# Skeleton-bridged Point Completion: From Global Inference to Local Adjustment

We thank all reviewers for the time and expertise they have invested in the comments. We really appreciate their recognition of our work on the method novelty, experiment performance and writing clarity. We sort out our responses (**R**) to the comments (**C**) as follows. Hope they can address reviewers' concerns.

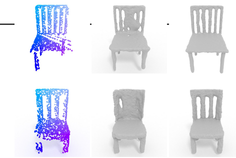

Figure 1: Comparing with coarse points (right: ours).

**C1.1:** *1) "Only using skeleton module (baseline of the ablation study) cannot beat MSN, etc." 2) "It is necessary to replace the skeleton generated with a coarse point cloud to see the difference."*

**R1.1:** Thanks for this advice. 1) It is because this baseline completes surface points only using predicted skeletons (see Sec 5.4). Details from input scans could be lost. We devise this baseline to instigate how much the other modules leverage the input to improve the results. 2) We made this ablation on 'chair' category (see Figure 1). The (CD$\downarrow$, Normal Cons.$\uparrow$) values are (2.96e-4, 0.81) and our (1.59e-4, 0.86). We think the reason could be that: coarse point cloud is still a type of surface points. While skeletal points keep compact topology of the shape without surface details. Using it as a bridge makes our method easier to recover complex structures. We will discuss it in detail.

**C1.2:** *"Please show the skeleton points result and the ground-truth and provide quantitative analysis."*

**R1.2:** Thanks for the suggestion. We output some skeleton results (2,048 points) in Figure 2 and will put them in the final version. The average CD and EMD values to the GT are 2.98e-4 and 1.44e-2. Our codes will also be released.

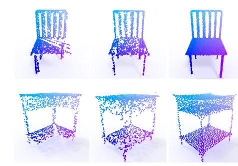

Figure 2: Skeleton results. l->r: input, prediction, GT

**C2.1:** *"Whether the baseline methods will improve to be comparable to this method if using the same losses? Also, does using normal loss means more supervision than the baseline methods?"*

**R2.1:** We keep their original loss because some methods (ONet, DMC) adopt implicit functions to represent shapes, which do not support point losses, and some methods use similar losses with us (PF-Net). In Sec 5.3 and supp. material, we have augmented the baseline P2PNet with our modules + repulsive and normal losses (P2PNet*) to see the difference with us. The normal loss indeed means an extra supervision but to supervise normal estimation. We augment P2PNet with our modules to (P2PNet+normal loss) and (P2PNet+normal&adversarial loss) on 'chair' category. Repulsive loss is added for each. The (CD$\times$e4$\downarrow$, EMD$\times$e2$\downarrow$) values are (2.94, 3.13), (2.98, 3.19), (2.76, 1.70) respectively, and ours are (2.55, 0.49). We cannot see improvements in CD/EMD involving the normal loss.

**C2.2:** *"Such a multi-stage pipeline may perform worse if the first-step prediction is of low quality. Is this true?..."*

**R2.2:** Indeed the skeleton quality would affect surface results, but since our network is trained jointly, subsequent modules can optimize the skeleton deviations and produce optimal results. We will put failure cases in the final version.

**C3.1:** *"In my opinion, the fact that this work learns skeleton from partial scan, compared to learning skeleton from other modality [31], does not form a strong contribution."*

**R3.1:** The meso-skeleton provides a topology-consistent shape abstraction that inspires us to learn surface completion bridged by skeletal points. As mentioned in other reviews, it is a novel attempt and achieves SOTA results. [31] also adopts skeletal points but for another task (single-view reconstruction). Their image input presents totally different modality with our sparse and irregular partial points. It thus requires us to tailor a unique method for partial-to-full point completion, which is inherently different from their image-skeleton-voxel reconstruction with an encoder-decoder network. Indeed we share an intuitive insight that skeleton can provide a global structure. For shape completion, however, the major bottleneck lies on inferring the complicated topology from irregular points, where the meso-skeleton presents a distinct advantage. The experiments also verified the worthiness of this first attempt. As mentioned by reviewer#4, we also believe this work can provide a new perspective of point completion guided by shape skeleton.

**C3.2:** *"It would be much better if a figure showing the whole pipeline is given. Fig 3, 4 are not easy to follow."*

**R3.2:** Thanks for this comment. The whole pipeline is illustrated in Figure 2 (in our paper), and the detailed layer information and data flow are demonstrated in the supplementary file. We will mention this in the paper to make it easier to follow. Figure 3 and 4 are explained in Section 3.2 and 3.3, which will be further detailed in the revised version.

**C3.3:** *"I would suggest to provide the performance of the model with each component removed..."*

**R3.3:** Sec 5.4 presents the ablation study on our main modules. We will further ablate the skeleton module as in R1.1.

**C3.4:** *"It is better to provide experiments on real dataset in order to understand the robustness against real data noise."*

**R3.4:** Thanks for the advice. We test our network on real scans in Figure 3. More results will be put in the final version.

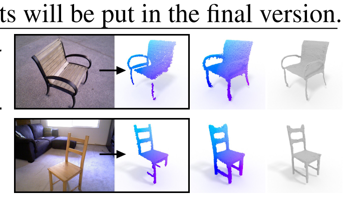

Figure 3: In-the-wild tests. l->r: input, points, mesh.

**C4.1:** *"Although paper claims the importance of preserve the geometry on the observable region, I do not see clear motivation using the adversarial training..."*

**R4.1:** Before considering the adversary loss, we actually refined the scan with observable points only using CD loss. However, we observed uneven point distribution on the observable area as discussed in PU-GAN. Thus we adopted the adversarial loss in PU-GAN to improve the visual quality on surfaces (see Section 3.3).

**C4.2:** *"More qualitative results of generated skeleton and discussions on its effects..."*

**R4.2:** Thanks for this suggestion. Here we list some results in Figure 2. For the page limit, we will analyze its effectiveness in the revised version. Our current skeleton generation is designed for supervised learning. We hope it can lay a foundation for the future work in unsupervised skeleton learning.