of Traffic Scenes

Controllable Scene Generation with Guidance

Spatial Region Constraint · Actor Attribute Constraint · Initial Scene Constraint · Collision Constraint · On-road Constraint

Experiments

Unconditional Traffic Scene Generation

	Model	Nearest Dist.	Distribution JSD						Common Sense	
			Lat. Dev.	Ang. Dev.	Length	Width	Speed	Collision %	Off-road %
AV2	ATISS [21]	0.26	0.26	0.75	0.49	0.20	0.24	22.59	49.09
	ATISS++ [21]	0.30	0.22	0.21	0.23	0.15	0.18	1.58	24.84
	SceneGen [29]	0.20	**0.08**	0.38	0.26	0.21	0.06	1.80	1.10
	SceneControl (Ours)	**0.11**	0.16	**0.18**	0.22	**0.09**	**0.05**	**1.37**	**0.00**
Highway	ATISS [21]	0.20	0.56	0.75	0.36	0.34	0.13	8.27	35.34
	ATISS++ [21]	0.24	0.37	0.11	0.34	**0.17**	0.12	9.10	3.84
	SceneGen [29]	0.13	0.19	0.29	0.36	0.21	**0.08**	1.06	**0.11**
	SceneControl (Ours)	**0.10**	**0.18**	**0.07**	0.31	**0.17**	0.11	**0.32**	1.73

SceneControl achieves SOTA realism on unconditional traffic scene generation for urban and highway data as measured by distribution matching and common sense metrics.

Controllable Traffic Scene Generation

	Model	Guide	Region Constraint			Size Constraint			Speed Constraint		
			Suc. %	JSD	Col. %	Suc. %	JSD	Col. %	Suc. %	JSD	Col. %
AV2	ATISS++ [21]		8.34	0.23	4.66	85.50	0.22	20.34	76.52	0.22	18.86
	SceneGen [29]		7.66	0.22	1.54	86.67	0.21	2.53	82.09	0.20	1.98
	SceneControl (Ours)		24.78	**0.17**	1.42	80.33	**0.14**	2.23	81.33	**0.15**	1.55
	SceneControl (Ours)	✓	**77.22**	0.19	**1.35**	**95.62**	**0.14**	2.17	**98.87**	**0.14**	**1.37**
Highway	ATISS++ [21]		17.86	0.47	5.91	78.48	0.40	19.80	59.52	0.36	5.91
	SceneGen [29]		27.66	0.22	0.77	75.86	0.20	0.97	64.93	0.21	0.77
	SceneControl (Ours)		35.69	**0.16**	0.79	79.46	**0.15**	0.68	71.56	**0.15**	**0.47**
	SceneControl (Ours)	✓	**86.60**	**0.16**	**0.15**	**94.66**	**0.14**	0.54	**90.84**	0.16	**0.47**

SceneControl achieves greater controllability and realism than existing methods for traffic scene generation.

Qualitative Results

In Region · In Region · In Occlusion · In Occlusion
Less Actors · More Actors · Low Density · High Density
Small Actors · Large Actors · Low Speed · High Speed