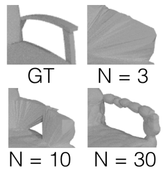

Figure 1: Reconstruction of a hole with varying # of primitives N.

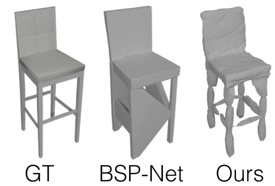

Figure 2: Reconstruction of a chair with holes with $N = 10$ primitives.

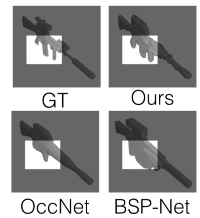

Figure 3: Reconstruction of a rifle.

We thank the reviewers for their thoughtful feedback. We are encouraged by the reviewers having identified our work to be novel (R2, R3), easy to follow (R2), well written (R2, R3), convincing in experiments (R2), and sufficient contribution to be an interesting work for the NeurIPS audience (R3). We are glad that they found our novel implicit-explicit primitive representation to be very exciting, technically interesting, and elegant (R2), contributing to more accurate shape reconstruction (R1, R2, R3) and improving surface reconstruction performance (R1, R2, R3). We address the reviewers' comments below.

R1 **Can the authors explain more clearly the problem setting?:** We believe our work is easy to understand overall, given that R2 and R3 evaluated our paper as "easy to follow" and "well written." We agree that there are some confusing points, which we will address in the camera ready. Given this, we would like to clarify the problem setting once again. The goal of this research is to learn a model that accurately reconstructs the target shape characterized by an indicator function O (L92) and a set of surface points P (L91), by predicting the corresponding approximation $\hat{O}$ and $\hat{P}$ (L96). Moreover, to understand the target shape structure, we reconstruct the shape by combining multiple semantic parts (primitives). To do so, we define a primitive characterized by an indicator function $\hat{O}_i$ and a surface point function $\hat{P}_i$ (Section 3.3). Note that $\hat{P}$ is a set, $\hat{P}_i$ is a function, and $\hat{P}_i(\mathbb{S}^2)$ is a set of surface points of a primitive. We also study how to combine $\hat{O}_i$ and $\hat{P}_i$ to represent $\hat{O}$ and $\hat{P}$ (Section 3.4). In the camera ready, we will address the confusing notation in Eq. 6 and L97, in which $\hat{P}$ takes arguments like a function, although it is a set.

R1 **How does the proposed method deal with complicated topologies?:** By increasing the number of primitives N, our model learns to handle complex topologies such as holes (see Figure 1). Note that even with a small number of primitives (parsimony is an essential criterion in the primitive based approaches), our approach can handle complicated topologies better than the leading method (BSP-Net), as shown in Figure 2. Although small holes are difficult to deal with, other high-frequency details such as small parts are reconstructed better by our approach. For example, in Figure 3, our method successfully reconstructs the rifle's three distinct handles while other methods fail. Our method works better because the explicit surface of NSD enables the optimization of shapes directly against the points sampled from the small parts, while implicit based methods tend to miss such small parts during sampling and training.

R3 **The paper mostly builds on existing ideas:** We agree that we strongly build our method in existing ideas, but we have developed on these ideas and made novel progress and several contributions. First, we propose a novel, **differentiable** implicit-explicit representation. BSP-Net realized the instant surface extraction during inference, but it needs a complex surface approximation scheme during training. We take a step further to realize the exact and differentiable surface extraction in a simple and novel manner (as R2 agrees), improving the reconstruction accuracy (Table 3 in the paper). Moreover, our proposed primitive representation is far more expressive than previous works. (see Figure 1 in the paper). Although previous works have gradually improved the primitives' expressivity, their low-dimensional parameter space still limits it. We propose NSD, whose expressivity is equivalent to a capacity of neural network (see supplementary Section B for proof), realizing far more expressive primitives. We believe these novel contributions make our work sufficient to be a good conference paper.

R2 **Too expressive primitive representation leads to less meaningful part decomposition:** We appreciate R2 for raising the concern around the critical question: how we should evaluate the quality of the decomposition result under self-supervised settings. Having the same concern, and following the previous works (BSP-Net, CVXNet), we evaluated our work based on the consistency with parts annotated by humans because we would like to know how meaningful the decomposition result is for **humans**. In Figure 5 in the paper, we show that the part decomposition of our method is semantically consistent with human annotations, comparable to the leading method in this task (BSP-Net).

R2 **Current composite indicator function $\hat{O}$ unfavorably encourages the overlapping primitives:** We appreciate R2's constructive suggestion; we also had the same concern. Actually, we considered $\mathrm{Sigmoid}(\sum_\mathrm{i} \mathrm{V_i})$. However, as the region of V_i includes both positive and negative domains, the summation can unfavorably cancel out each other terms. We tried ReLU instead of sigmoid for less overlap in Eq. 3, but we experimentally found sigmoid works only slightly better in terms of overlap by 6%. In the camera ready, we will report the overlap regularizer result.

R2 **Eq. 6 needs the double-checking:** We appreciate R2 again for pointing this out. we will fix the Eq. 6 in the camera ready as follows: $\hat{P} = \bigcup_i \{\hat{P}_i(\mathbf{d};\mathbf{t}_i) | \forall j \in [N \setminus i], \hat{O}_j(\hat{P}_i(\mathbf{d};\mathbf{t}_i);\mathbf{t}_i) < \tau_s, \mathbf{d} \in \{\mathbf{d}_k\}_{k=1}^K\}$.