Shape Completion with Meso-Skeleton Learning

Dongyue Lu
Technical University of Munich
dongyue.lu@tum.de

Zhisheng Zheng
Technical University of Munich
Finn.zheng@tum.de

Yinyu Nie
Technical University of Munich
yinyu.nie@tum.de

Abstract

Point cloud completion is a significant topic in computer vision. Many deep-learning-based methods have been proposed to solve this problem directly based on an encoder-decoder structure. However, these architectures heavily rely on the representation ability of the encoded global feature. Some researchers try to leverage Meso-Skeleton to explicitly learn the global structure of objects. In this project we propose two architectures respectively backboned by PU-Net and PF-Net to study the effect of Meso-Skeleton in improving point cloud completion. Experiments show that a high-quality skeleton largely boosts the shape completion performance in both CD and EMD scores.

1. Introduction

Point cloud can capture rich 3D shape information, *e.g.* 3D point location, RGB/intensity and semantics. The most common method to obtain point cloud data is to use a 3D laser scanner. However, due to the limitation of view angles and occlusions of devices, the collected raw point sets are usually incomplete. Hence, a 3D point cloud completion method is necessary for the downstream applications.

Many deep learning approaches have been proposed to solve problems in point cloud. **PointNet** [1] is an inspiring and pioneering network, which proves that deep learning model can directly operate on point set data for several applications. Based on that, Charles *et al.* [2] propose an improved architecture **PointNet++** which applies hierarchical feature learning to obtain both local and global features and can achieve better results. Generally speaking, most existing point cloud completion methods try to recover objects by an encoder-decoder framework [3]. However, Nie *et al.* [4] propose **SK-PCN** to leverage a novel intermediate modality Meso-Skeleton to better capture the global features of original point cloud which outperforms the auto-encoder-based method. This method inspires us to leverage Meso-Skeleton learning to complete partial point cloud and study the role that skeleton plays in this task.

2. Methodology

We propose two different neural network architectures based on PU-Net and PF-Net to study the impact of Meso-skeleton learning in 3D point cloud completion.

2.1. PU-Net

PU-Net [5] inherits the hierarchical feature learning introduced by PointNet++ [2]. This model consists of three modules: point feature embedding, feature expansion and coordinate reconstruction. In the first module, the input point cloud is transformed to multi-level feature maps with different resolutions and then aggregated to one feature map. After that, the feature map is expanded by using several 1×1 convolution layers and fed into a MLP to generate the final 3D coordinate. Figure 1 shows the structure of PU-Net.

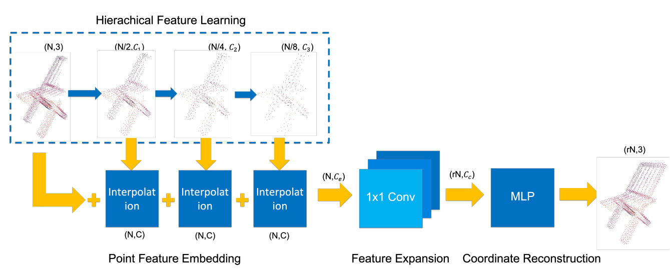

Figure 1. Brief Structure of PU-Net

2.2. PF-Net

PF-Net [6] is more like a traditional encoder-decoder-based framework. In the encoder part, a novel feature extractor CMLP transforms three input point clouds with different resolutions to three latent factors. Then, these latent

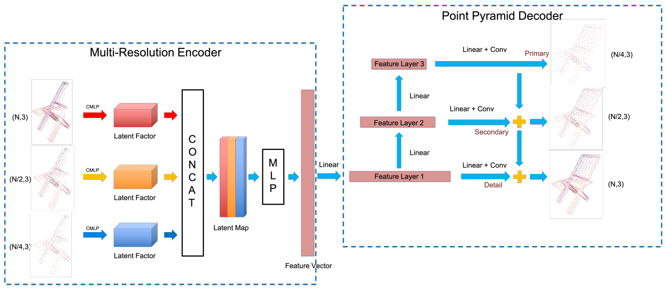

Figure 2. Brief Structure of PF-Net

factors are concatenated and fed into a MLP to generate the final feature vector. In the decoder part, three different feature layers are obtained by passing feature vectors/layers through different full-connected layers. Each feature layer is responsible for generating point cloud under one resolution. By combining these point clouds we can obtain our final prediction. Detailed workflow can be seen in Figure 2.

2.3. Point2Skeleton

Point2Skeleton [7] is the model we implement in this project to generate Meso-Skeleton for further study.

2.4. Proposed architecture

To further study the effect of Meso-Skeleton in 3D point cloud completion, we propose the following architecture. Our model consists of skeleton learning module and displacement learning module, and each module consists of a PU-Net/PF-Net framework. Partial point cloud is inputted into skeleton learning module supervised by the full point cloud skeleton to generate a skeleton with the same number of input points, which can learn the global feature of the complete point cloud. The same partial point cloud is inputted to the displacement learning module to predict the displacement of each point from the skeleton to the complete point cloud. Then we add the predictions of these two modules together to obtain the final result. Figure 3 demonstrates the architecture of our proposed model.

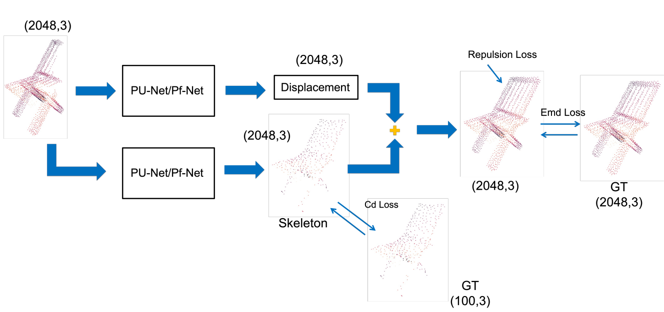

Figure 3. Proposed Architecture

2.5. Loss Function

The loss function of our model consists of three parts: Chamfer Distance loss for skeleton prediction, Earth Movers Distance loss for final prediction and Repulsion loss.

Chamfer Distance for Skeleton Prediction: Due to the limitation of device and the capability of Point2Skeleton, we can only obtain skeleton with less than 1200 points, which results in the inequality between generated skeleton points and ground truth skeleton, therefore we select Chamfer Distance to supervise skeleton generation:

$$\mathcal{L}_{\mathcal{CD}} = \frac{\sum_x \min y||x-y||_2}{|P|} + \frac{\sum_y \min x||y-x||_2}{|Q|} \quad (1)$$

Earth Movers Distance loss for final prediction: Earth Movers Distance is a classic approach to evaluate the similarity between two point clouds. We choose EMD to supervise our final prediction:

$$\mathcal{L}_{\mathcal{EMD}} = \min_{\phi:P\to Q} \sum_{x\in P} ||x-\phi(x)||_2 \quad (2)$$

Repulsion Loss for final prediction: Inspired by PU-Net, we apply repulsion loss [5] for the final prediction to improve the uniformity of distribution:

$$\mathcal{L}_{Repulsion} = \sum_{i=0}^{N} \sum_{i'(i)} \eta(||x_{i'} - x_i||_2)\omega(||x_{i'} - x_i||_2) \quad (3)$$

Altogether, we build an end-to-end model by minimizing the following joint loss function.

$$\mathcal{L} = \alpha\mathcal{L}_{\mathcal{CD}} + \beta\mathcal{L}_{EMD} + \gamma\mathcal{L}_{Repulsion} \quad (4)$$

where α,β and γ balance each sub-loss.

3. Experiment

3.1. Dataset

We use the method from PF-Net [6] to directly dig points on the original complete data from ShapeNet [8] to obtain the partial point cloud. We randomly select a point from the normalized complete point cloud and delete the points in a sphere with a certain radius centered on this point. Then we use iterative farthest point sampling (IFPS) to make the number of complete points and partial points the same as 2048. We randomly select 1600 samples and 400 samples in the chair category as training set and test set. In order to explore the influence of the number of skeleton points on the network, we input the complete point cloud into Point2Skeleton, and obtain four groups of different number of skeleton points: 100, 200, 400, and 1200.

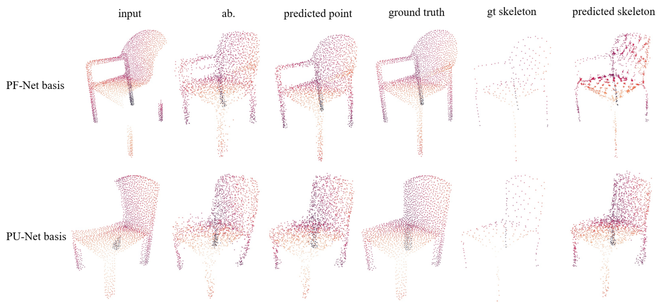

Figure 4. We compared the results of PU-Net-Basis and PF-Net-Basis, where 'ab.' represents the ablation study of skeleton.

3.2. Implementation details

We train the dataset with different number of skeletons on the PU-Net-Basis and PF-Net-Basis network. In order to explore the influence of skeleton on the network, we set different ratios of full point loss and skeleton loss in joint loss optimization. For PF-Net-Basis, the joint loss also includes full point loss and skeleton loss under different scales, so the weights of different losses in the two networks are not exactly the same.

The networks are based on Pytorch. We train 300 epochs using the Adam algorithm with a learning rate of 0.0005, and the training time is about 8 hours.

ratio	0	0.1	0.3	0.5	0.7	0.9
loss	2.950	**2.554**	2.920	3.201	3.225	3.598

Table 1. PU-Net-Basis: Results on different skeleton weights. We use EMD loss and show the loss under different weight in joint optimization, where "0" represents the ablation study for not using skeleton in completion. The results are multiplied by 1000 for readability and comparison convenience.

ratio	0	0.3	0.5	0.6	1	2
loss	1.685	**1.629**	1.636	1.640	1.691	1.742

Table 2. PF-Net-Basis: Results on different skeleton weights. The EMD loss is multiplied by 1000.

4. Results

The overall comparison result of PU-Net-Basis and PF-Net-Basis is shown in the figure 4. It can be seen that the result of PF-Net-Basis is better than PU-Net-Basis.

4.1. Different skeleton weights

For PU-Net-Basis, We ablate the ratio of skeleton loss in full point loss to 0.1, 0.3, 0.5, 0.7, 0.9 respectively. For PF-Net-Basis, we set to 0, 0.3, 0.5, 0.6, 1 and 2. The quantitative results are shown in the table 1 and 2. The quantitative results are shown in figure 5 and 8. It can be seen full

Point number	100	200	400	1200
PU-Net-Basis full point loss	2.554	**2.472**	2.684	3.236
PF-Net-Basis full point loss	1.629	**1.603**	1.650	1.655
PU-Net-Basis skeleton loss	0.803	**0.756**	0.890	1.260
PF-Net-Basis skeleton loss	0.325	**0.210**	0.364	0.388

Table 3. Results on different skeleton point numbers. CD loss is used to evaluate the quality of the predicted skeleton, which is multiplied by 100 for the convenience of comparison.

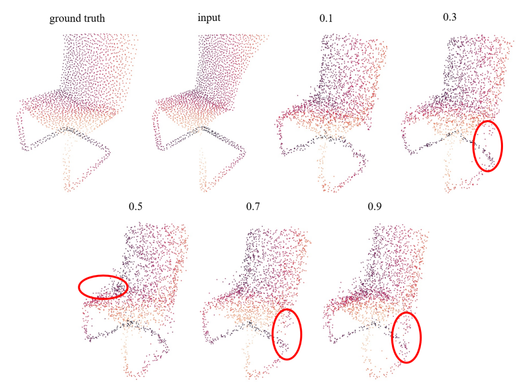

Figure 5. PU-Net-Basis experiment on different weights. The number on the sample indicates the respective weight in joint loss. There is more noise in larger weight cases.

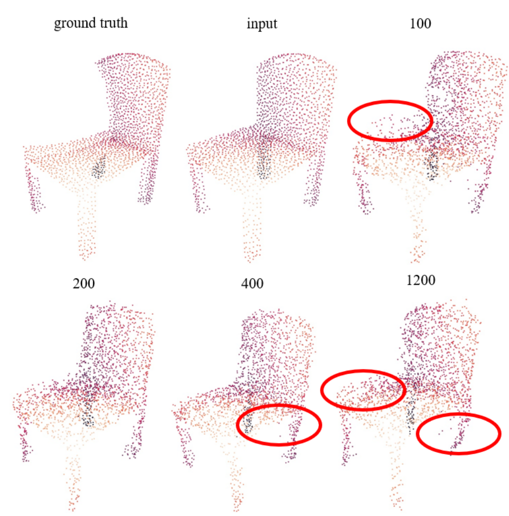

Figure 6. PU-Net-Basis experiment on different number of skeleton. The number on the sample indicates the number of skeleton used. Skeleton with 200 points achieves best results.

point loss decreases as the proportion of skeleton loss decreases. The explanation is that, first, in actual experiments, the original magnitude of the two losses is different. In order to balance these two in joint optimization, we increase the weight of full point loss. Secondly, it is obvious that the importance of full point loss in optimization increases as its weight increases, and the smaller the final loss will be.

For PU-Net-Basis, we use the best ratio: 0.1 to carry out the next step of the comparative test of different skeletons. For PF-Net-Basis, we choose 0.6 to take into account better skeleton loss and full point loss at the same time.

4.2. Different number of skeleton points

We explore the influence of different number of skeleton points on the final result. The quantitative result is shown in table 3 and quantitative results are in figure 6 and 9. It can be seen that the result of 100 and 200 cases are better than 400 and 1200 cases. This is contrary to our intuition that larger number of skeletons can represent richer features and improve the result. We think this may be related to the quality of skeleton generation. The quality of the skeleton generated by Point2Skeleton in the case of a small number of output skeletons is better than that of a large output number. As shown in figure 7, in the case of 1200 skeletons, we can easily observe that a large number of points gather together, which has a negative effect on the subsequent displacement learning. Skeleton with 200 points combines the advantages of good quality of skeleton ground truth and appropriate skeleton number at the same time, and the more evenly distributed skeletons effectively guide the optimization of the entire network. It is also expected that the predicted full point loss is synchronized with the change of skeleton loss, which can be seen by comparing the loss.

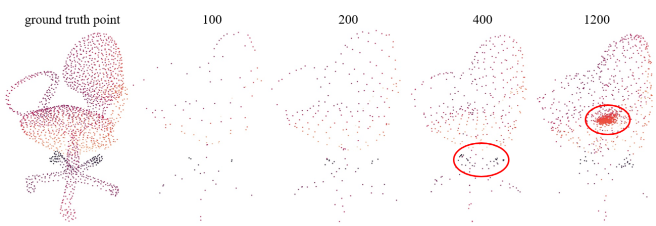

Figure 7. We show the comparison result of different number of skeleton points generated by Point2Skeleton. The cases of 400 and 1200 have more noise and clustering.

4.3. Analysis

In the ablation study for skeleton, we remove the skeleton from the dataset and input our network with only the partial point. It can be seen in table 1 and 2 that the full point loss is acceptable, but the effect is relatively poor, which proves that a high-quality skeleton can effectively improve the prediction results.

Failure cases are also analysed. There are some complex features which skeleton cannot effectively extract and represent. For example, samples which have features that vary with thickness, and skeleton tends to ignore these features and extract the sample into a plane. Due to the inherent limitations of skeleton and the limited application range of the skeleton generation module we use, the current model can only predict simple samples well, which can be improved in future work.

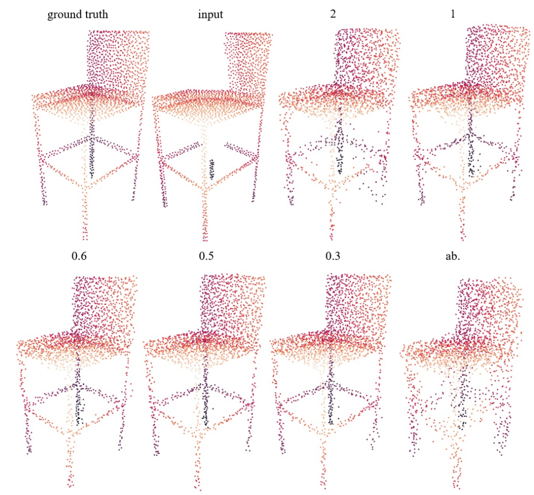

Figure 8. PF-Net-Basis experiment on different weights. The number above the sample represents the respective weight, and 'ab.' represents ablation study for not using skeleton.

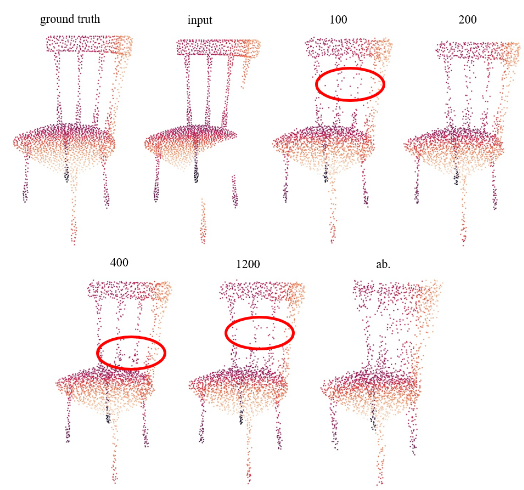

Figure 9. PF-Net-Basis experiment on different number of skeleton. The number above the sample represents the respective used skeleton, and 'ab.' represents ablation study. In the 100, 400 and 1200 cases, there is noise on the back of the chair.

5. Conclusion

Skeleton can effectively extract the feature of point cloud, which brings new idea for solving point cloud related problem. We hope to use skeleton as an intermediate medium to provide effective guidance for point cloud completion. Based on the basic framework of PU-Net and PF-Net, we build a point cloud completion model. The result shows that compared with direct prediction without skeleton, the features provided by high-quality skeleton can effectively improve the quality of the predicted point cloud. Our work can be further improved by improving the quality and applicability of skeleton generation, and adding more rigorous skeleton generation quality evaluation.

References

[1] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 1

[2] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*, 2017. 1

[3] Angela Dai, Charles Ruizhongtai Qi, and Matthias Nießner. Shape completion using 3d-encoder-predictor cnns and shape synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5868–5877, 2017. 1

[4] Yinyu Nie, Yiqun Lin, Xiaoguang Han, Shihui Guo, Jian Chang, Shuguang Cui, and Jian Jun Zhang. Skeleton-bridged point completion: From global inference to local adjustment. *arXiv preprint arXiv:2010.07428*, 2020. 1

[5] Lequan Yu, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. Pu-net: Point cloud upsampling network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2790–2799, 2018. 1, 2

[6] Zitian Huang, Yikuan Yu, Jiawen Xu, Feng Ni, and Xinyi Le. Pf-net: Point fractal network for 3d point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7662–7670, 2020. 1, 2

[7] Cheng Lin, Changjian Li, Yuan Liu, Nenglun Chen, Yi-King Choi, and Wenping Wang. Point2skeleton: Learning skeletal representations from point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4277–4286, 2021. 2

[8] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository, 2015. 2