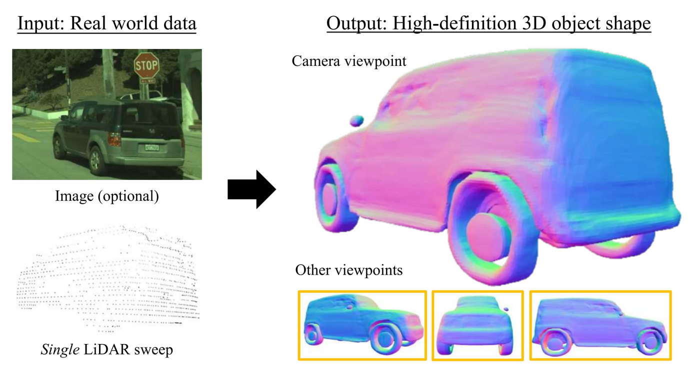
Mending Neural Implicit Modeling for 3D Vehicle Reconstruction in the Wild

Shivam Duggal*, Zihao Wang*, Wei-Chiu Ma, Sivabalan Manivasagam,
Justin Liang, Shenlong Wang, Raquel Urtasun

Uber ATG · University of Toronto · MIT

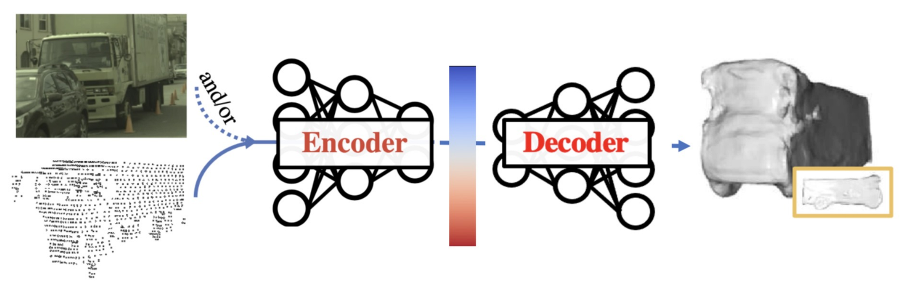
Introduction

- **High-fidelity 3D assets are core components of 3D simulation** softwares like – Unreal Engine, Carla, GeoSIM etc.
- Current deep implicit modeling approaches are:
  - Expressive
  - Easy to learn
  - Generate high-resolution reconstructions
  - Do not perform well on real-world sparse observations.
- Our approach learns strong shape priors from synthetic data and adapts them to generate **high-quality shapes in the wild**.



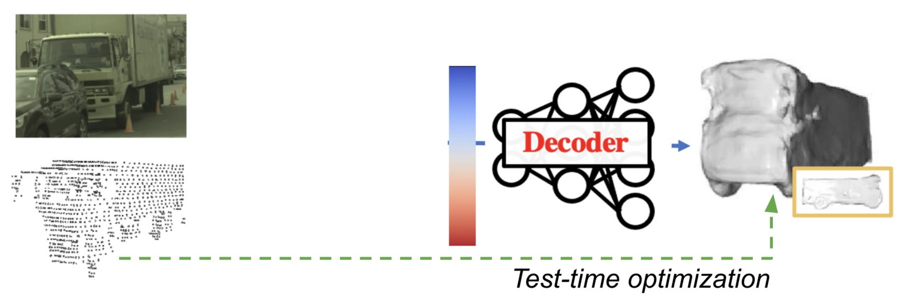
Related Work

Feed forward approaches (ONet, IMNet, etc.)



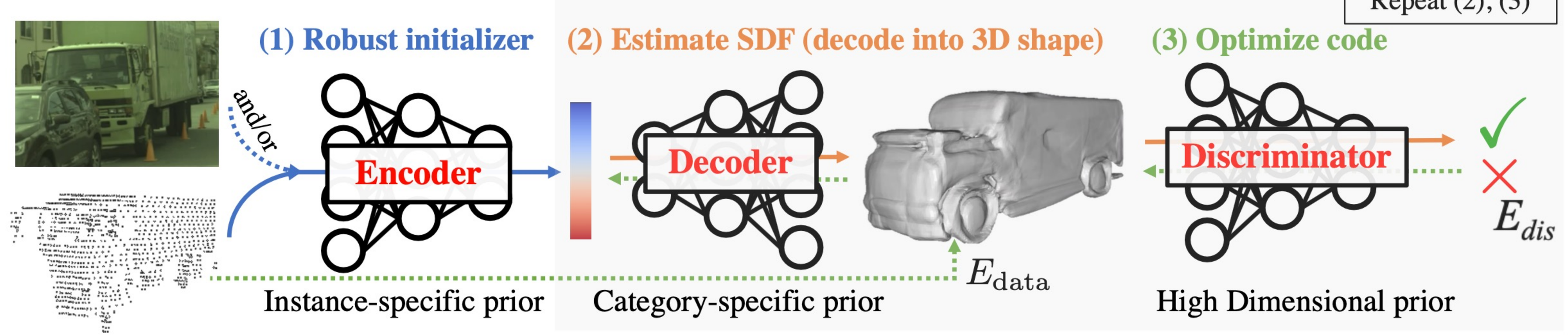
- ✅ Robust to noise
- ❌ Low-fidelity with GT at test-time, leading to smoother shapes

Deep Optimization approaches (DeepSDF, DIST etc.)

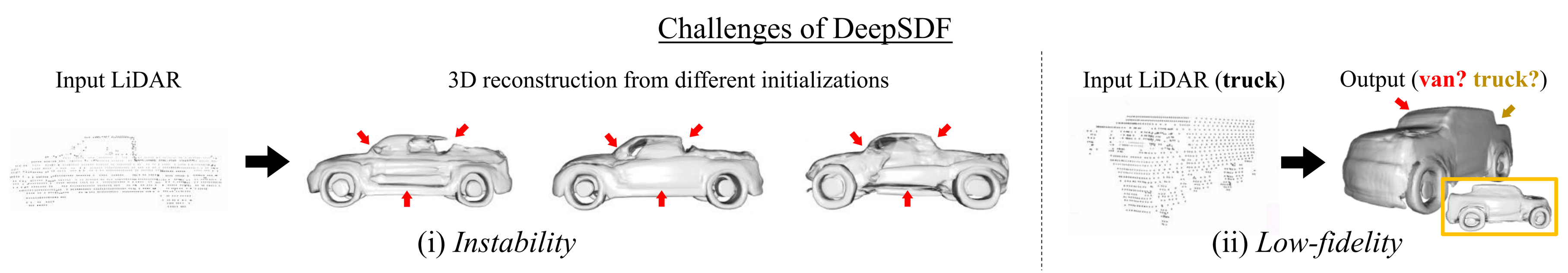


- ✅ Robust to noise
- ❌ Low-quality reconstructions to unseen inputs
- ❌ Unstable to latent-code initializations
- ❌ Lack of global shape consistency

Ours



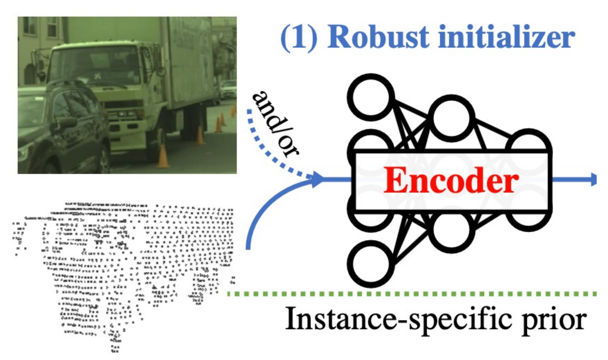
Issues with Prior work



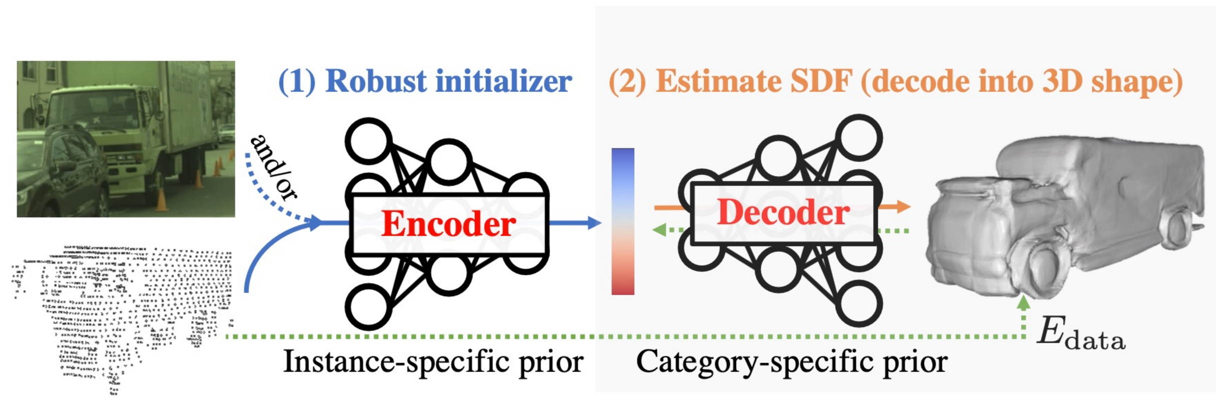
Mending Neural Implicit Approaches

Given raw sensory data, we first utilize deep encoder as a robust initializer for the shape code. The shape-code is then optimized through the auto-decoder framework, in presence of discriminator-induced high-dimensional shape prior.

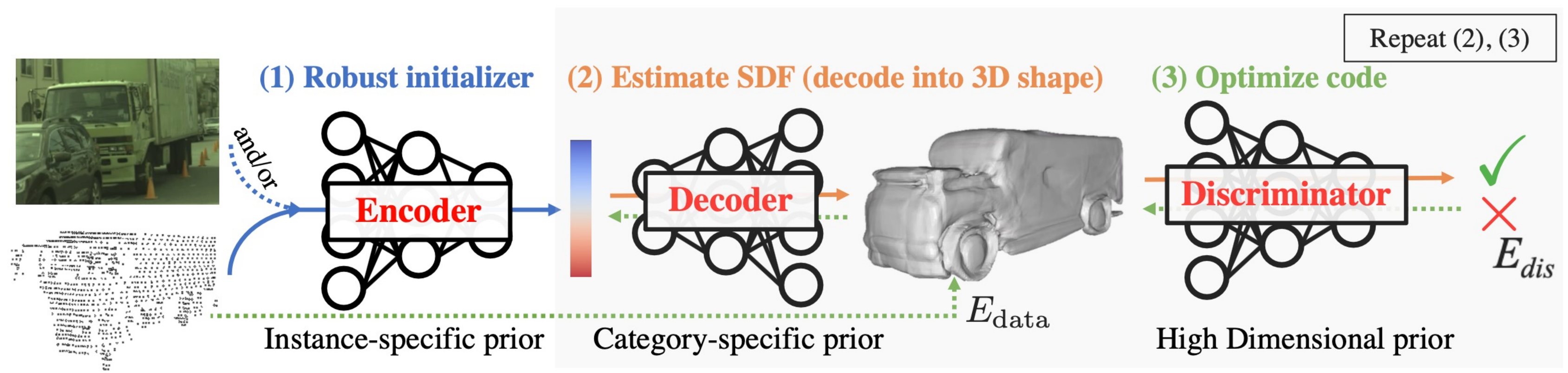
1. **Robust Initialization** of shape latent-code.



2. **Test-time Optimization** of shape latent-code.

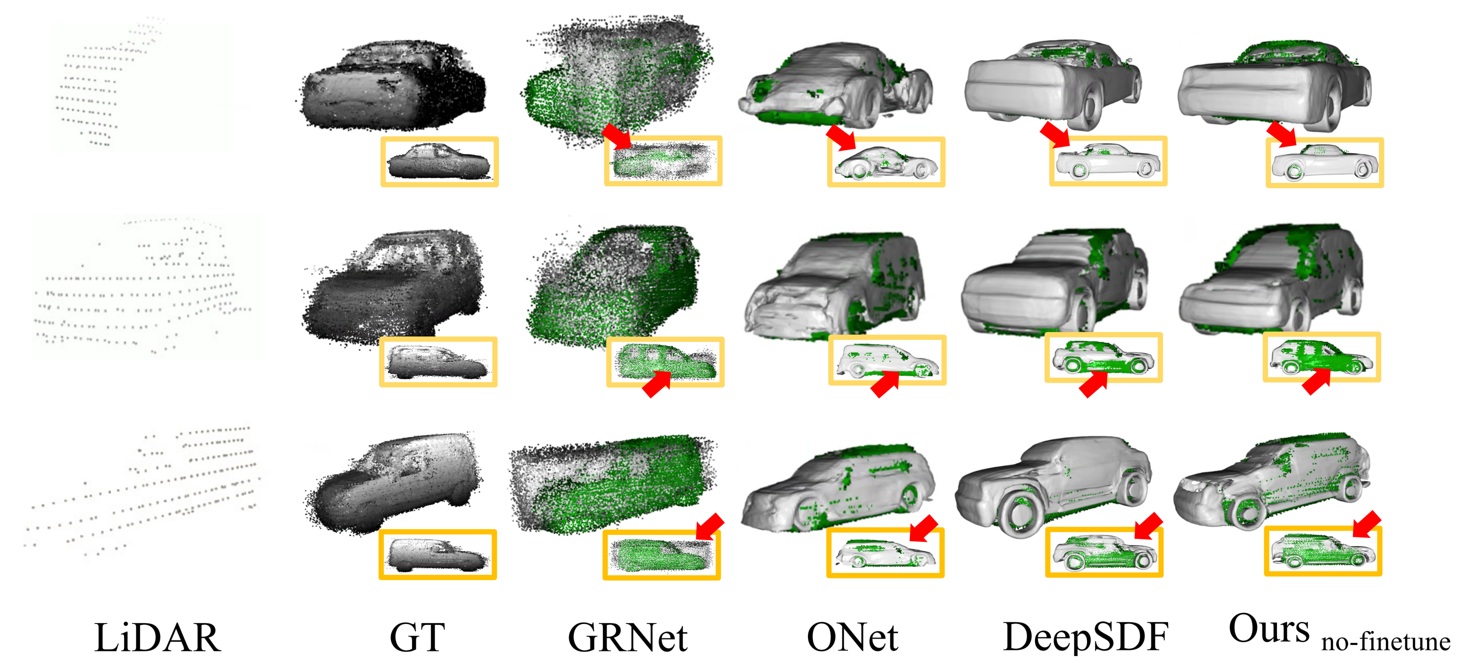


3. **High-dimensional learned shape prior** during training and optimization.



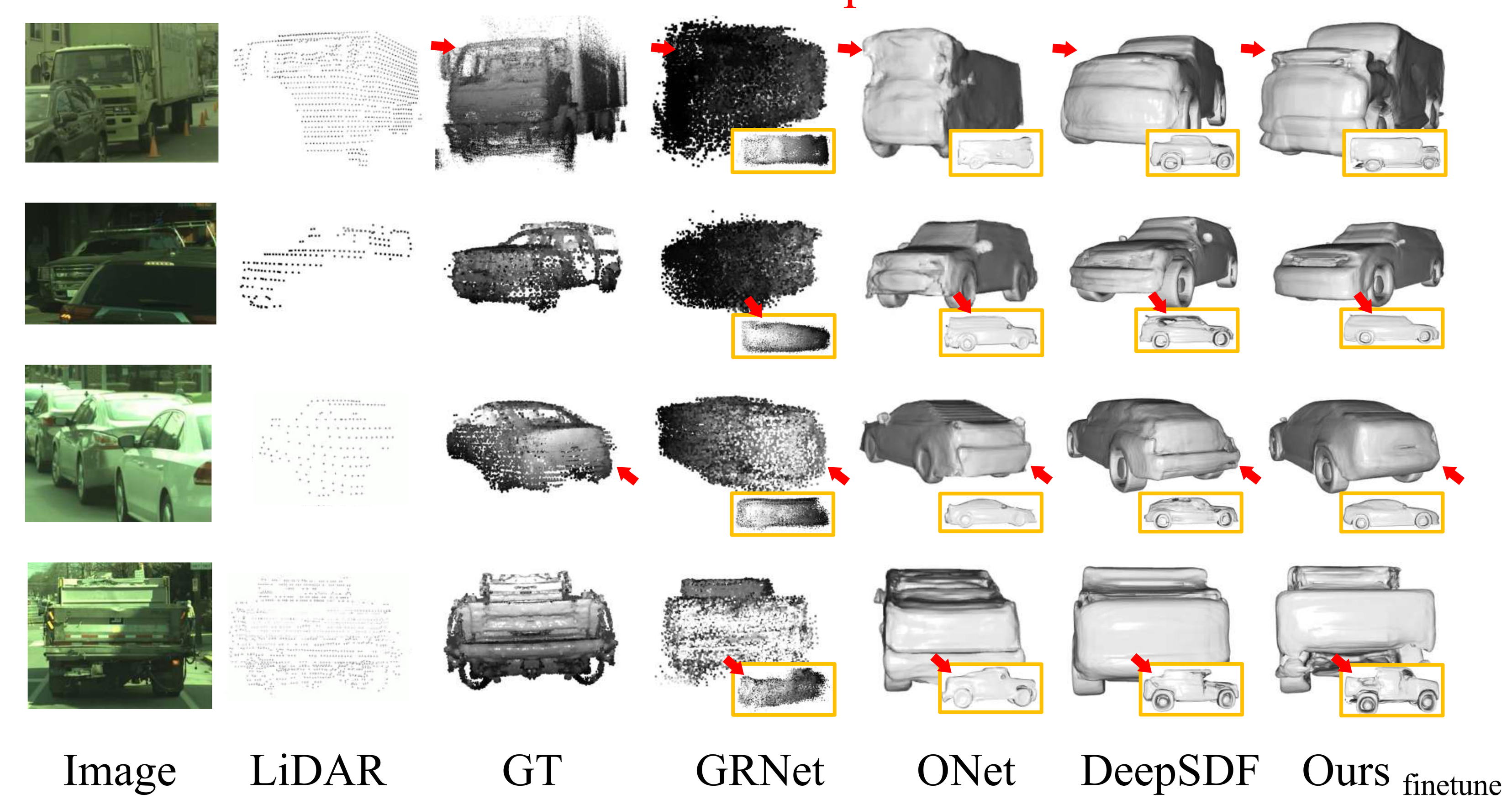
4. **Advesarial Curriculum Learning Strategy:** To allow each component encode rich shape priors, we proposed a multi-stage learning strategy.

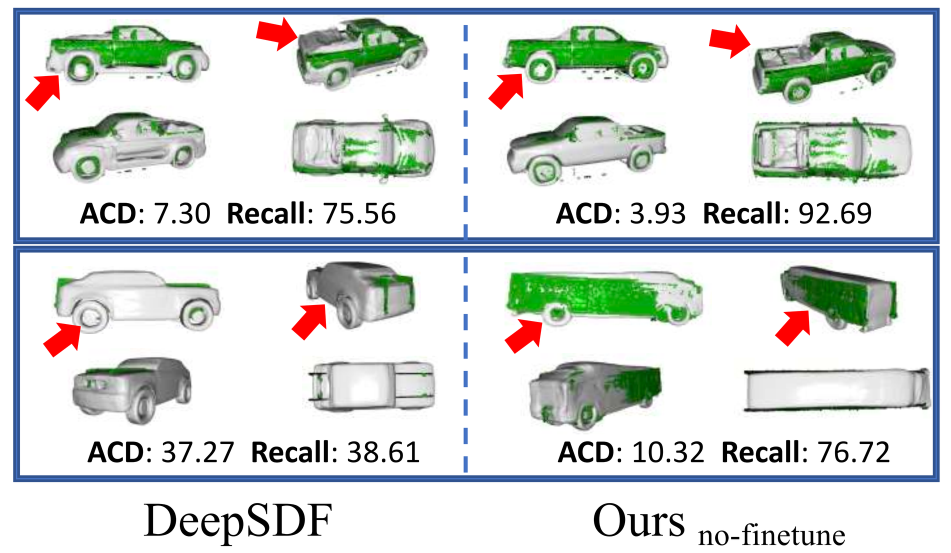
Evaluation on KITTI Dataset



Evaluation on NorthAmerica Dataset

LiDAR Completion



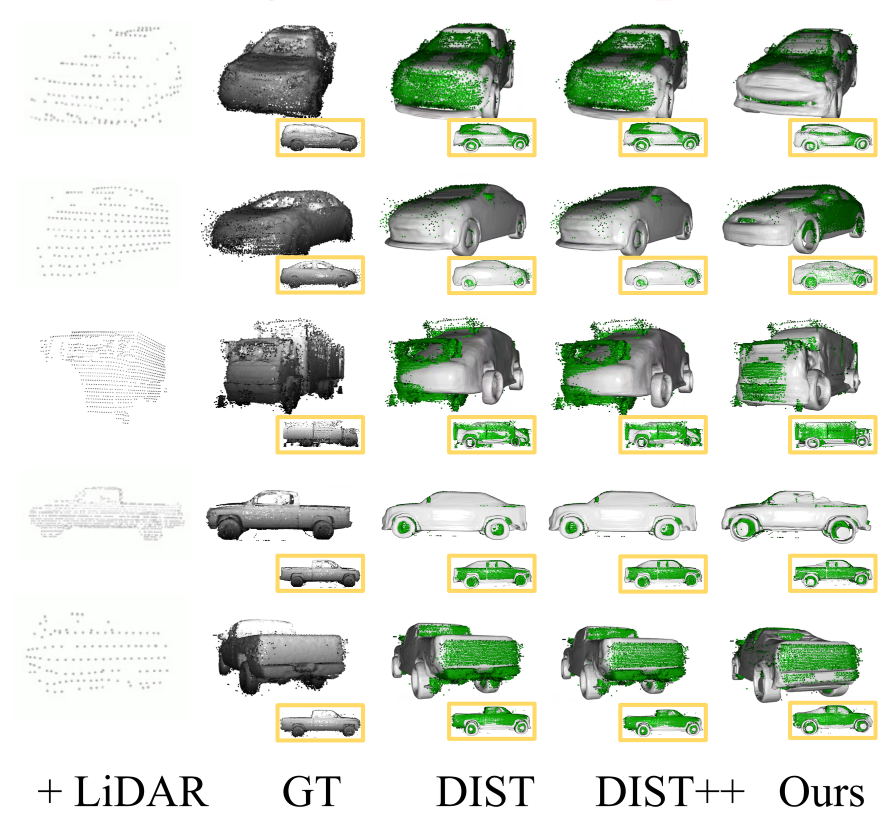


Method	ACD (mm) ↓	Recall (%) ↑
ONet	22.76	49.56
GRNet	12.70	77.59
SAMP	176.42	65.58
DIST	19.55	71.54
DIST++	17.29	72.50
DeepSDF	8.34	84.71
$\text{Ours}_{\text{no-finetune}}$	7.02	86.48
$\text{Ours}_{\text{finetune}}$	**5.93**	**88.18**

Image-based Reconstruction

Method	ACD (mm) ↓	Recall (%) ↑
DIST	62.97	48.82
Ours	**8.89**	**84.32**

Image + Lidar Completion



Method	ACD (mm) ↓	Recall (%) ↑
DIST	23.40	71.99
DIST++	17.52	72.65
Ours	**5.36**	**89.05**