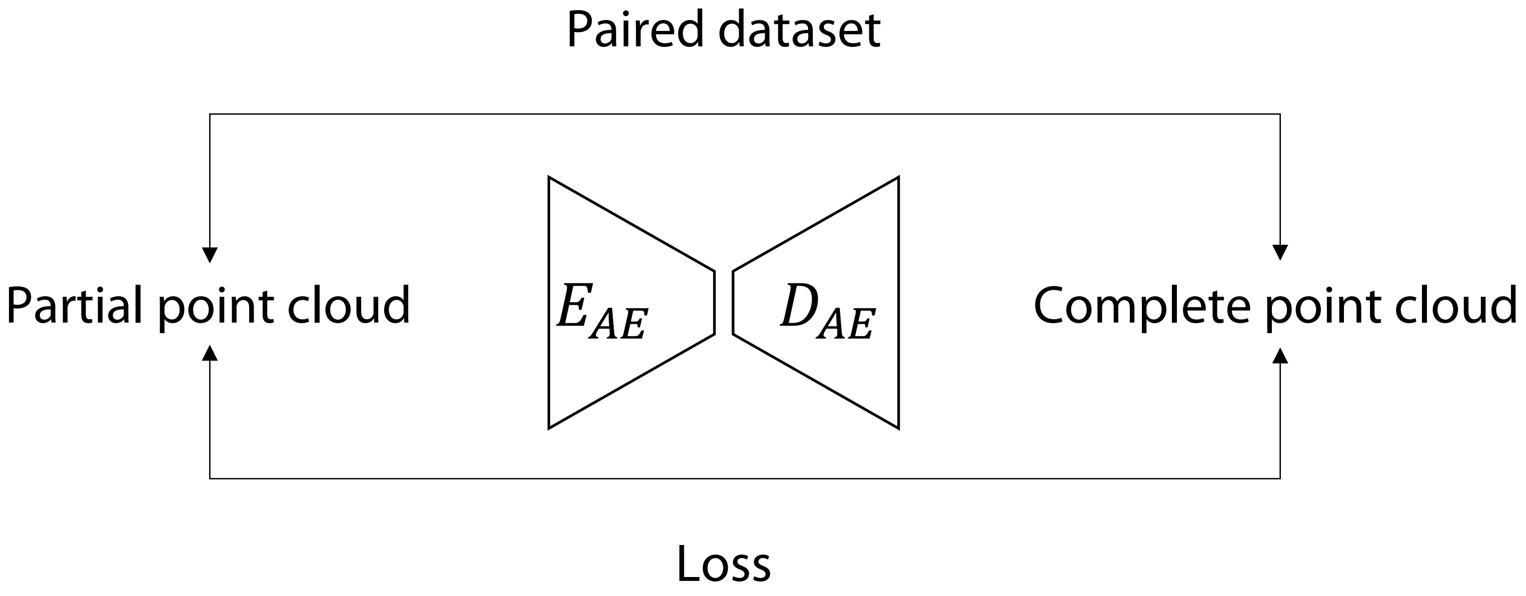
Multimodal Shape Completion via Conditional Generative Adversarial Networks

Paper by Wu et al. (2020)

Presented by Yongsoo Park

The task of shape completion (a naive approach)

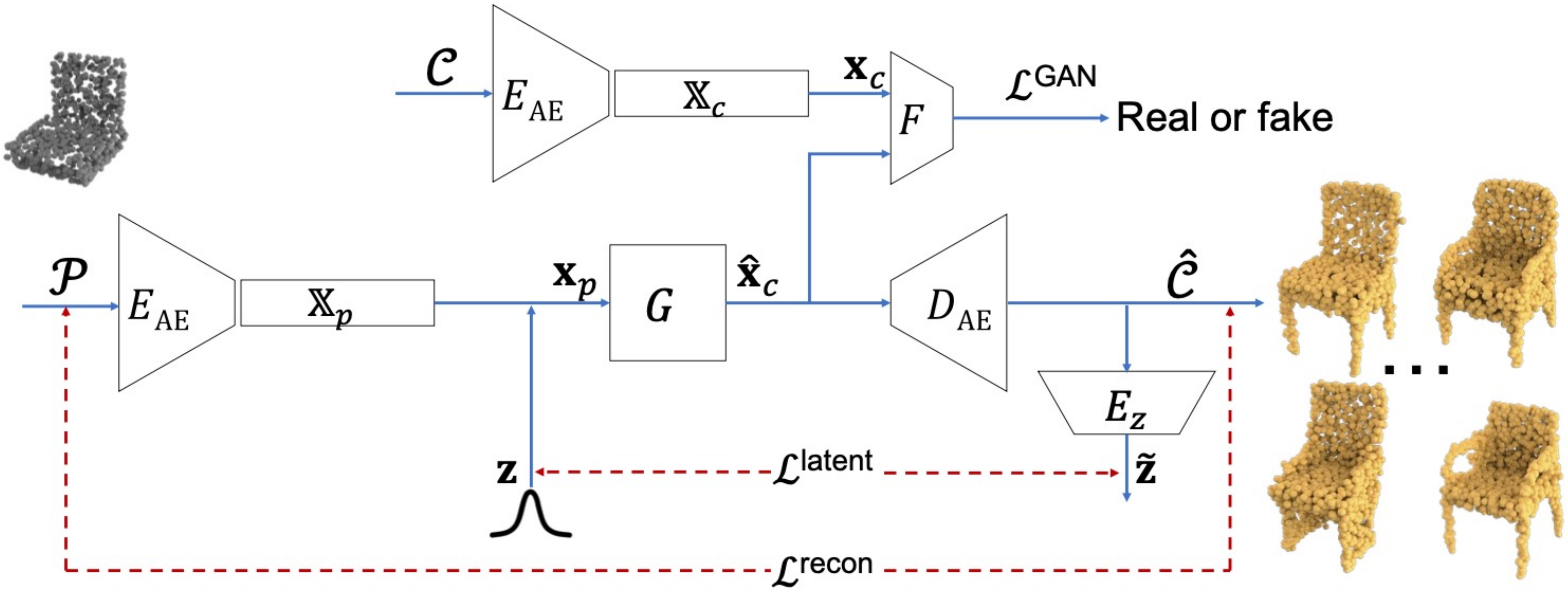


Motivation

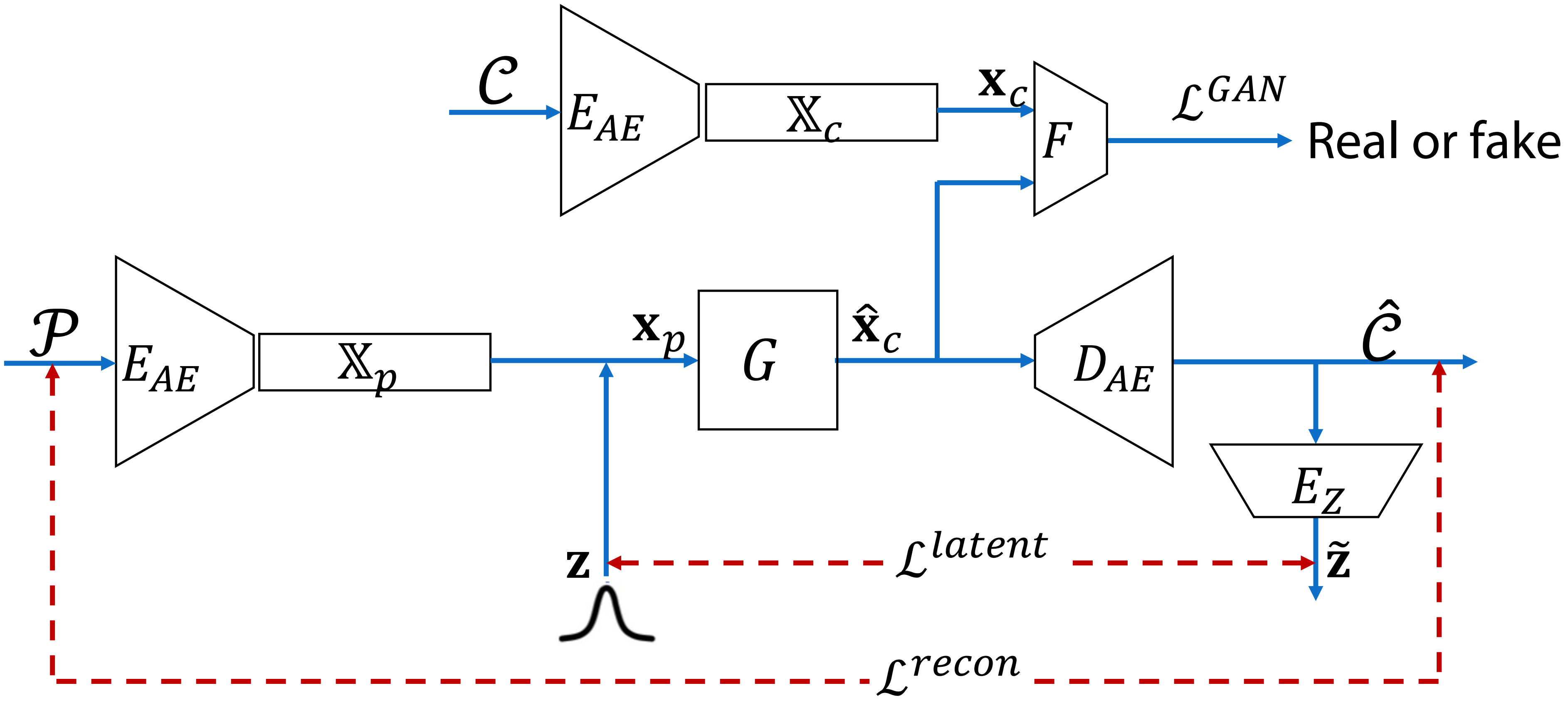
Given a partial point cloud, generate ~~a unique~~ multiple completed point clouds

and train without paired dataset

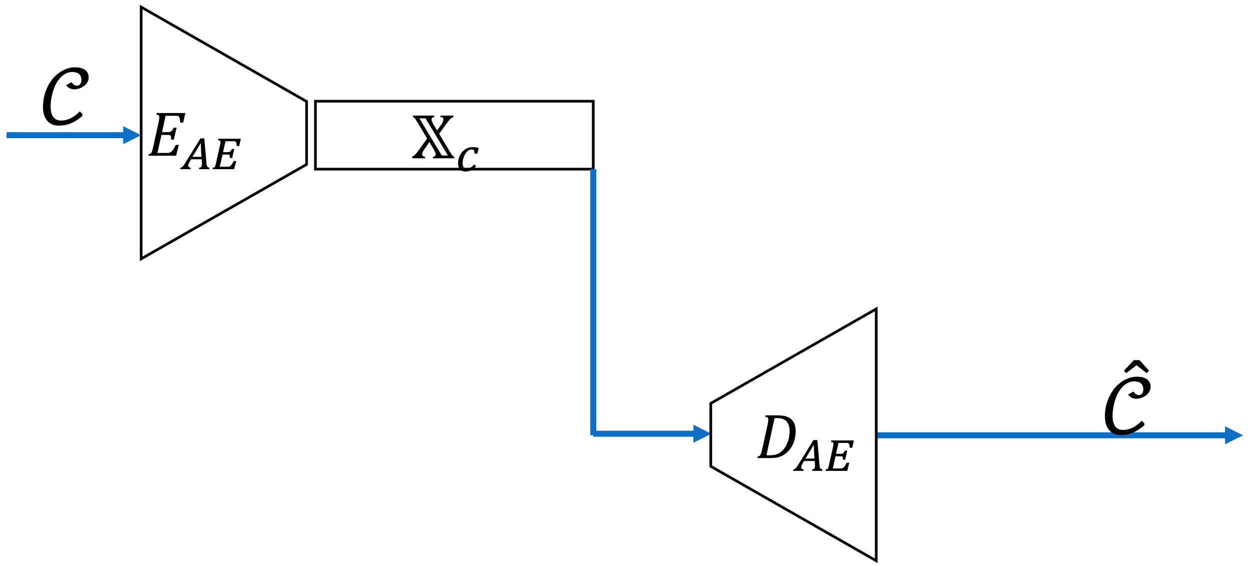
Network architecture



Network architecture

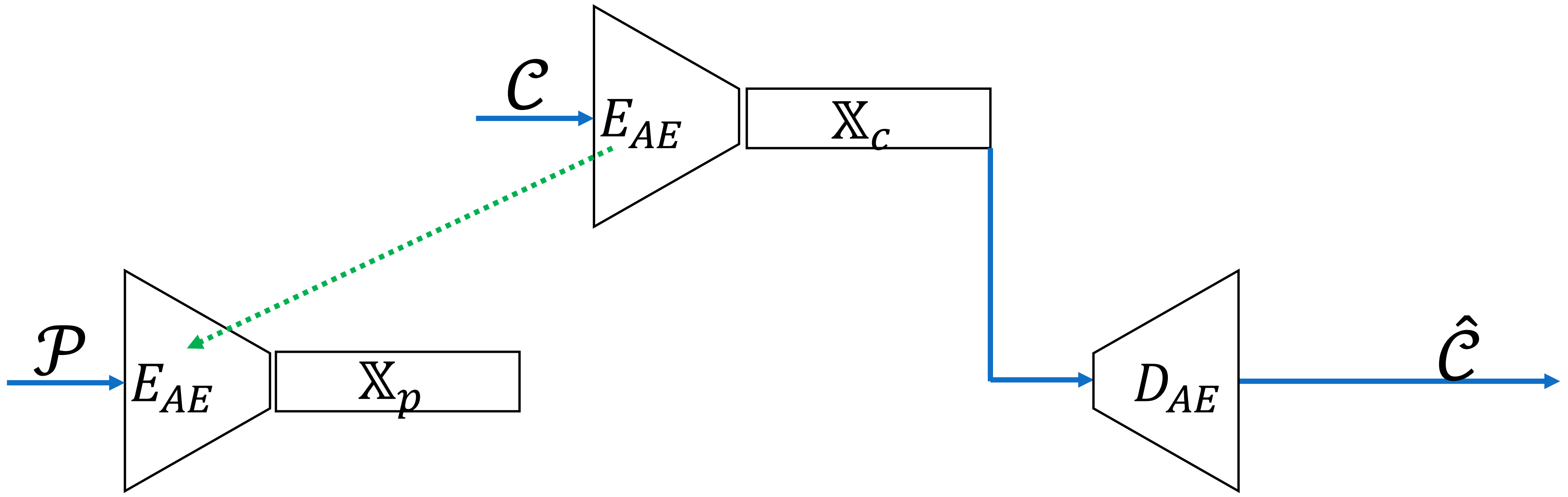


Network architecture



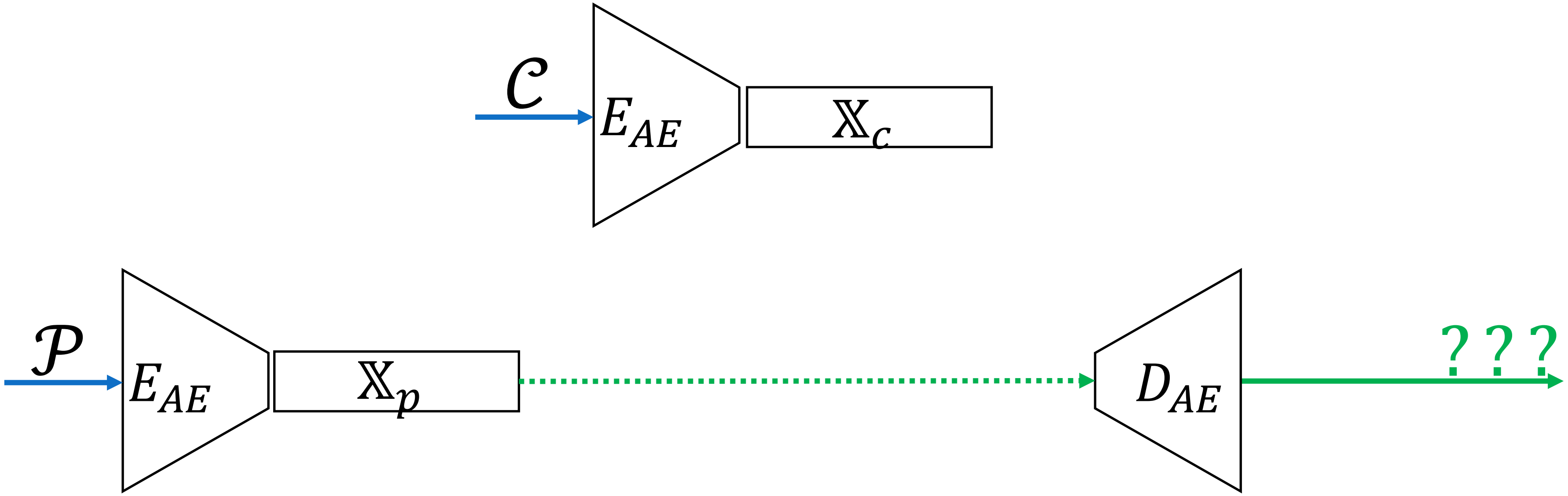
Train an auto-encoder with complete point clouds

Network architecture



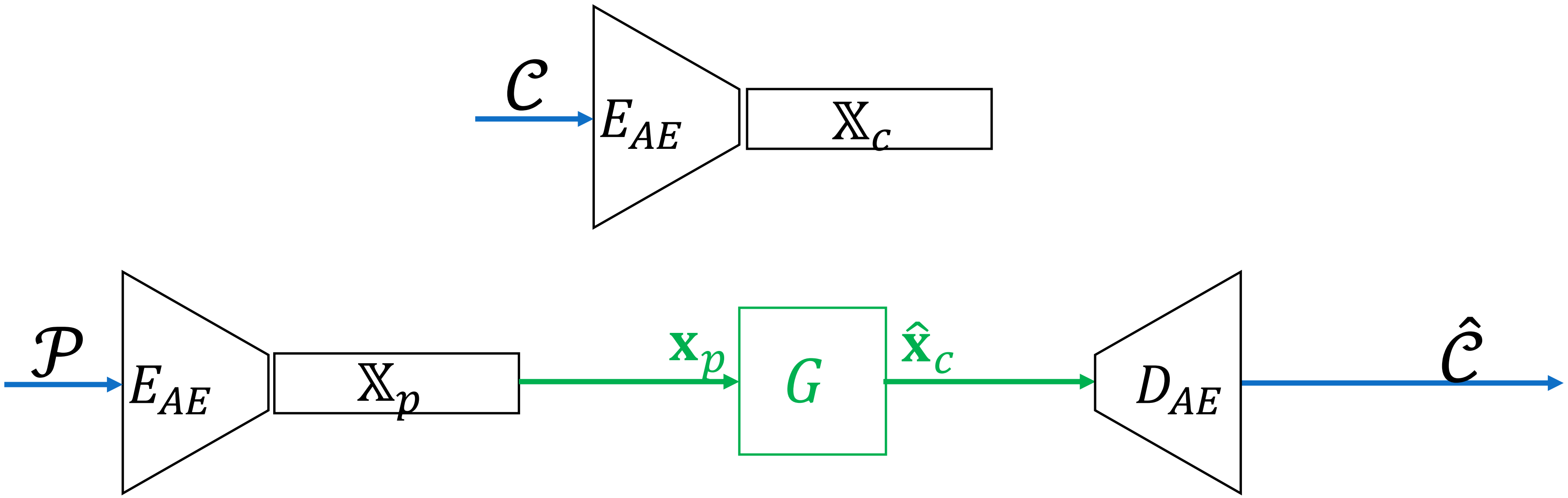
Can encode partial point clouds with the pre-trained auto-encoder

Network architecture



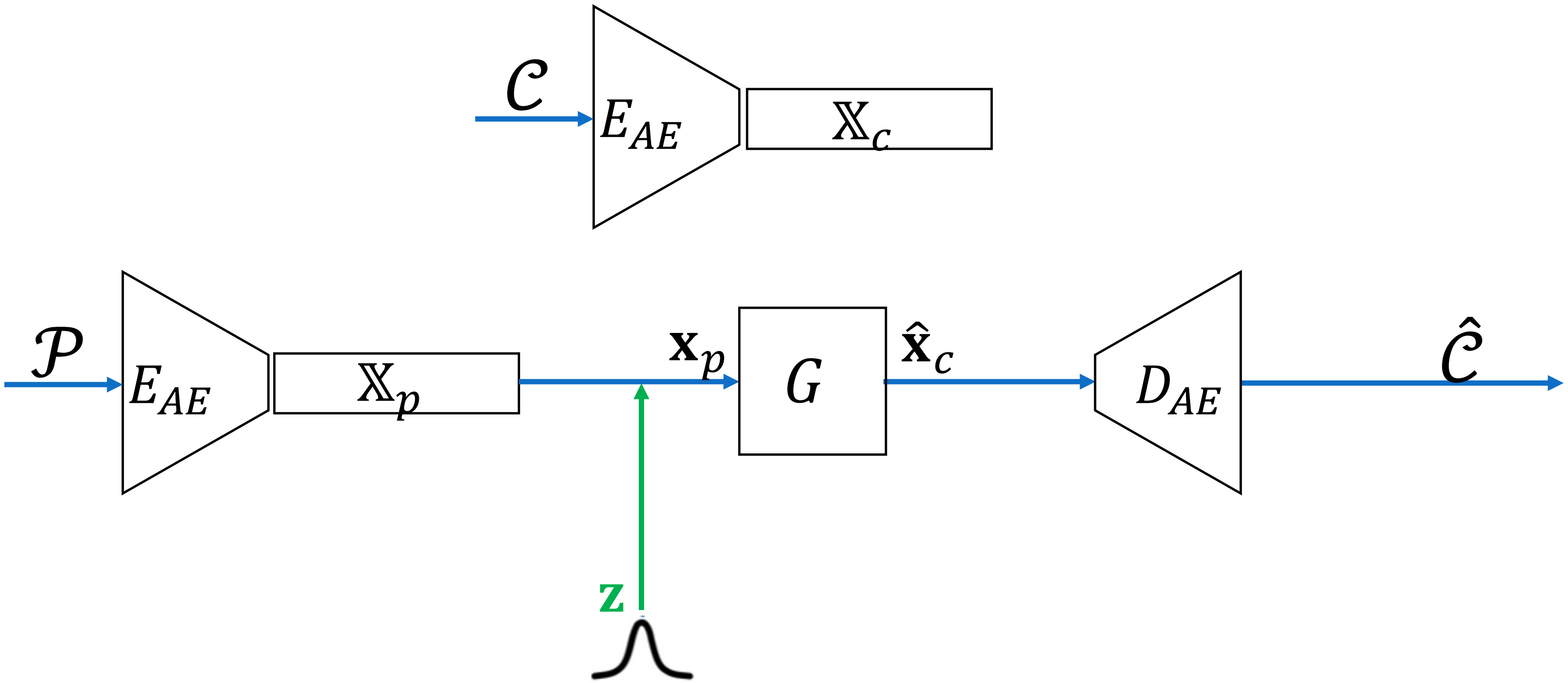
This won't complete the partial point clouds

Network architecture



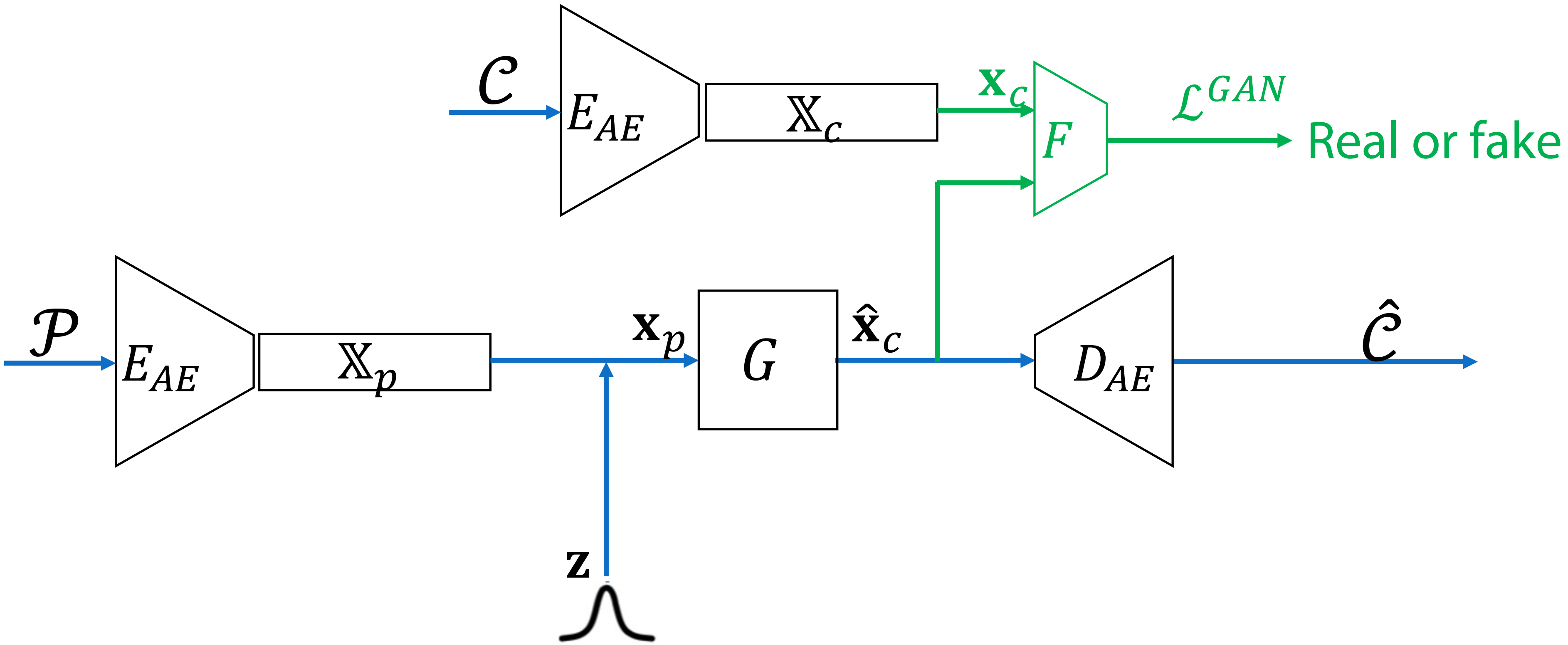
We want a generator (G) that can generate latent code for complete point cloud ($\hat{\mathbf{x}}_c$) given the latent code of a partial point cloud ($\mathbf{x}_p$)

Network architecture



And we want this generator to generate various shapes rather than a unique shape aka. multimodal aka. conditional

Network architecture

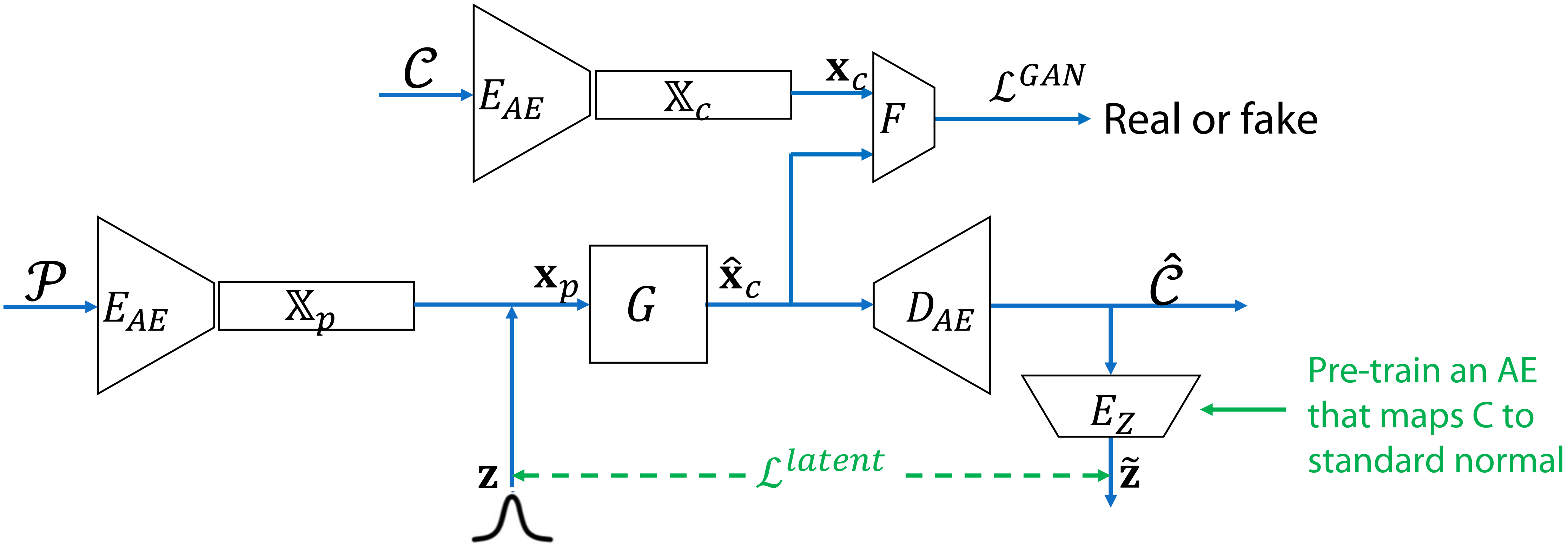


How do we train such a generator? → GAN

$$\mathcal{L}_F^{\text{GAN}} = \mathbb{E}_{\mathbf{C}\sim p(\mathbf{C})}[F(E_{\text{AE}}(\mathbf{C})) - 1]^2 + \mathbb{E}_{\mathbf{P}\sim p(\mathbf{P}),\mathbf{z}\sim p(\mathbf{z})}[F(G(E_{\text{AE}}(\mathbf{P}),\mathbf{z}))]^2$$

$$\mathcal{L}_G^{\text{GAN}} = \mathbb{E}_{\mathbf{P}\sim p(\mathbf{P}),\mathbf{z}\sim p(\mathbf{z})}[F(G(E_{\text{AE}}(\mathbf{P}),\mathbf{z})) - 1]^2,$$

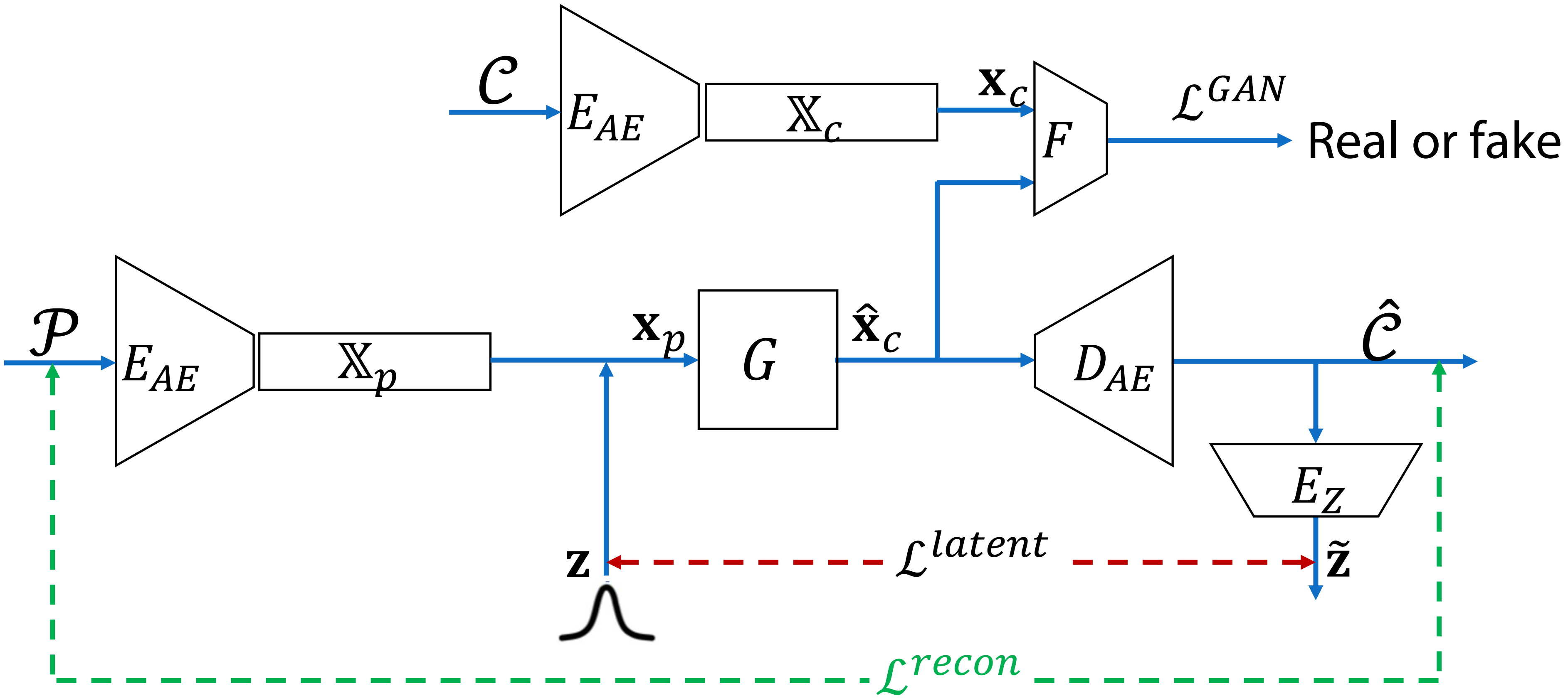
Network architecture



To encourage the generator to use the conditional mode vector **z**

$$\mathcal{L}^{\text{latent}}_{G,E_z} = \mathbb{E}_{\mathbf{P}\sim p(\mathbf{P}),\mathbf{z}\sim p(\mathbf{z})}[\|\mathbf{z}, E_z(D_{\text{AE}}(G(E_{\text{AE}}(\mathbf{P}),\mathbf{z})))\|_1],$$

Network architecture



To encourage the generator to partially reconstruct a partial input

$$\mathcal{L}_G^{\text{recon}} = \mathbb{E}_{\mathbf{P}\sim p(\mathbf{P}),\mathbf{z}\sim p(\mathbf{z})}\left[d^{\text{HL}}(\mathbf{P}, D_{\text{AE}}(G(E_{\text{AE}}(\mathbf{P}), \mathbf{z})))\right]),$$

Total loss

$$\operatorname*{argmin}_{(G,E_z)} \operatorname*{argmax}_{F} \mathcal{L}_F^{\mathrm{GAN}} + \mathcal{L}_G^{\mathrm{GAN}} + \alpha \mathcal{L}_G^{\mathrm{recon}} + \beta \mathcal{L}_{G,E_z}^{\mathrm{latent}},$$

Qualitative results

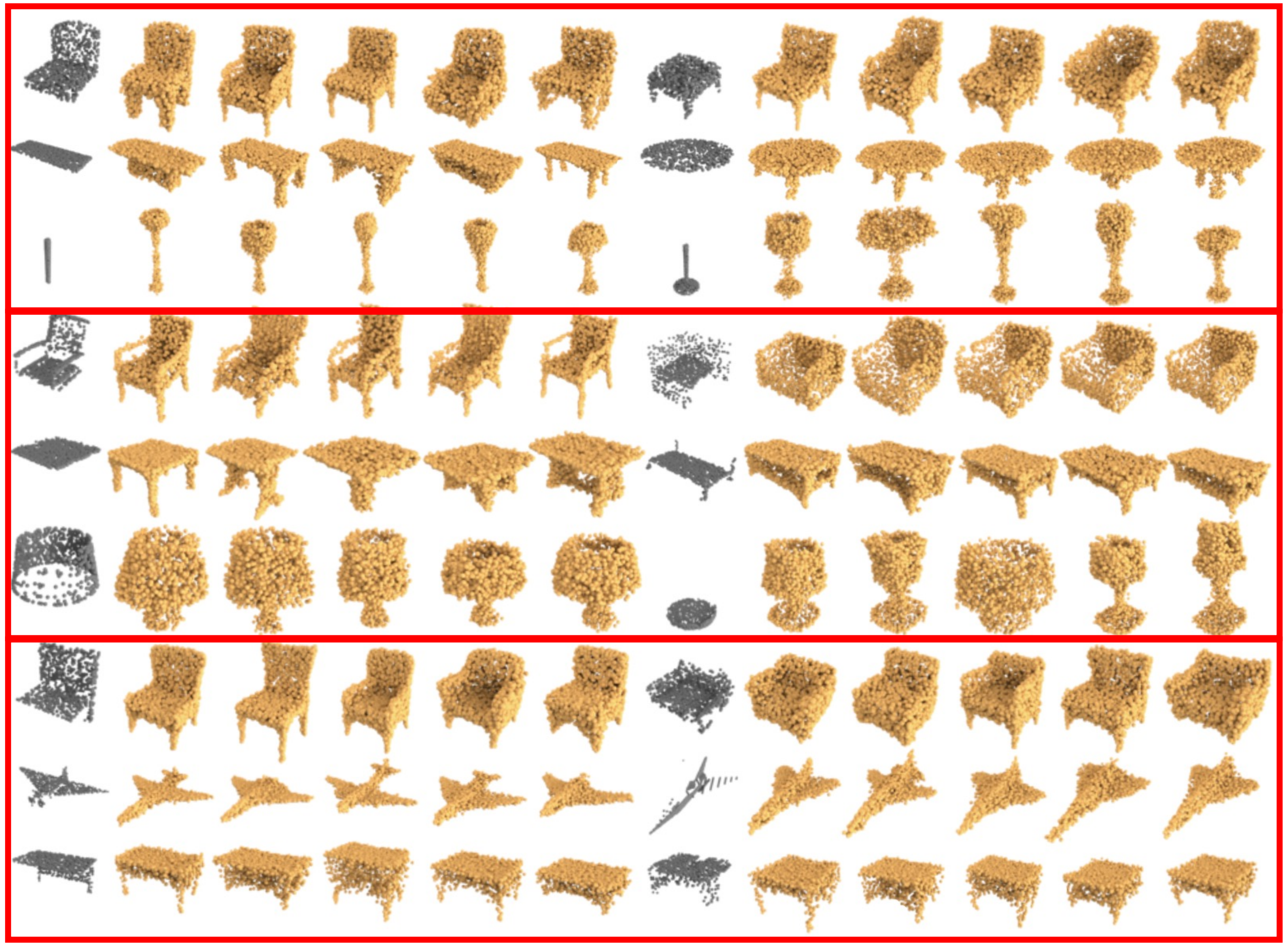


Fig. 3. Our multimodal shape completion results. We show result examples, where the input partial shape is colored in grey and is followed by five different completions in yellow. From top to bottom: `PartNet` (rows 1-3), `PartNet-Scan` (rows 4-6), and `3D-EPN` (rows 7-9).

- PartNet: Leave out part(s)
- PartNet-Scan: Leave out part(s) + incomplete scan
- 3D-EPN: Arbitrary incompleteness

Qualitative results

Encode a reference shape to get the latent code **z** and use it to complete the partial input

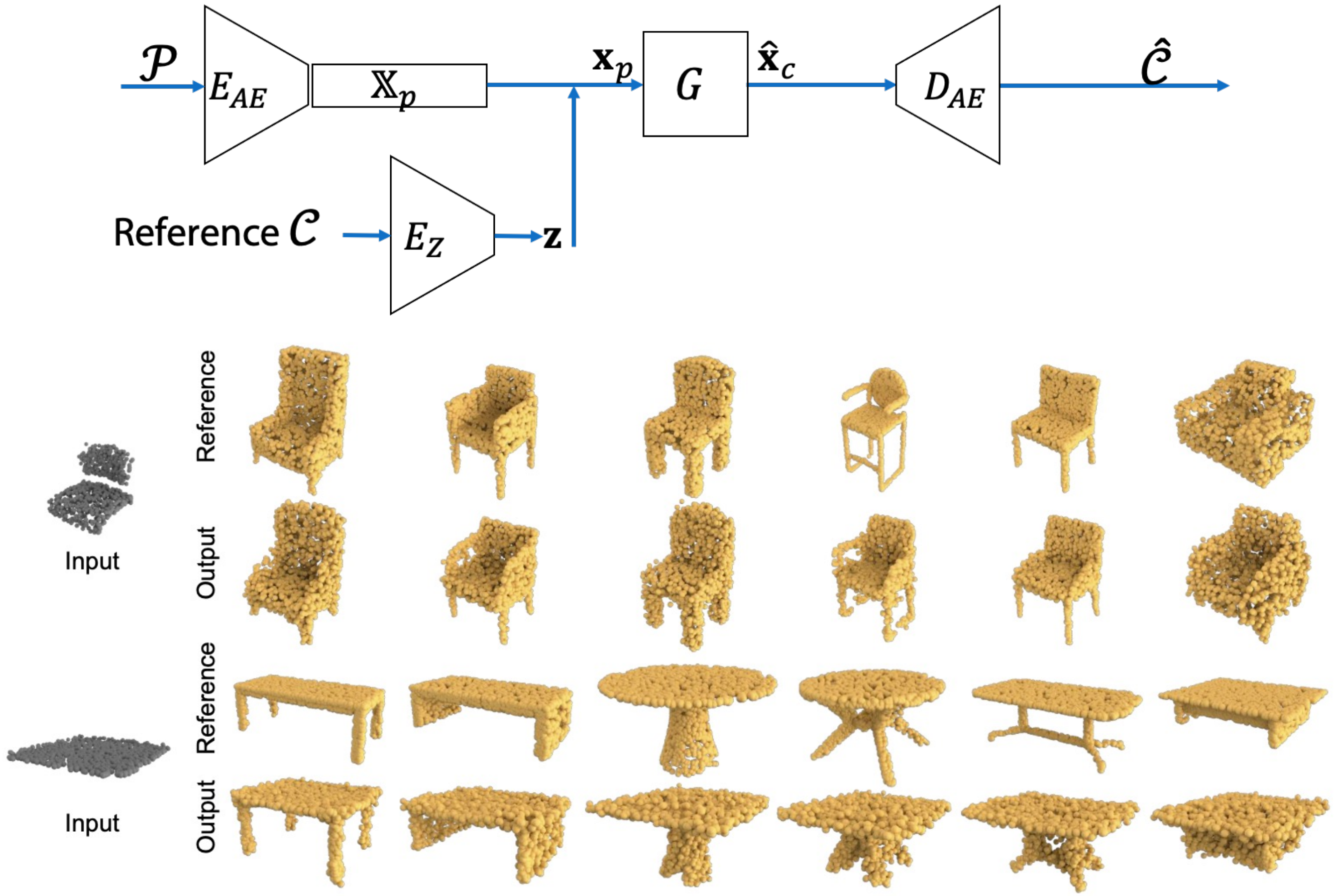
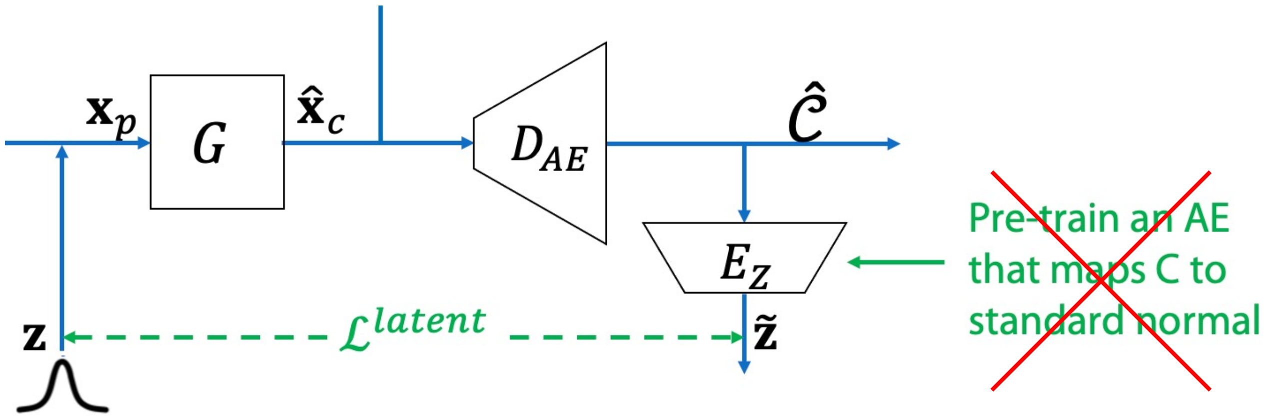


Fig. 4. Shape completion guided by reference shapes. The completion result varies accordingly when the reference shape changes.

Comparing with baselines

Baselines

- pcl2pcl: Also uses GAN & unpaired dataset but NO conditional (multimodal) generation
- KNN-latent: Encode → Latent code → Search for KNN latent codes → Decode
- Ours-im-l2z: Simultaneously train E_Z by feeding $\hat{\mathbf{x}}_c$ into E_Z
- Ours-im-pc2z: Simultaneously train E_Z by feeding $\hat{\mathcal{C}}$ into E_Z



Jointly train

Quantitative comparisons

Evaluation metric: Minimum Matching Distance (MMD)

<u>Meaning</u>

Measures the quality of the completed samples (w.r.t the test set) → The lower the better

<u>How it's calculated</u>

Essentially one-way CD or EMD

Quantitative comparisons

Evaluation metric: Total Mutual Difference (TMD)

<u>Meaning</u>

Measures the diversity of the completed samples → The higher the better

<u>How it's calculated</u>

1. Partial input → Generate ***k*** completed samples

2. For each completed sample, calculate the Chamfer Distance to the other ***k* – *1*** samples and get the mean

3. Sum ***k*** of these

Quantitative comparisons

Evaluation metric: Unidirectional Hausdorff Distance (UHD)

<u>Meaning</u>

Measures the fidelity of the completed samples (w.r.t the partial inputs) → Lower the better

<u>How it's calculated</u>

Calculate the average unidirectional HD from a partial input to ***k*** completed samples

Quantitative comparisons

PartNet	MMD (lower is better)				TMD (higher is better)				UHD (lower is better)			
Method	Chair	Lamp	Table	Avg.	Chair	Lamp	Table	Avg.	Chair	Lamp	Table	Avg.
pcl2pcl	1.90	2.50	1.90	2.10	0.00	0.00	0.00	0.00	4.88	4.64	4.78	**4.77**
KNN-latent	1.39	1.72	1.30	**1.47**	2.28	4.18	2.36	2.94	8.58	8.47	7.61	8.22
Ours-im-l2z	1.74	2.36	1.68	1.93	3.74	2.68	3.59	**3.34**	8.41	6.37	7.21	7.33
Ours-im-pc2z	1.90	2.55	1.54	2.00	1.01	0.56	0.51	0.69	6.65	5.40	5.38	**5.81**
Ours	1.52	1.97	1.46	**1.65**	2.75	3.31	3.30	**3.12**	6.89	5.72	5.56	6.06

PartNet-Scan	MMD (lower is better)				TMD (higher is better)				UHD (lower is better)			
Method	Chair	Lamp	Table	Avg.	Chair	Lamp	Table	Avg.	Chair	Lamp	Table	Avg.
pcl2pcl	1.96	2.36	2.09	2.14	0.00	0.00	0.00	0.00	5.20	5.34	4.73	**5.09**
KNN-latent	1.40	1.80	1.39	**1.53**	3.09	4.47	2.85	3.47	8.79	8.41	7.50	8.23
Ours-im-l2z	1.79	2.58	1.92	2.10	3.85	3.18	4.75	**3.93**	7.88	6.39	7.40	7.22
Ours-im-pc2z	1.65	2.75	1.84	2.08	1.91	0.50	1.86	1.42	7.50	5.36	5.68	**6.18**
Ours	1.53	2.15	1.58	**1.75**	2.91	4.16	3.88	**3.65**	6.93	5.74	6.24	6.30

3D-EPN	MMD (lower is better)				TMD (higher is better)				UHD (lower is better)			
Method	Chair	Plane	Table	Avg.	Chair	Plane	Table	Avg.	Chair	Plane	Table	Avg.
pcl2pcl	1.81	1.01	3.12	1.98	0.00	0.00	0.00	0.00	5.31	9.71	9.03	**8.02**
KNN-latent	1.45	0.93	2.25	**1.54**	2.24	1.13	3.25	2.21	8.94	9.54	12.70	10.40
Ours-im-l2z	1.91	0.86	2.78	1.80	3.84	2.17	4.27	**3.43**	9.53	10.60	9.36	9.83
Ours-im-pc2z	1.61	0.91	3.19	1.90	1.51	0.82	1.67	1.33	8.18	9.55	8.50	**8.74**
Ours	1.61	0.82	2.57	**1.67**	2.56	2.03	4.49	**3.03**	8.33	9.59	9.03	8.98

Rank #2 Rank #2 Rank #2

Quantitative comparisons

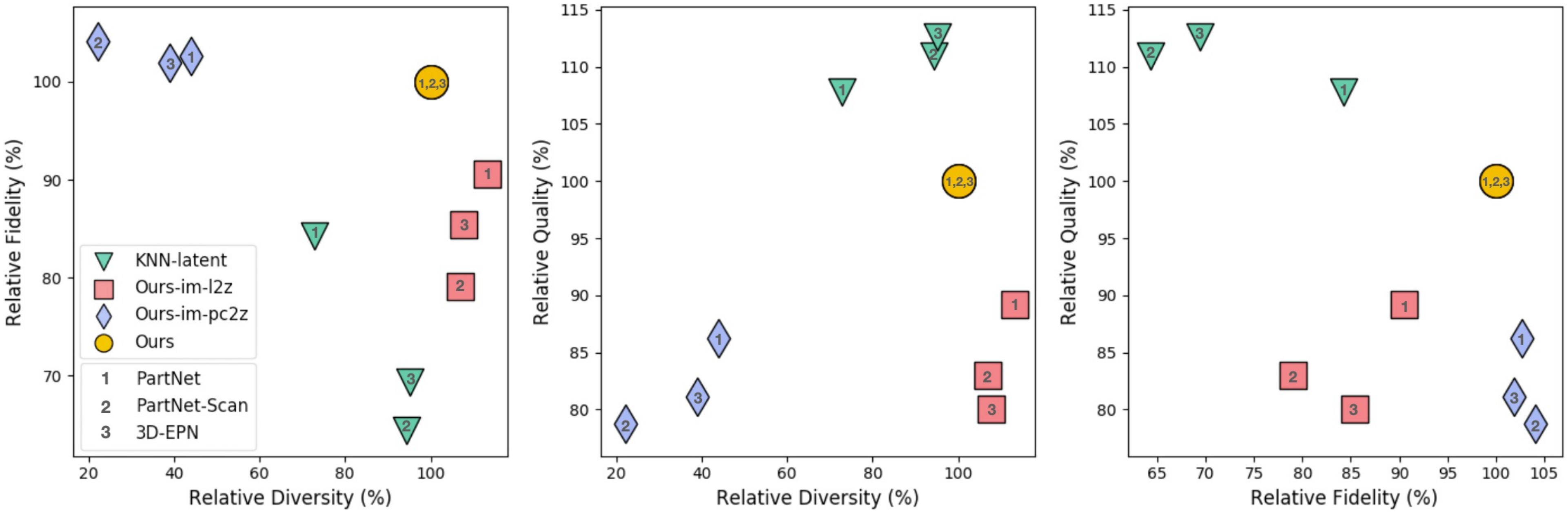
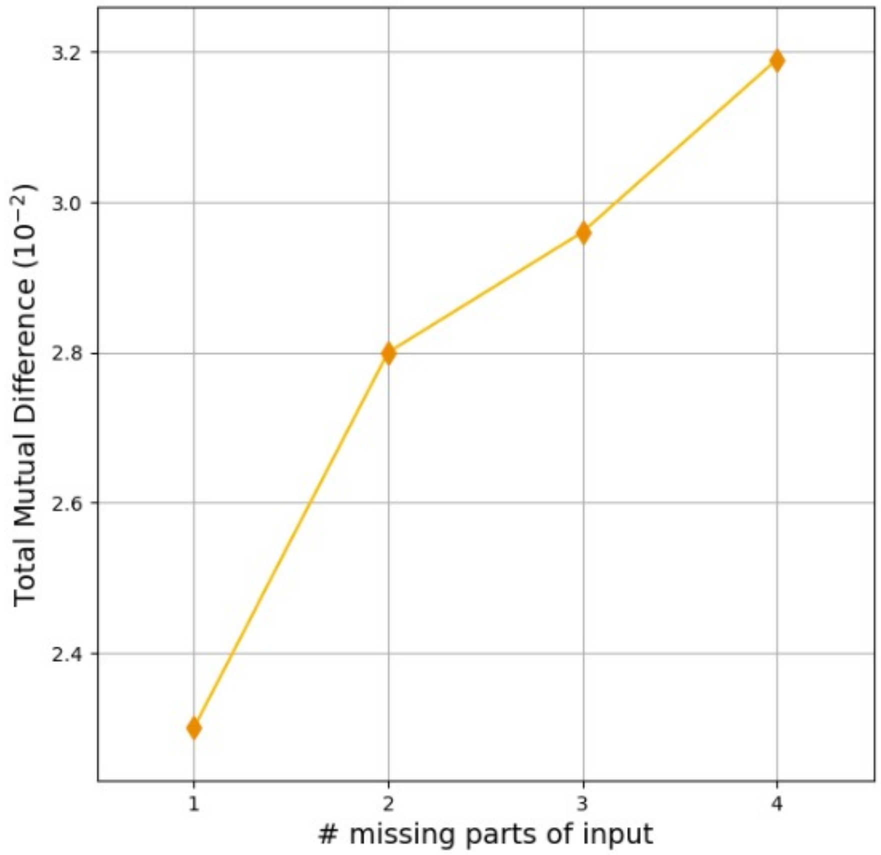


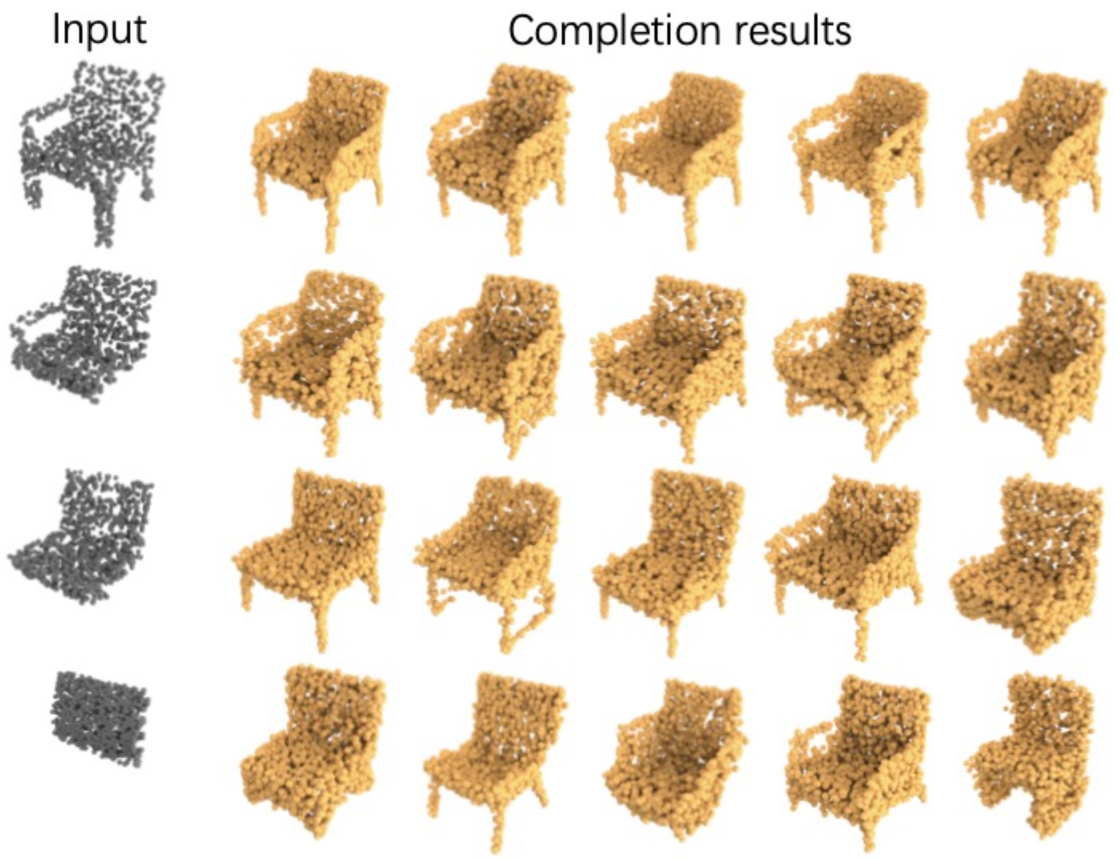
Fig. 6. Comparisons using metrics combinations. Our results present high diversity, quality and fidelity in comparisons using combinations of metrics. *Relative* performance is plotted, and **pcl2plc** is excluded as it fails to present completion diversity.

More experiments

Effect of input incompleteness



More missing parts → More diverse



Summary

Novelty

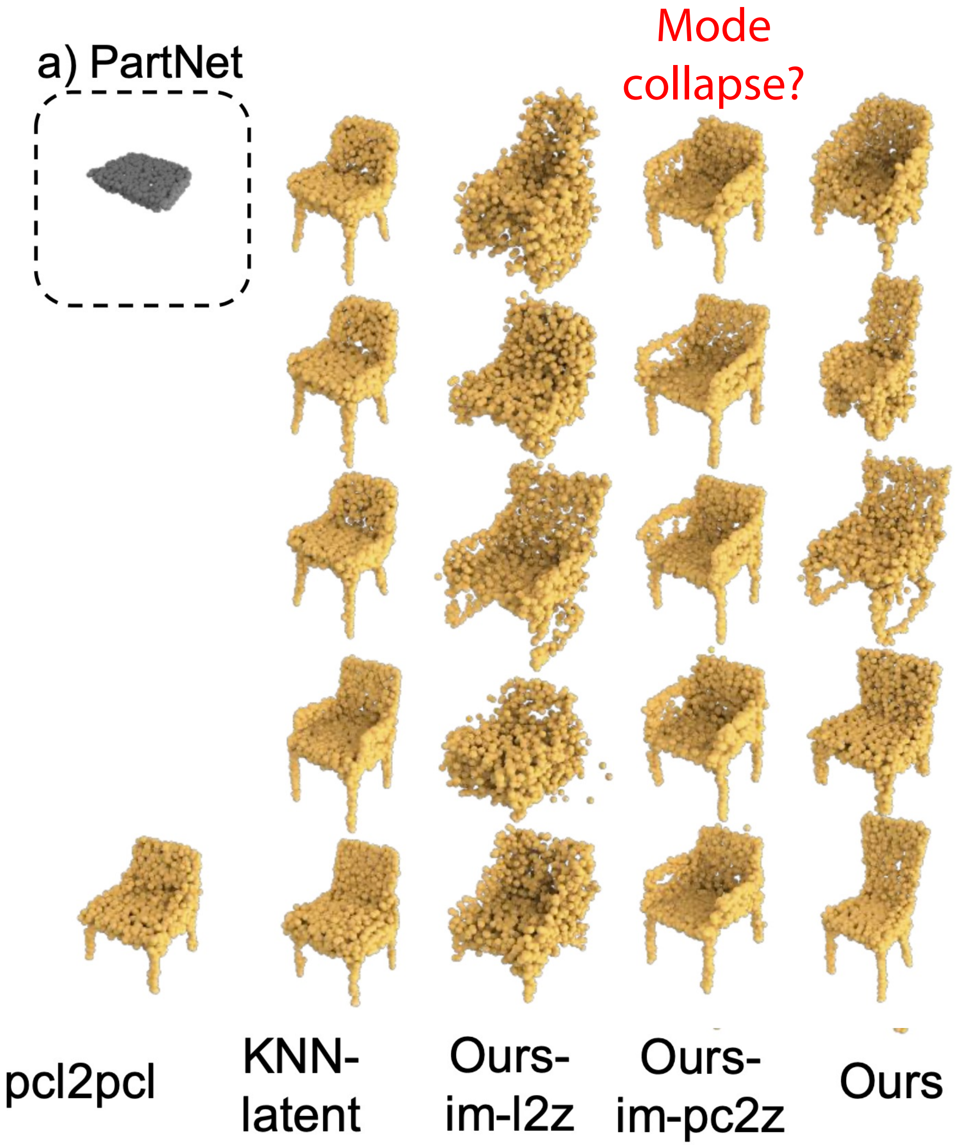
- Used conditional GAN to for multimodal shape completion.
- Found that jointly training E_z tend to produce worse results.
- Can complete into a shape we want to simulate + more missing part → more diversity.

Shortcomings (common issues)

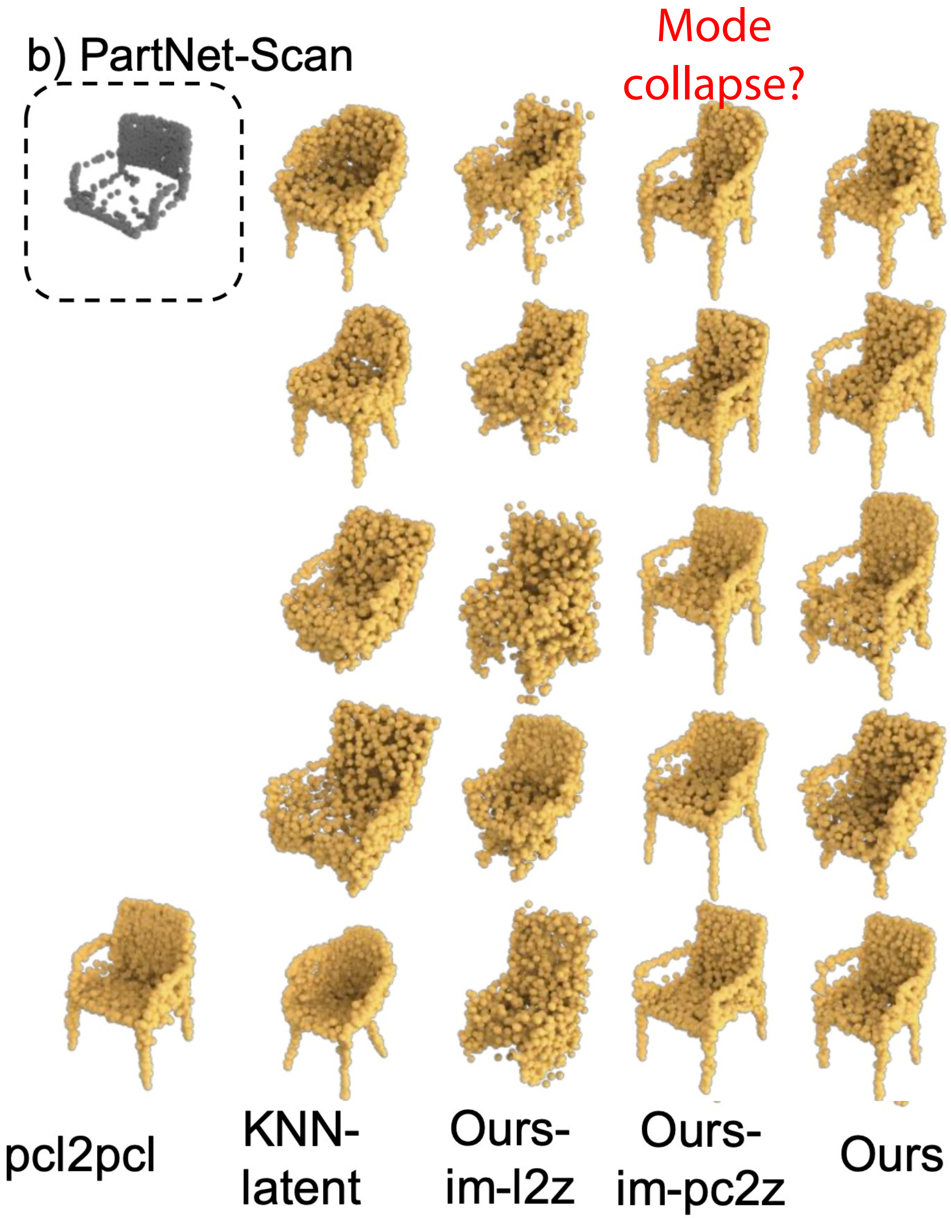
- Not producing samples with fine-scale details.
- Canonical orientation.

Backup slides

Qualitative comparisons



Qualitative comparisons



Perhaps KNN-latent

more diverse on this one?

Qualitative comparisons