SPoVT: Semantic-Prototype Variational Transformer for Dense Point Cloud Semantic Completion

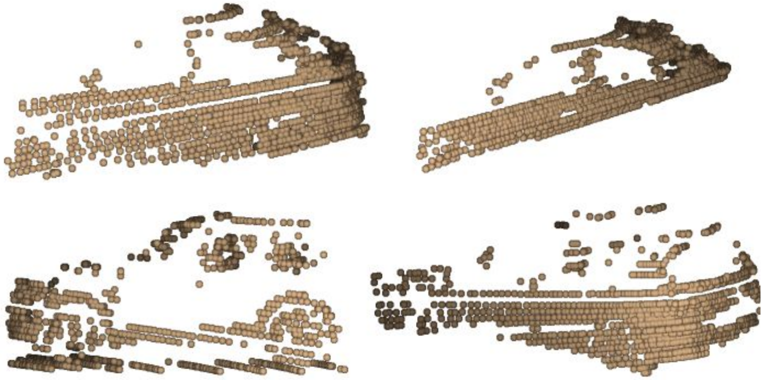
Sheng-Yu Huang[1*] Hao-Yu Hsu[1*] Yu-Chiang Frank Wang[1,2]

[1]Graduate Institute of Communication Engineering, National Taiwan University
[2]NVIDIA

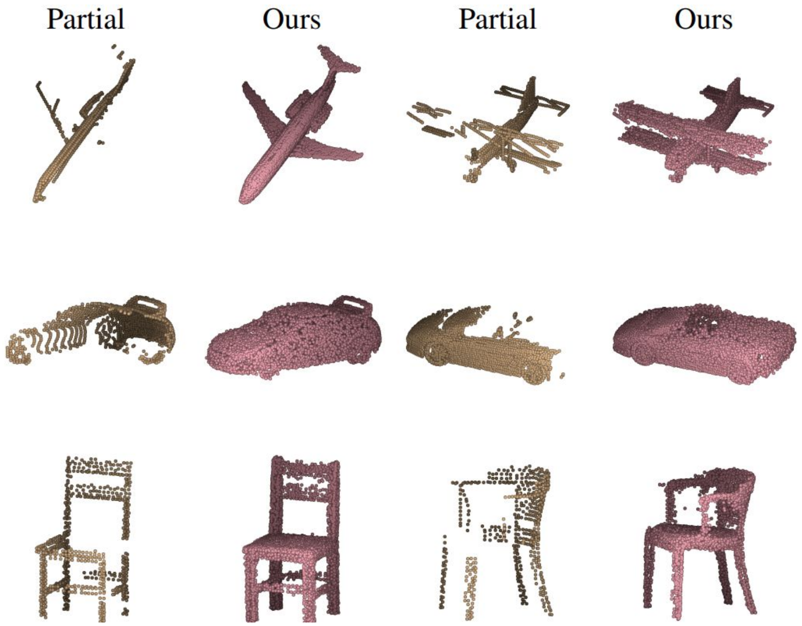
NeurIPS 2022

Point Cloud Completion

- Point clouds from sensors are usually occluded or broken
- Point cloud completion：**repairing** the incomplete point cloud
- Making further applications easier (e.g. Object detection, Semantic segmentation)



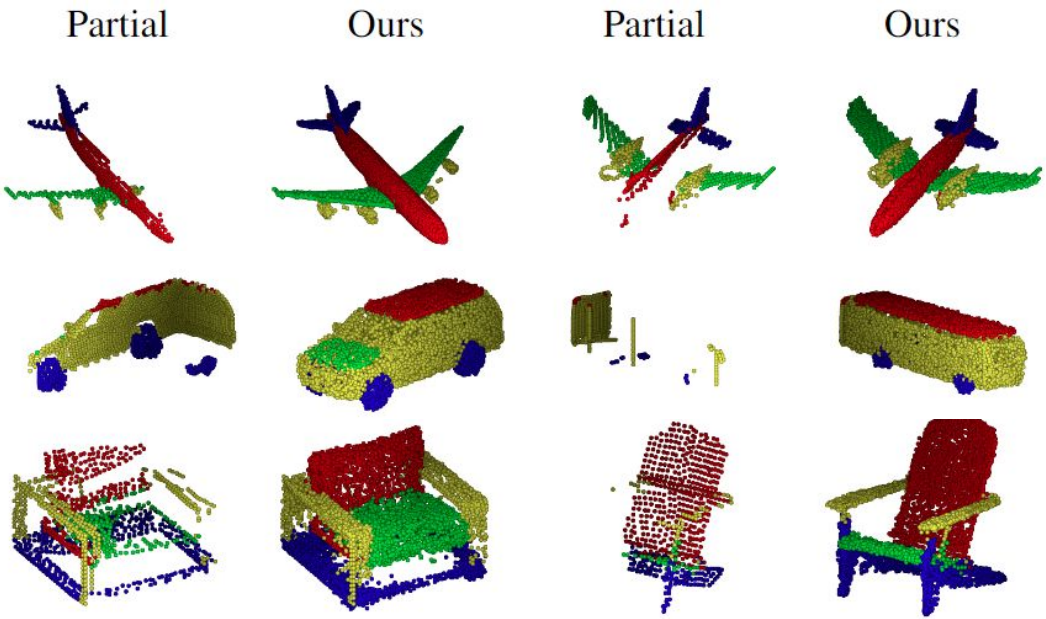
“car” objects cropped from LiDAR point cloud



Point cloud completion

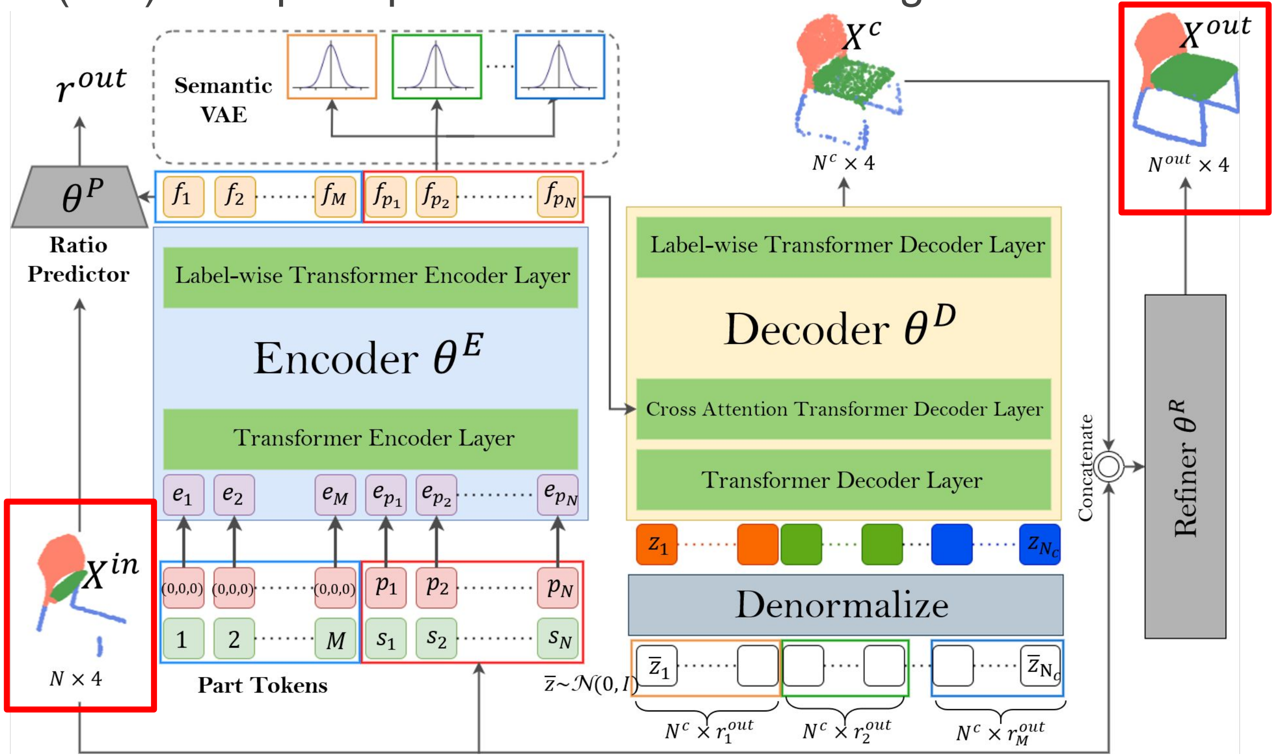
Goal

- Point cloud semantic completion
  - Input both point cloud coordinates and per-point semantic part label
  - Predict both complete point cloud and semantic segmentation
  - Semantic Part：back, seat, handle, and leg for chairs, etc.



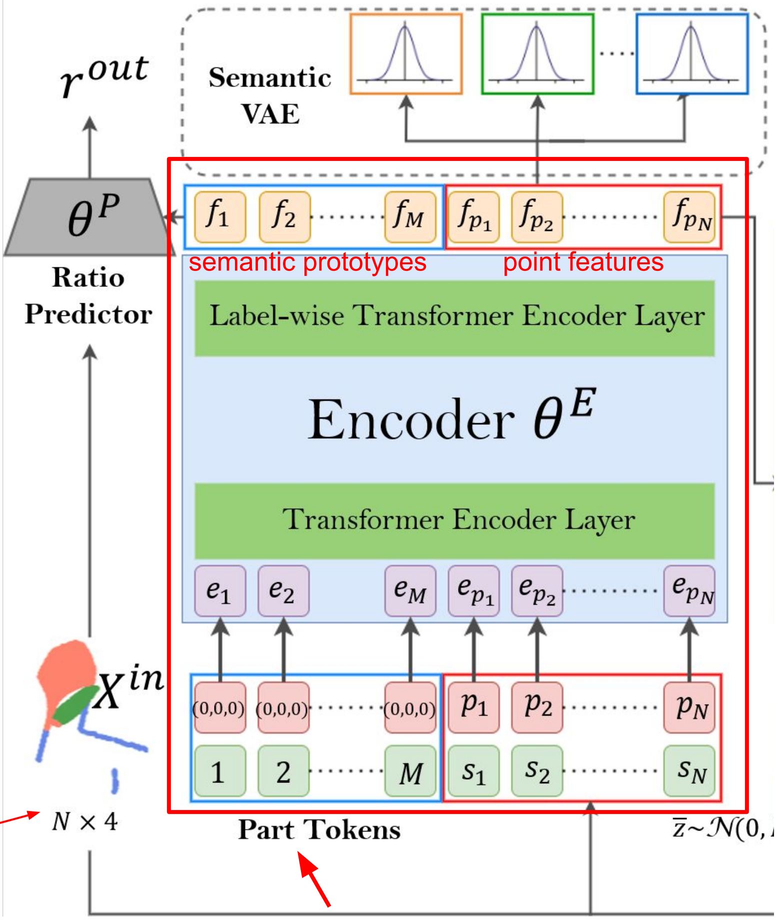
Method -- Overview

- Transformer-based Encoder-Decoder structure
  - Input(X^{in}) : 3-D coordinate + 1-D semantic part label of partial point cloud
  - Output(X^{out}) : Complete point cloud + semantic segmentation



Method -- Encoding

- Encoder：Encode input point cloud as Semantic Prototypes and Point Features

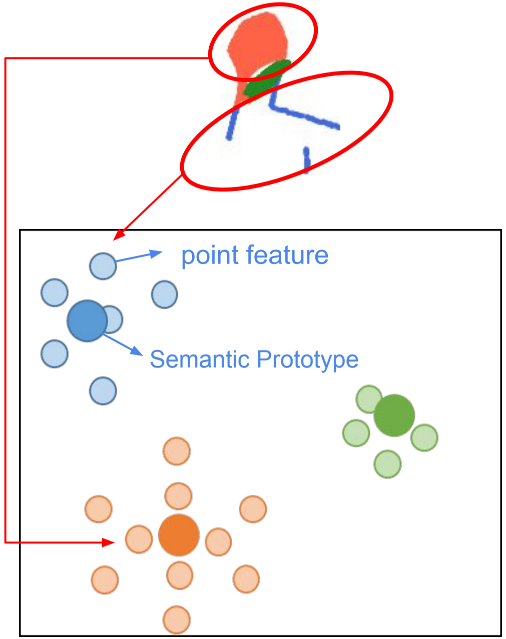


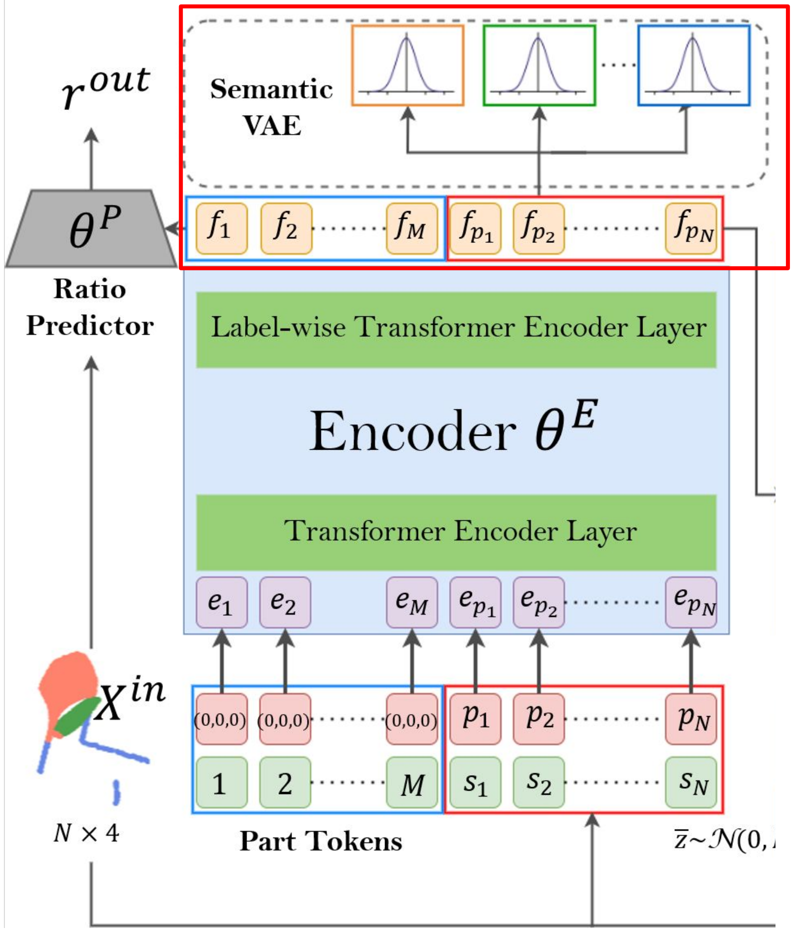
3D coordinates+ 1D semantic label

for learning Semantic Prototypes

Method -- Encoding

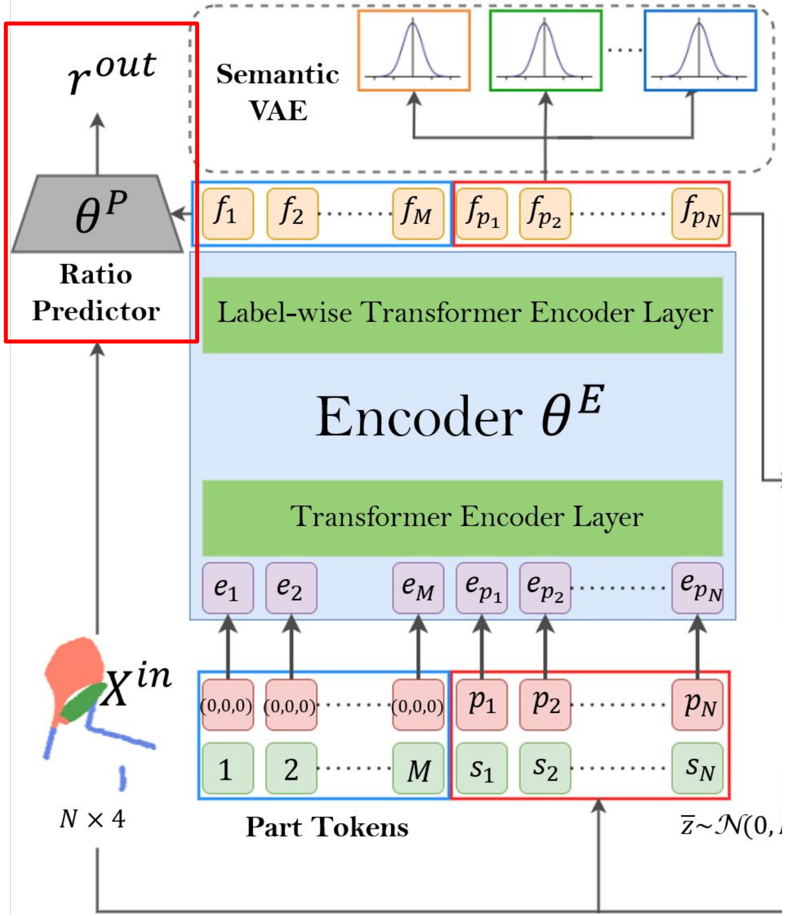
- Encoder：Encode input point cloud as Semantic Prototypes and Point Features
- Semantic VAE: Learns **feature distribution** for each semantic part





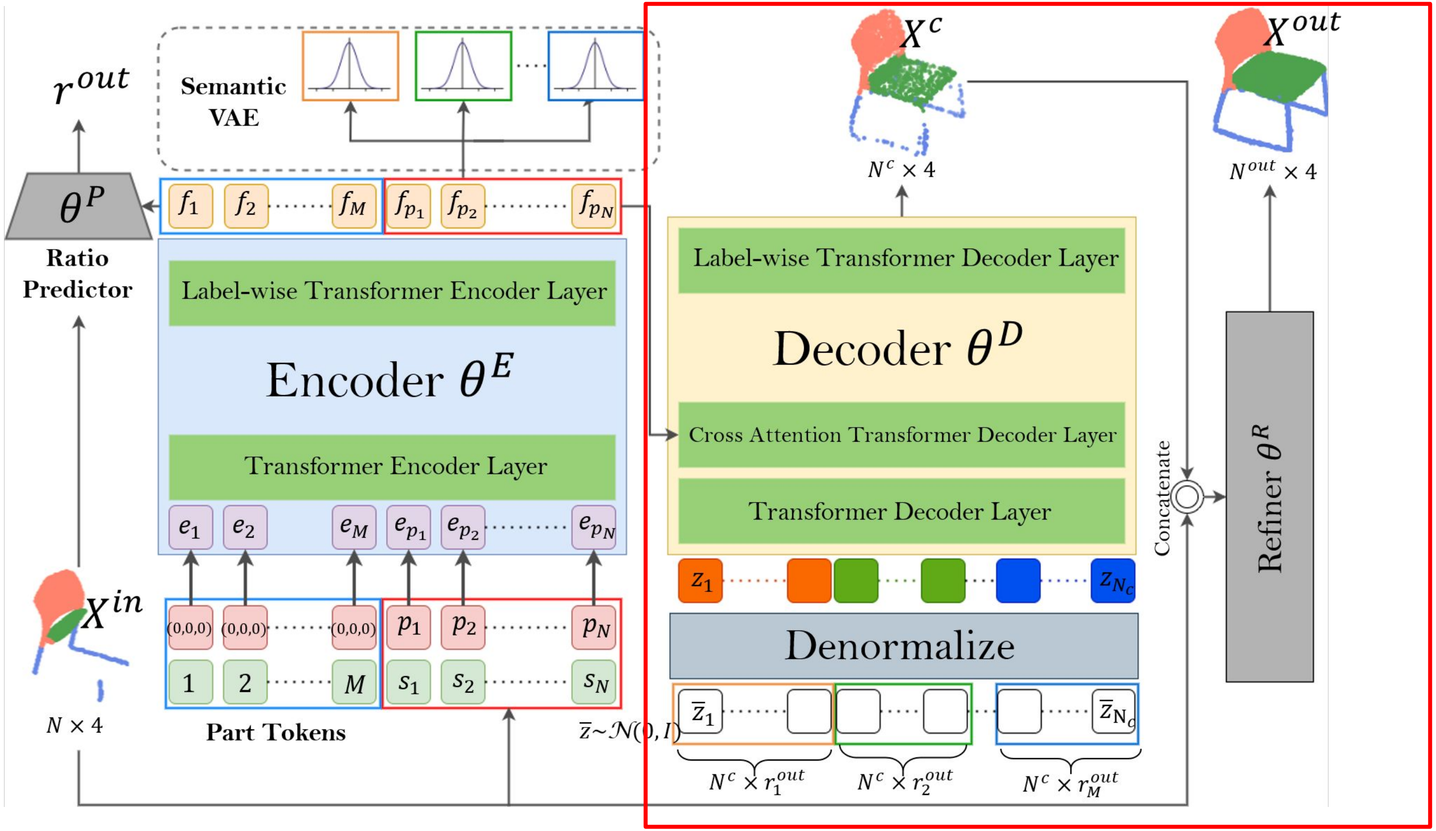
Method -- Encoding

- Encoder：Encode input point cloud as Semantic Prototypes and Point Features
- Semantic VAE: Learns **feature distribution** for each semantic part
- Ratio Predictor：
  predict r^{out} to prevent **over-dense** or **sparse** semantic parts



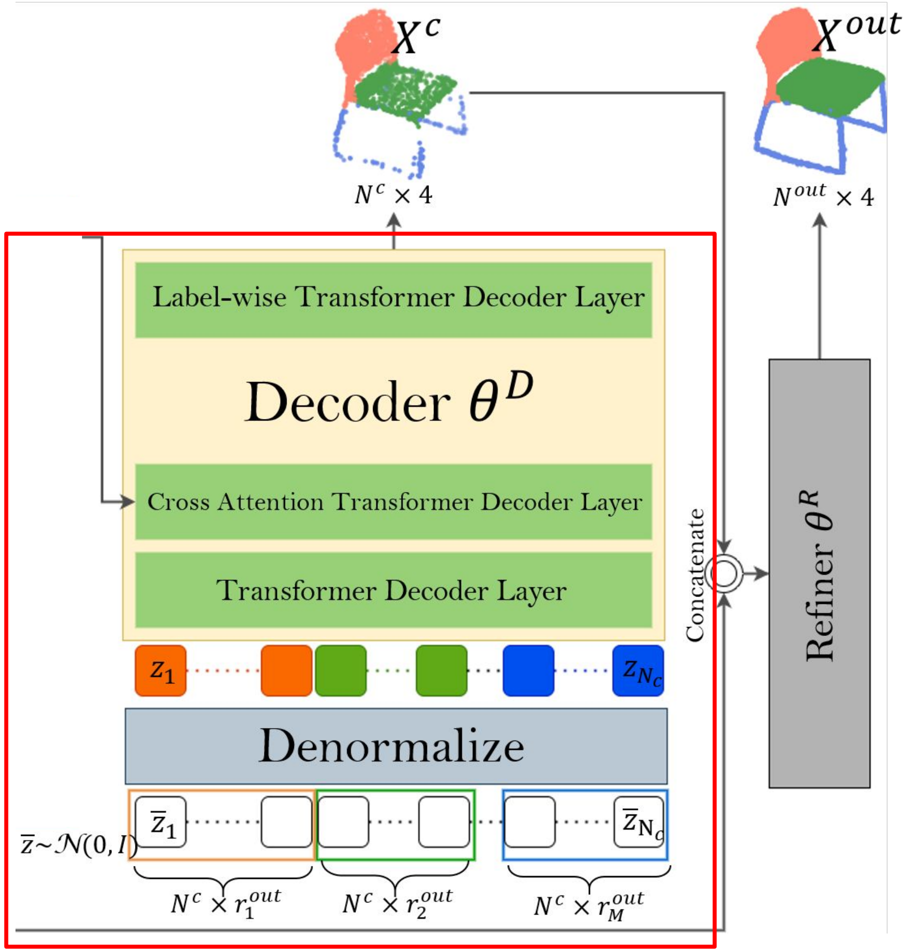
Method -- Decoding

- Decoding：using point features, semantic prototypes and r^{out} to generate X^{out}



Method -- Decoding

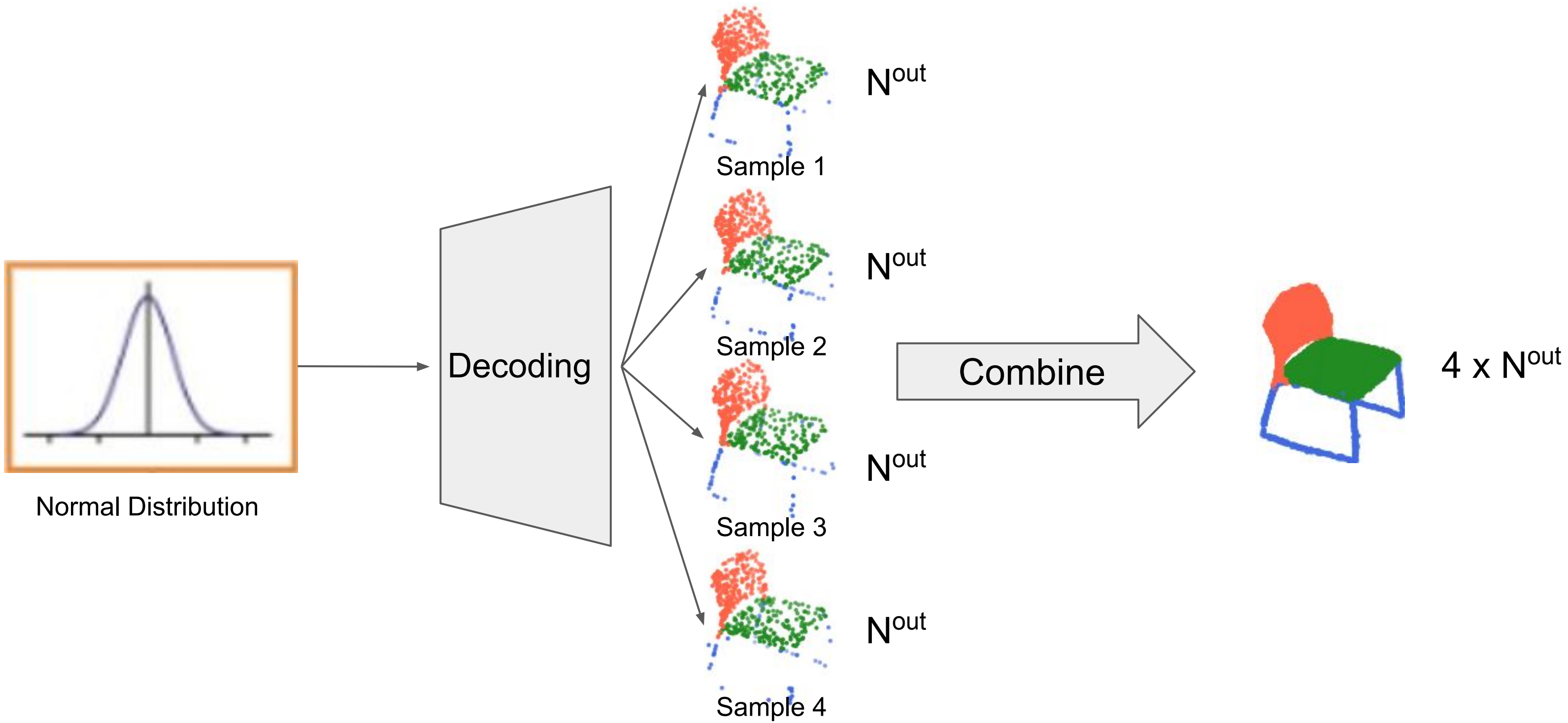
- Decoder：**random sample** noise from N~(0,I), denormalized by **Semantic Prototype** and generate a coarse completion (X^c)



Method -- Special Property

Varying resolution

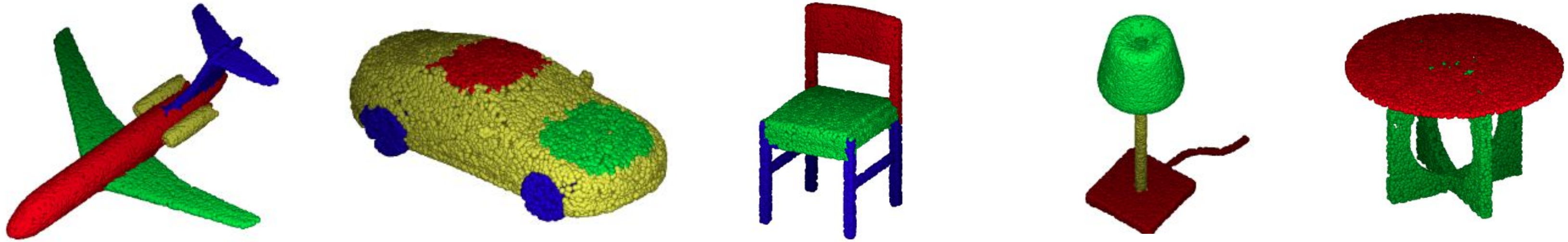
- Repeating the sampling and Decoding stage can directly increase the resolution of output point cloud, achieving **varying resolution**



Point Cloud Completion Dataset

- Subset of ShapeNet：
  - **5** different categories, each with **3~4** of semantic parts
  - Partial point cloud：rendered from eight different views
  - Complete point cloud：each with **16384** points

Aiplane Car Chair Lamp Table



Dataset: Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012, 2015. https://arxiv.org/abs/1512.03012

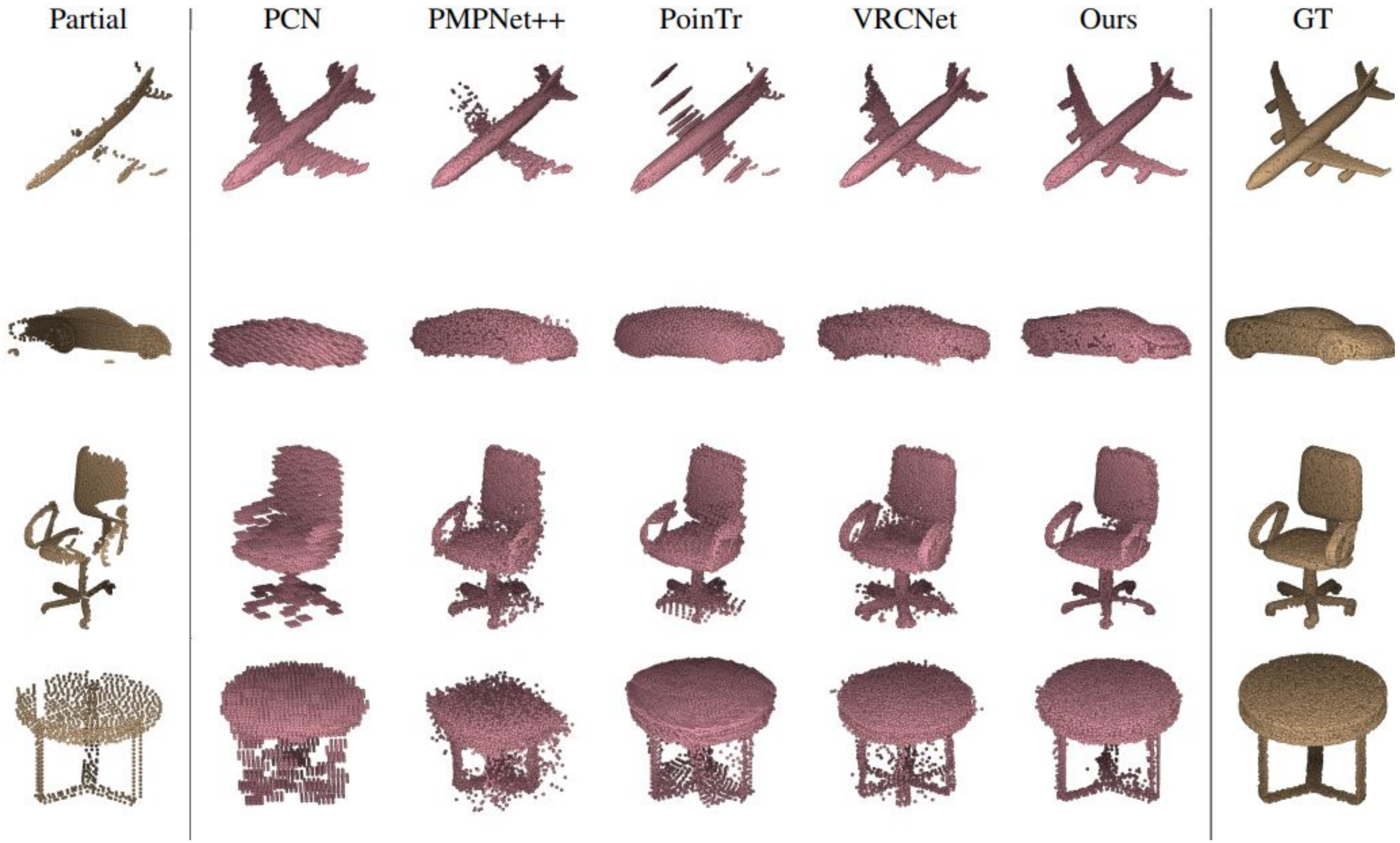
Quantitative Comparison

- Compare with current state-of-the-arts

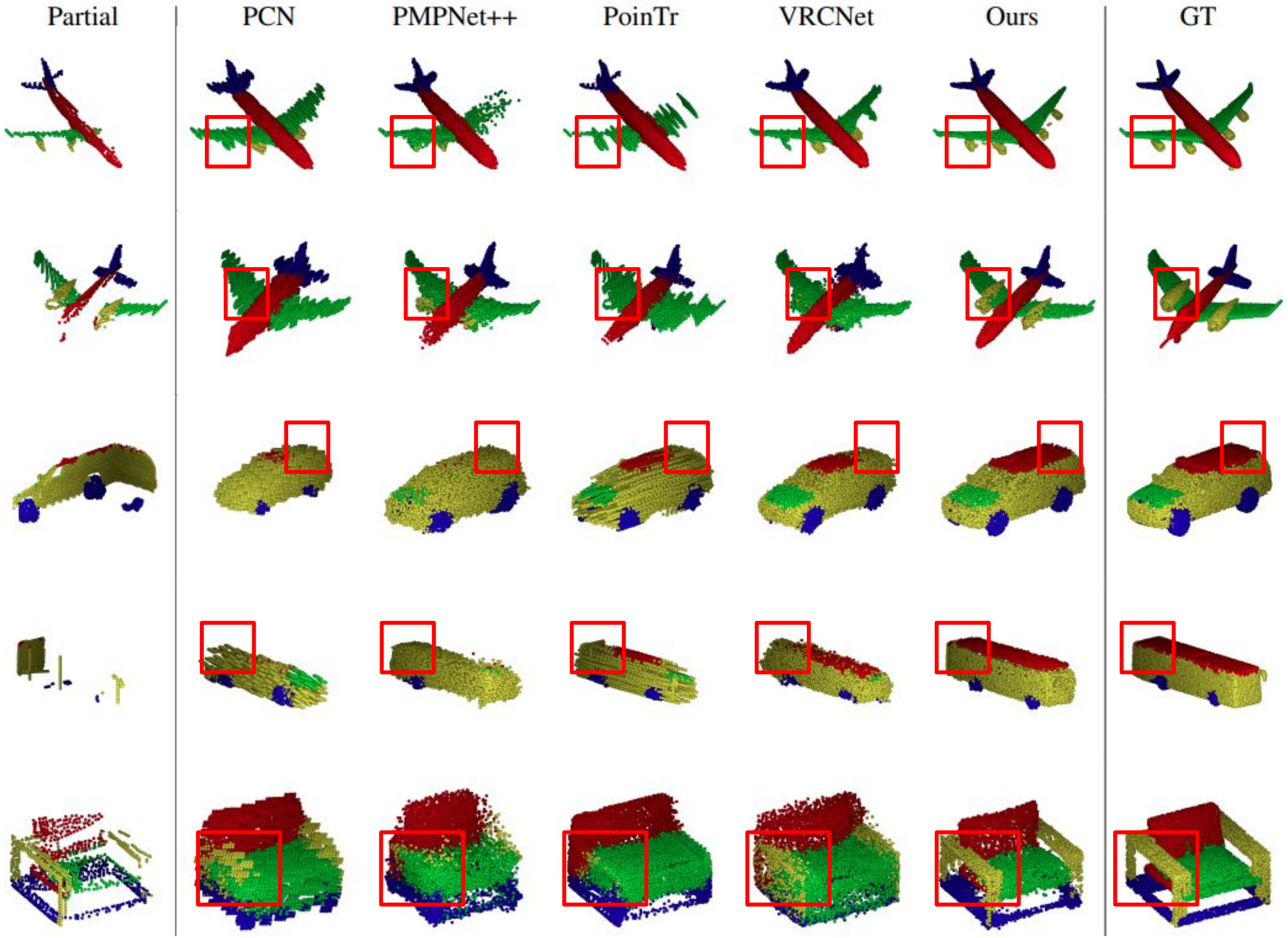
Method	Airplane		Car		Chair		Lamp		Table		Avg.	
	CD	mIoU	CD	mIoU	CD	mIoU	CD	mIoU	CD	mIoU	CD	mIoU
PCN [1]	1.26	67.4	10.8	38.1	5.77	79.3	11.4	62.1	5.22	76.6	6.88	64.7
PMP-Net++ [11]	1.80	70.3	3.82	48.6	3.42	75.3	7.93	66.3	7.87	59.3	4.97	64.0
VRC-Net [12]	0.84	69.7	3.15	60.6	3.50	82.2	4.90	75.5	4.76	74.1	3.43	72.4
PoinTr [14]	1.88	53.6	3.73	50.8	3.01	79.2	4.55	60.5	2.97	76.1	3.23	64.0
SPoVT*	0.75	82.1	2.99	76.9	2.97	77.0	4.50	86.1	3.04	84.1	2.85	81.2
SPoVT (Ours)	**0.73**	**82.6**	**2.86**	**82.5**	**2.36**	**85.2**	**4.12**	**91.5**	**2.50**	**86.5**	**2.51**	**85.7**

[1] Yuan, W., Khot, T., Held, D., Mertz, C., & Hebert, M. Pcn: Point completion network. In Proc. 3DV, 2018.
[2] Xin Wen, Peng Xiang, Zhizhong Han, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Yu-Shen Liu. Pmpnet++: Point cloud completion by transformer-enhanced multi-step point moving paths. IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), 2022.
[3] Pan, L., Chen, X., Cai, Z., Zhang, J., Zhao, H., Yi, S., & Liu, Z. Variational Relational Point Completion Network. In Proc. **CVPR**, 2021.
[4] Yu, X., Rao, Y., Wang, Z., Liu, Z., Lu, J., & Zhou, J. Pointr: Diverse point cloud completion with geometry-aware transformers. In Proc. **ICCV,** 2021.

Qualitative Comparison of Completion Results

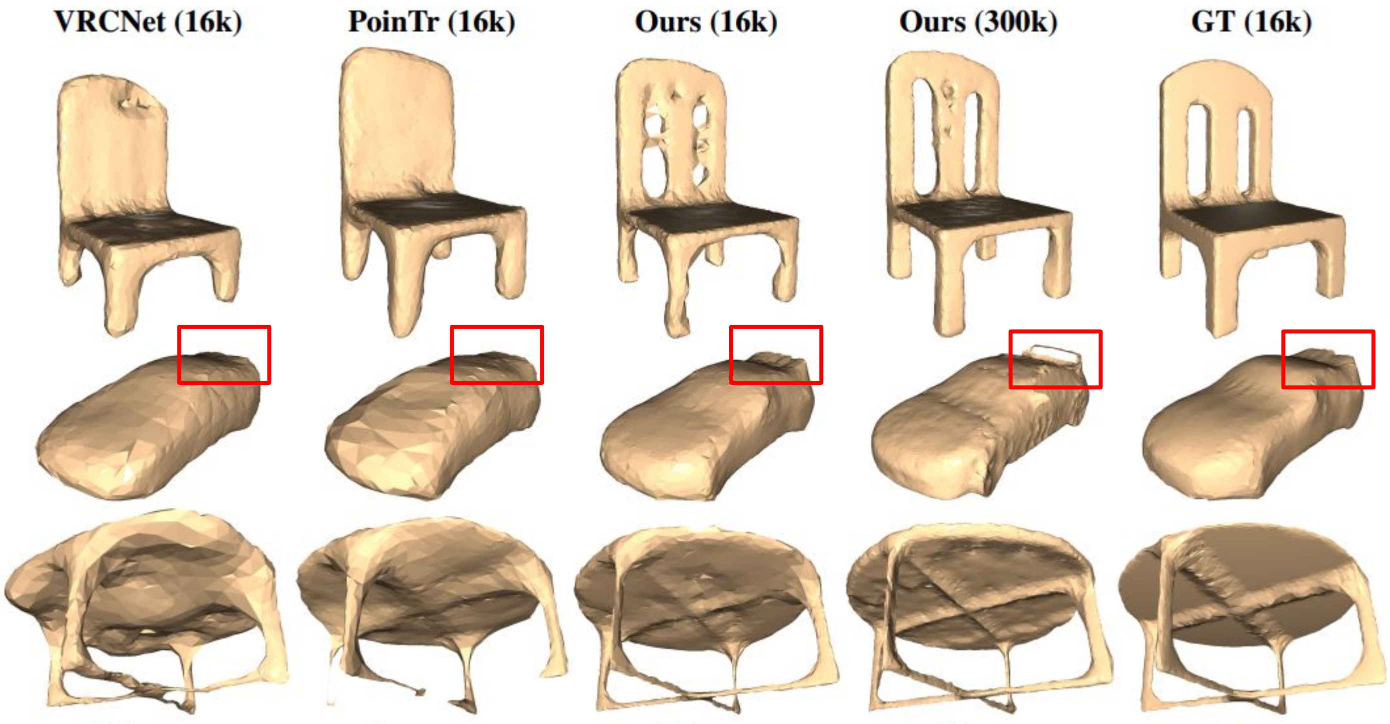


Qualitative Comparison of Segmentation Results



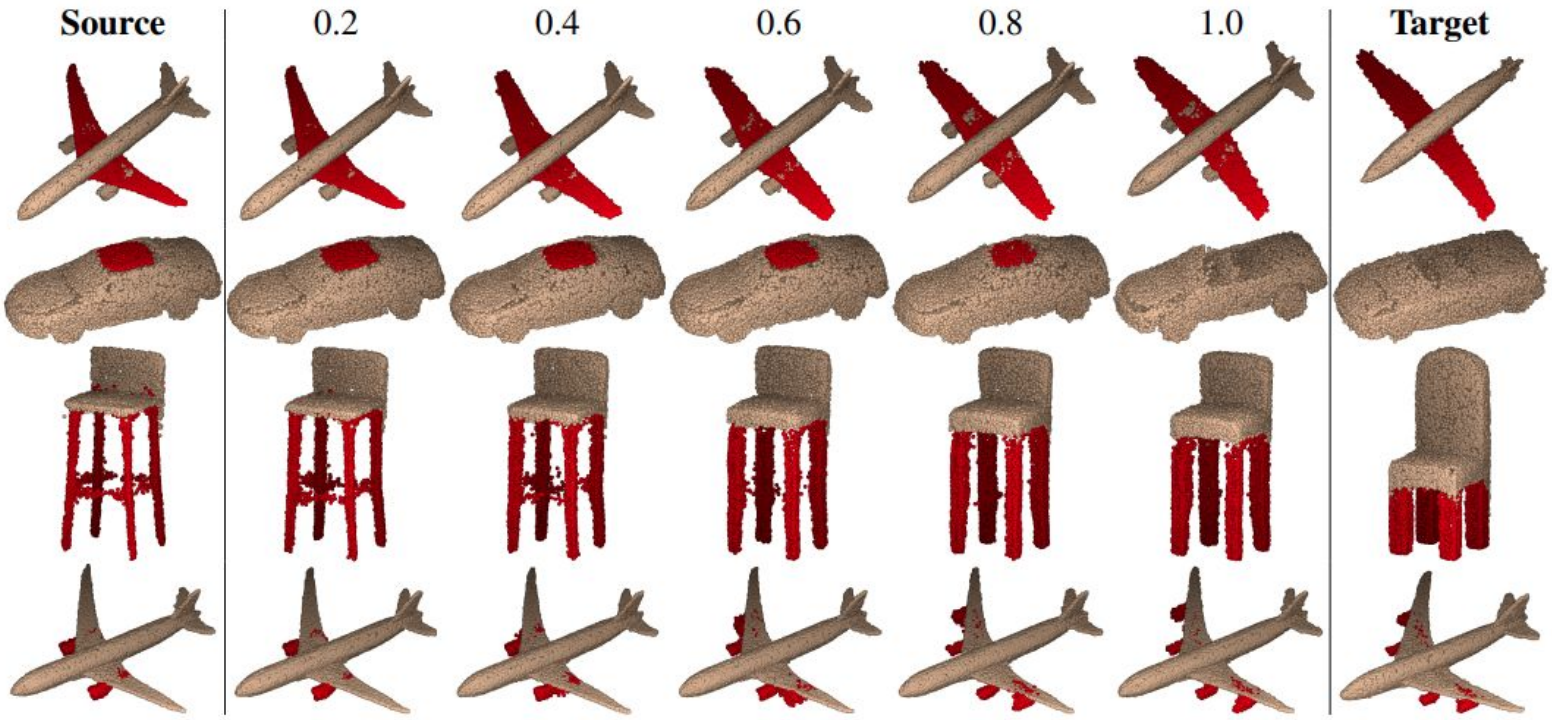
Surface Reconstruction (Mesh)

- Generate high resolution results via the **Varying Resolution** property
- Preserve more details for the reconstructed mesh objects



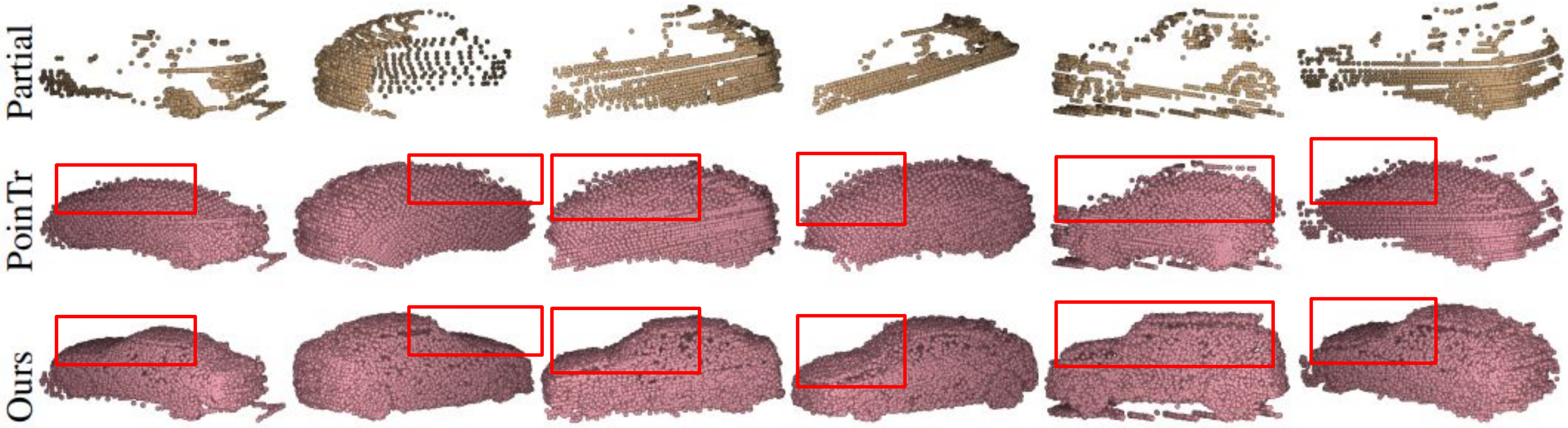
Part-wise Manipulation

- Achieve part-wise manipulation by interpolating between specific semantic prototypes of two different objects.



Real-world LiDAR Point Cloud Completion (KITTI dataset)

- **Semantic labels** of real-world LiDAR point cloud are not available
- Complete the point cloud after passing through a pre-trained segmentor



Ref: Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. The International Journal of Robotics Research, 2013.

Thanks