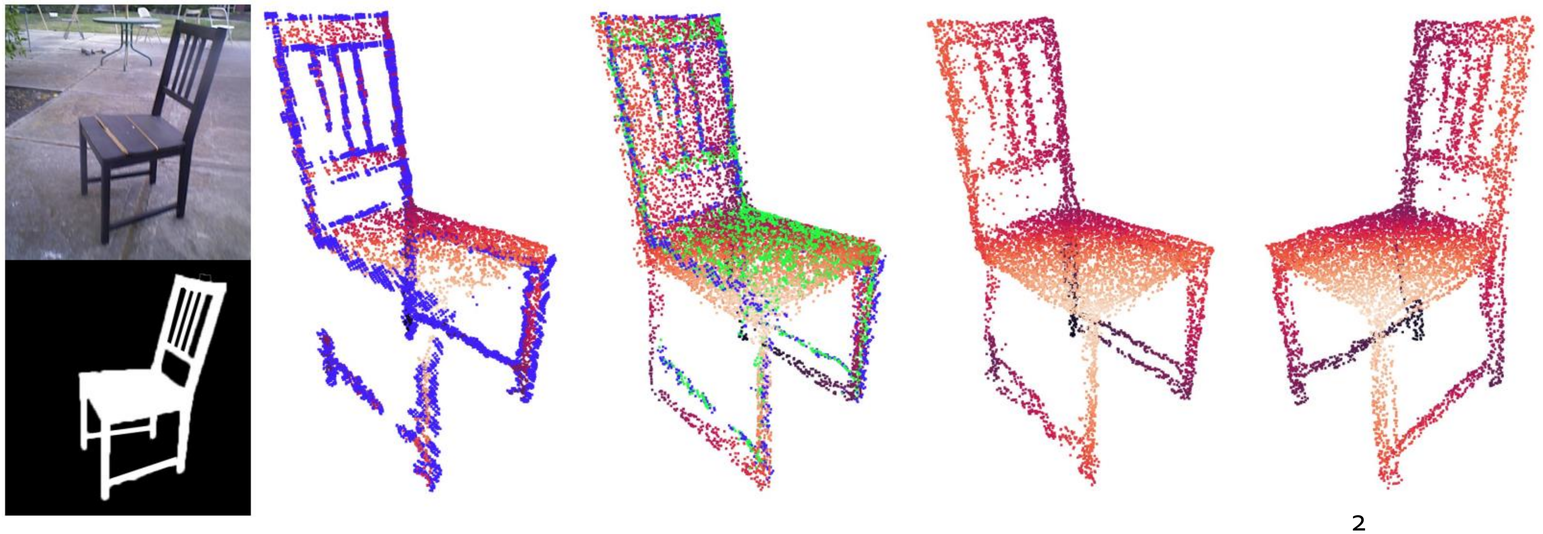
ME-PCN: POINT COMPLETION CONDITIONED ON MASK EMPTINESS

Bingchen Gong, Yinyu Nie, Yiqun Lin, Xiaoguang Han, and Yizhou Yu

香 港 大 學
THE UNIVERSITY OF HONG KONG

Point Cloud Completion

- Point completion refers to completing the missing geometries of an object from incomplete **observations**.
- **e.g.**, data from 3D scanning sensors like LIDAR and structured light depth cameras.
- Point cloud is a more compact, scalable and computation-efficient representation of 3D shapes.



Core Ideas

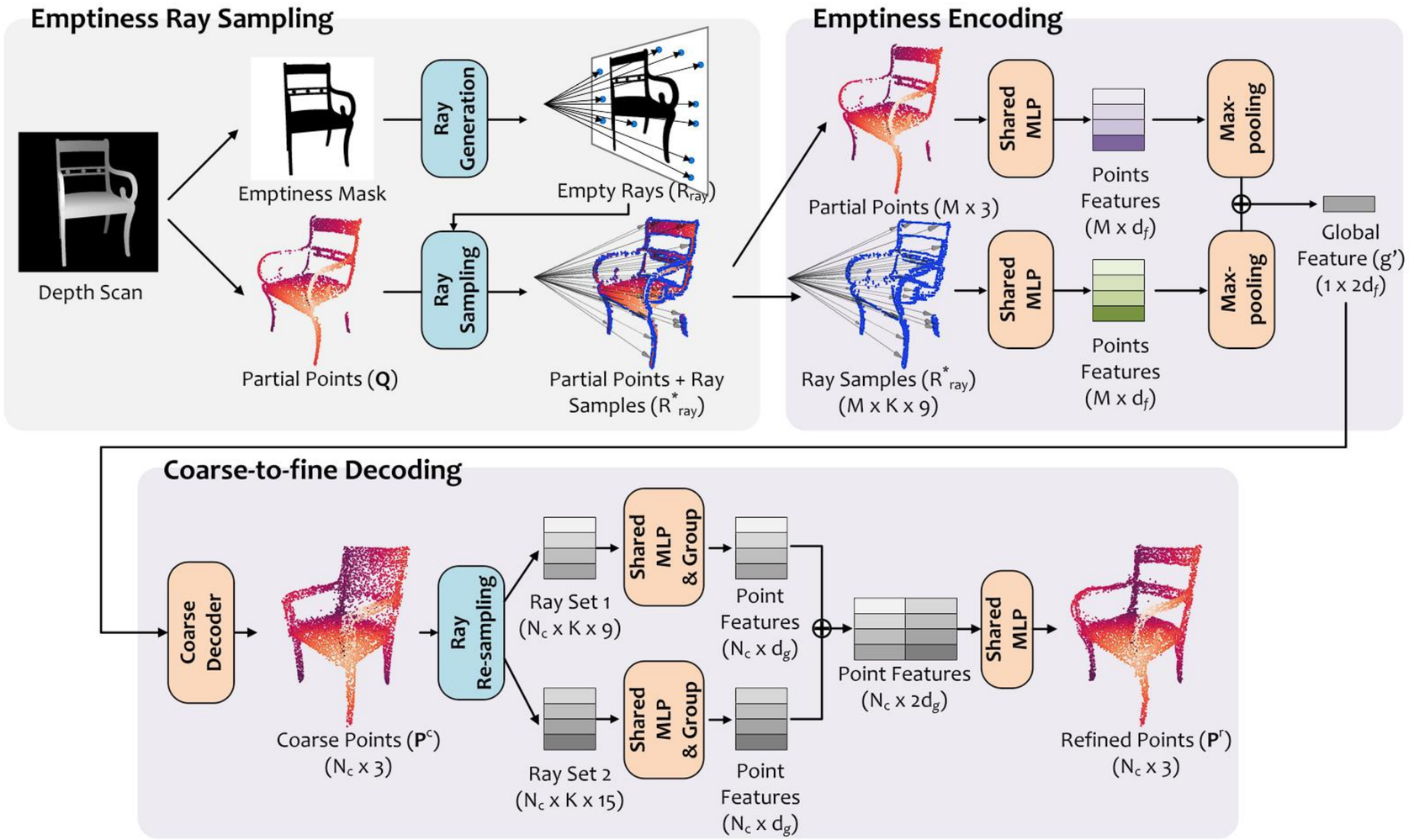
- Encoder-decoder architecture followed by an optional refining process.
- The completed point cloud is generated from an encoded global feature.
- **Directly** predict the complete points from the ***visible***, occupied input points.
- The **unoccupied** regions are the complement of shape occupancy, thus also indicating the topology of 3D objects.

Core Ideas

(a) Input (b) Ours (c) MSN (d) GT

- Learning the ***emptiness*** presents extra significance, especially for complex shapes, such as non-convex surfaces with holes.
- The emptiness in the input can tells ‘where should not be occupied’
- To encode the emptiness clues on a mask, 3D rays are radiated from the viewpoint towards the empty regions of the mask.
- All points along the rays will be encoded as empty points.

Architecture of ME-PCN



Method: Ray Points Sampling

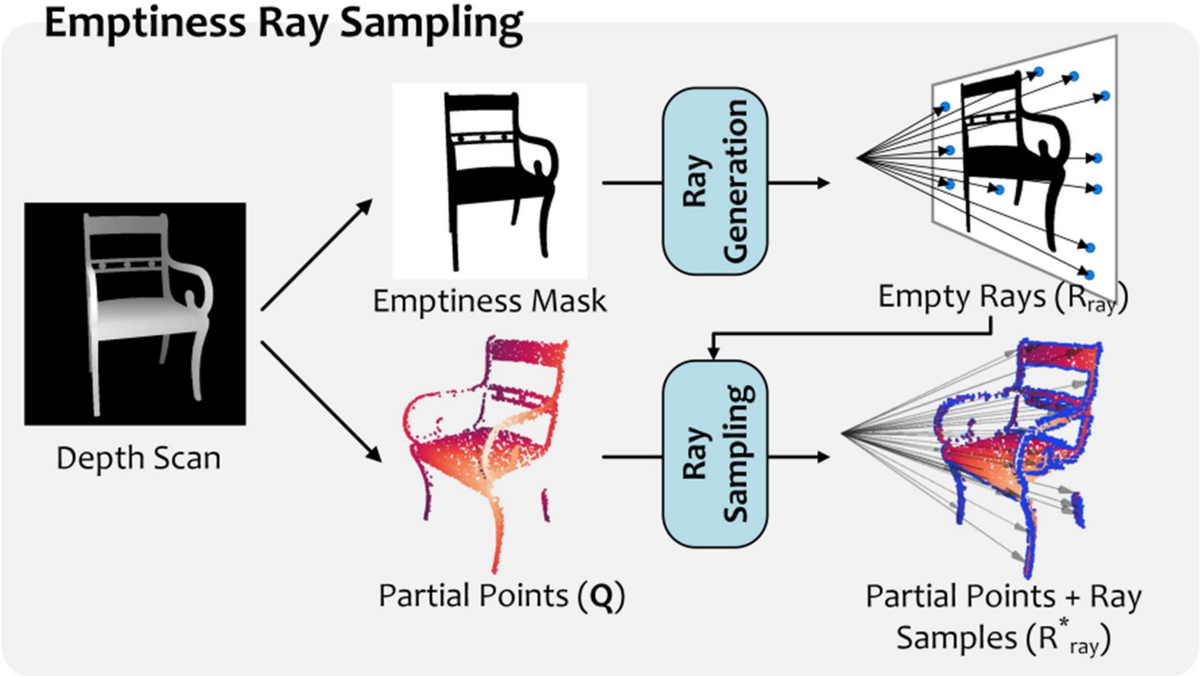
- We sample K nearest rays for each visible point $q \in R^3$. The Euclidean distance $\|D_r, q\|$ between a ray $r = (p, v) \in R_{ray}$ and a visible point q is defined by:

$$p^e = p - [(p - q) \cdot v]\, v,$$

$$D_{\mathbf{r},q} = p^e - q,$$

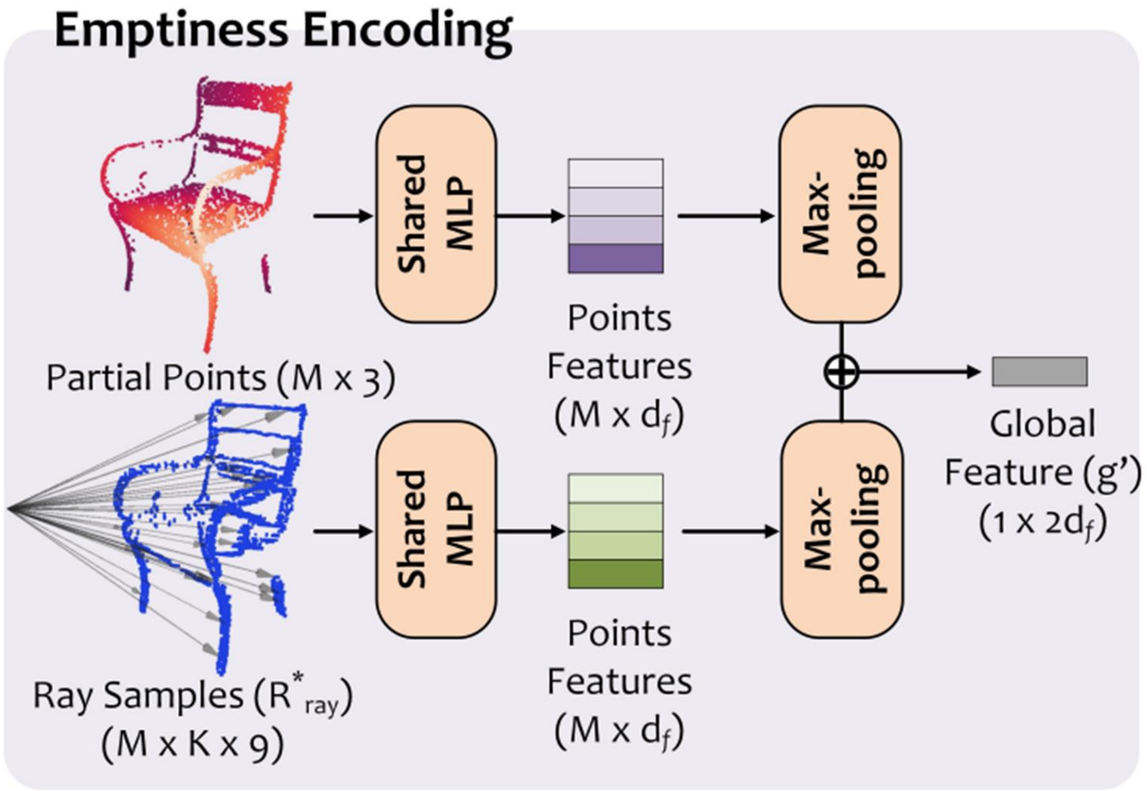
- After sampling, for each visible point q, we combine its K nearest empty points

$$\mathcal{R}^*_{ray} = \{\{p^e_k\}, \{D_k\}, \{v_k\}\} \in \mathbf{R}^{M \times K \times 9}$$



Method: Emptiness Encoding

- The encoder part consists of two Feature Encoding (FE) layers to respectively process visible points Q and sampled rays R^*_{ray} .
- The two FE layers output two feature matrices $F = \{f_i\}, G = \{g_i\}$. A point-wise max-pooling is respectively performed on F, G to obtain d_f-dimensional global features f and g.
- Lastly, f and g are concatenated together to form a single global feature vector $g' = [f, g] \in R^{2d_f}$



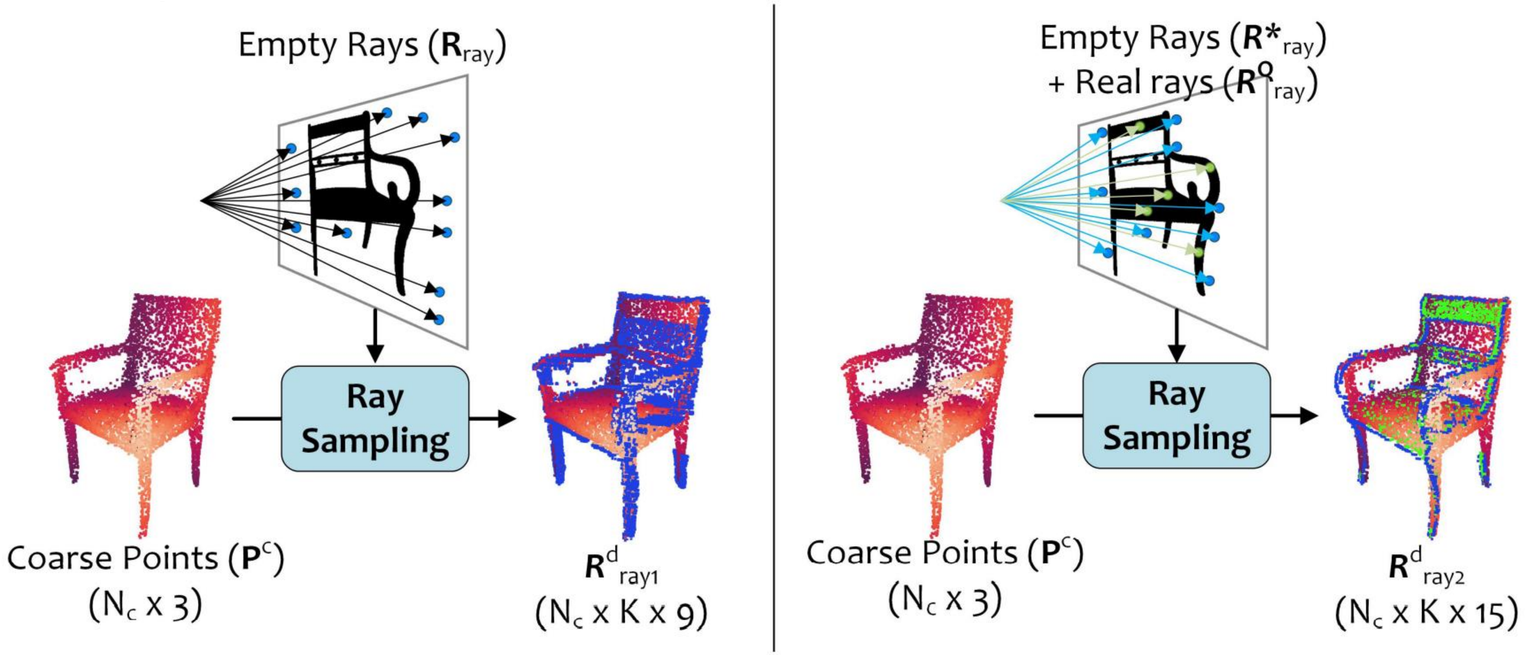
Results: Effectiveness of Emptiness Encoding

Category	vanilla MSN	vanilla MSN+Ray
Faucet	8.61 / 5.03	6.59 / 3.63
Cabinet	7.17 / 6.07	6.13 / 5.24
Table	8.59 / 5.75	6.10 / 4.66
Chair	7.51 / 5.59	5.84 / 4.50
Vase	8.67 / 6.75	6.63 / 5.62
Lamp	8.59 / 5.42	7.84 / 4.38
Average	8.19 / 5.77	6.52 / 4.67

(a) Input **(b)** v. MSN+ray **(c)** v. MSN **(d)** GT

Method: Coarse-to-Fine Decoding

- Coarse points P^c decoded from a global feature are still not accurate to preserve a consistent boundary compared to the ground-truth due to its roughness;
- The point information in R^*_{ray} conveys the surface clues that can improve shape detail recovery.

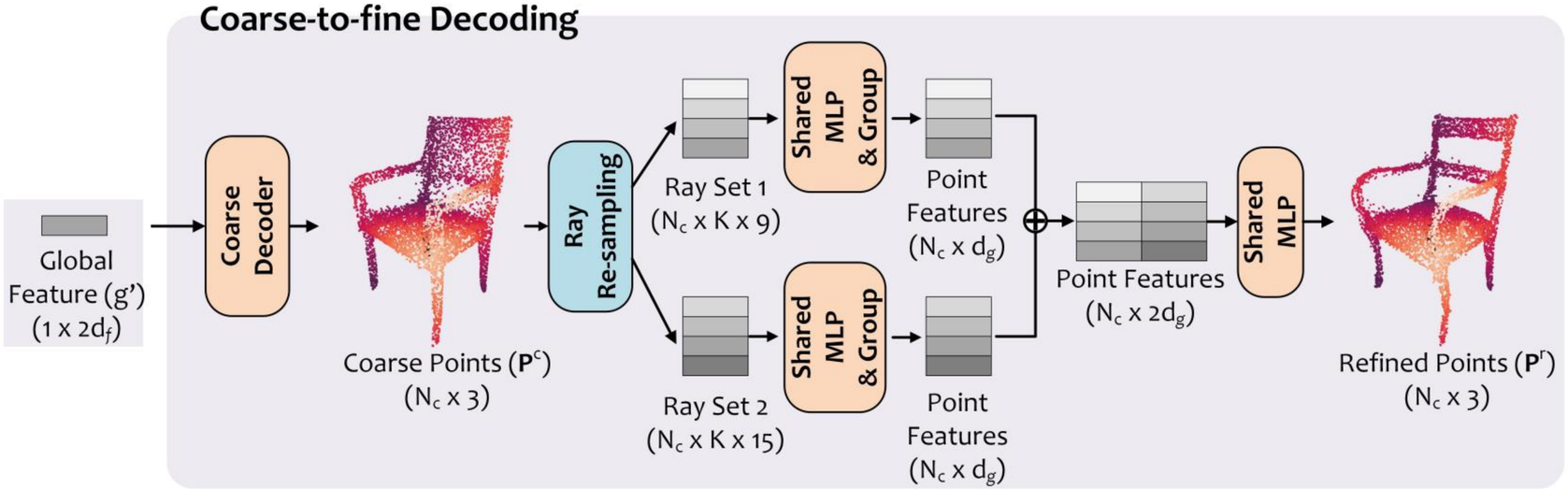


Method: Decoding Refined Shape

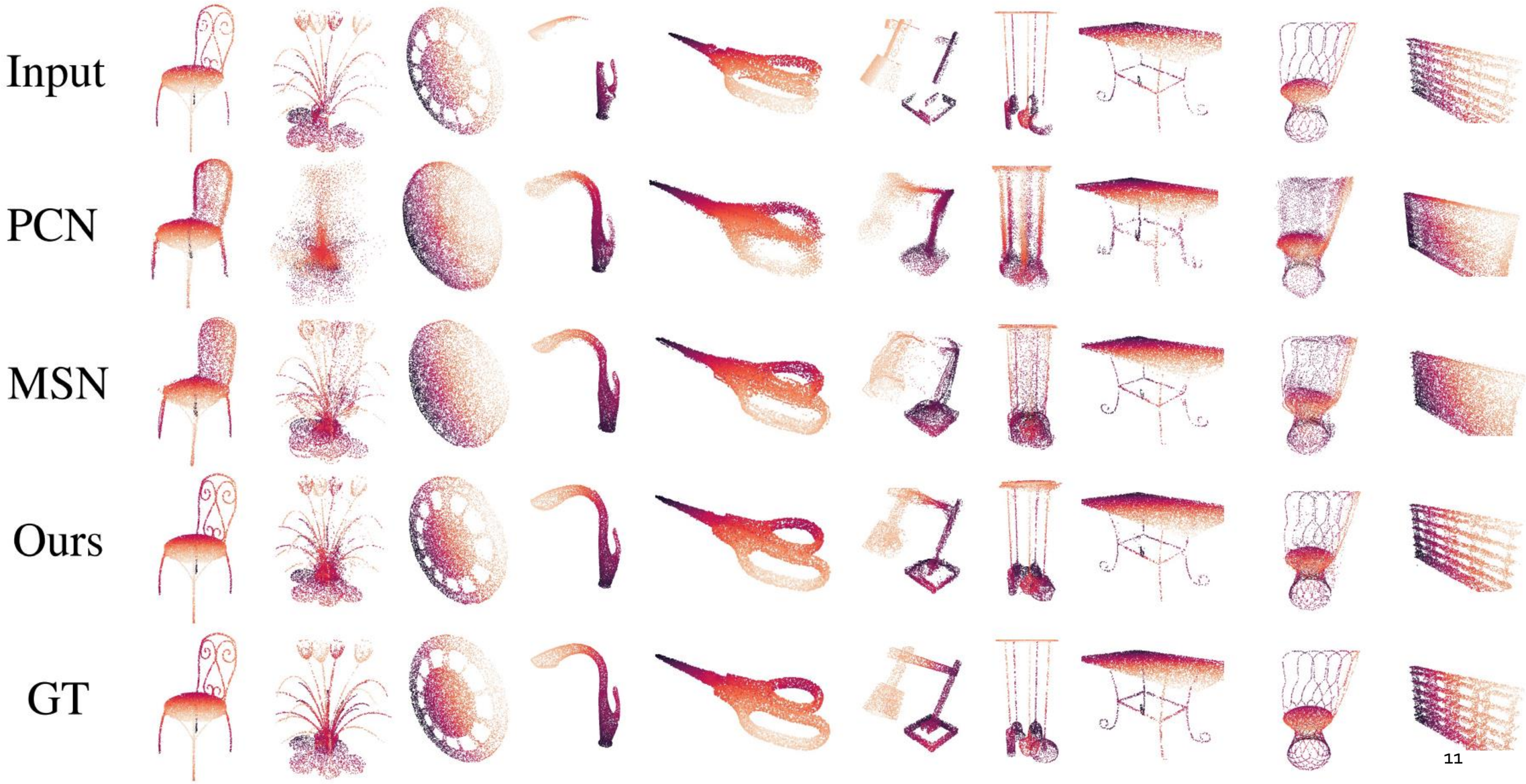
- Two FE layers are respectively used to encode R^d_{ray1} and R^d_{ray2}.
- Two shared MLPs are used to transform points in R^d_{ray1} and R^d_{ray2} into grouped point feature vectors $\{f_i^d\}$ and $\{g_i^d\} \in R^{N_c \times d_g}$:

$$f_i^d = \sum_{k=1}^{K} \sum_{d=1}^{d_g} w(k,d)\mathbf{r}(i,k,d), \mathbf{r} \in \mathcal{R}^d_{ray1},$$

- We concatenate $\{f_i^d\}$ and $\{g_i^d\}$ to regress the coordinates of complete surface points as the refined output P^r.



Results: Qualitative Comparison



Results: Evaluation on EMD and CD

methods	faucet	cabinet	table	chair	vase	lamp	average
PCN	16.81	10.47	12.22	11.81	13.25	16.67	13.54
PCN+Ray	16.13	10.18	11.68	10.61	11.13	14.90	12.44
PF-Net	16.11	10.04	9.97	10.61	11.50	14.07	12.05
P2P-Net	16.09	11.64	10.73	12.29	16.36	13.52	13.44
SoftPoolNet	15.03	14.30	11.28	14.05	17.63	15.89	14.70
CRN	14.00	11.00	9.09	9.70	13.32	12.09	11.53
GRNet	11.30	9.16	8.61	8.82	12.27	11.28	10.24
MSN	8.52	8.19	7.82	7.82	8.36	8.51	8.20
Ours	**6.89**	**7.48**	**6.63**	**6.63**	**7.16**	**7.48**	**7.05**

Table 3.3: Evaluation on EMD ($\times 10^2$) with Res.=2,048

methods	faucet	cabinet	table	chair	vase	lamp	average
PCN	5.62	7.28	5.95	6.14	8.71	5.15	6.48
PCN+Ray	4.35	7.14	5.19	5.98	7.19	4.61	5.74
PF-Net	8.96	8.15	6.94	7.48	10.10	7.56	8.20
P2P-Net	4.47	7.21	**5.49**	5.92	7.62	4.41	5.85
SoftPoolNet	5.54	7.85	6.41	6.59	8.27	5.56	6.70
CRN	5.14	7.13	5.59	5.94	7.96	4.63	6.06
GRNet	4.72	7.21	5.77	6.00	7.90	4.92	6.08
MSN	5.25	8.06	6.50	6.70	7.92	5.66	6.68
Ours	**3.90**	**7.01**	5.65	**5.61**	**6.68**	**4.26**	**5.51**

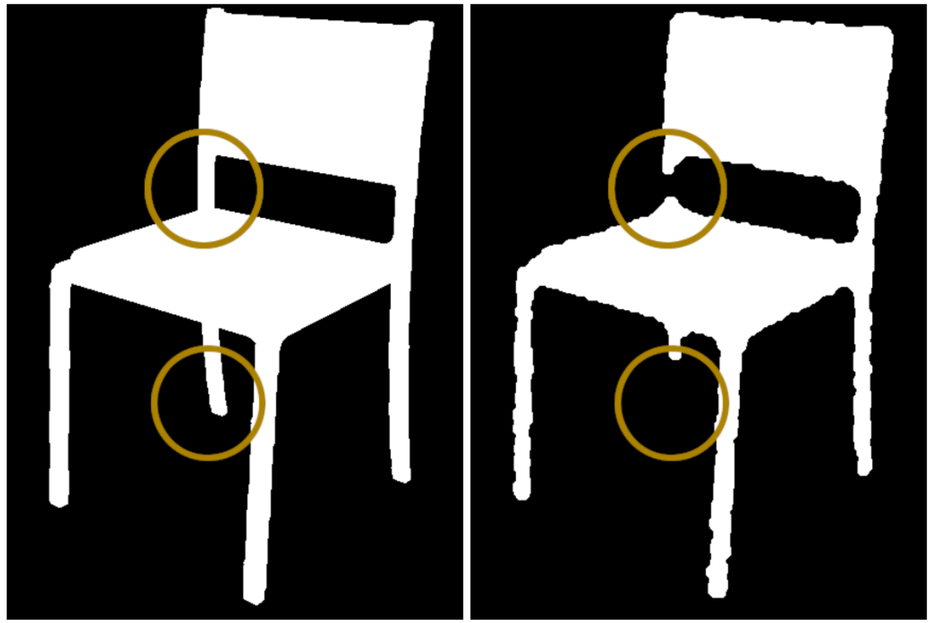
Table 3.4: Evaluation on CD ($\times 10^2$) with Res.=2,048

Results: Robustness of Emptiness Encoding

- To verify our robustness to noisy masks, we simulate the masks from real-world depth/RGB data, and add strong Gaussian noise to the boundaries of masks.
- We fine-tune our model on chair and table categories using the noisy mask for 2000 iterations
- EMD score increase for chair/table:

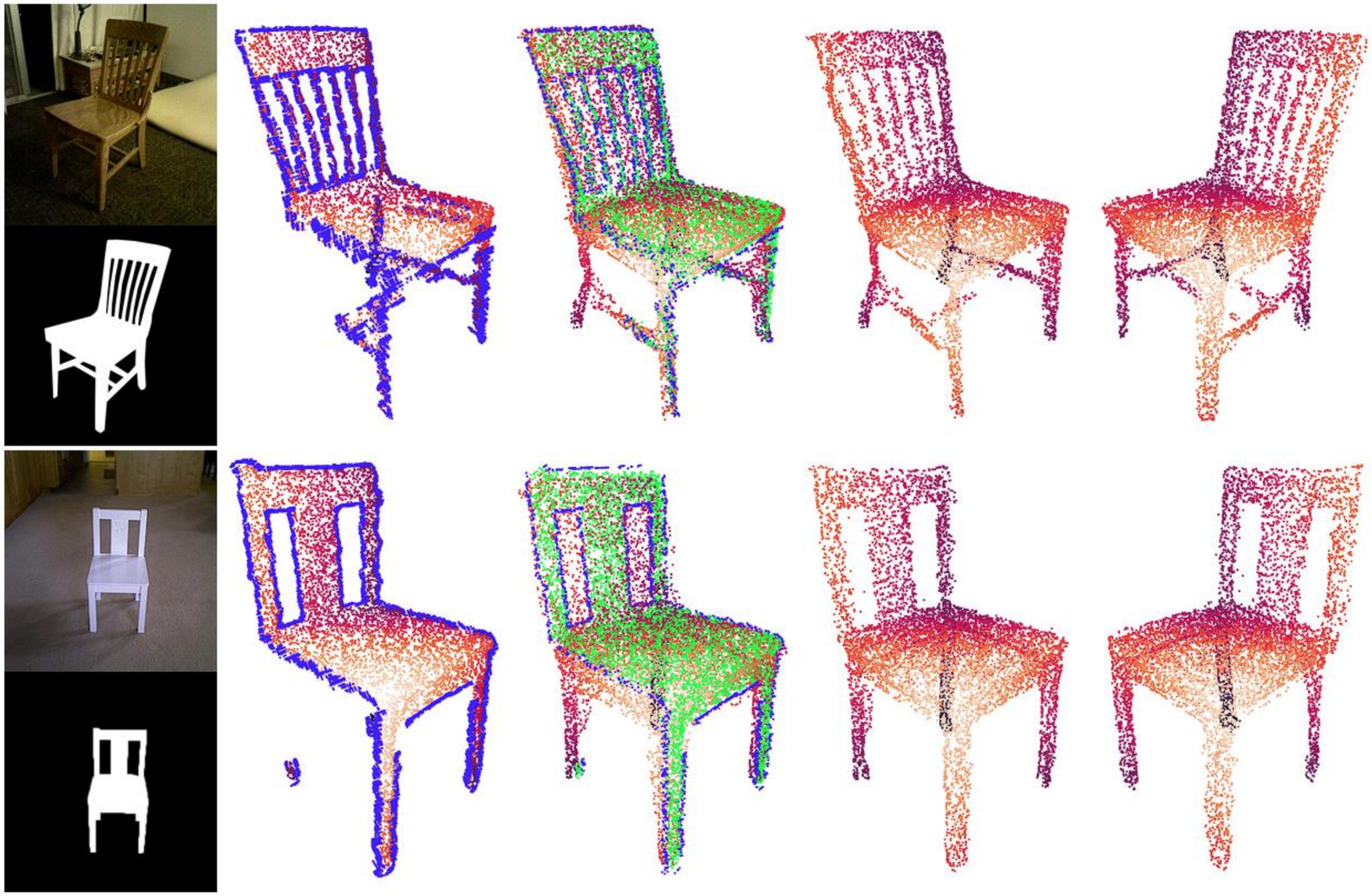
  from 5.12 / 5.33 to 5.26 / 5.56

(a) Before **(b)** After



(a) Before **(b)** After

Results: Tests on Real Scans



(a) Scans (b) Input (c) Coarse (d) Output (e) Output

Thank you!
- Q&A