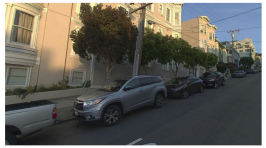
CADSim: Robust and Scalable in-the-wild 3D Reconstruction for Controllable Sensor Simulation

Jingkang Wang, Sivabalan Manivasagam, Yun Chen, Ze Yang, Ioan Andrei Bârsan, Anqi Yang, Wei-Chiu Ma, Raquel Urtasun
{wangjk,manivasagam,zeyang,yun,iab,ajyang,urtasun}@cs.toronto.edu weichium@mit.edu

Motivation: Realistic Sensor Simulation

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


Novel View Synthesis

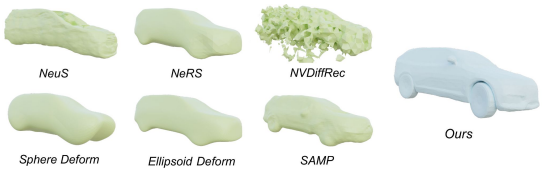

Mixed Reality (Actor Insertion)


Vehicle Re-Animation


Texture Transfer

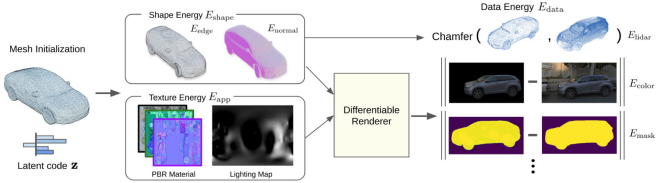
Building Assets from In-the-Wild Data

- **Building digital twins from the real world:**
  - **scalable**: data collection platform drives anywhere to collect data
  - **diverse**: different types of actors observed under different conditions
  - **realistic**: same operational area and smaller sim-real domain gap
- **Existing methods:**
  - poor underlying geometries under sparse and noisy observations
  - generated rigid mesh cannot be articulated
  - training is computationally expensive (>hours)
  - non real-time rendering (<30 FPS)



CADSim

- **CADSim:** (a) accurate shape and appearance, (b) editable & controllable, (c) fast and robust 3D reconstruction (d) real-time rendering



Munkberg et al., 2022

- **Energy minimization:**

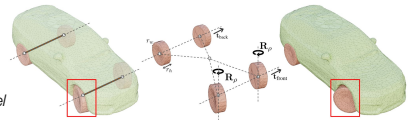
$$\underset{\mathcal{M},\Pi,\mathcal{A}}{\operatorname{argmin}}\{E_{\text{data}}(\mathcal{M},\Pi,\mathcal{A};\mathcal{I},\mathcal{P}) + \lambda_{\text{shape}}E_{\text{shape}}(\mathcal{M}) + \lambda_{\text{app}}E_{\text{app}}(\mathcal{M},\Pi,\mathcal{A};\mathcal{I},\mathcal{P})\}$$

Matching sensor observation | Shape priors | Material and light priors

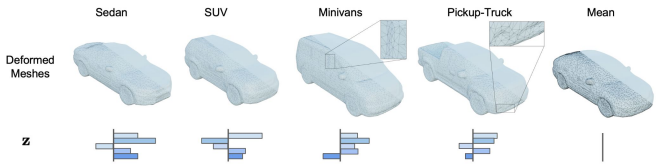
$\mathcal{M}$: mesh $\mathcal{A}$: appearance Π: sensor intrinsic/extrinsic $\mathcal{I}$: images $\mathcal{P}$: point cloud

- **Vehicle parameterization:**

(a) scale (tire radius and thickness)
(b) front wheel rotation / translation
(c) back wheel translation
— *physically plausible model*

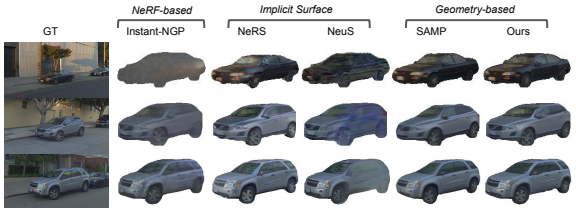


- **Learning shape priors from a CAD library:**



Results

- **Qualitative comparison with SOTA approaches:**



- **Quantitative comparison with SOTA approaches:**

Method	PSNR ↑	SSIM ↑	LPIPS ↓	T (hour)	FPS
Instant-NGP [Müller et al., 2022]	21.68	0.641	0.319	0.05	1.14
NeRS [Zhang et al., 2021]	18.49	0.562	0.265	1.37	3.23
NeuS [Wang et al., 2021]	21.37	0.640	0.247	6.25	0.02
SAMP [Engelmann et al., 2017]	19.52	0.628	0.283	0.09	71.4*
CADSim (ours)	**21.72**	**0.674**	**0.220**	0.13	49.6*

* using differentiable render nvdiffrast. Faster rendering (>100 FPS) is expected with modern graphics engines

- **Texture transfer in the real world:**

- **Mixed reality camera simulation for safety-critical scenarios:**

- **Limitations:** (a) fixed topology, (b) limited inpainting capacity, (c) requires segm. masks and camera parameters, (d) limited quality when topology is complex