



UNIVERSITY
OF WOLLONGONG
AUSTRALIA

BILATERAL COARSE-TO-FINE NETWORK FOR POINT CLOUD COMPLETION

Tran Thanh Phong Nguyen, Son Lam Phung, Vinod Gopaldasani,
Jane Whitelaw

Contents

1

Introduction

- Background
- Research gaps
- Fill the gaps

4

Experiments and analysis

- Datasets and experimental methods
- Evaluations on ShapeNet
- Evaluations on KITTI

2

Related work

- Point cloud completion
- View-point guidance on point clouds

5

Conclusion

- Contributions
- Limitations

3

Proposed method

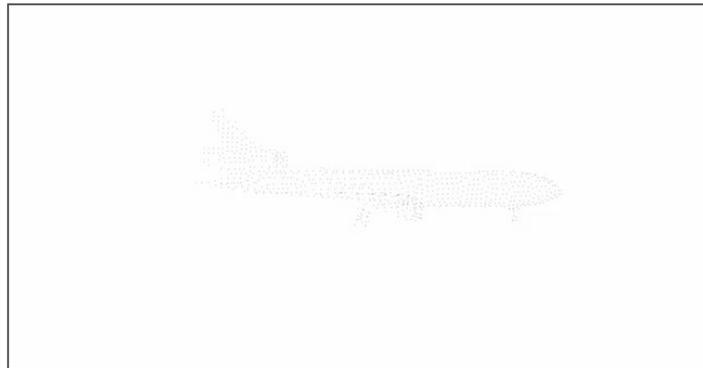
- Network architecture
- 2D modules
- 3D modules

Introduction

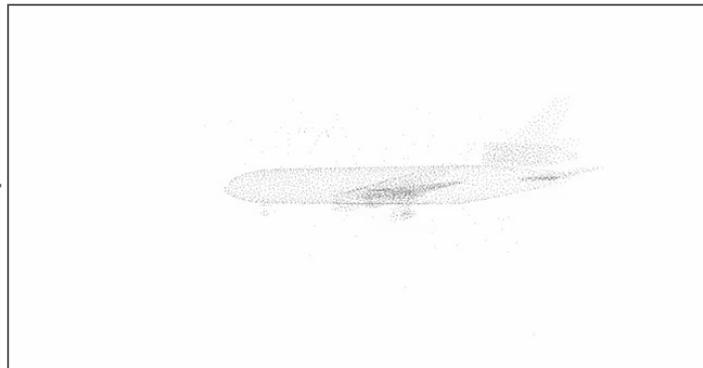
Background

- Point cloud completion predicts the complete 3D shape of an object from partial observations.
- Point cloud completion is important for improving the quality of 3D data, especially when capturing occluded or constrained views.
- Accurate and complete 3D data is crucial for the increasing number of 3D applications.

Partial



Complete



- Predicting complete point clouds from **partial inputs** leads to **crucial information loss** and **uncertain predictions**.
- Applying convolution operations to **irregular and unordered point clouds** is difficult. Converting point clouds to voxels is **computationally expensive** and results in **geometric information loss**.

- We propose a novel auto-encoder architecture that combines **2D and 3D modules** to address the structural loss of incomplete point clouds.
- We introduce a **multi-level codeword combination** that functions as a multi-scale skip-connection operation to predict and maintain structural details.
- We present **experimental results** that demonstrate improved completion outcomes compared to existing approaches on both **synthetic and real-world data**.

Related work

Point cloud completion

Papers	Methods
PCN [1]	Utilize folding-based decoders to create a universal 2D-to-3D mapping and generate complete point clouds in a coarse-to-fine manner.
FoldingNet [2]	
TopNet [3]	Utilize a free-structured decoder to improve structure-aware point cloud generation.
GR-Net [4]	Utilize 3D grids and skip-connections to learn context-aware and spatially-aware features.
PoinTr [5]	Utilize Transformers to learn structural information and global correlations for point cloud completion.

[1] W. Yuan, T. Khot, D. Held, C. Mertz, and M. Hebert, "PCN: Point completion network," in International Conference on 3D Vision, 2018, pp. 728–737.

[2] Y. Yang, C. Feng, Y. Shen, and D. Tian, "Foldingnet: Point cloud auto-encoder via deep grid deformation," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 206–215.

[3] L. P. Tchapmi, V. Kosaraju, H. Rezatofighi, I. Reid, and S. Savarese, "Topnet: Structural point cloud decoder," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 383–392.

[4] H. Xie, H. Yao, S. Zhou, J. Mao, S. Zhang, and W. Sun, "Grnet: Gridding residual network for dense point cloud completion," in European Conference on Computer Vision, 2020, pp. 365–381.

[5] X. Yu, Y. Rao, Z. Wang, Z. Liu, J. Lu, and J. Zhou, "Pointr: Diverse point cloud completion with geometry-aware transformers," in Proc. IEEE/CVF International Conference on Computer Vision, 2021, pp. 12498–12507.

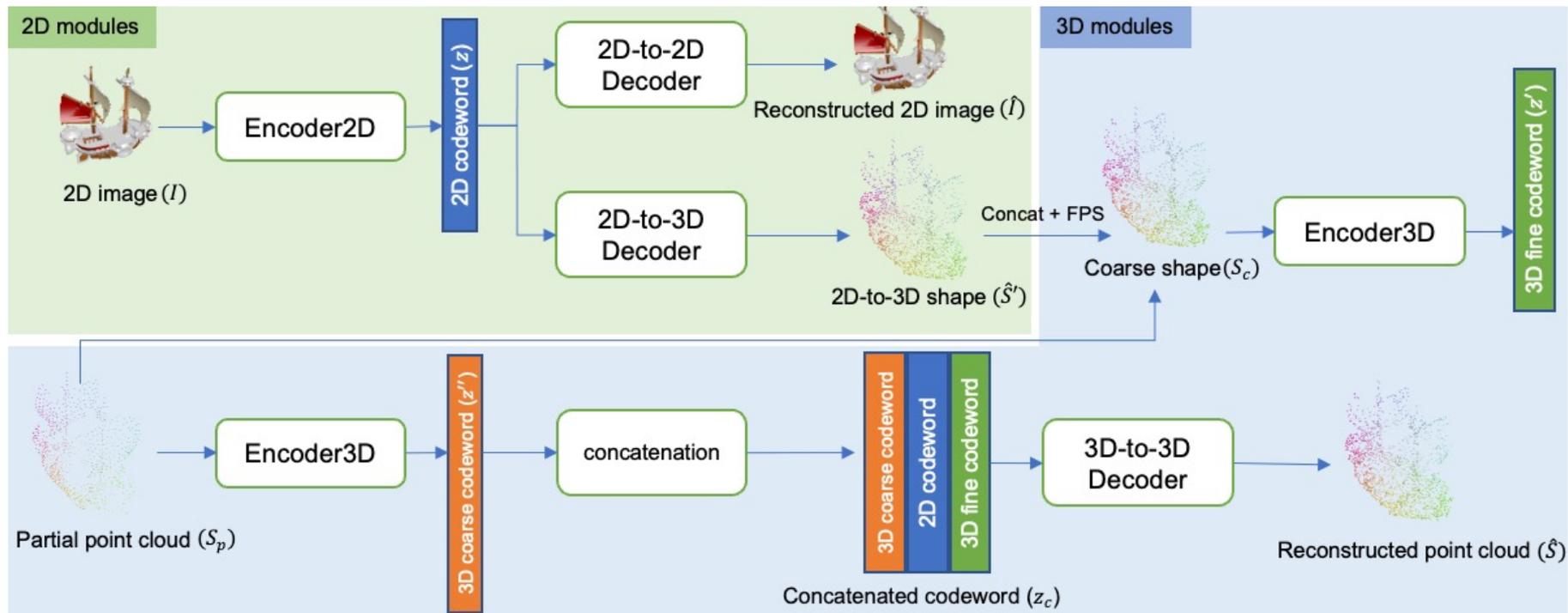
- [Bilateral filtering](#) [13] uses the image as guidance for target reconstruction.
- Su et al. [14] apply this idea to solve the [segmentation task](#).
- Our method also incorporates [point-based and image-based](#) representations to solve the point cloud completion task.

[13] K. He, J. Sun, and X. Tang, "Guided image filtering," in European Conference on Computer Vision, 2010, pp. 1–14.

[14] H. Su, V. Jampani, D. Sun, S. Maji, E. Kalogerakis, M. H. Yang, and J. Kautz, "Splatnet: Sparse lattice networks for point cloud processing," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 2530–2539.

Proposed method

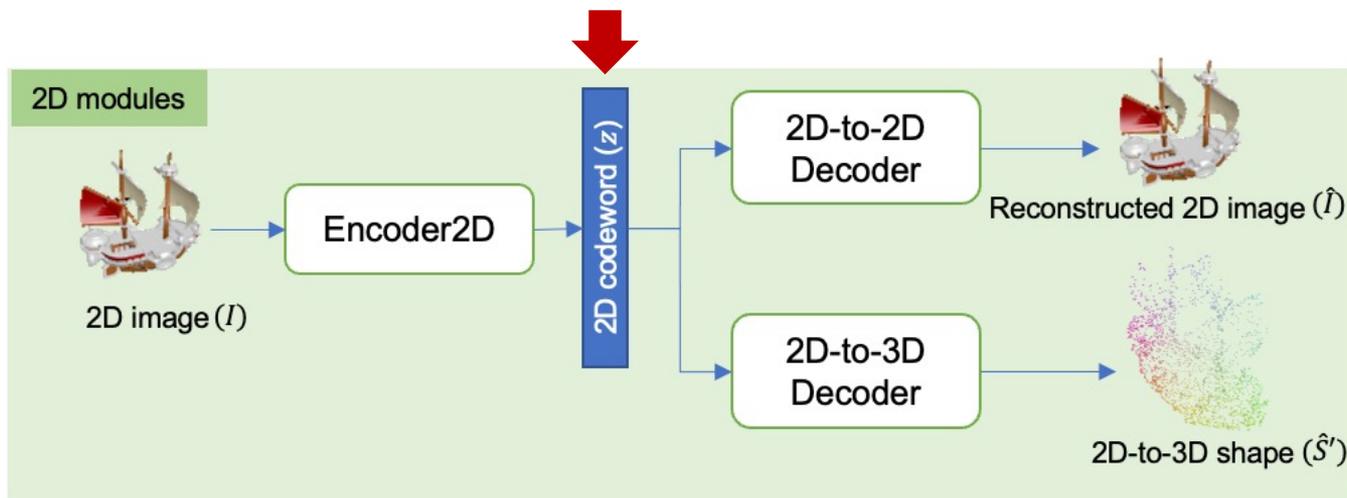
Network architecture



The network architecture of BCF-Net for point cloud completion.

Proposed method

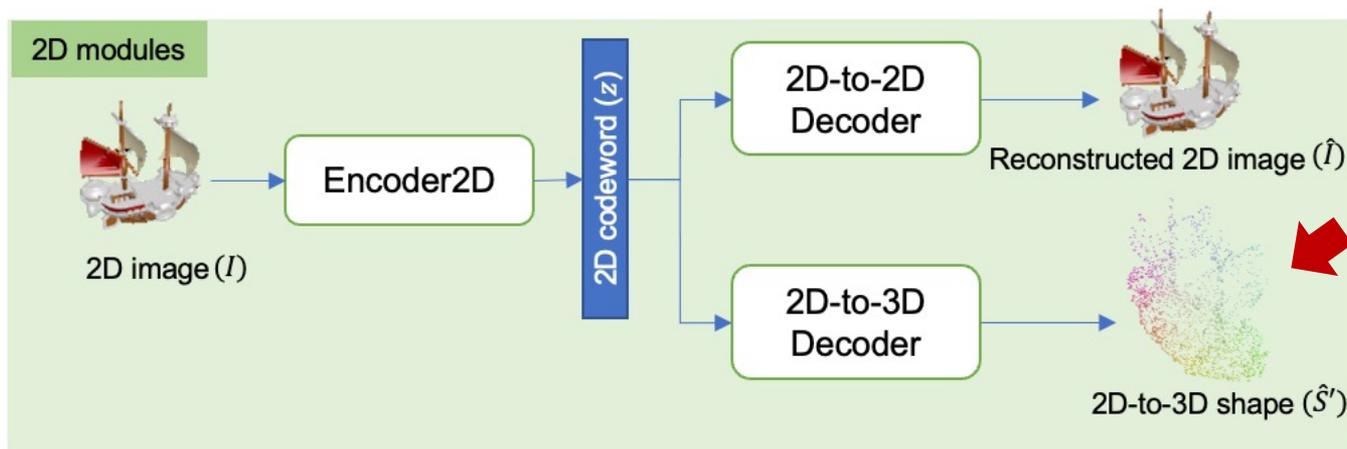
2D modules



- **Encoder2D** and **2D-to-2D Decoder** forms a variational auto-encoder to regulate **2D codeword z** .

Proposed method

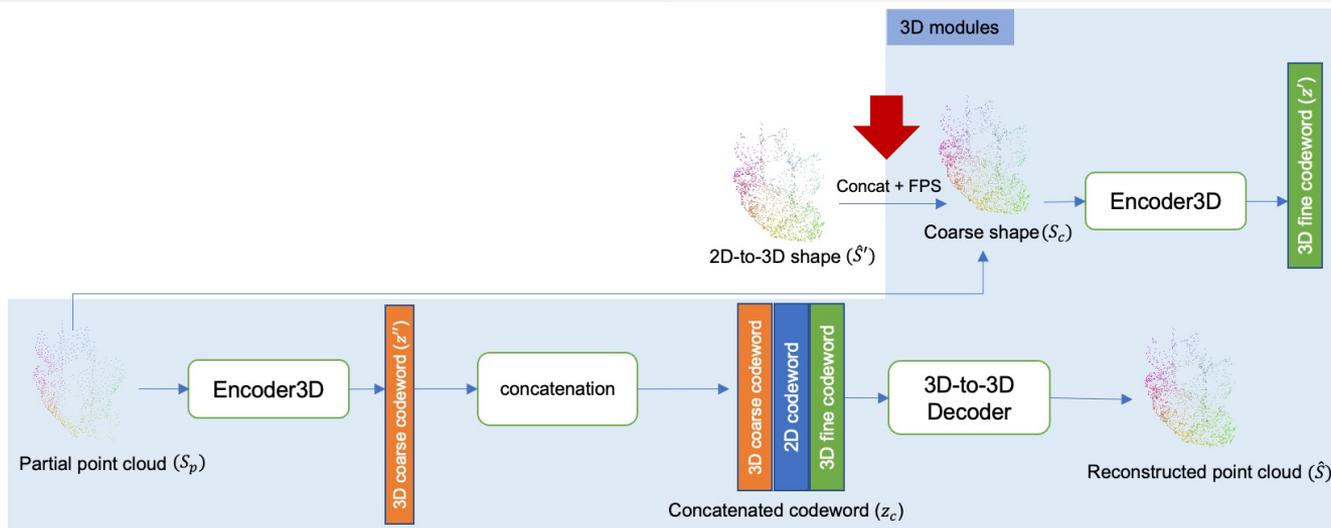
2D modules



- Encoder2D and 2D-to-2D Decoder forms a variational auto-encoder to regulate 2D codeword z .
- 2D-to-3D Decoder uses z to reconstruct the 2D-to-3D shape \hat{S}' .

Proposed method

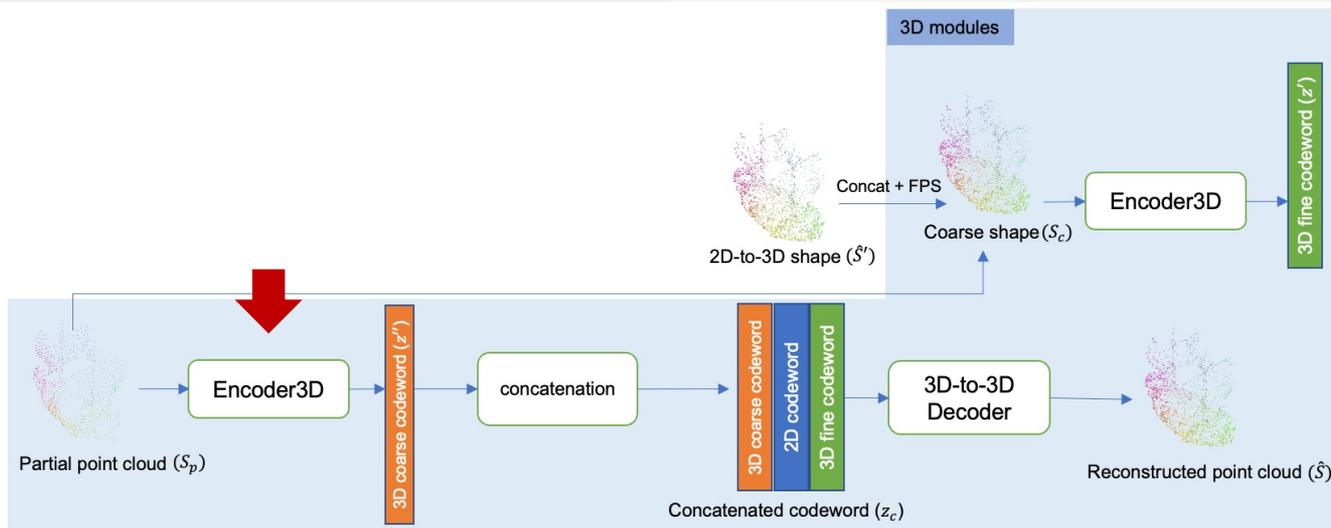
3D modules



- We concat the 2D-to-3D shape \hat{S}^t and the partial shape S_p to the coarse shape S_c .

Proposed method

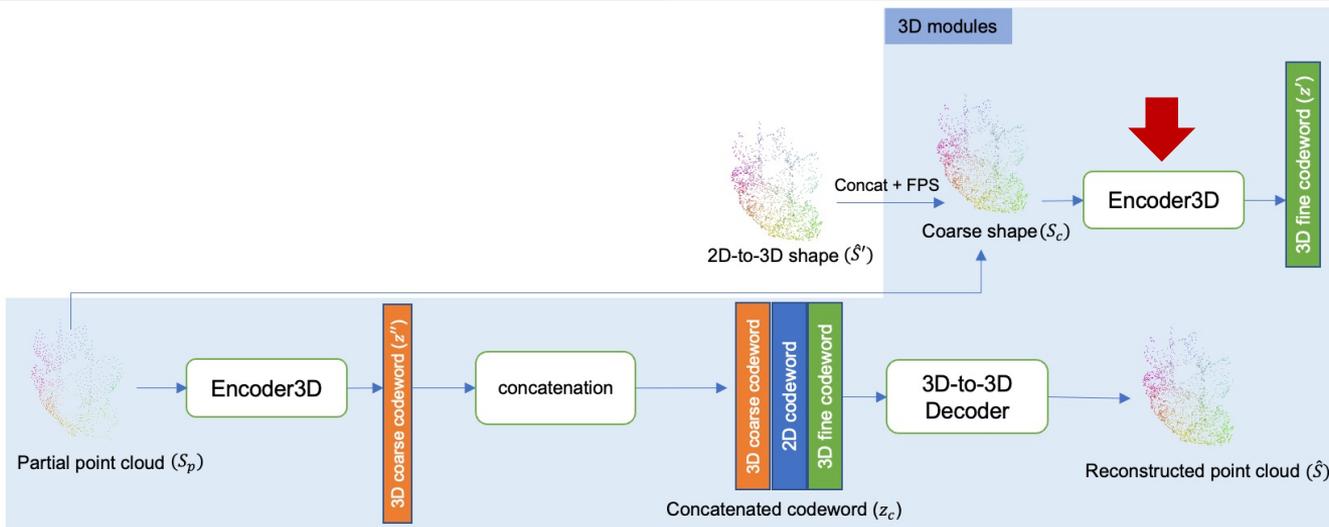
3D modules



- We concat the 2D-to-3D shape \hat{S}^t and the partial shape S_p to the coarse shape S_c .
- Partial shape S_p goes to coarse encoder to obtain the coarse codeword z^c .

Proposed method

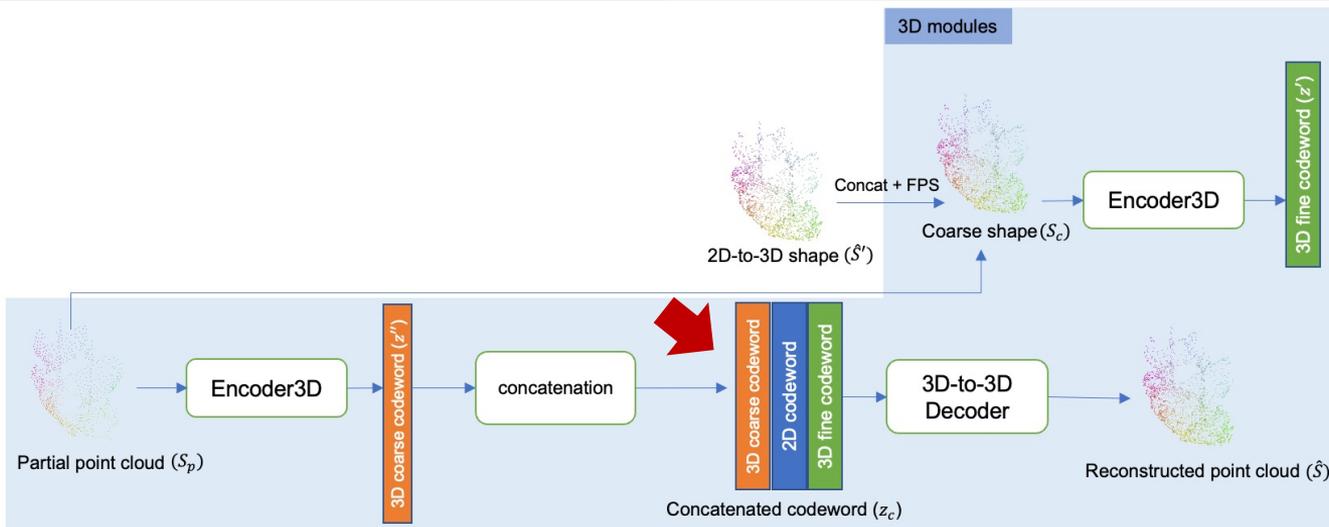
3D modules



- We concat the 2D-to-3D shape \hat{S}^t and the partial shape S_p to the coarse shape S_c .
- Partial shape S_p goes to coarse encoder to obtain the coarse codeword z^c .
- Coarse shape S_c goes into fine encoder to obtain the fine codeword z^f .

Proposed method

3D modules



- We concat the 2D-to-3D shape \hat{S}' and the partial shape S_p to the coarse shape S_c .
- Partial shape S_p goes to coarse encoder to obtain the coarse codeword z'' .
- Coarse shape S_c goes into fine encoder to obtain the fine codeword z' .
- All the codewords z, z', z'' go to the decoder to reconstruct the complete point cloud \hat{S} .

- ➔ • 2D image **reconstruction** loss: $L_{2D}(\mathbf{I}, \hat{\mathbf{I}}) = \sum (\mathbf{I} - \hat{\mathbf{I}})^2 + L_{\text{KL}}$

- 2D image **reconstruction** loss: $L_{2D}(\mathbf{I}, \hat{\mathbf{I}}) = \sum (\mathbf{I} - \hat{\mathbf{I}})^2 + L_{\text{KL}}$

- ➔ • 3D point cloud **coarse** loss: $L_{\text{coarse}}(\mathbf{S}, \hat{\mathbf{S}}') = \frac{1}{|\mathbf{S}|} \sum_{\mathbf{x} \in \mathbf{S}} \min \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 + \frac{1}{|\hat{\mathbf{S}}'|} \sum_{\hat{\mathbf{x}} \in \hat{\mathbf{S}}'} \min \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2$

- 2D image **reconstruction** loss: $L_{2D}(\mathbf{I}, \hat{\mathbf{I}}) = \sum (\mathbf{I} - \hat{\mathbf{I}})^2 + L_{\text{KL}}$.

- 3D point cloud **coarse** loss: $L_{\text{coarse}}(\mathbf{S}, \hat{\mathbf{S}}') = \frac{1}{|\mathbf{S}|} \sum_{\mathbf{x} \in \mathbf{S}} \min \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 + \frac{1}{|\hat{\mathbf{S}}'|} \sum_{\hat{\mathbf{x}} \in \hat{\mathbf{S}}'} \min \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2$.

-  • 3D point cloud **fine** loss: $L_{\text{fine}}(\mathbf{S}, \hat{\mathbf{S}}) = \frac{1}{|\mathbf{S}|} \sum_{\mathbf{x} \in \mathbf{S}} \min \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 + \frac{1}{|\hat{\mathbf{S}}|} \sum_{\hat{\mathbf{x}} \in \hat{\mathbf{S}}} \min \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2$.

- 2D image **reconstruction** loss: $L_{2D}(\mathbf{I}, \hat{\mathbf{I}}) = \sum (\mathbf{I} - \hat{\mathbf{I}})^2 + L_{\text{KL}}$.

- 3D point cloud **coarse** loss: $L_{\text{coarse}}(\mathbf{S}, \hat{\mathbf{S}}') = \frac{1}{|\mathbf{S}|} \sum_{\mathbf{x} \in \mathbf{S}} \min \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 + \frac{1}{|\hat{\mathbf{S}}'|} \sum_{\hat{\mathbf{x}} \in \hat{\mathbf{S}}'} \min \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2$.

- 3D point cloud **fine** loss: $L_{\text{fine}}(\mathbf{S}, \hat{\mathbf{S}}) = \frac{1}{|\mathbf{S}|} \sum_{\mathbf{x} \in \mathbf{S}} \min \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 + \frac{1}{|\hat{\mathbf{S}}|} \sum_{\hat{\mathbf{x}} \in \hat{\mathbf{S}}} \min \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2$.

 • Total loss: $L_{\text{training}} = \alpha L_{\text{CD}}(\mathbf{S}, \hat{\mathbf{S}}) + \beta L_{\text{CD}}(\mathbf{S}, \hat{\mathbf{S}}') + \gamma L_{2D}(\mathbf{I}, \hat{\mathbf{I}})$.

Datasets	ShapeNet	KITTI
Number of categories	8	1
Synthetic	Yes	No
Real-world	No	Yes



- **ShapeNet** is a **synthetic** dataset consisting of objects from **eight categories** including airplane, cabinet, car, chair, lamp, sofa, table, and watercraft.
- **KITTI** is a **real-world** dataset consisting of **real cars**.

Quantitative results on ShapeNet using the Chamfer Distance metric.

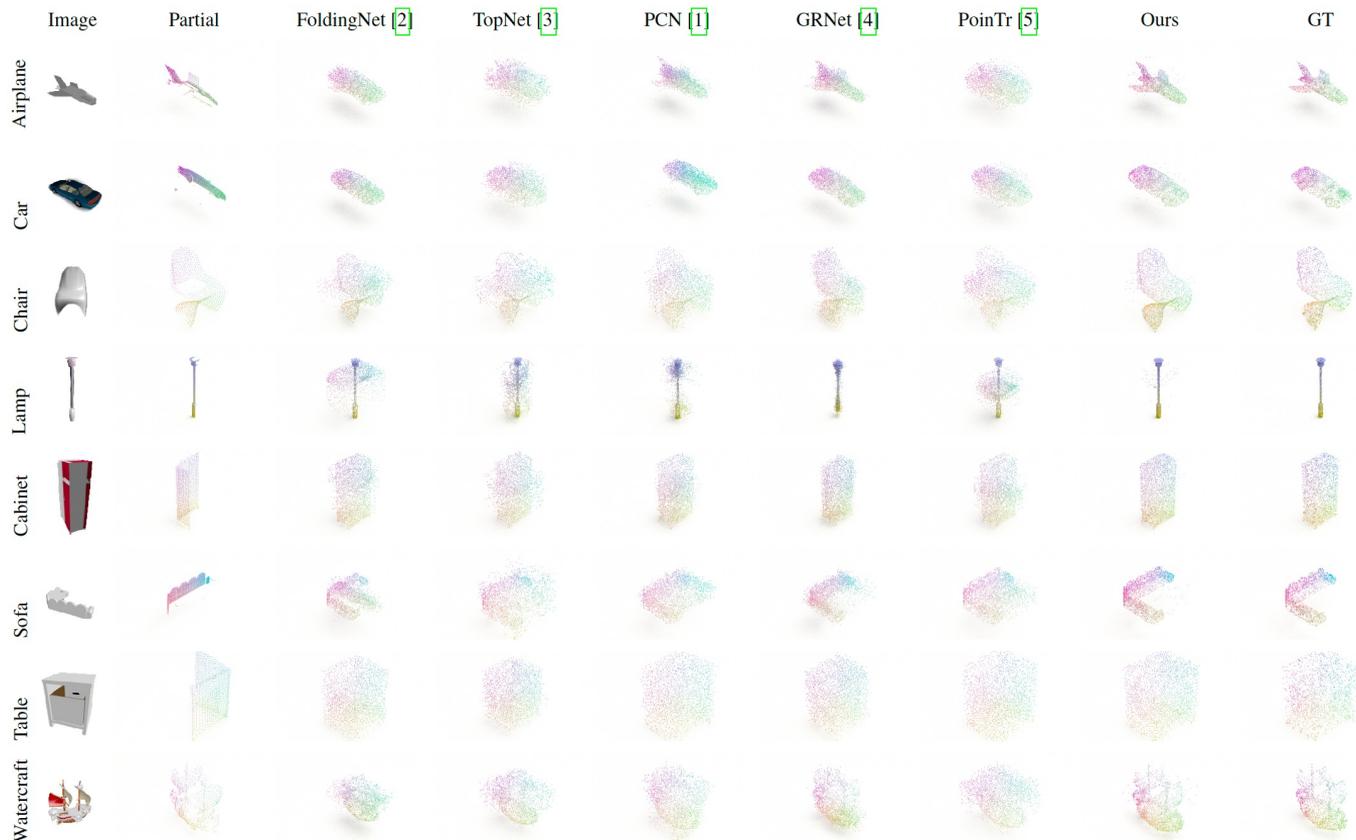
	Avg	Airplane	Cabinet	Car	Chair	Lamp	Sofa	Table	Watercraft
PCN [1]	1.112	0.496	1.278	0.662	1.172	1.513	1.923	1.063	0.788
FoldingNet [2]	1.262	0.554	1.626	0.595	1.803	1.620	1.641	1.318	0.937
TopNet [3]	2.080	0.694	2.451	1.801	2.145	2.381	3.804	1.807	1.556
GRNet [4]	1.026	0.666	1.091	0.567	1.105	1.649	1.345	1.124	0.665
PoinTr [5]	5.014	3.010	5.709	4.427	6.016	4.304	9.781	4.040	2.828
Ours	0.913	0.432	1.002	0.473	1.153	1.385	1.057	1.055	0.745



The results show that our method **outperforms** other methods in most categories on the **Chamfer Distance** metric.

Experiments and analysis

Evaluations on ShapeNet



Visual results of point cloud completion methods on the ShapeNet dataset.

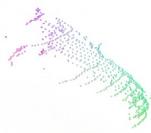
Experiments and analysis

Evaluations on KITTI

Image



Partial



FoldingNet [2]



TopNet [3]



PCN [1]



GRNet [4]



PoinTr [5]



Ours



Visual results of point cloud completion methods on the KITTI dataset.

- We conduct ablation experiments to study the **contributions** of the **2D modules** in our method.
- We generate a **variant** of our method **by removing** the 2D modules from the original network.



For quantitative evaluation, we compare the performance of the variant method with the original network on ShapeNet on the Chamfer Distance metric.

	Avg	Airplane	Cabinet	Car	Chair	Lamp	Sofa	Table	Watercraft
Variant	1.265	0.533	1.319	0.728	1.307	1.540	2.290	1.530	0.876
Ours	0.913	0.432	1.002	0.473	1.153	1.385	1.057	1.055	0.745



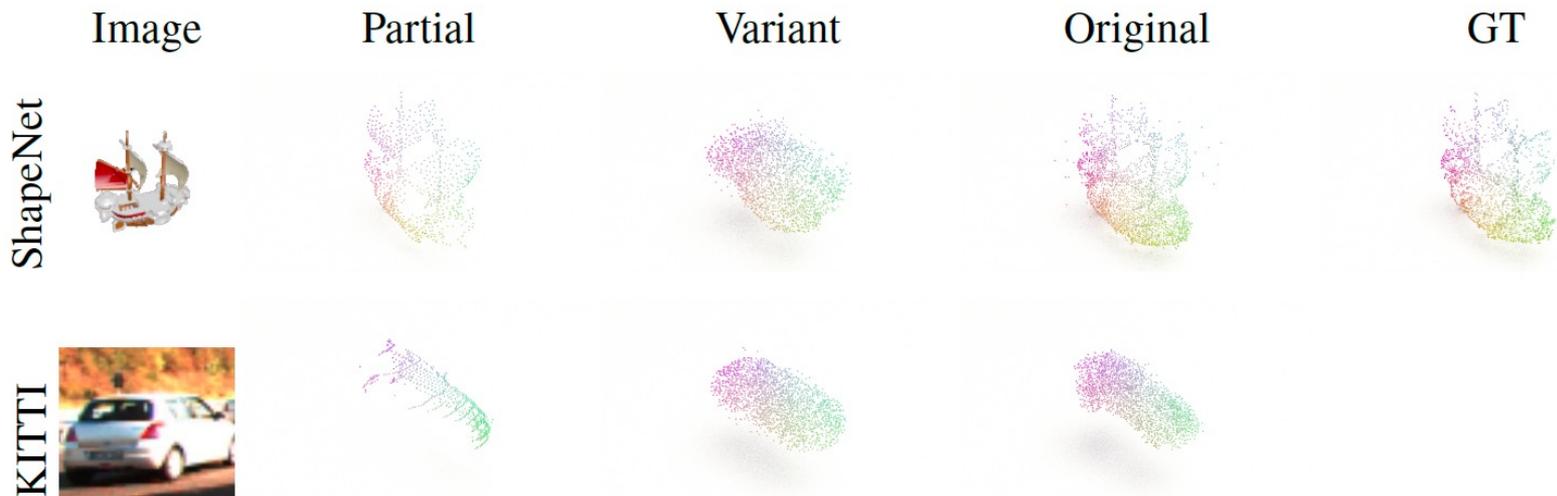
The performances in all categories **drop** when we remove the **2D modules**, which shows the importance of the **image guidance** to our network.

Experiments and analysis

Ablation studies



For qualitative evaluation, we visualize the results produced by the variant method on ShapeNet and KITTI datasets.



Contributions

- The proposed BCF-Net successfully **addresses** the **structural loss problem** in point cloud completion by leveraging both point-based and image-based representations.
- The multi-level skip-connection codeword enables the **preservation** of **non-missing regions** and the **estimation** of **local structural details**.
- Our BCF-Net **outperforms** existing SOTA methods on both **synthetic** and **real-world** datasets.

Future works

- A future research direction is to explore the use of **multiple 2D views** to further improve the performance of point cloud completion.

“We appreciate the opportunity to share our ideas
with you.”



- From authors