ICCV23

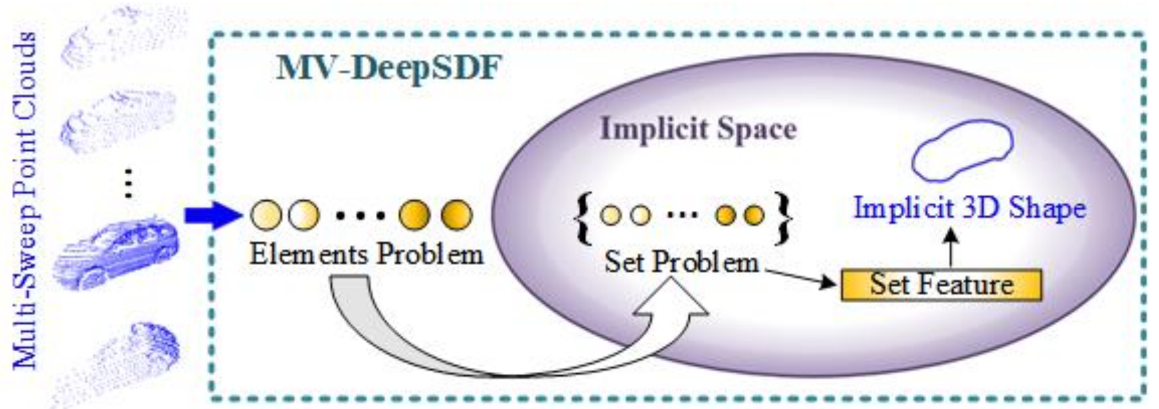
MV-DeepSDF: Implicit Modeling with Multi-Sweep Point Clouds for 3D Vehicle Reconstruction in Autonomous Driving

Yibo Liu, Kelly Zhu, Guile Wu, Yuan Ren, Bingbing Liu, Yang Liu, Jinjun Shan

HUAWEI · YORK UNIVERSITÉ UNIVERSITY · UNIVERSITY OF TORONTO

Introduction

Motivation



- Implicit modeling in-the-wild 3D vehicles from noisy and sparse multi-sweep partial point clouds.
- We simplify this task into an element-to-set feature extraction problem.
- We infer an optimal estimation of the 3D shape described in the abstract implicit space.

Contributions

- Consistency and complementarity analysis of multi-sweep in the latent feature space.
- Proposal of a novel MV-DeepSDF framework for implicit modeling.
- SOTA performance on KITTI and Waymo.

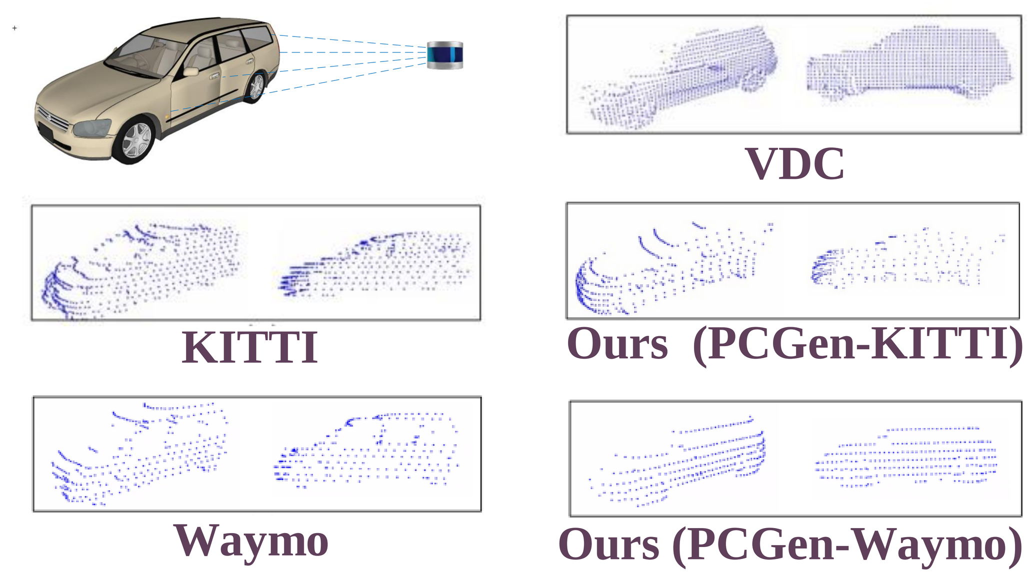
Model Training

Stage 1:

$$\underset{\theta,\{z_j\}_{j=1}^{J}}{\arg\min} \sum_{j=1}^{J}\left(\frac{1}{\sigma^2}\|z_j\|_2^2+\sum_{k=1}^{K}\mathcal{L}_1\left(f_\theta\left(z_j, x_k\right), s_k\right)\right)$$

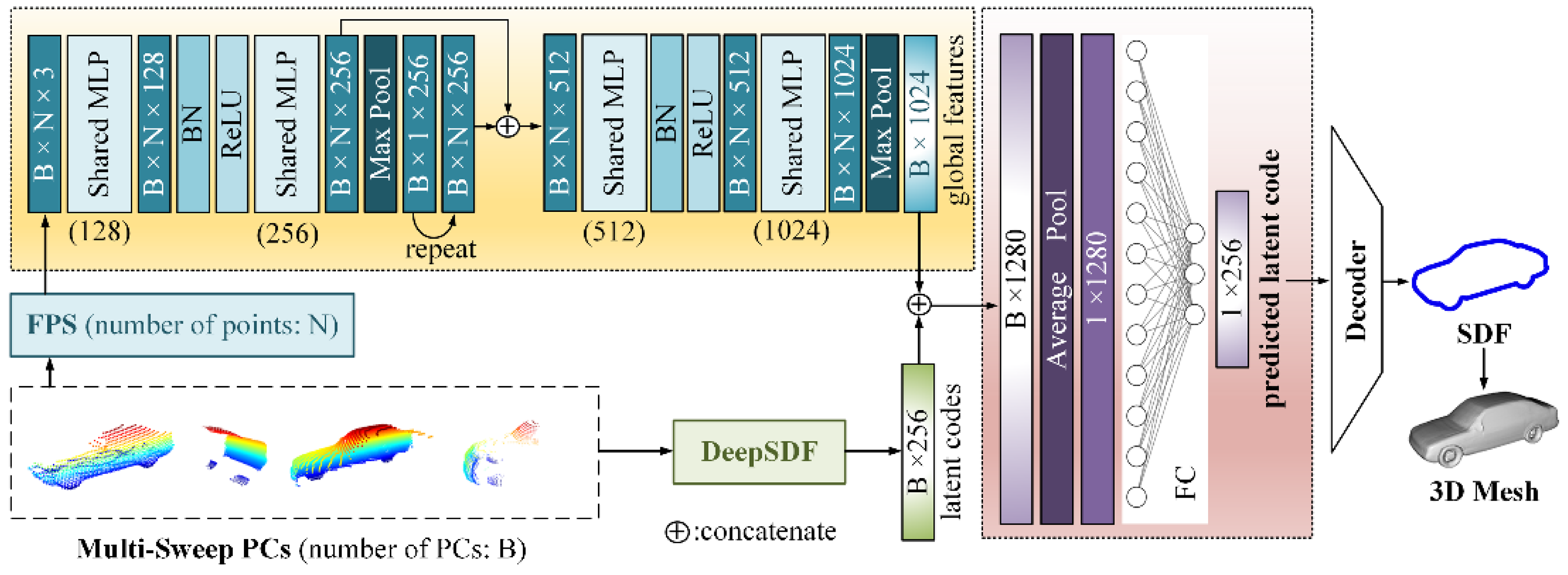
Stage 2:

$$\underset{\alpha}{\arg\min}\sum_{c=1}^{C}\mathcal{L}_2\left(g_\alpha(Z_{B,c}, P_{B,c}), z_{gt,c}\right)$$



Method

Framework

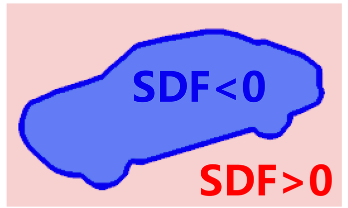


1. Preprocessing 2. Optimal Latent Code Prediction 3. 3D Mesh Extraction

Preliminaries

DeepSDF (baseline):

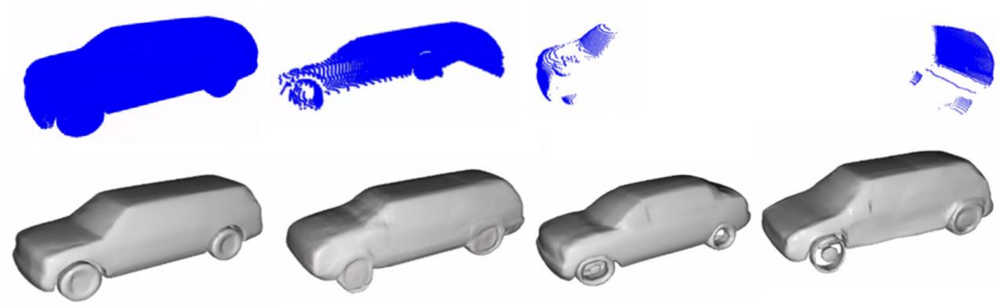
$$\hat{s} = f_\theta\left(\hat{z}, x\right)$$



f_θ : DeepSDF decoder $\hat{z}$: latent code

x : 3D querying position $\hat{s}$: SDF value

Consistency and Complementarity Analysis



Case 1 Case 2 Case 3 Case 4

- $z_i = z_{i,c} + z_{i,s} + e_i$ where $z_{i,c} = z_i \cap z_{gt}$ and $(z_{i,s} + e_i) = z_{gt} - z_i$

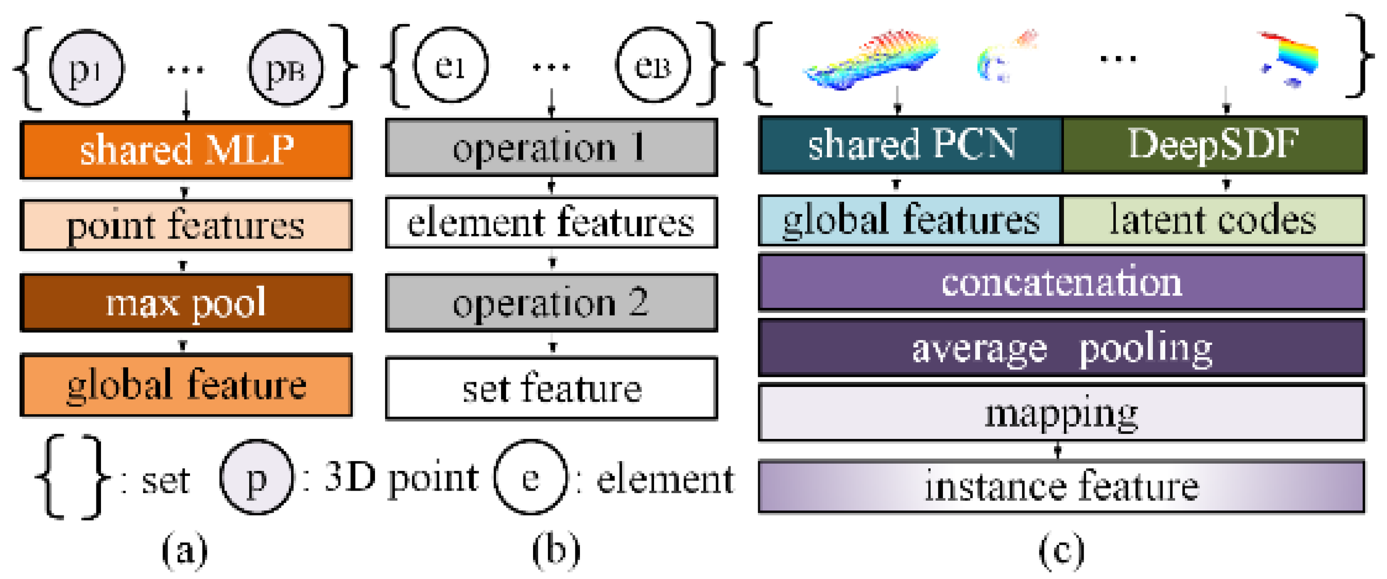
Situation 1: $z_{i,c} \cap z_{j,c} = \varnothing$

Situation 2: $z_{i,c} \cap z_{j,c} \neq \varnothing$, but $z_{i,c} \nsubseteq z_{j,c}$ and $z_{j,c} \nsubseteq z_{i,c}$

Situation 3: $z_{i,c} \cap z_{j,c} \neq \varnothing$, but $z_{i,c} \subseteq z_{j,c}$ or $z_{j,c} \subseteq z_{i,c}$

- $\hat{z} = z_{1,c} \cup z_{2,c} \cdots \cup z_{B,c} + p$

Architecture Design



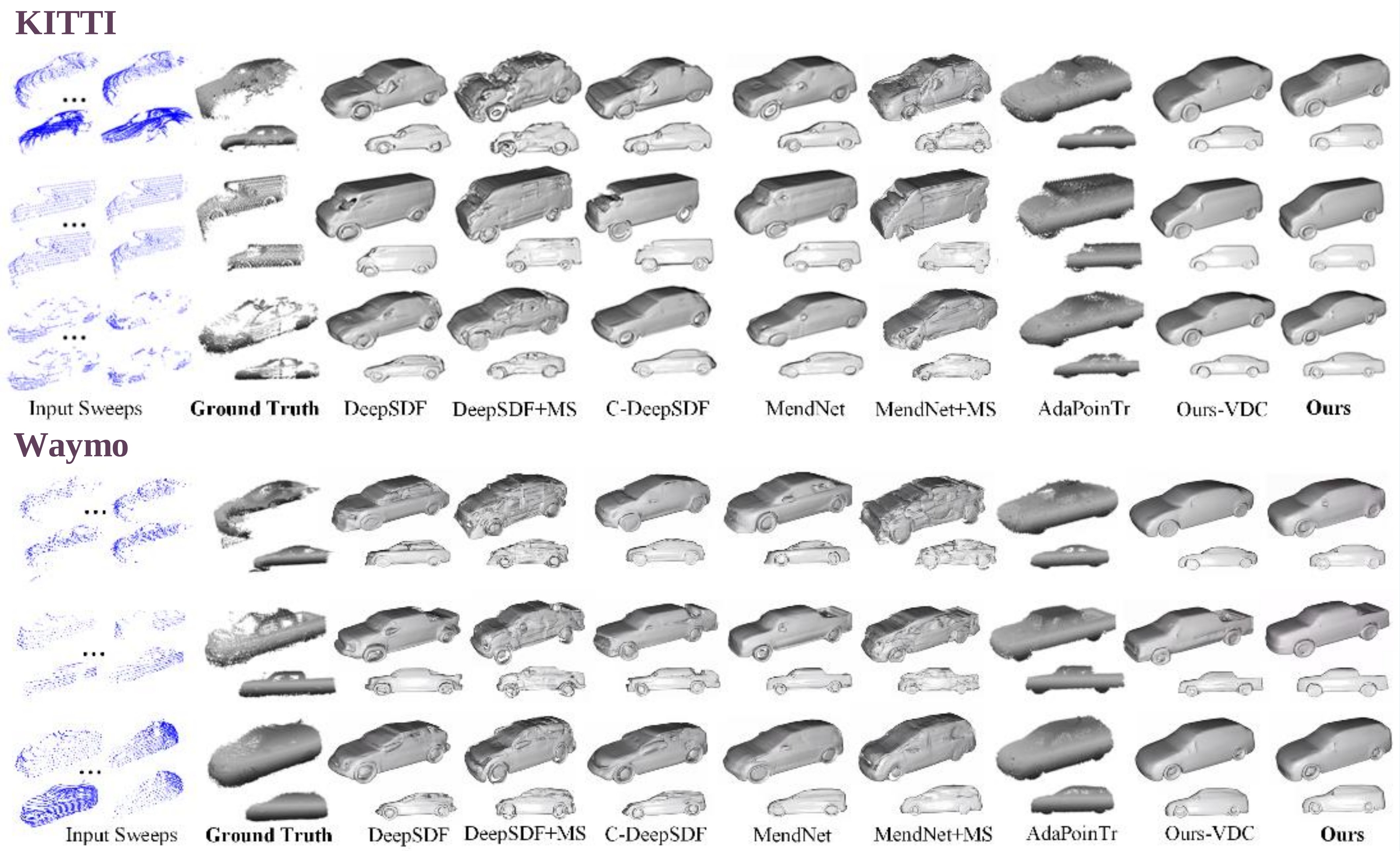
$$P=\{\mathcal{P}_1, \cdots, \mathcal{P}_B\}$$

$$Z=\{z_1, \cdots, z_B\}$$

$$g_\alpha\left(Z, P\right) = \hat{z} \rightarrow z_{gt}$$

Results

Qualitative Results



Qualitative Results

Waymo

Method \ Metric	$ACD_{mean}\downarrow$	$ACD_{median}\downarrow$	Recall $\uparrow$
DeepSDF [32]	6.26	5.81	93.51
DeepSDF+MS	5.12	5.09	95.57
C-DeepSDF [11]	6.21	5.64	93.98
MendNet [12]	4.92	4.79	95.39
MendNet+MS	4.85	4.77	95.76
AdaPoinTr* [49]	4.79	4.74	95.95
Ours-VDC	4.76	4.55	96.05
Ours	**3.36**	**2.26**	**96.84**

KITTI

Method \ Metric	$ACD_{mean}\downarrow$	$ACD_{median}\downarrow$	Recall $\uparrow$
DeepSDF [32]	6.81	6.17	80.65
DeepSDF+MS	6.11	5.83	82.73
C-DeepSDF [11]	6.73	5.99	80.77
MendNet [12]	5.94	5.64	83.84
MendNet+MS	5.83	5.61	84.26
AdaPoinTr [49]	5.89	5.67	84.20
Ours-VDC	5.75	5.24	84.39
Ours	**4.27**	**3.01**	**85.88**

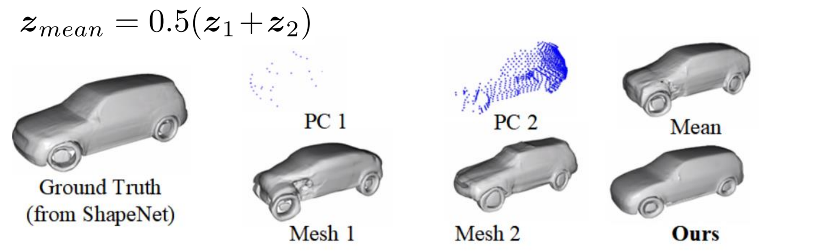
Ablation Study

Operation 1		Operation 2			Metric	
Enc.	Dep.	Mer.	Pool	Map.	ACD $\downarrow$	Recall $\uparrow$
✓			Avg.		4.89	95.50
✓	✓	Mul.	Avg.		7.33	84.93
✓	✓	Con.	Max.	✓	4.35	96.61
✓	✓	Con.	Avg.	✓	3.36	96.84

Extension to Other Taxonomies

Comparison to a Non-Learning Approach

$z_{mean} = 0.5(z_1+z_2)$



Additional Results

Effect of Number of Point Clouds

Num of PCs	$ACD_{mean}\downarrow$	$ACD_{median}\downarrow$
3	3.47	2.44
6	3.36	2.26
9	3.32	2.21

Effect of Number of Points Per Point Cloud

Num of Points	256		128	
Metric	$ACD_{mean}\downarrow$	$ACD_{median}\downarrow$	$ACD_{mean}\downarrow$	$ACD_{median}\downarrow$
DeepSDF	6.26	5.81	12.52	8.51
Ours	**3.36**	**2.26**	**3.47**	**2.64**

Video