CoRL 2022

CADSim: Robust and Scalable in-the-wild 3D Reconstruction for Controllable Sensor Simulation

Jingkang Wang[1,2], Sivabalan Manivasagam[1,2], Yun Chen[1,2], Ze Yang[1,2],
Ioan Andrei Bârsan[1,2], Anqi Joyce Yang[1,2], Wei-Chiu Ma[1,3], Raquel Urtasun[1,2]

1 Waabi

2 University of Toronto

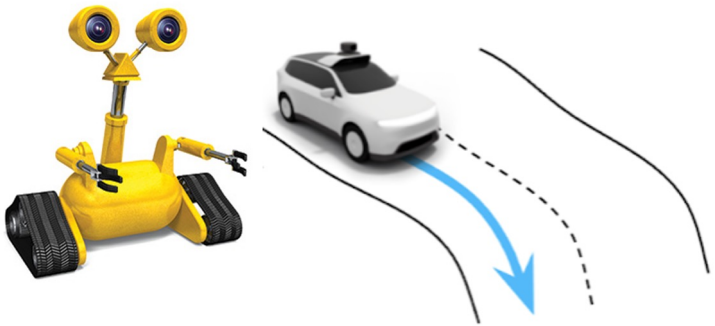
3 MIT

Outline

- Motivation
- Methodology
- Additional Experiment Results
- Applications – Mixed Reality Simulation, Texture Transfer

Simulation for Robot Learning and Testing

- Long-tail scenarios are critical for robot learning and evaluation
- We need scalable and affordable way to generate experiences - Simulation!
- Realistic sensor simulation is key for running the full autonomy system



Autonomy testing with sensor simulation

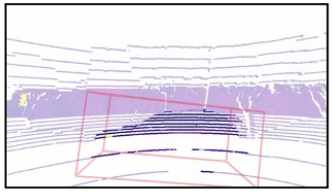
Existing Simulators Lack Scale and Diversity

- Standard game engines for simulation - (a) not scalable: artists create assets manually + simple automation; (b) lacking diversity; (c) not realistic
- We need to cover the full space in the real world for small domain gap

[1] CARLA: An open urban driving simulator. [Dosovitskiy, et al. CoRL 2017]

Building Assets from In-the-Wild Data for Diversity

LiDAR Points



Camera Images

CADSim

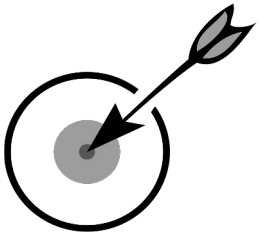
Fast Recon.

360° Textured
Vehicle Assets

Real-time rendering

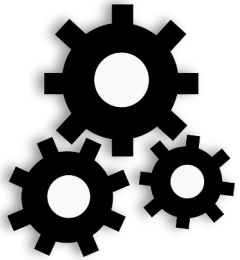
Realistic and Controllable Simulation

Assets for Self-driving



Accurate shape and appearance

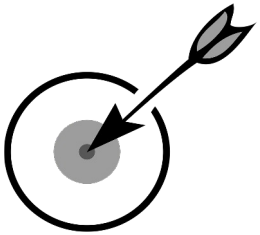
Editable



Controllable

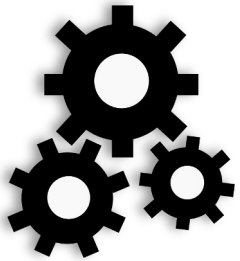
Real-time rendering

Assets for Self-driving



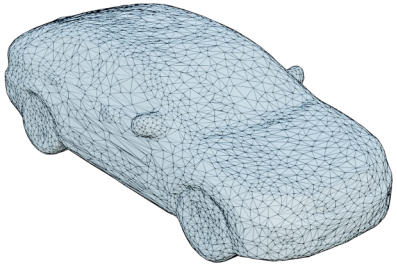
Accurate shape and appearance

Editable



Controllable

Real-time rendering



Mesh Representation

Current Mesh Approaches do not work in the Wild

- Underlying geometry is poor

Recon.

NeuS [1]

NeRS [2]

NVDiffRec [3]

Sphere Deform [4]

Ellipsoid Deform [5]

SAMP [6]

[1] Wang et al. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. NeurIPS 2021
[2] Zhang et al. NeRS: Neural Reflectance Surfaces for Sparse-view 3D Reconstruction in the Wild. NeurIPS 2022
[3] Munkberg et al. Extracting Triangular 3D Models, Materials, and Lighting From Images. CVPR 2022.
[4] Kanazawa et. al. Learning Category-Specific Mesh Reconstruction from Image Collections. ECCV 2018.
[5] Wang et. al. Pixel2mesh: Generating 3d mesh models from single rgb images. ECCV 2018.
[6] Engelmann et al. SAMP: Shape and Motion Priors for 4D Vehicle Reconstruction. WACV 2017.

Current Mesh Approaches do not work in the Wild

- Underlying geometry is poor

Recon.

NeuS [1]

NeRS [2]

NVDiffRec [3]

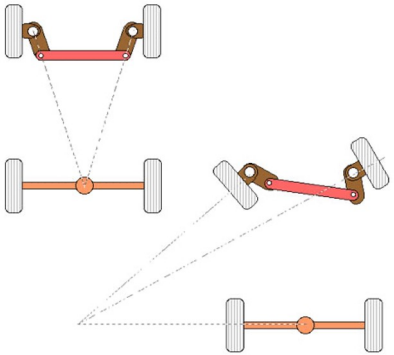
Sphere Deform [4]

Ellipsoid Deform [5]

SAMP [6]

- Generates rigid mesh that cannot be articulated

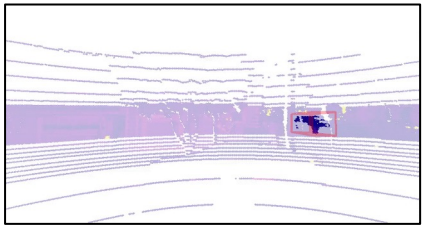
Needs to be articulated



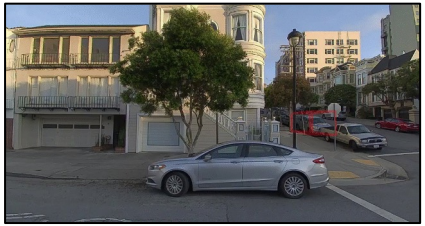
[1] Wang et al. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. NeurIPS 2021
[2] Zhang et al. NeRS: Neural Reflectance Surfaces for Sparse-view 3D Reconstruction in the Wild. NeurIPS 2022
[3] Munkberg et al. Extracting Triangular 3D Models, Materials, and Lighting From Images. CVPR 2022.
[4] Kanazawa et. al. Learning Category-Specific Mesh Reconstruction from Image Collections. ECCV 2018.
[5] Wang et. al. Pixel2mesh: Generating 3d mesh models from single rgb images. ECCV 2018.
[6] Engelmann et al. SAMP: Shape and Motion Priors for 4D Vehicle Reconstruction. WACV 2017.

CADSim

LiDAR Points



Camera Images

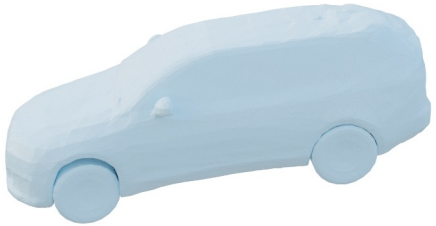


Generic CAD Models

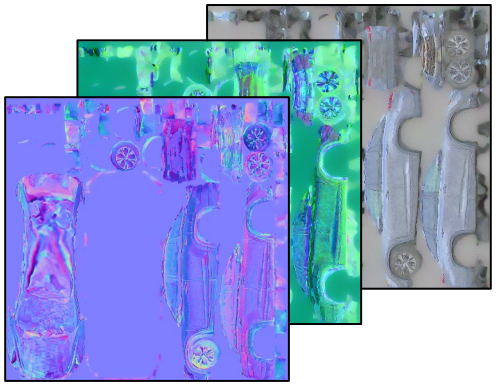
CADSim

Fast Recon.

Geometry



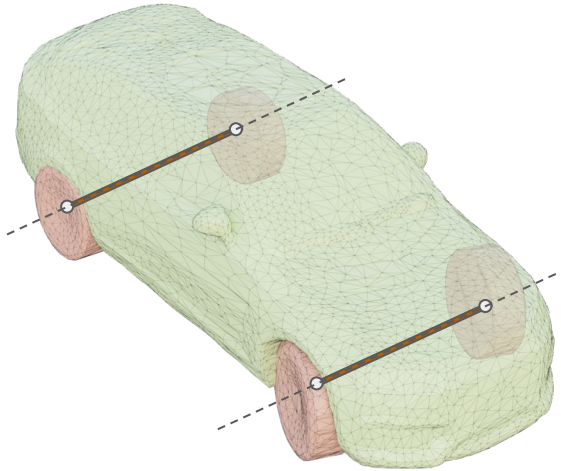
Material



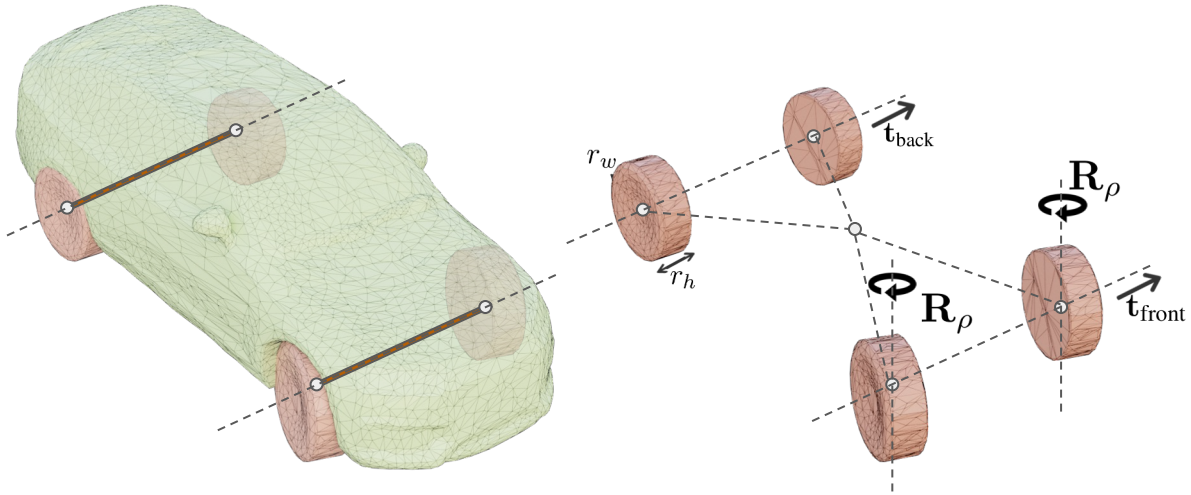
Appearance

Lighting

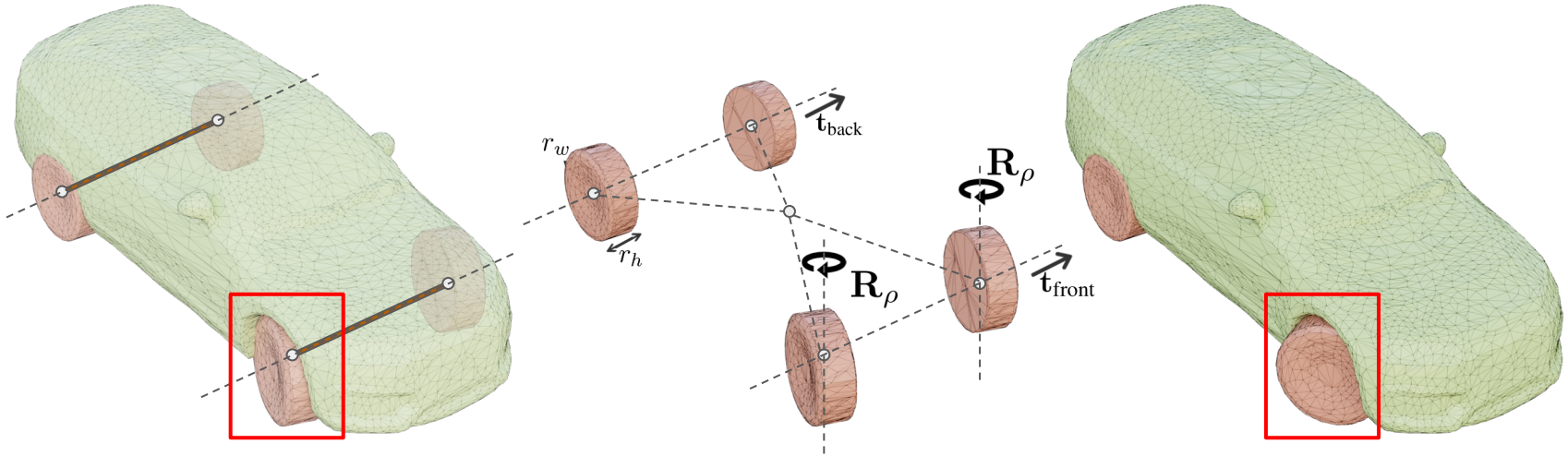
CADSim - Vehicle Parameterization



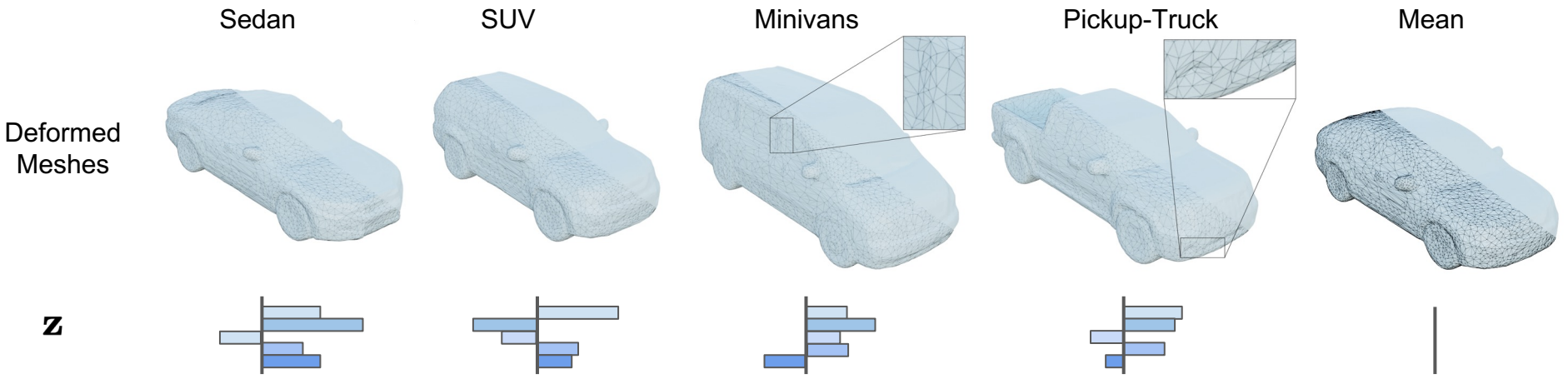
CADSim - Vehicle Parameterization



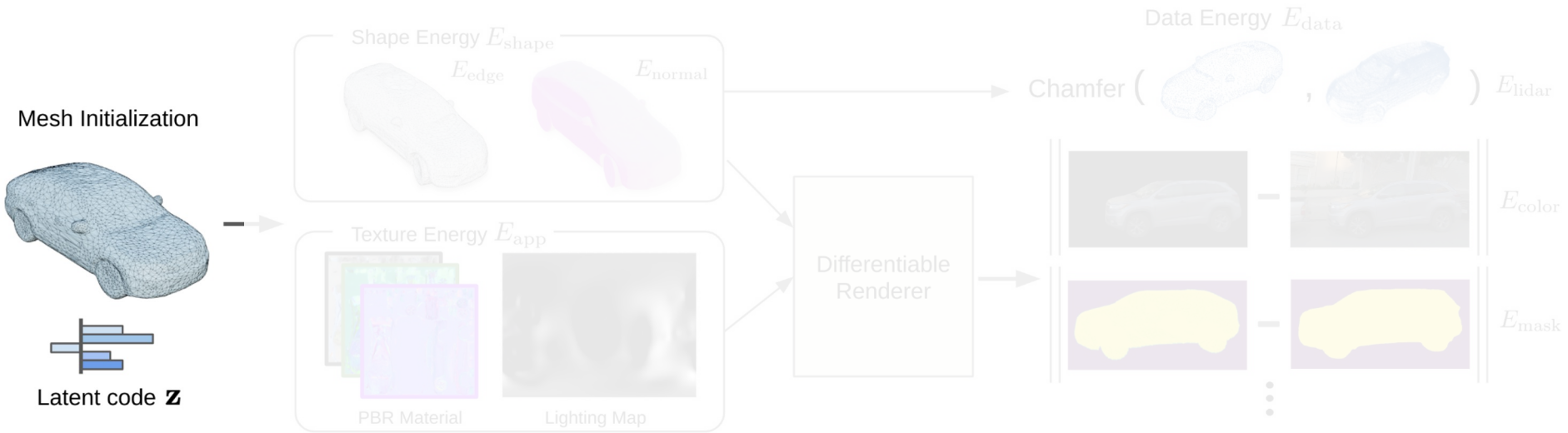
CADSim - Vehicle Parameterization



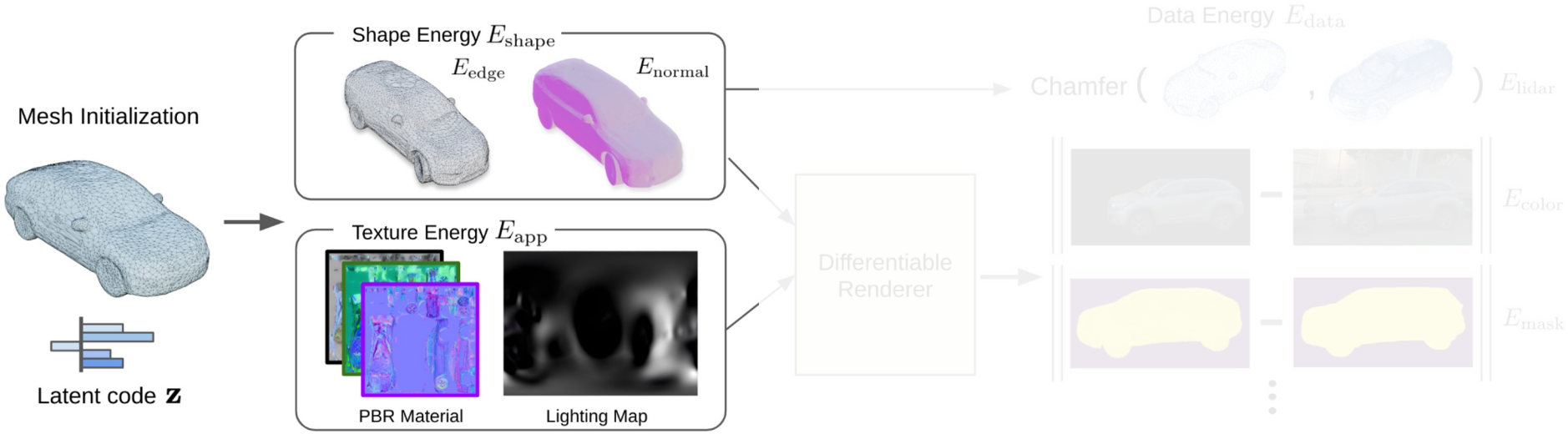
Learning a Shape Prior over a CAD Library



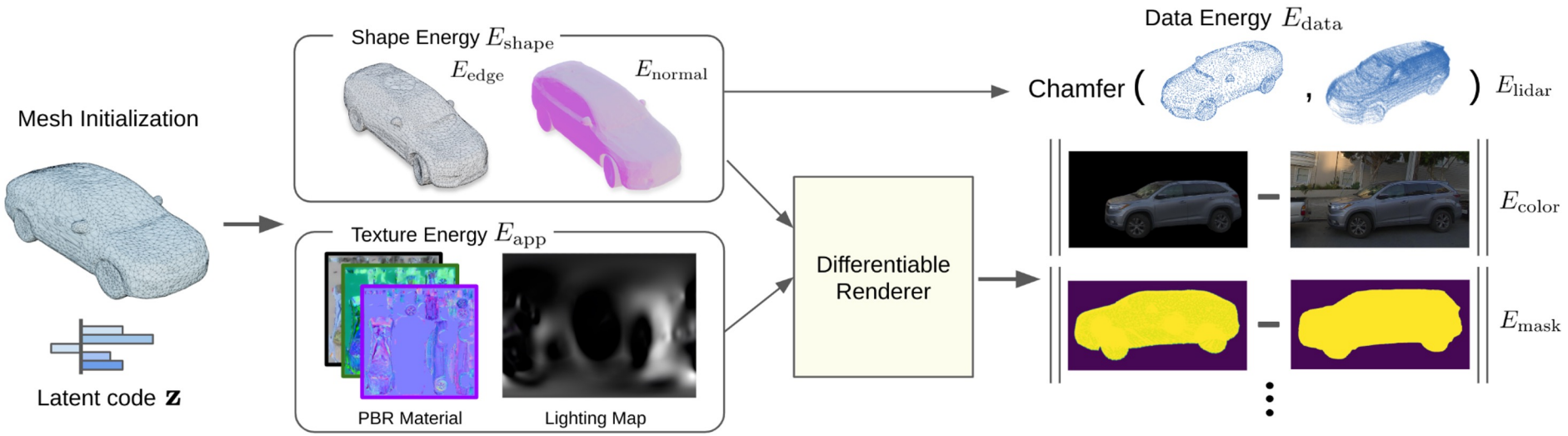
CADSim - Differentiable Rendering



CADSim - Differentiable Rendering



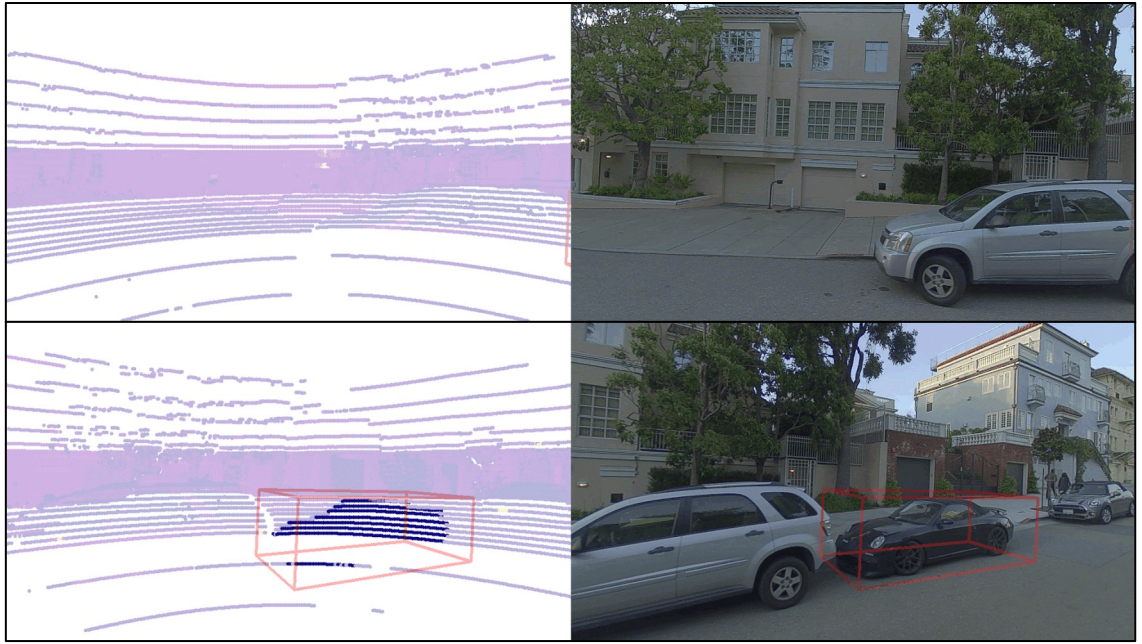
CADSim - Differentiable Rendering



PandaVehicle Dataset

Training frames (left camera)

Testing frames (front-left camera)



LiDAR sensor

Camera sensor

Novel View Synthesis SOTA Comparison

Ground-Truth

Instant-NGP [1]

NeuS [2]

SAMP [3]

CADSim (Ours)

[1] Müller et al. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding SIGGRAPH 2022

[2] Wang et al. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. NeurIPS 2021

[3] Engelmann et al. SAMP: Shape and Motion Priors for 4D Vehicle Reconstruction. WACV 2017.

Novel View Synthesis SOTA Comparison

Ground-Truth

Instant-NGP [1]

NeuS [2]

SAMP [3]

CADSim (Ours)

[1] Müller et al. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding SIGGRAPH 2022

[2] Wang et al. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. NeurIPS 2021

[3] Engelmann et al. SAMP: Shape and Motion Priors for 4D Vehicle Reconstruction. WACV 2017.

Novel View Synthesis SOTA Comparison

Ground-Truth

Instant-NGP [1]

NeuS [2]

SAMP [3]

CADSim (Ours)

[1] Müller et al. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding SIGGRAPH 2022

[2] Wang et al. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. NeurIPS 2021

[3] Engelmann et al. SAMP: Shape and Motion Priors for 4D Vehicle Reconstruction. WACV 2017.

Novel View Synthesis SOTA Comparison

Ground-Truth

Instant-NGP [1]

NeuS [2]

SAMP [3]

CADSim (Ours)

[1] Müller et al. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding SIGGRAPH 2022

[2] Wang et al. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. NeurIPS 2021

[3] Engelmann et al. SAMP: Shape and Motion Priors for 4D Vehicle Reconstruction. WACV 2017.

Quantitative Comparison

- CADSim produces the best performance on all metrics

Method	SSIM ↑	LPIPS ↓	T (hour)	FPS
NeRF++ [Zhang et al., 2020]	0.611	0.300	4.70	0.05
Instant-NGP [Müller et al., 2022]	0.641	0.319	0.05	1.14
NeRS [Zhang et al., 2021]	0.562	0.265	1.37	3.23
NVDiffRec [Munkberg et al., 2021]	0.593	0.396	1.07	51.2
NeuS [Wang et al., 2021]	0.640	0.247	6.25	0.02
SI-ViewWarp [Tulsiani et al., 2018]	0.514	0.371	–	1.67
SAMP [Engelmann et al., 2017]	0.628	0.283	0.09	71.4
CADSim (ours)	**0.674**	**0.220**	0.13	49.6

Quantitative Comparison

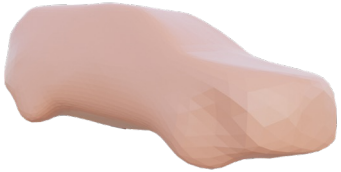
- CADSim produces the best performance on all metrics
- CADSim results in fast reconstruction and real-time rendering

Method	SSIM ↑	LPIPS ↓	T (hour)	FPS
NeRF++ [Zhang et al., 2020]	0.611	0.300	4.70	0.05
Instant-NGP [Müller et al., 2022]	0.641	0.319	0.05	1.14
NeRS [Zhang et al., 2021]	0.562	0.265	1.37	3.23
NVDiffRec [Munkberg et al., 2021]	0.593	0.396	1.07	51.2
NeuS [Wang et al., 2021]	0.640	0.247	6.25	0.02
SI-ViewWarp [Tulsiani et al., 2018]	0.514	0.371	–	1.67
SAMP [Engelmann et al., 2017]	0.628	0.283	0.09	71.4
CADSim (ours)	**0.674**	**0.220**	0.13	49.6

Leveraging CAD Improves Reconstruction

Unit Sphere

Ellipsoid

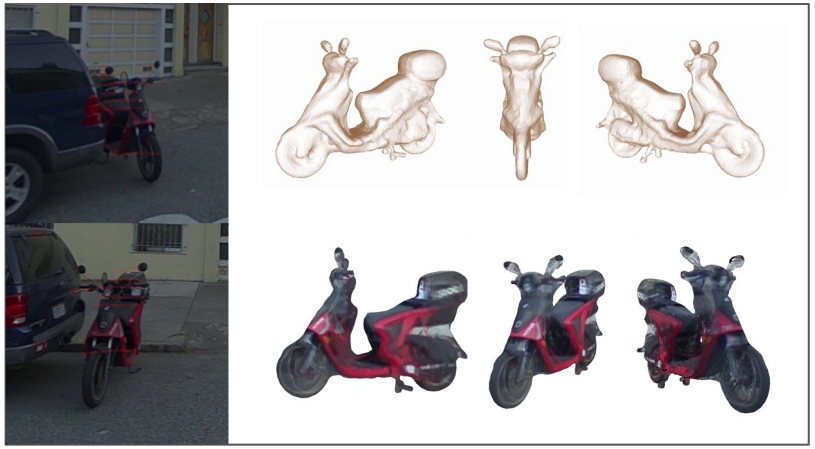


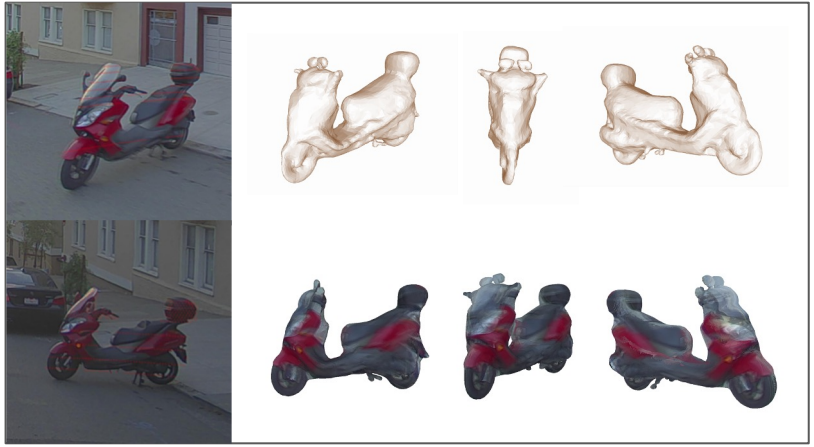
NeRS

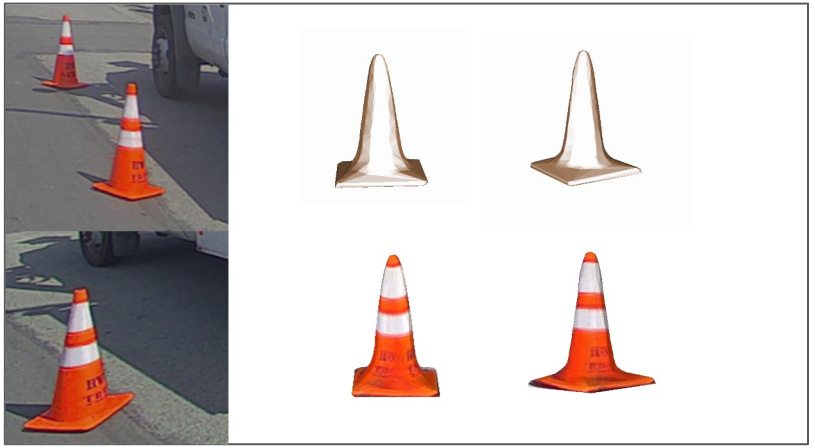
SAMP

Ours

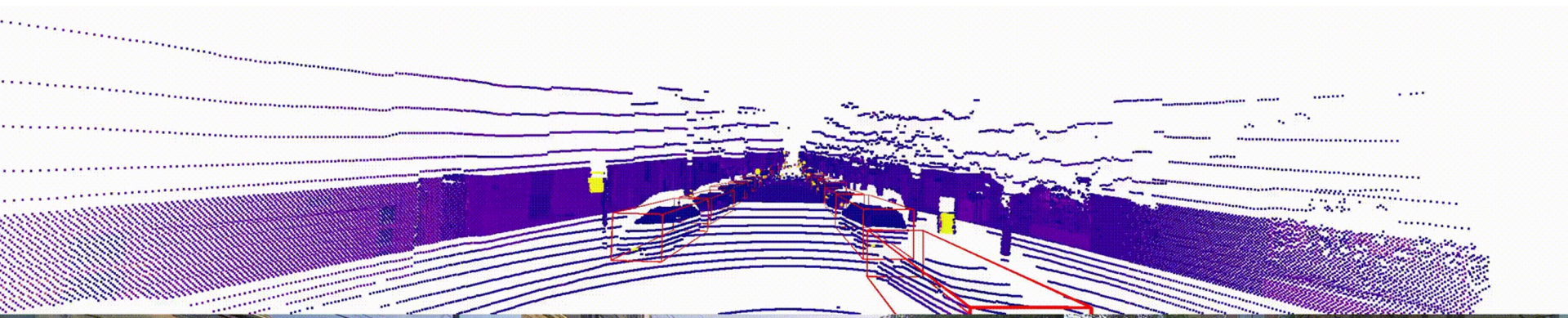
Results on Non-vehicle Objects





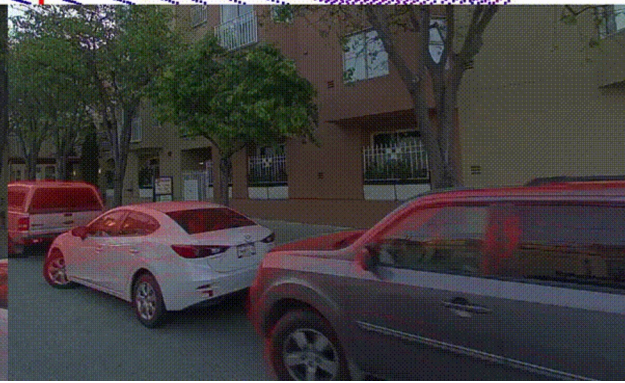


Sensor Observations (Log 028)



Left Camera

Front Camera

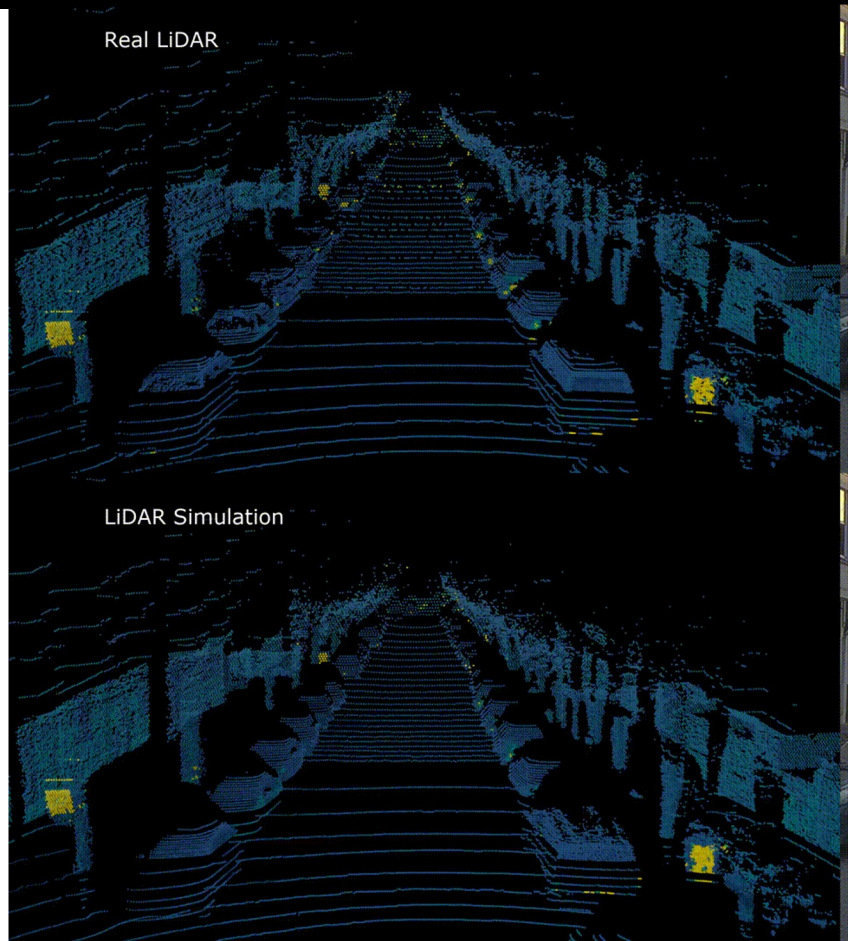


Right Camera

Reconstructing Nearby Vehicles (Log 028)

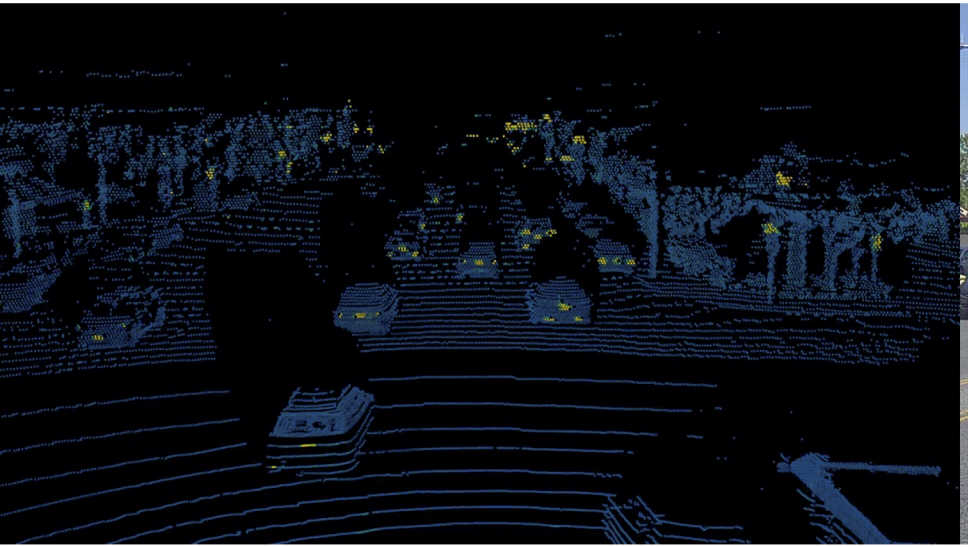
Reconstructing Nearby Vehicles (Log 028)

Log-Replay Simulation (Log 028)

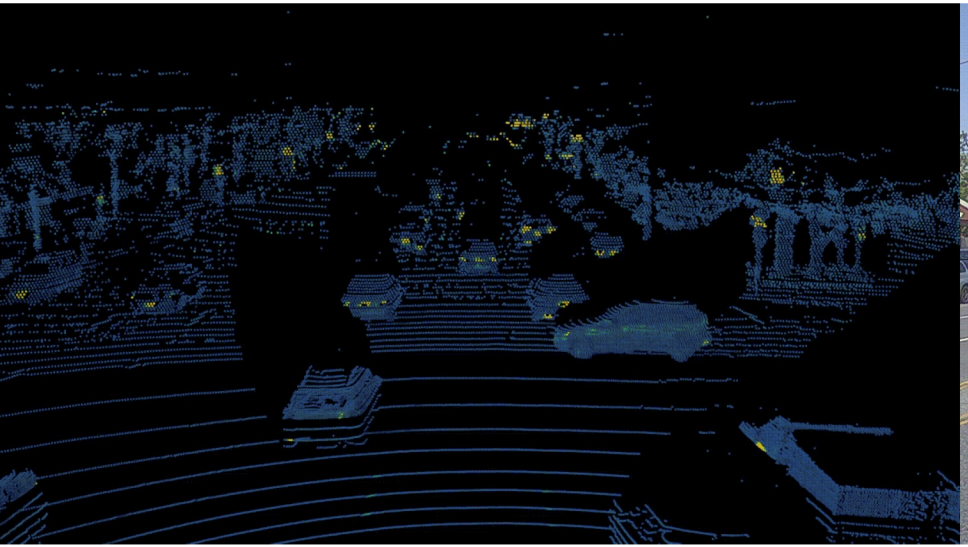


Log-Replay Simulation (Log 028) - Side Camera

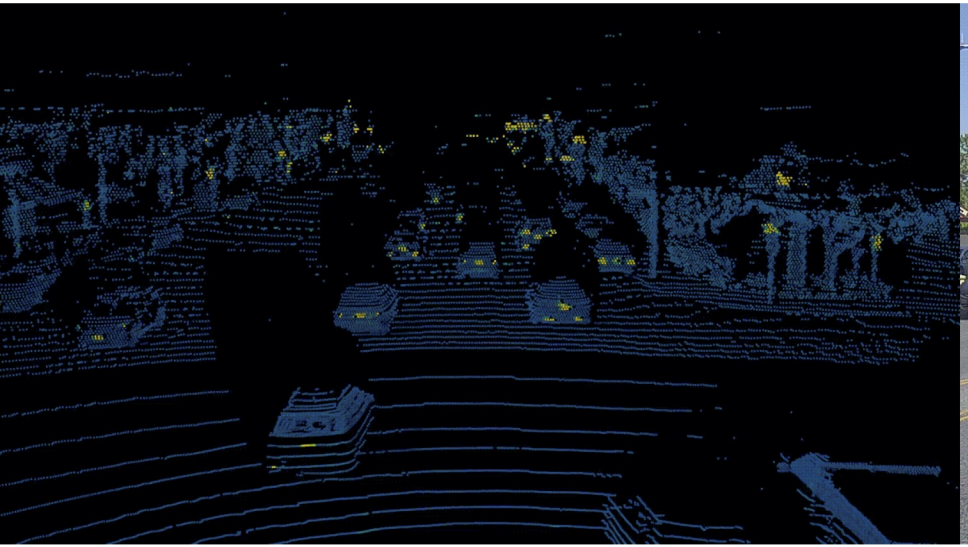
Mixed Reality: Actor Manipulation



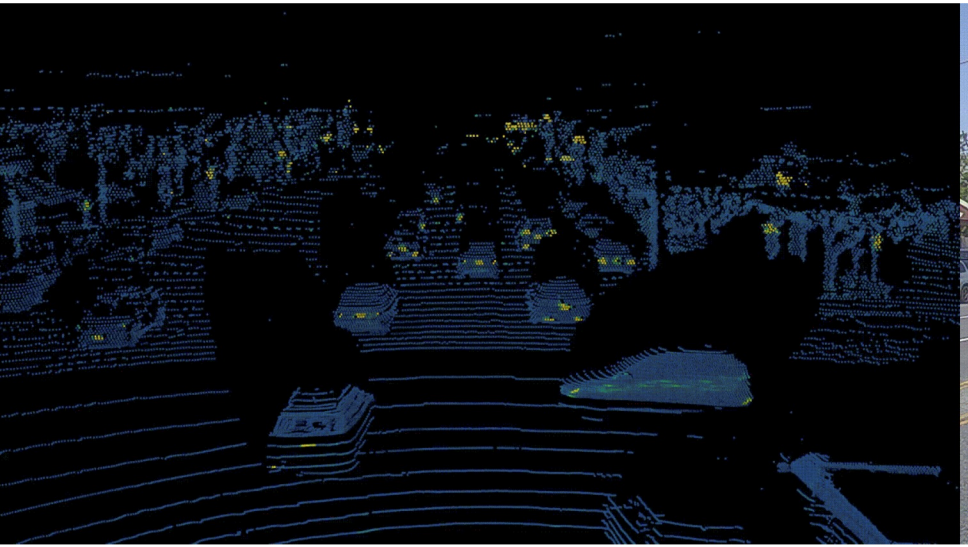
Mixed Reality: Actor Manipulation



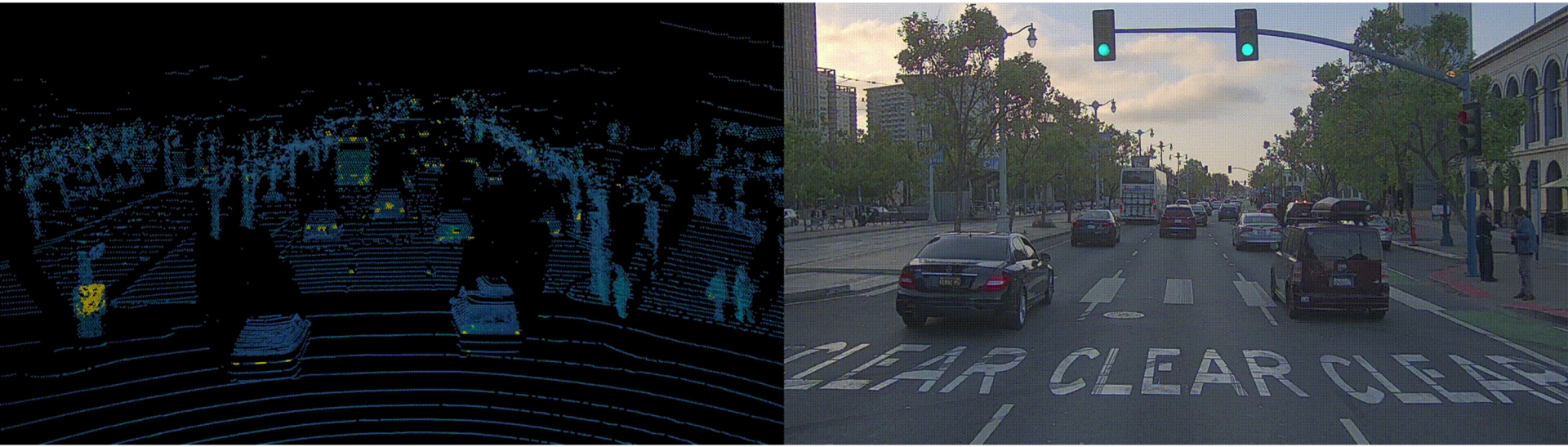
Mixed Reality: Safety-Critical Scenario



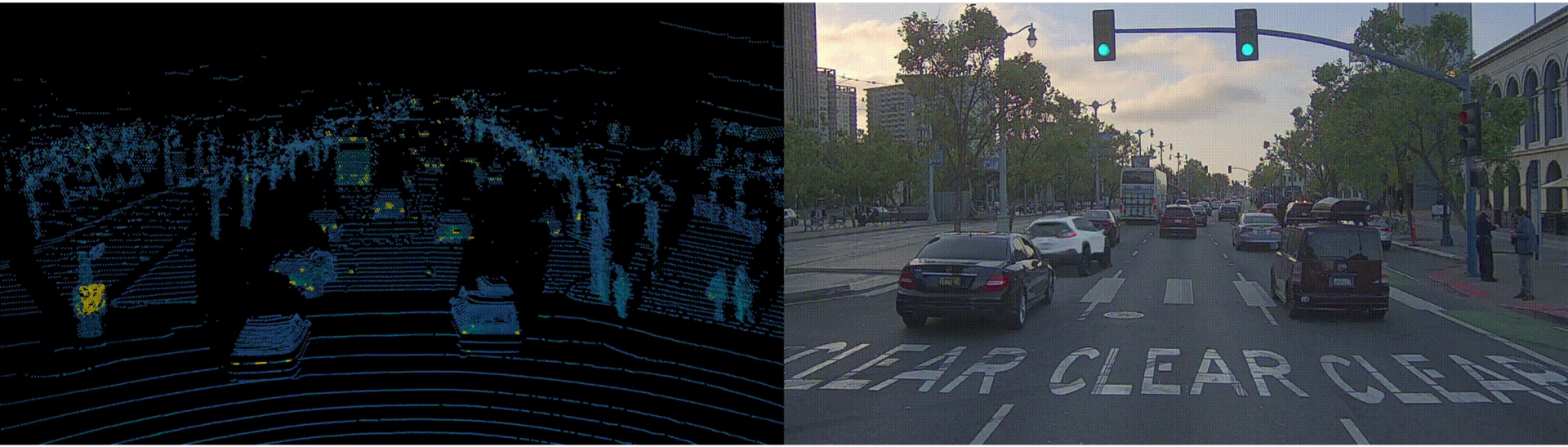
Mixed Reality: Safety-Critical Scenario



Mixed Reality: Safety-Critical Scenario



Mixed Reality: Safety-Critical Scenario



Swapping Texture in the Real World

Swapping Texture in the Real World

Swapping Texture in the Real World

Swapping Texture in the Real World

Swapping Texture in the Real World

Swapping Texture in the Real World

Swapping Texture in the Real World

Swapping Texture in the Real World

Swapping Texture in the Real World

Swapping Texture in the Real World

Thank you!

Robust and Scalable in-the-wild 3D Reconstruction for Realistic and Controllable Sensor Simulation

CoRL Paper ID 56

Supplementary Video

This video contains audio.