CVPR
JUNE 17-21, 2024
SEATTLE, WA

Project Page Github arXiv

LiDAR Diffusion

Towards Realistic Scene Generation with LiDAR Diffusion Models

Haoxi Ran Vitor Guizilini Yue Wang

Carnegie Mellon University
TOYOTA RESEARCH INSTITUTE
University of Southern California

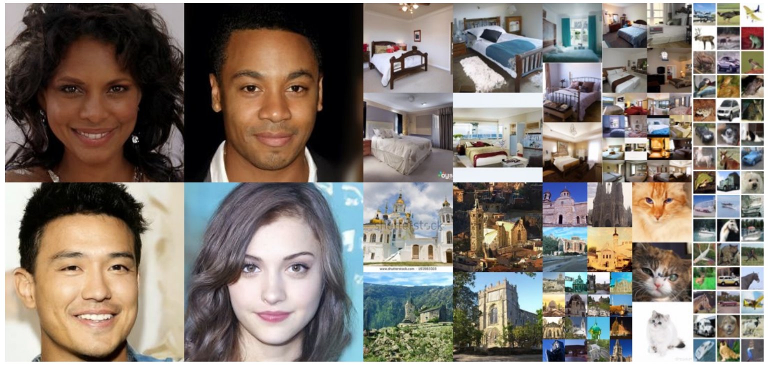
Motivation

Can we apply the successful controllable DMs to LiDAR scene simulation?

- Why?
  - (Layout-to-LiDAR) To synthesize corresponding LiDAR scenes given some bounding boxes, thus turning them into labeled data
  - (Camera-to-LiDAR) To generate a 3D scene from a set of images
  - (Text-to-LiDAR) To control LiDAR simulation driven by language

Related Work

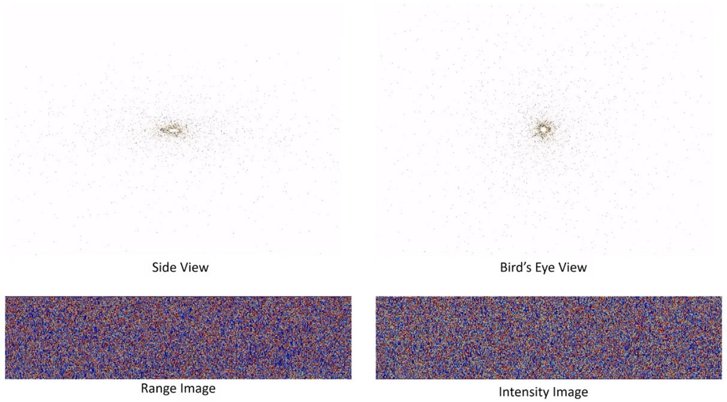
Diffusion Model (DM)
DDPM (NeurIPS'20)



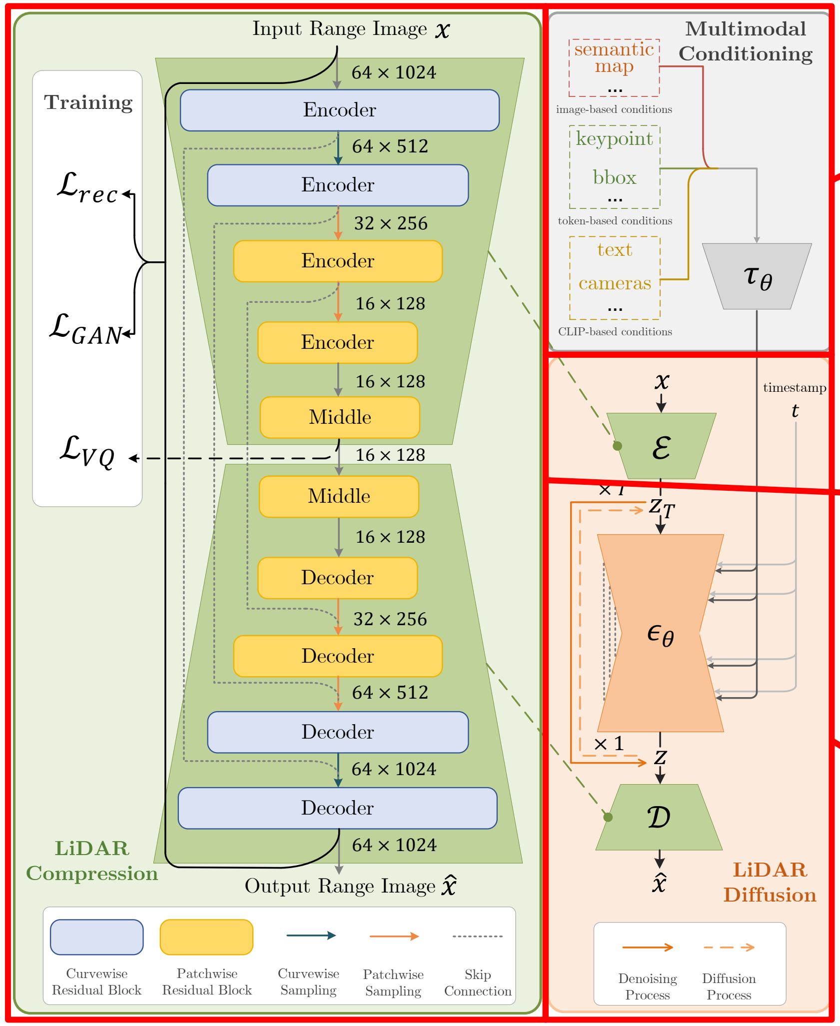
Controllable DM
Latent Diffusion (CVPR'22)

LiDAR Simulation
CORLA (CoRL'17)

LiDAR Generation
LiDARGen (ECCV'22)

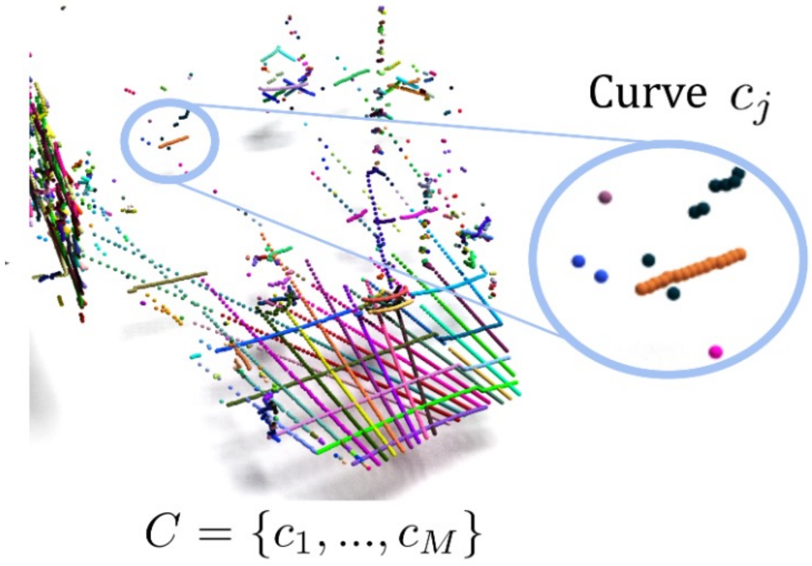


Problem Formulation



- **Input & Output: range images** $x, \hat{x} \in \mathbb{R}^{H \times W}$
- **Input Condition**:
  - Image-based (e.g., semantic maps)
  - Token-based (e.g., bbox)
  - CLIP-based (e.g., text prompts)
- **LiDAR Compression**:
  - Encoding $\mathcal{E} : x \to z$
  - Decoding $\mathcal{D} : z \to \hat{x}$
- **LiDAR Diffusion**:
  - Unconditional $z_{t-1} = \epsilon_\theta(z_t, t)$
  - Conditional $z_{t-1} = \epsilon_\theta(z_t, t, \tau_\theta(y))$

Towards LiDAR-Realistic Generation

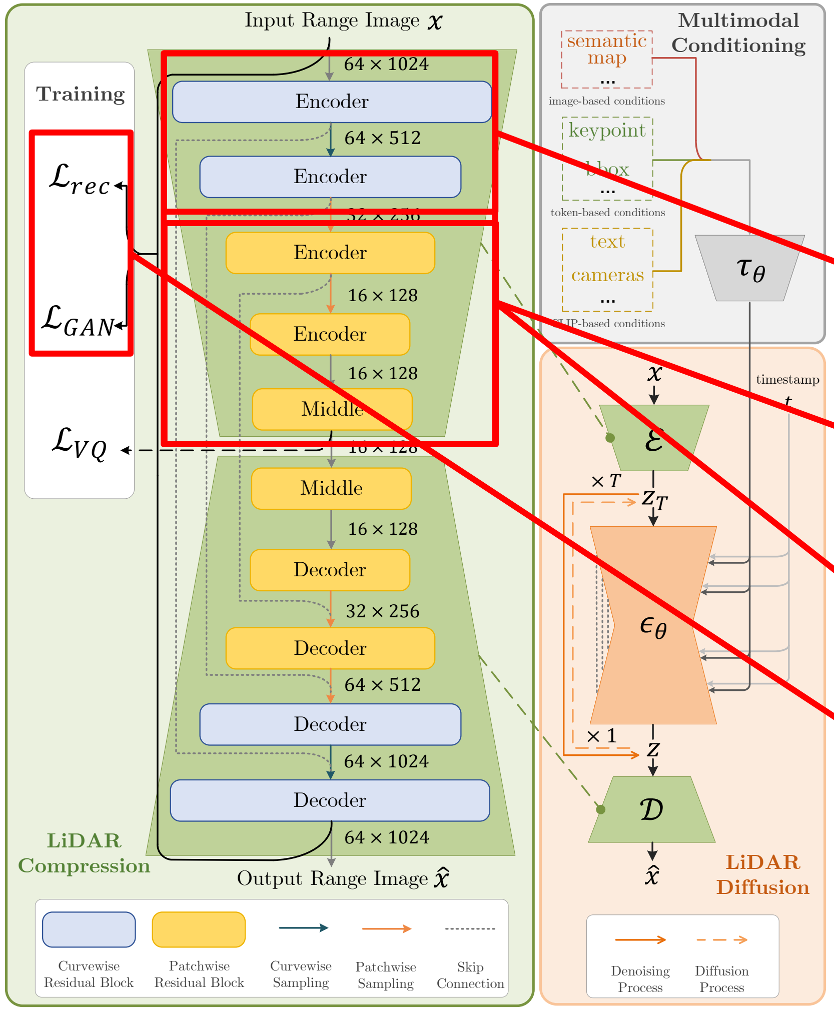


Curve Cloud [1]:
A sequence of points where consecutive point pairs are connected by a line segment, i.e., a polyline

- **Pattern Realism**: Compression with Curve-wise Encoding
- **Geometry Realism**: Point-wise Coordinate Supervision
- **Object Realism**: Incorporating Patches for Hybrid Encoding

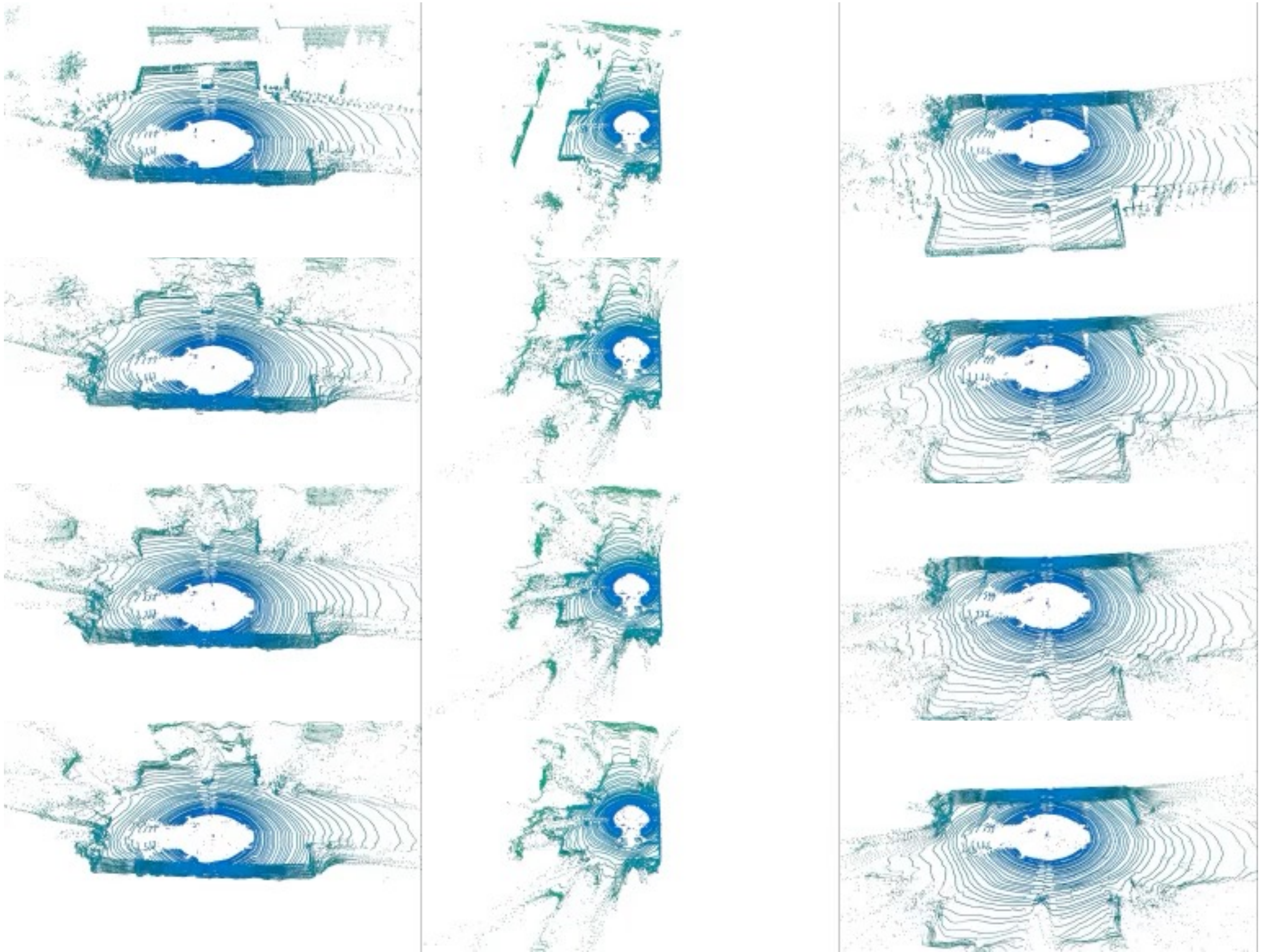
[1] Stearns, C., Liu, J., Rempe, D., Paschalidou, D., Park, J.J., Mascha, S. and Guibas, L.J., 2023. CurveCloudNet: Processing Point Clouds with 1D Structure. arXiv preprint arXiv:2303.12050.

Towards LiDAR-Realistic Generation



- **Pattern Realism**: Compression with Curve-wise Encoding
- **Object Realism**: Incorporating Patch-wise Encoding
- **Hybrid Encoding**: Curve-wise + Patch-wise Encoding
- **Geometry Realism**: Point-wise Coordinate Supervision

Hybrid Encoding boosts LiDAR Diffusion



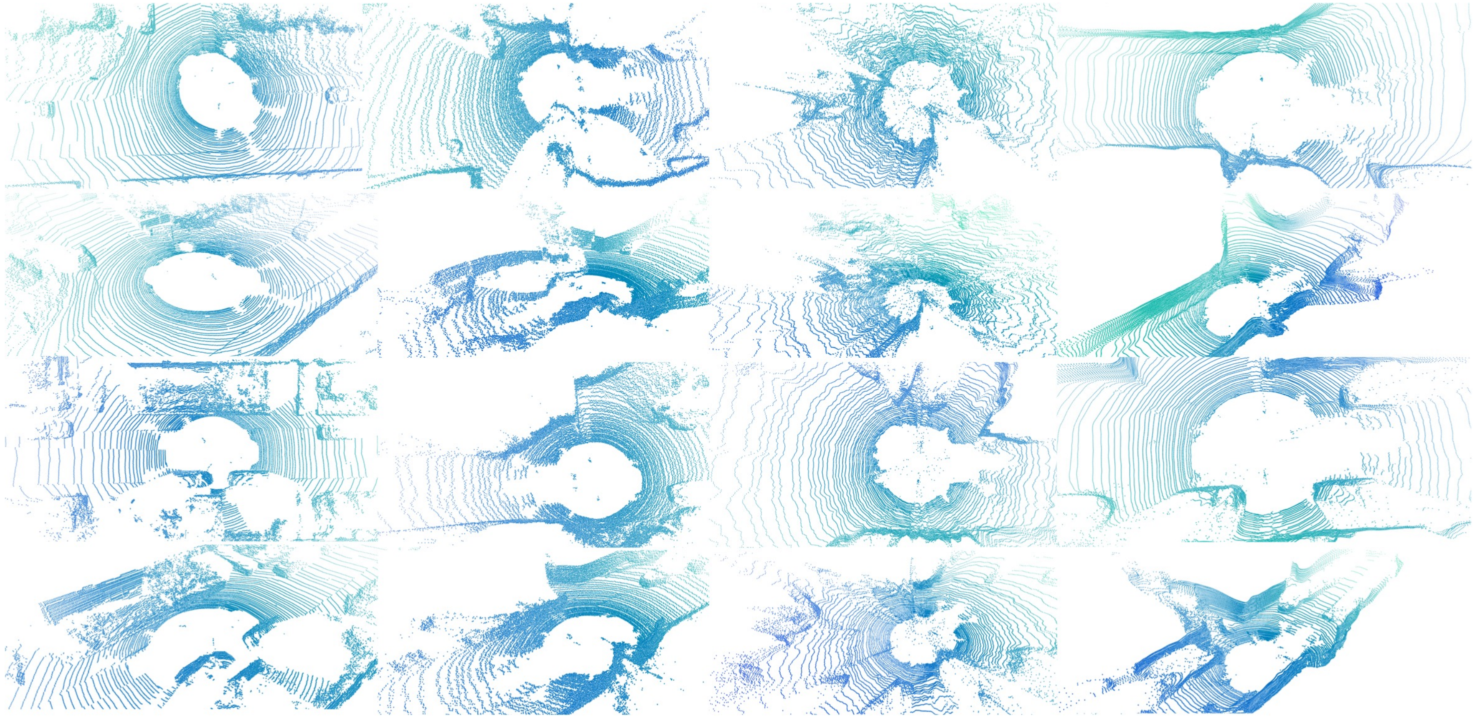
Ground Truth

Curvewise Encoding
Compression Rate: 16
FSVD(↓): 116 | JSD (↓): 0.232

Patchwise Encoding
Compression Rate: 16
FSVD(↓): 60 | JSD (↓): 0.230

Hybrid Encoding
Compression Rate: 32
FSVD(↓): 56 | JSD (↓): 0.228

Unconditional LiDAR Generation



Reference LiDARGen Latent Diffusion Ours

Visualization

Method	Perceptual			Statistical	
	FRID ↓	FSVD ↓	FPVD ↓	JSD ↓	MMD ↓ (10^{-4})
Noise	3277	497.1	336.2	0.360	32.09
LiDARGAN [7]	1222	183.4	168.1	0.272	4.74
LiDARVAE [7]	199.1	129.9	105.8	0.237	7.07
ProjectedGAN [54]	149.7	44.7	33.4	0.188	2.88
UltraLiDAR [67]	370.0	72.1	66.6	0.747	17.12
LiDARGen [75] (1160s)	129.1	39.2	33.4	**0.188**	**2.88**
LiDARGen [75] (50s)	2051	480.6	400.7	0.506	9.91
LDM [51] (50s)†	199.5	70.7	61.9	0.236	5.06
LiDM (ours, 50s)†	158.8	53.7	42.7	0.213	4.46
Δ *Improv.*	20.4%	24.0%	31.0%	9.7%	11.9%
LiDM (ours, 50s)	**125.1**	**38.8**	**29.0**	0.211	3.84
Δ *Improv.*	37.3%	45.1%	53.2%	10.6%	24.1%

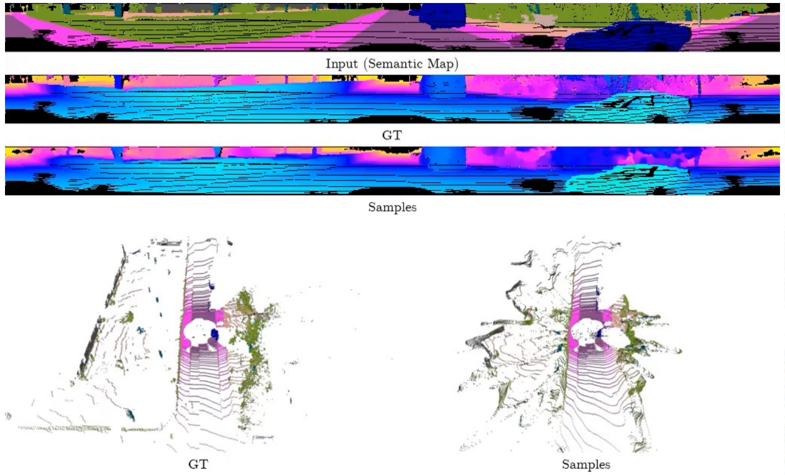
Quantitative Performance

Method	Diffusion Size	Throughput↑	Infer.Speed↑
LiDARGen [75]	64×1024	0.015	17.5
LiDM (ours)	16×128	**1.603**	**80.2**

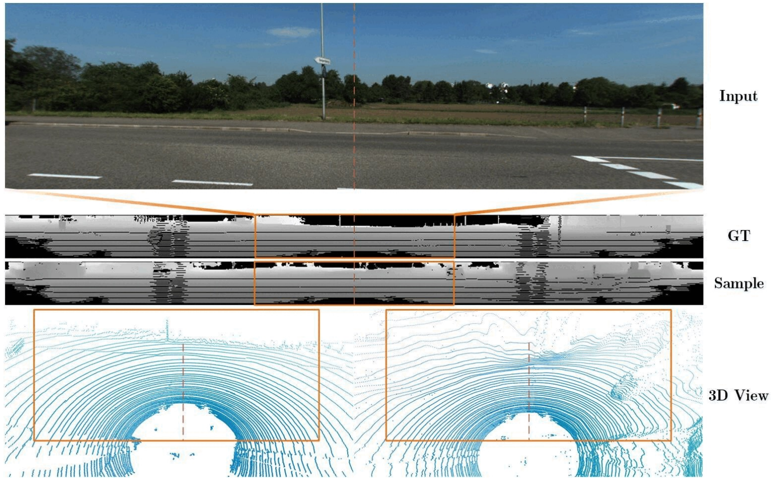
Overhead Analysis

- **Performance:** 10.6–53.2% improvements over the baseline
- **Efficiency:** 116× faster than previous SOTA

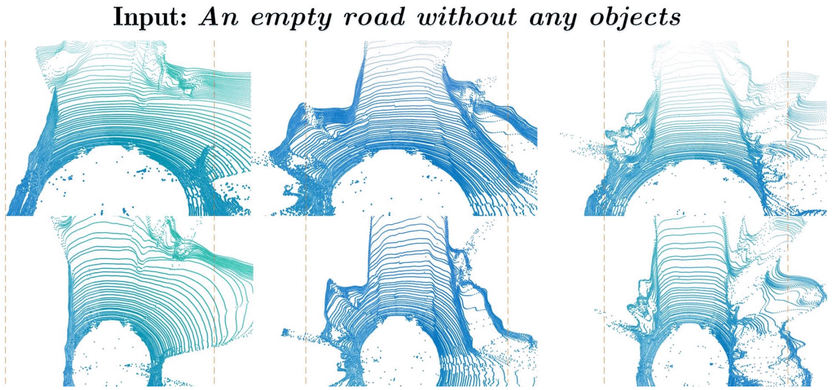
More Applications with LiDAR Diffusion



Semantic-Map-to-LiDAR



Camera-to-LiDAR



Text-to-LiDAR

Thanks!

Project Page

Github

arXiv