Sign-Agnostic Implicit Learning of Surface Self-Similarities for Shape Modeling and Reconstruction from Raw Point Clouds
– *Supplemental Material* –

A. Least Squares Fitting of A Sphere to Observed Points

We introduce how to estimate the center $\bar{\boldsymbol{t}}$ and the radius $\bar{\boldsymbol{r}}$ given the observed points $\{\boldsymbol{p}\}$ using least squares, by solving the following problem. Its solution can be easily obtained in a closed form.

$$
\begin{aligned}
&(\bar{r}, \bar{\boldsymbol{t}}) = \arg\min_{\bar{r}', \bar{\boldsymbol{t}}'} \|\boldsymbol{A}\boldsymbol{b} - \boldsymbol{y}\|_2, \\
&\boldsymbol{y} = \begin{bmatrix} \|\boldsymbol{p}^1\|_2^2 \\ \vdots \\ \|\boldsymbol{p}^n\|_2^2 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} \bar{t}'_1 \\ \bar{t}'_2 \\ \bar{t}'_3 \\ \bar{r}'^2 - \|\bar{\boldsymbol{t}}'\|_2^2 \end{bmatrix}, \\
&\boldsymbol{A} = \begin{bmatrix} 2p_1^1 & 2p_2^1 & 2p_3^1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 2p_1^n & 2p_2^n & 2p_3^n & 1 \end{bmatrix}.
\end{aligned} \tag{i}
$$

Its closed-form solution is given by

$$
\begin{aligned}
&\bar{r} = \sqrt{b_4 + \|\boldsymbol{b}_{[1:3]}\|_2^2}, \\
&\bar{\boldsymbol{t}} = \boldsymbol{b}_{[1:3]},
\end{aligned} \tag{ii}
$$

where $\boldsymbol{b}$ can be computed by $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{y}$.

B. Minimum Spanning Tree for a Global Post-Optimization of Local Sign Flipping

I explain here how we have used minimum spanning tree (MST) in Section 4.3 for a global post-optimization of local sign flipping. Simply put, MST grows a set $\mathcal{V}_{\text{MST}}$ as a tree by selecting the vertices from $\mathcal{V}$. Starting from $\mathcal{V}_{\text{MST}} = \emptyset$, MST randomly selects a vertex from $\mathcal{V}$, denoted as v_1, and assigns $h(v_1) = 1$. The tree then grows iteratively by selecting from $\mathcal{V}/\mathcal{V}_{\text{MST}}$ the vertex that has the lowest edge weight in $\mathcal{W}$ to connect with any vertex in $\mathcal{V}_{\text{MST}}$, and ensures not to form a closed loop with those previous selected edges. In any iteration, denote the selected edge as $e_{i,j}$, $v_j \in \mathcal{V}/\mathcal{V}_{\text{MST}}$ is the selected vertex and v_i is the corresponding vertex already in $\mathcal{V}_{\text{MST}}$; we set $h(v_j) = h(v_i)$ when $w^1_{i,j}(e_{i,j}) < w^0_{i,j}(e_{i,j})$, and $h(v_j) = -h(v_i)$ otherwise. The tree is spanned until $|\mathcal{V}_{\text{MST}}| = |\mathcal{V}|$, and we have the signs $\{h(v_i)\}_{i=1}^N$ determined for all the vertices by then.

C. Proof of Corollary 4.1

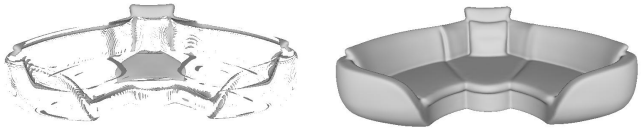
Corollary 4.1. *Let $f : \mathbb{R}^{3+d} \to \mathbb{R}$ be an l-layer MLP with ReLU activation ν. That is, $f(\boldsymbol{p}, \boldsymbol{z}) = \boldsymbol{w}^T\nu(\boldsymbol{W}^l(\cdots\nu(\boldsymbol{W}^1_{\boldsymbol{p}}\boldsymbol{p} + \boldsymbol{W}^1_{\boldsymbol{z}}\boldsymbol{z} + \boldsymbol{b}^1)) + \boldsymbol{b}^l) + c$, where $\boldsymbol{W}^1_{\boldsymbol{p}} \in \mathbb{R}^{d^1_{out}\times 3}$ and $\boldsymbol{W}^1_{\boldsymbol{z}} \in \mathbb{R}^{d^1_{out}\times d}$ denote the weight matrices of the first layer, and $\boldsymbol{b}^1 \in \mathbb{R}^{d^1_{out}}$ denotes the bias; $\boldsymbol{W}^i \in \mathbb{R}^{d^i_{out}\times d^{i-1}_{out}}$ and $\boldsymbol{b}^i \in \mathbb{R}^{d^i_{out}}$ denote parameters of the i^{th} layer; $\boldsymbol{w} \in \mathbb{R}^{d^l_{out}}$ and $c \in \mathbb{R}$ are parameters of the last layer; $\boldsymbol{p} \in \mathbb{R}^3$ is the input point, and $\boldsymbol{z} \in \mathbb{R}^d$ is the latent code, whose elements follow the i.i.d. normal $\mathcal{N}(0, \sigma_z^2)$. Let $\boldsymbol{w} = \sqrt{\frac{\pi}{d^l_{out}}}\boldsymbol{1}$, $c = -\bar{r}$, $\bar{r} > 0$, let all entries of $\boldsymbol{W}^i$ $(2 \le i \le l)$ follow i.i.d. normal $\mathcal{N}(0, \frac{2}{d^i_{out}})$, let entries of $\boldsymbol{W}^1_{\boldsymbol{p}}$ follow i.i.d. normal $\mathcal{N}(0, \frac{2}{d^1_{out}})$, and let $\boldsymbol{b}^i = \boldsymbol{0}$ $(2 \le i \le l)$. If $\boldsymbol{W}^1_{\boldsymbol{z}} = \boldsymbol{W}^1_{\boldsymbol{p}}[\boldsymbol{I} \in \mathbb{R}^{3\times 3}, \boldsymbol{0} \in \mathbb{R}^{3\times(d-3)}]$ and $\boldsymbol{b}^1 = -\boldsymbol{W}^1_{\boldsymbol{p}}\bar{\boldsymbol{t}}$, then $\lim_{\sigma_z\to 0} f(\boldsymbol{p}, \boldsymbol{z}) = \|\boldsymbol{p} - \bar{\boldsymbol{t}}\| - \bar{r}$. That is, f is approximately the signed distance function to a 3D sphere of radius $\bar{r}$ centered at $\bar{\boldsymbol{t}}$.*

Proof. To prove this theorem, we reduce the problem to a single hidden layer network. By plugging $\boldsymbol{W}^1_{\boldsymbol{z}}, \boldsymbol{b}^1$ in f we get $f(\boldsymbol{p}, \boldsymbol{z}) = \boldsymbol{w}^T\nu(\boldsymbol{W}^1_{\boldsymbol{p}}(\boldsymbol{p} + \boldsymbol{z}_{[1:3]} - \bar{\boldsymbol{t}})) + c$. Let $\boldsymbol{x} \in \mathbb{R}^3 = \boldsymbol{p} + \boldsymbol{z}_{[1:3]} - \bar{\boldsymbol{t}}$ and further plug $\boldsymbol{w}$, c in f, we get $f(\boldsymbol{x}) = \sqrt{\frac{\pi}{d^1_{\text{out}}}}\sum_{i=1}^{d^1_{\text{out}}}\nu(\boldsymbol{w}^1_{i,:}\cdot\boldsymbol{x}) - \bar{r}$, where $\boldsymbol{w}^1_{i,:}$ is the i^{th} row of $\boldsymbol{W}^1_{\boldsymbol{p}}$. Let μ denotes the density of multivariate normal distribution $\mathcal{N}(0, \frac{2}{d^1_{\text{out}}}\boldsymbol{I}_{d^1_{\text{out}}})$. The first term of $f(\boldsymbol{x})$ converges to $\|\boldsymbol{x}\|$, which is a direct consequence following Theorem 2 in [2]. In other words, $f(\boldsymbol{p}, \boldsymbol{z}) \approx \|\boldsymbol{p} - \bar{\boldsymbol{t}} + \boldsymbol{z}_{[1:3]}\| - \bar{r} \approx \|\boldsymbol{p} - \bar{\boldsymbol{t}}\| - \bar{r}$ when $\sigma_{\boldsymbol{z}} \to 0$. Note that to make the assumption of $\boldsymbol{b}^1$ and $\boldsymbol{W}^1_{\boldsymbol{z}}$ to be true, d^1_{out} should satisfy $d^1_{\text{out}} \ge 3$. □

D. Ablation Studies

Initialization for Signed Solutions – To evaluate the advantages of our proposed initialization for signed solution

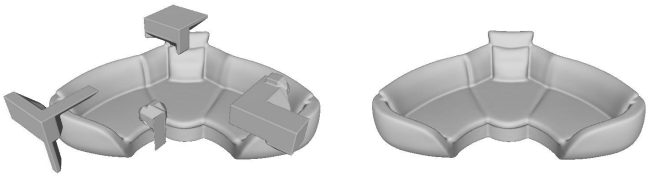
presented in Section 4.2.2, we conduct experiments that optimize SAIL-S3 with or without using the proposed initialization. Figure a shows that replacing our geometric initialization with random initialization will generate results of isolated surface stripes, which are extracted from the incorrect signed solutions of implicit fields.



W/O proper initialization SAIL-S3

Figure a: Without proper initialization, the learned local implicit function outputs $+/-$ boundaries only at isolated field regions (i.e. grey stripes in the left figure); note that only these boundaries can be extracted as surfaces via Marching Cubes [8].

Local Sign Flipping – Fig. b shows that switching off the local sign flipping presented in Section 4.3 would make the signs of different subfields inconsistent, ultimately resulting in surfaces with artifacts.



W/O proper sign flipping SAIL-S3

Figure b: Without proper sign flipping, the resulting mesh may have blocky defects.

Interpolation of Local Fields – The validity of the proposed interpolation of local subfields can be checked by substituting with interpolation by max pooling or average pooling, as shown in Fig. c.

Max Average SAIL-S3

Figure c: Example results of ablation study among interpolation by max pooling, interpolation by average pooling and our weighted interpolation.

E. Additional Qualitative Results

More qualitative results are given in Figure d, Figure e and Figure f. Our results are better than existing ones in terms of recovering smoother surfaces with more details.

F. Quantitative Results for Robustness Evaluation

Table 1 shows the quantitative results for our robustness evaluation presented in Section 5.2. The CD results confirm that our method is more robust against noisy observations.

Methods	CD@0.01 ↓	CD@0.005 ↓	CD@0.001 ↓
SPSR [7]	0.023	0.020	**0.003**
IGR [5]	0.035	0.018	0.008
SAL [2]	0.014	0.012	0.010
LIG [6]	0.013	0.008	0.007
CON [9]	0.016	0.014	0.011
P2S [4]	0.008	0.008	0.005
SAIL-S3	**0.004**	**0.003**	**0.003**

Table 1: Quantitative results for noisy point clouds of sculptures in ThreeDScans [1]. For CD, the smaller, the better. @0.01, @0.005, and @0.001 denote the levels (standard deviations) of Gaussian noise.

References

[1] Vienna Albertina, Vienna Kunsthistorisches Museum, Vienna Theater Museum, Paris Musée Guimet, Paris Musée des Monuments français, Cité de l'architecture et du patrimoine, Dépôt des sculptures de la Ville de Paris, Paris Musée Carnavalet, Lincoln The Collection, Lincoln Usher Gallery, Museo Archeologico Nazionale di Firenze, and Bergen KODE Artmuseums. Three d scans. https://threedscans.com. Accessed: 2020. 2, 3, 4

[2] Matan Atzmon and Yaron Lipman. Sal: Sign agnostic learning of shapes from raw data. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1, 2, 3, 4

[3] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 3

[4] Philipp Erler, Paul Guerrero, Stefan Ohrhallinger, Michael Wimmer, and Niloy J Mitra. Points2surf: Learning implicit surfaces from point cloud patches. *arXiv preprint arXiv:2007.10453*, 2020. 2, 3, 4

[5] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. *arXiv preprint arXiv:2002.10099*, 2020. 2, 3, 4

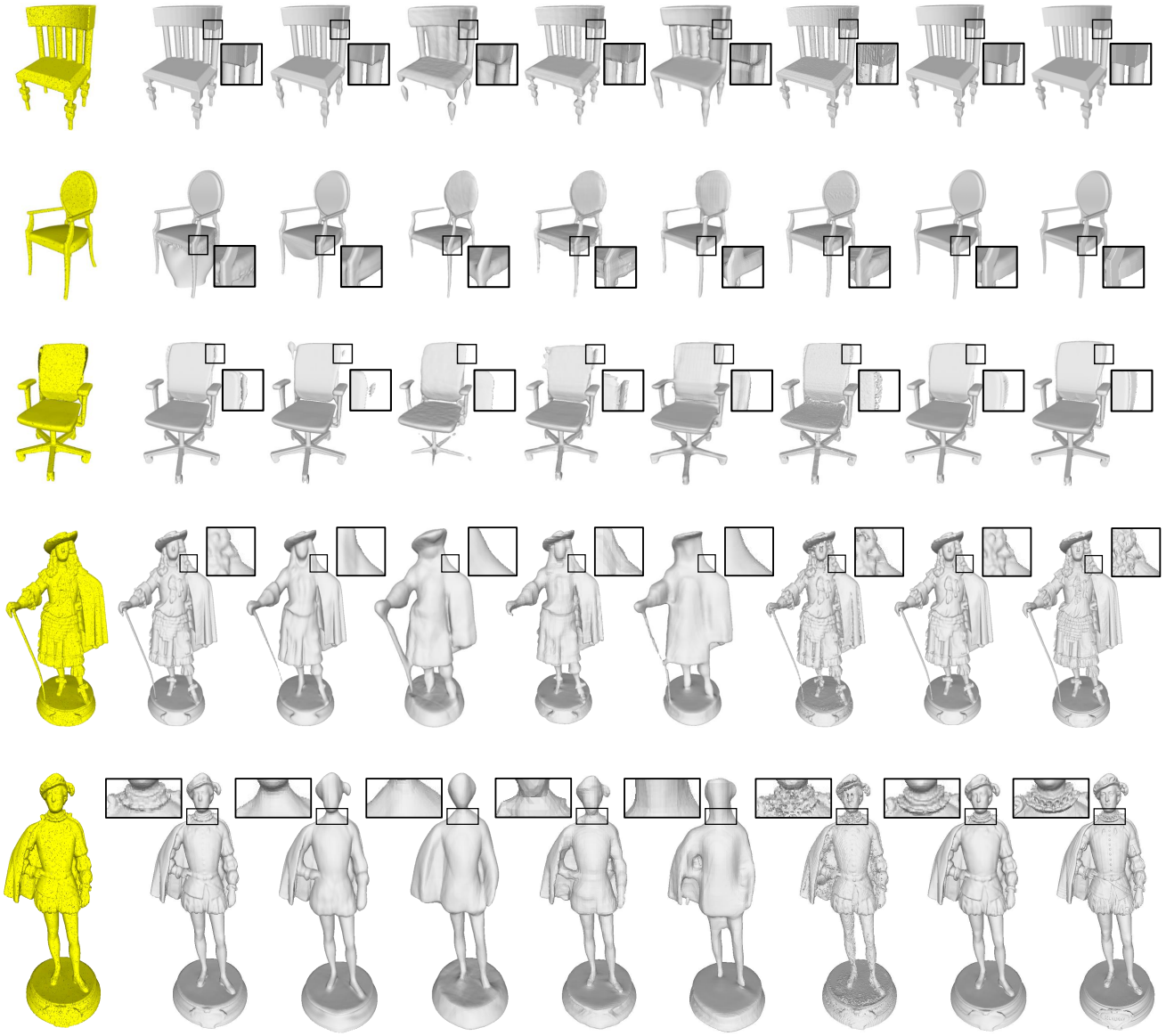


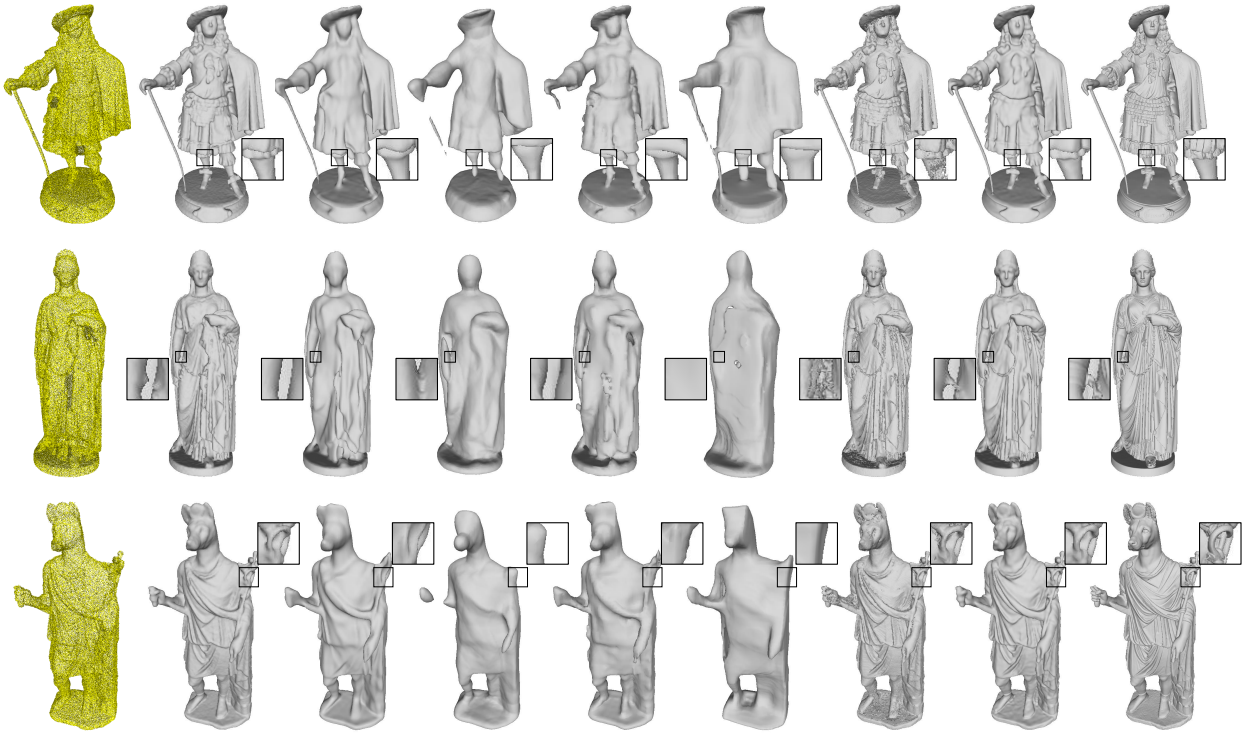
Figure d: Additional qualitative reconstruction results of ShapeNet [3] (top three rows) and ThreeDScans [1] (bottom two rows), black regions of the input point cloud denote incorrect normal orientation. Note that IGR [5] and SAL [2] belong to global fitting methods; while LIG [6], CON [9] and P2S [4] belong to the local learning methods. Lens are used to highlight the differences among the comparative methods.

[6] Chiyu Jiang, Avneesh Sud, Ameesh Makadia, Jingwei Huang, Matthias Nießner, and Thomas Funkhouser. Local implicit grid representations for 3d scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6001–6010, 2020. 2, 3, 4

[7] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013. 2, 3, 4

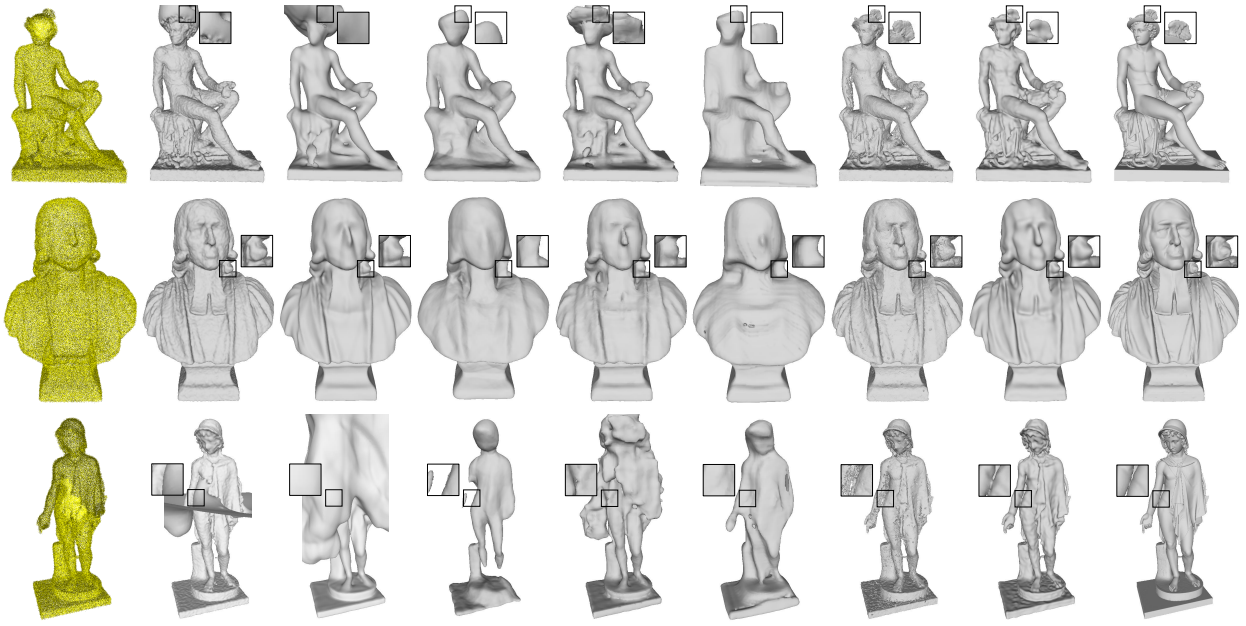
[8] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. *ACM siggraph computer graphics*, 21(4):163–169, 1987. 2

[9] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *European Conference on Computer Vision*, 2020. 2, 3, 4



Input PC SPSR [7] IGR [5] SAL [2] LIG [6] CON [9] P2S [4] SAIL-S3 GT

Figure e: Qualitative results of different methods when adding point-wise Gaussian noise of standard deviation 0.001 to input points of sculptures in ThreeDScans [1]. Black points on the three inputs denote incorrect estimations of normal orientations. Note that IGR and SAL belong to global fitting methods, and LIG, CON, and P2S belong to locally learned methods.



Input PC SPSR [7] IGR [5] SAL [2] LIG [6] CON [9] P2S [4] SAIL-S3 GT

Figure f: Qualitative results of different methods when adding point-wise Gaussian noise of standard deviation 0.005 to input points of sculptures in ThreeDScans [1]. Black points on the three inputs denote incorrect estimations of normal orientations. Note that IGR and SAL belong to global fitting methods, and LIG, CON, and P2S belong to locally learned methods.