

SSP: Semi-signed prioritized neural fitting for surface reconstruction from unoriented point clouds

Runsong Zhu^{1*} Di Kang² Ka-Hei Hui¹ Yue Qian² Shi Qiu¹

Zhen Dong^{3†} Linchao Bao² Pheng-Ann Heng¹ Chi-Wing Fu¹

¹The Chinese University of Hong Kong ²Tencent AI Lab ³Wuhan University

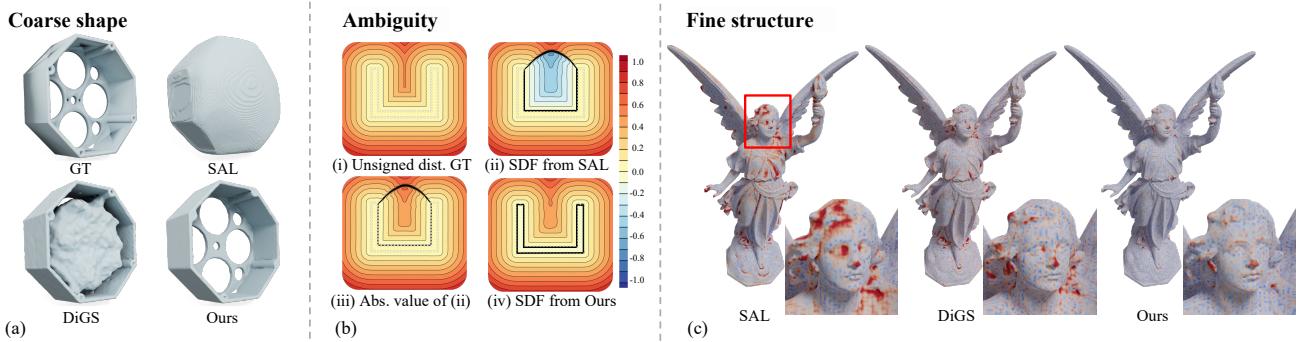


Figure 1. **(A):** Existing methods occasionally suffer from finding the coarse shape. **(B):** A 2D example to illustrate the “ghost” surface issue. Using GT unsigned distance field B-i as optimization objective, SAL [2] yields an SDF with undesired surfaces B-ii. However, its corresponding absolute-valued distance field B-iii is very close to its GT B-i, which means there is (almost) no loss to minimize. Our method does not suffer from this issue and produces a correct SDF B-iv due to the help of signed supervision in the “outside” region. **(C):** The error maps of our method and baselines [2, 14]. The red color denotes a larger one-way Chamfer distance. Thanks to the proposed loss-based per-region sampling and progressive position encoding, our method reconstructs more accurate surfaces for complicated regions.

Abstract

Reconstructing 3D geometry from unoriented point clouds can benefit many downstream tasks. Recent shape modeling methods mostly adopt implicit neural representation to fit a signed distance field (SDF) and optimize the network by unsigned supervision. However, these methods occasionally have difficulty in finding the coarse shape for complicated objects, especially suffering from the “ghost” surfaces (i.e., fake surfaces that should not exist). To guide the network quickly fit the coarse shape, we propose to utilize the signed supervision in regions that are obviously outside the object and can be easily determined, resulting in our semi-signed supervision. To better recover high-fidelity details, a novel loss-based region sampling strategy and a progressive positional encoding (PE) method are applied to prioritize the optimization towards underfitting and complicated regions. Specifically, we voxelize and partition the object space into sign-known and sign-uncertain regions, in which different supervisions are applied. Besides, we adaptively adjust the

sampling rate of each voxel according to the tracked reconstruction loss, so that the network can focus more on the complicated under-fitting regions. We conduct extensive experiments to demonstrate that our method achieves state-of-the-art performance compared to the existing fitting-based methods and comparable performance to learning-based methods on multiple datasets. The code is publicly available at <https://github.com/Runsong123/SSP>.

1. Introduction

Surface reconstruction from *unoriented* point clouds is a long-standing fundamental task for many downstream applications in computer vision, computer graphics, and AR/VR. However, due to the unstructured data format of point clouds, it remains challenging to reconstruct accurate surfaces for complicated topology-agnostic objects.

Among various approaches, implicit methods have gained increasing interest as they can reconstruct smooth and high-fidelity surfaces. Traditional implicit methods reconstruct surfaces by calculating global (e.g., RBF [10], SPSR [25]) or local (e.g., IMLS [40]) implicit functions. However, these methods suffer from cumbersome pre-processing (e.g., de-

*Work partially done during an internship at Tencent AI Lab.

†Corresponding author.

noising, upsampling, normal estimation, normal orientation, *etc.*), among which accurately oriented normals contribute a lot to high-quality reconstructions. Unfortunately, it is notoriously hard to compute orientation information from point clouds [7], thus limiting these methods’ applicability.

Recently, significant progress [1–3, 5, 14, 17, 37] has been made in directly optimizing an implicit function (*e.g.*, SDF) from unoriented point clouds. For example, [1–3, 5, 14, 17] explore unsigned supervision (due to lack of GT except the points) to optimize the neural network and demonstrate promising reconstruction results. If the optimization process goes well, the neural network will fit a coarse shape (*e.g.*, an overly-smoothed hull of the object) at the early stage and then recover the fine structures. Although substantial improvements have been achieved by existing methods [2, 3, 14, 17], there remain several challenges that prevent them from producing high-quality reconstructions.

First, most existing methods occasionally have difficulty in finding coarse shapes for complicated objects, leading to the production of “ghost” surfaces (*i.e.*, fake surfaces that should not exist) in undesired locations and large errors (see Fig. 1(a)). This implies that the existing unsigned supervision may not be able to provide sufficient guidance, so as the network occasionally gets stuck at bad local minimums and generates ghost structures. A 2D synthetic example is shown in Fig. 1(b). It satisfies the unsigned distance supervision [2] very well but a fake surface still appears, which means using only unsigned distance supervision is insufficient for certain cases. The second issue we observe is that most existing methods tend to reconstruct over-smoothing surfaces in complicated regions (*e.g.*, containing details in different levels) and ignore some fine structures (see, *e.g.*, Fig. 1(c)).

In this paper, we make a step toward overcoming those two problems. For the “ghost-surface” problem, we propose a simple yet effective solution by introducing an extra coarse signed supervision. The insight is that signed supervision is more informative and we can apply signed supervision to regions that are apparently “outside” the target object. Specifically, we propose a novel *semi-signed* fitting module, which simultaneously provides coarse signed supervisions and unsigned supervisions for different regions determined by our automatic *space partitioning* algorithm. With the additional signed guidance, the network can quickly fit a coarse shape to the given point cloud in the early stage, thus having higher chance in avoiding potential sub-optimal local minimums (within the computation budget). For the lack of details, we propose a new importance sampling strategy to increase the optimization efficiency and utilize progressive positional encoding to better recover the details. Specifically, we design a new loss-based region sampling (LRS) strategy that tracks the losses in the full 3D space and adaptively increases the sampling density of regions with larger losses. As the semi-signed fitting module can effectively avoid bad

local minimums, LRS helps the network focus on the details without suffering from severe ghost surface issues. Considering that MLP may have difficulty in fitting high-frequency signals [42], we explore progressive positional encoding (PE) to further improve the reconstruction details.

Overall, we propose semi-signed prioritized (SSP) neural fitting for more stable and accurate surface reconstruction from raw point clouds. To evaluate its effectiveness, we conduct extensive experiments. Compared to existing neural fitting methods, our SSP achieves state-of-the-art accuracy on multiple datasets, including the ABC subset [15] and various challenging data (*i.e.*, objects with complicated structures [44], objects with varying sampling density [18], and objects with sampling noise [18]). We also conduct several experiments to show that our proposed signed supervision can be adopted in existing methods (*e.g.*, IGR [17] and DiGS [6]) to boost their accuracy. In addition, we provide experiments to reveal the difficulty of optimization from unoriented point clouds.

Our contributions are summarized as follows:

- We propose a new semi-signed fitting module that provides additional signed supervision, which significantly alleviates the difficulty in finding coarse shapes for complicated objects.
- We introduce a loss-based per-region sampling and progressive PE, resulting in accurate surfaces with more details while generating fewer artifacts.
- We propose semi-signed prioritized (SSP) neural fitting, achieving improved performances compared to existing neural fitting methods on multiple datasets, especially with significant $CD-L_1$ reduction (*e.g.*, 20% improvement over previous state-of-the-art (*i.e.*, DiGS [6]) on ABC subset [15]).

2. Related Work

Reconstructing surfaces from unoriented point clouds is a long-standing problem. Here, we review the existing methods from traditional methods to learning-based and fitting-based neural methods.

Traditional methods. Early methods address the reconstruction task based on either handcrafted heuristics or numerical optimizations. In particular, some adopt heuristic guidance to progressively build and refine the reconstructed surface, such as growing triangulation [8, 38] and deforming an initial template mesh [27, 39]. Yet, these methods are sensitive to hyperparameters and initialization, requiring careful and time-consuming tuning on *each* point cloud. Some other methods [10, 23–25, 31, 40] assume that the point cloud comes with a consistently-aligned normal field. However, it is highly non-trivial to obtain accurate orientation information for point clouds [21, 37]. The final surface can be reconstructed by solving a Poisson equation [24, 25] or calculating a signed distance function using RBF [10], moving

least squares [40], Fourier coefficients [23], or wavelets [31].

Learning-based neural implicit methods. Recently, neural implicit function [12, 32, 35] has shown its superiority in representing 3D shapes. With the access to large shape data, such as ShapeNet [11] and Thingi10k [44], [4, 15, 16, 22, 28, 29, 32, 33, 35] propose to learn a data prior encoded in a neural implicit function for surface reconstruction from a point cloud. The reconstructed surface can be obtained by applying the trained model to a new point cloud. To enhance the generalization of the learned prior, [4, 16, 22, 29] propose to learn local priors. One major advantage of these learning-based methods (compared to fitting-based methods) is the fast inference speed. However, their accuracy may degrade severely if some characteristic (*e.g.*, shape, sampling density, total number of points, *etc.*) of test point cloud is obviously different from the training samples.

Fitting-based neural implicit methods. Instead of utilizing a data-driven prior, another stream of works attempts to leverage the neural network as a universal function approximator to solve an optimization for each point cloud input. These methods optimize *one network per object* to implicitly encode a signed distance field (SDF) whose zero-level set represents the reconstructed surface. Some recent works propose to utilize various unsigned supervision to optimize the SDF [2, 3, 14, 17], such as unsigned distance, unsigned normal, *etc.* Other methods [1, 5] project samples in the free space to a zero-level set and compute the distance metric with the input point cloud as the loss. Note that we do not classify Neural-Pull [5] as an unsigned method, since the surface cannot be represented as a zero-level set, which does not satisfy the basic property of SDF. Our work falls into this neural optimization category and extends previous methods with a novel semi-signed supervision, a novel loss-based region sampling strategy and progressive PE, so that more complex shapes could be better reconstructed.

3. Method

Given an *unoriented* point cloud \mathcal{P} , our objective is to optimize a network f_θ (parameterized by θ) to reconstruct the underlying surface represented with a signed distance field (SDF). Then, we can obtain the explicit surface S_θ by extracting the zero-level set from f_θ :

$$S_\theta = \{p \in \mathbb{R}^3 \mid f_\theta(p) = 0\}. \quad (1)$$

To obtain an accurate surface, *appropriate supervisions* are required to guide the optimization for the network f_θ . Specifically, the designed supervisions are applied either on the *on-surface* samples (denoted as $p \in \mathcal{P}$) or on the *off-surface* samples (denoted as $q \in \mathbb{R}^3 - \mathcal{P}$) during the optimization process. The on-surface losses are used to encourage faithful reconstruction on the sampled surface points, while the off-surface losses are used as regularization to suppress the

existence of degenerated structures.

3.1. Revisiting unsigned supervision

Prior to our improvements in Sec. 3.2-3.4, we first revisit existing supervisions (losses) and discuss their limitations. The existing supervisions can be categorized into on-surface and off-surface supervisions.

On-surface supervision. On-surface distance loss $\mathcal{L}_{\text{dist}}^{\text{on}}$ [17] encourages the extracted zero-level set to contain the existing points in given point clouds:

$$\mathcal{L}_{\text{dist}}^{\text{on}} = \sum_{p \in \mathcal{P}} \|f_\theta(p)\|. \quad (2)$$

The on-surface unoriented derivative loss $\mathcal{L}_{\text{grad}}^{\text{on}}$ [3] constrains $\nabla f_\theta(p)$ with the given unoriented surface normal n_p at p :

$$\mathcal{L}_{\text{grad}}^{\text{on}} = \sum_{p \in \mathcal{P}} \min\{\|\nabla f_\theta(p) - n_p\|, \|\nabla f_\theta(p) + n_p\|\}, \quad (3)$$

where $\nabla f_\theta(p)$ is the derivative of the network in point p .

Off-surface supervision. SAL [2] supervises the SDF predictions $f_\theta(q)$ with unsigned distance d , approximated with unsigned distance from q to its closest point in \mathcal{P} :

$$\mathcal{L}_{\text{dist}}^{\text{free}} = \sum_{q \in \mathbb{R}^3 \setminus \mathcal{P}} \min\{\|f_\theta(q) - d\|, \|f_\theta(q) + d\|\}. \quad (4)$$

Besides, IGR [17] utilizes the Eikonal regularization [13], which encourages SDF to maintain unit-length gradients in the whole space to produce a valid SDF.

$$\mathcal{L}_{\text{E}} = \sum_{q \in \mathbb{R}^3 \setminus \mathcal{P}} (\|\nabla f_\theta(q) - 1\|)^2. \quad (5)$$

Discussion. Although methods using the existing unsigned supervision demonstrate promising reconstructions from raw point clouds, we notice that their success is highly dependent on whether the sphere initialization [17] is a good “coarse shape” of the target point clouds. For example, the initial sphere is not a good approximation of the target point clouds when there exists a huge volume difference between them (even if the point cloud is enclosed in the sphere). In this case, most existing methods, *e.g.*, IGR [17], SAL [2], SALD [3] and DiGS [6], tend to fail for these shapes as shown in Fig. 1(a) and Fig. 5. More failed cases for current methods are provided in the Supp. This observation motivates us to determine a rough *outside* region and apply more informative signed supervision on it, so that the network can better avoid bad local minimums.

3.2. Semi-signed optimization module

To alleviate the difficulty in finding a coarse shape, we propose a simple yet effective solution that utilizes more informative signed supervision in regions apparently *outside* of the object. More concretely, the underlying surface of a point cloud is bounded, and the sign of regions outside

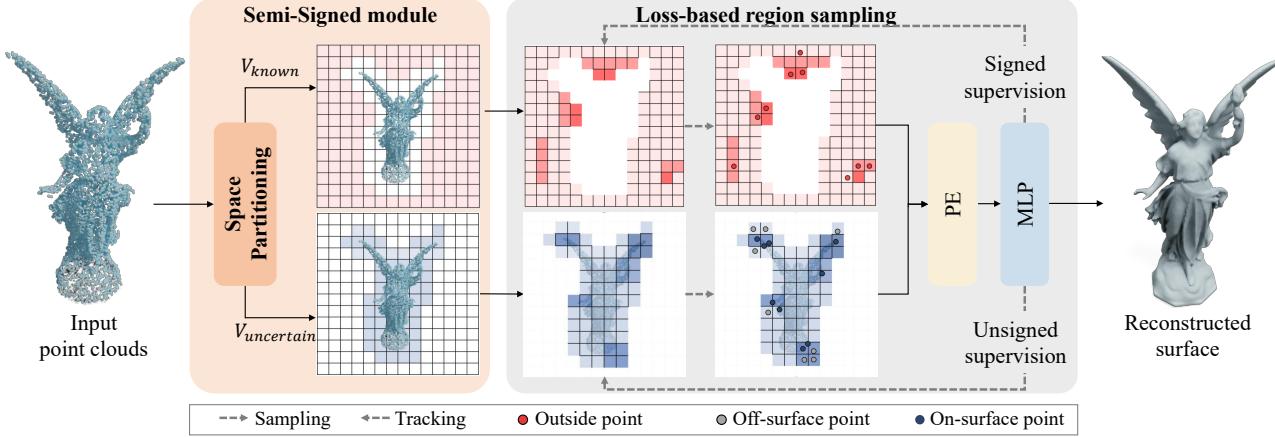


Figure 2. **Overview of our SSP method.** First, we partition the object space into “outside” (sign-known) regions $\mathcal{V}_{\text{known}}$ and “uncertain” (sign-uncertain) regions $\mathcal{V}_{\text{uncertain}}$, such that we can safely impose signed supervision on the outside regions. This simple signed supervision can effectively avoid reconstructing undesired ghost surfaces, complementing the existing unsigned supervisions (Sec. 3.2). Second, to reconstruct fine structures better, we propose a loss-based region sampling strategy (Sec. 3.3) to adaptively increase the sampling frequency in complicated regions with larger losses and utilize progressive positional encoding (Sec. 3.4) for fitting high-frequency signal.

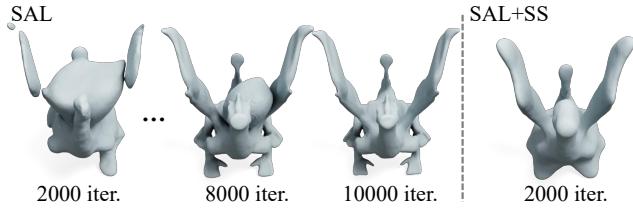


Figure 3. Visual comparison of the intermediate results optimized by SAL [2] and “SAL + SS”, where SS means signed supervision.

this bounding hull should be positive. We propose the *semi-signed optimization module*, that automatically finds outside regions and applies more informative signed supervision.

Space partitioning. To apply signed supervision, we first need to find the apparently “outside” region $\mathcal{V}_{\text{known}}$. The complement region of $\mathcal{V}_{\text{known}}$ is the *sign-uncertain* region, in which the signs of its points are not known beforehand. Note that this partition does not need to be precise to help the network quickly find a coarse shape during the optimization.

More concretely, the input point cloud is first normalized to a cube ranged $[-0.9, +0.9]^3$, following the common practice in [2, 3, 17]. Then, we voxelize the space $[-1.0, +1.0]^3$ into an N^3 grid, where N is the resolution calculated based on the density of input point clouds (see Supp. for the details). For the boundary voxels, it is easy to determine whether they belong to the *outside* region by simply testing if it contains any point in the given cloud points. Starting from the *outside* boundary voxels, we recursively use a breadth-first search (BFS) to find outside voxels connected to them. The search stops if a voxel or any of its neighboring voxels contains a point in the given point cloud

(i.e., approaching the neighboring of object boundary). The set of empty voxels forms $\mathcal{V}_{\text{known}}$ after the recursive search. The pseudo code for this procedure is provided in Supp.

Signed supervision in sign-known regions. We enforce the signs predicted by $f_\theta(q), q \in \mathcal{V}_{\text{known}}$ to be positive. Mathematically, we propose the following loss function:

$$\mathcal{L}_{\text{signed}} = \sum_{q \in \mathcal{V}_{\text{known}}} \tau(\epsilon - f_\theta(q))_+, \quad (6)$$

where $\epsilon = \frac{1}{N}$ is a positive margin distance and $\tau(x)_+ := \max(x, 0)$. According to Eq. (6), this loss only imposes the penalization when the predicted SDF value in $\mathcal{V}_{\text{known}}$ is smaller than ϵ . Although we do not use the exact signed distance as supervision, we observe that this signed supervision can help the network quickly learn a coarse shape of the target point cloud, compared with existing unsigned methods; see an example in Fig. 3.

Unsigned supervision in sign-uncertain region. For the $\mathcal{V}_{\text{uncertain}}$ region, only unsigned supervisions [2, 3, 17] can be used. Specifically, we adopt both distance loss $\mathcal{L}_{\text{dist}}$ (Eq. (2)) and derivative loss $\mathcal{L}_{\text{grad}}$ (Eq. (3)) for on-surface points, $\mathcal{L}_{\text{free}}$ (Eq. (4)), and Eikonal regularization \mathcal{L}_{E} (Eq. (5)) for off-surface points. To add more constraints on the $\mathcal{V}_{\text{uncertain}}$ region, we introduce an extra first-order guidance (i.e., unsigned derivative loss) for q as follows:

$$\mathcal{L}_{\text{grad}}^{\text{free}} = \sum_{q \in \mathcal{V}_{\text{uncertain}} \setminus \mathcal{P}} \min\{||\nabla f_\theta(q) - n_p||, ||\nabla f_\theta(q) + n_p||\}, \quad (7)$$

where $p \in \mathcal{P}$ is the nearest point to q . Here, we utilize the unoriented normal of p to approximate the derivative on its

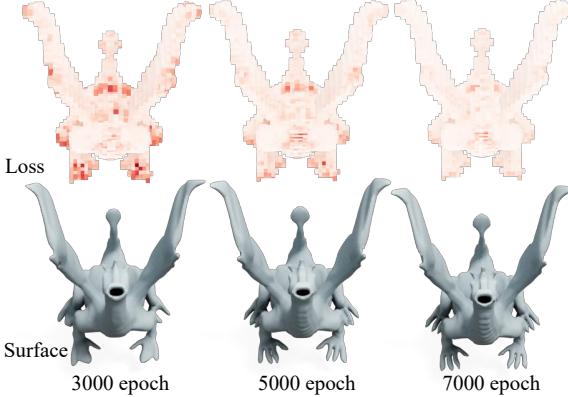


Figure 4. Visualizing the tracked moving average loss ($\mathcal{M}_{\text{grad}}^{j-\text{on}}$) in every voxel (top) and reconstructed surface (bottom) at different iterations. Deeper red colors indicate voxels of higher losses.

neighboring point q as a regularization. Please refer to Supp. for our motivation and more discussion.

Overall optimization objective. The overall loss is the weighted sum of the aforementioned loss terms:

$$\mathcal{L} = w_1 \mathcal{L}_{\text{dist}}^{\text{on}} + w_2 \mathcal{L}_{\text{dist}}^{\text{free}} + w_3 \mathcal{L}_{\text{grad}}^{\text{on}} + w_4 \mathcal{L}_{\text{grad}}^{\text{free}} + w_5 \mathcal{L}_{\text{E}} + w_6 \mathcal{L}_{\text{signed}}, \quad (8)$$

where $\{w_i\}$ are weights of each loss term.

3.3. Loss-based per-region sampling

We propose to sample more points for optimization, if a region has a large tracked loss. This strategy can be seen as a new variant of importance sampling and can bring two main benefits: (i) facilitates the reconstruction of fine details in complicated regions, as regions with more details are normally harder to fit; and (ii) helps avoid creating degenerated surfaces in free space when applied together with the signed supervision. Note that our sampling strategy is also used in the outside region. In practice, it has two main steps: region-wise loss tracking and adaptive sampling.

Region-wise loss tracking. We consider all five major loss terms: $\mathcal{L}_{\text{dist}}^{\text{on}}$, $\mathcal{L}_{\text{dist}}^{\text{free}}$, $\mathcal{L}_{\text{grad}}^{\text{on}}$, $\mathcal{L}_{\text{grad}}^{\text{free}}$, $\mathcal{L}_{\text{signed}}$. Taking $\mathcal{L}_{\text{dist}}^{\text{on}}$ as an example, we track the running mean loss of all applicable voxels \mathcal{V}^i (occupied voxels for $\mathcal{L}_{\text{dist}}^{\text{on}}$) as follows:

$$\mathcal{M}_{\text{dist}}^{i-\text{on}} = (1 - \alpha) \times \mathcal{M}_{\text{dist}}^{i-\text{on}} + \alpha \times \mathcal{L}_{\text{dist}}^{i-\text{on}}, \quad (9)$$

where $\mathcal{L}_{\text{dist}}^{i-\text{on}}$ is the loss of voxel i in the current iteration, $\mathcal{M}_{\text{dist}}^{i-\text{on}}$ is the tracked running mean loss, and α is the momentum empirically set to 0.1. Similarly, we can track the other losses denoted as $\mathcal{M}_{\text{grad}}^{i-\text{on}}$, $\mathcal{M}_{\text{dist}}^{i-\text{free}}$, $\mathcal{M}_{\text{grad}}^{i-\text{free}}$, and $\mathcal{M}_{\text{signed}}^i$. Note that different losses are applied to different regions (voxels). Please refer to Supp. for more details.

Adaptive sampling. Next, we perform a two-step sampling for each loss: (i) adaptively sample a set of voxels

based on the previous region losses, and (ii) sample points within each voxel. In the first step, we sample a set of voxels, where each voxel \mathcal{V}^i is assigned with a sampling probability $p_{\text{type}-i}^i$ proportional to its tracked moving average loss:

$$p_{\text{type}-i} = \frac{\mathcal{M}_{\text{type}-i}}{\sum_j \mathcal{M}_{\text{type}-j}}, \quad (10)$$

where $\mathcal{M}_{\text{type}-j} \in \{\mathcal{M}_{\text{dist}}^{j-\text{on}}, \mathcal{M}_{\text{grad}}^{j-\text{on}}, \mathcal{M}_{\text{dist}}^{j-\text{off}}, \mathcal{M}_{\text{grad}}^{j-\text{off}}, \mathcal{M}_{\text{signed}}^j\}$. Then, we apply different procedures to obtain the final point samples according to the loss type as follows.

For off-surface losses $\{\mathcal{L}_{\text{dist}}^{\text{off}}, \mathcal{L}_{\text{grad}}^{\text{off}}\}$ and the signed loss $\mathcal{L}_{\text{signed}}$, we use a uniform random sampling strategy to sample the points within the voxel. For the on-surface losses $\{\mathcal{L}_{\text{dist}}^{\text{on}}, \mathcal{L}_{\text{grad}}^{\text{on}}\}$, we query the k-nearest neighbor (k-NN) input points to the voxel center. Note that a small Gaussian noise is added to the center location before k-NN, so that different points can be sampled in different iterations. Fig. 4 shows a visualization of the sampling probabilities of the voxels. Particularly, our strategy tends to sample regions with complicated structures and fine details. The ablation study in Sec. 4.4 also shows that our proposed sampling strategy helps the network better fit the complicated regions. More details are provided in Supp.

3.4. Progressive Positional Encoding

Inspired by the recent works [19, 34, 36, 43], we also adopt a kind of progressive positional encoding (PE) to gradually introduce high-frequency components. Compared to regular positional encoding, the progressive strategy can better enhance the optimization quality by avoiding overfitting the noise in the early stage, especially when processing complicated structures and details. Specifically, the progressive positional encoding can be represented by

$$\gamma(\mathbf{p}) = (M_0 \sin(2^0 \pi \mathbf{p}), M_0 \cos(2^0 \pi \mathbf{p}), \dots, M_{L-1} \sin(2^{L-1} \pi \mathbf{p}), M_{L-1} \cos(2^{L-1} \pi \mathbf{p})), \quad (11)$$

where \mathbf{p} is a point coordinate, L is the frequency band, and $\mathbf{M} = \{M_0, \dots, M_{L-1}\}$ is a mask vector. We use the following progressive rule for M_i at iteration n :

$$M_i = \begin{cases} 0, & \text{if } i > L_0 + \frac{n}{K} \\ 1, & \text{otherwise,} \end{cases} \quad (12)$$

where $i \in \{0, 1, \dots, L-1\}$, K is a parameter controlling the increasing speed, and L_0 is the initial frequency band. We empirically set $L_0 = 3$, $L = 6$, and $K = 1000$.

4. Results

We conducted extensive experiments on the ABC subset (100 objects) [15], and three types of challenging data that contain objects with fine details [44] (5 objects), varying sampling density (32 objects) [18], and noises (57 objects) [18].

Methods	F-score \uparrow	CD- $L_1 \downarrow$ ($\times 100$)	NC \uparrow
SPSR [25]	0.557	2.774	0.904
SAP [37]	0.660	1.368	0.915
IMLS [29]	0.626	1.245	0.923
POCO [9]	0.670	1.148	0.943
N-P [5]	0.370	2.071	0.912
SAL [2]	0.407	4.676	0.870
SALD [3]	0.560	1.719	0.919
IGR [17]	0.551	4.429	0.891
DiGS [6]	0.657	1.540	0.936
Ours	0.675	<u>1.225</u>	<u>0.938</u>

Table 1. Quantitative comparisons on the ABC subset [15]. We **bold** the best and underline the second best results.

Methods	F-score \uparrow	CD- $L_1 \downarrow$ ($\times 100$)	NC \uparrow
SPSR [25]	0.787	2.230	0.896
SAP [37]	0.940	0.540	0.947
IMLS [29]	0.793	0.759	0.882
POCO [9]	0.902	0.610	0.939
N-P [5]	0.627	0.934	0.927
SAL [2]	0.884	0.779	0.925
SALD [3]	0.730	1.187	0.891
IGR [17]	0.308	6.471	0.631
DiGS [6]	<u>0.942</u>	<u>0.529</u>	<u>0.954</u>
Ours	0.943	0.520	0.960

Table 2. Quantitative comparisons on complex shapes in Thingi10K [44].

4.1. Experiment setting

Evaluation metrics. We report the L1-based Chamfer Distance (denoted as CD- L_1), Normal Consistency (NC), and F1-score with default threshold 1% following SAP [37] (F-score) between reconstructed surface and GT. Note that different papers (*e.g.*, NP [5], DiGS [6] and SAP [37]) use different evaluation protocols. For *fairer* comparisons, we use the same procedure as SAP [37] to calculate the metrics (details provided in Supp.).

Baselines. We compare with three types of methods: (i) non-neural implicit methods including Screened Poisson Surface Reconstruction (SPSR) [25] and SAP [37]; (ii) fitting-based neural implicit methods, including IGR [17], SAL [2], SALD [3], Neural-Pull [5], and DiSG [6]; and (iii) learning-based neural implicit methods, POCO [9] and IMLSNet [29]. The results of POCO and IMLSNet (denoted as IMLS) are calculated based on their officially-released pre-trained model. For the other baselines (IGR, SAL, Neural-Pull (denoted as N-P) and SAP), we report the results optimized by their officially-released code. The normal inputs of our method, SALD, and SPSR are estimated from the state-of-the-art method AdaFit [46]. Since SPSR requires consistently oriented normals, we use minimum spanning tree [45] to propagate the normal orientations [20].

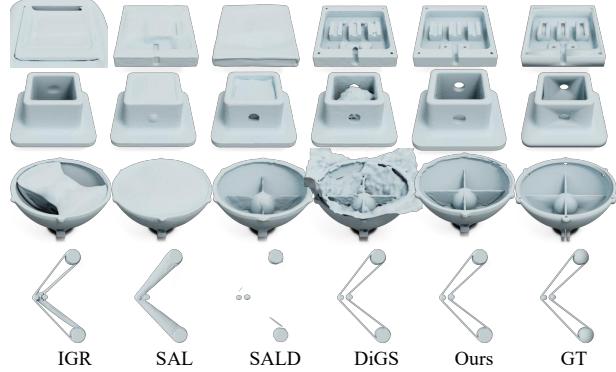


Figure 5. Visual comparisons on the ABC subset [15].

Implementation details. We use the same neural architecture and the same number of sampling points as IGR [17]. Following [2, 3, 5, 17], we overfit the network for every input point cloud and set the full iteration as 10,000 for the whole experiments. We adopt the Marching cubes [30] algorithm to extract the mesh from its implicit field, and the default resolution is 256^3 for all the neural implicit methods. All the experiments are conducted on the GeForce RTX 2080 Ti GPU. Details and architectures can be found in Supp.

4.2. Result on ABC dataset

To evaluate the stability of our method, we compare our method in a subset of ABC [26] (noise-free version) released by Points2surf [15]. This subset contains 100 3D CAD models with various topologies and we use the provided sampled point clouds as input. Quantitative results of different methods are shown in Tab. 1. Compared with existing fitting-based methods, our method achieves the best results on all the metrics (*i.e.*, F-score, CD- L_1 , NC) and achieves comparable performance with SOTA learning-based method POCO [9] trained on ABC dataset. As shown in Fig. 5, our method could significantly alleviate the ghost surface problem and reconstruct more accurate surfaces. In contrast, most other fitting-based methods (IGR [17], SAL [2], SALD [3] and DiGS [6]) occasionally create ghost surfaces in undesired locations, thereby severely worsening the CD- L_1 value. More visual comparisons can be found in Supp.

4.3. Result on challenging data

Data with high-level details We further conduct experiments on Thingi10k [44]. Similar to prior work [37], we use the same five challenging shapes with complex topology and high-level fine details. For SAP [37] and SPSR [24] results in Tab. 2, we use the numbers reported in SAP [37]. As shown in Tab. 2, we achieve the best performance on all metrics on Thingi10k [44]. As shown in Fig. 6, our method could preserve more details compared with existing methods while having fewer artifacts. Although DiGS [6] uses the SIRENs [41] neural structure for high-frequency representa-

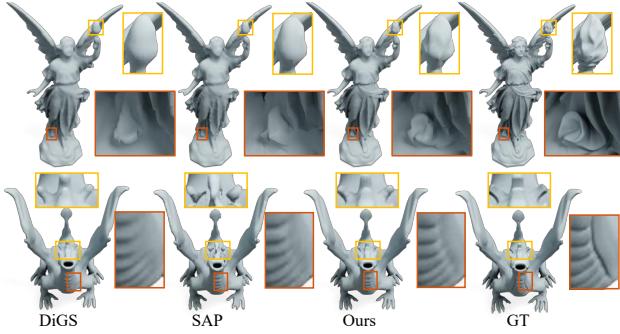


Figure 6. Visual comparisons on Thingi10K [44].

Methods	Density-variation			Noise		
	F-score \uparrow	CD- L_1 ↓ ($\times 100$)	NC \uparrow	F-score \uparrow	CD- L_1 ↓ ($\times 100$)	NC \uparrow
SPSR [25]	0.789	2.007	0.938	0.723	2.216	0.833
SAP [37]	0.889	<u>0.658</u>	0.932	0.580	<u>1.128</u>	0.693
IMLS [29]	0.830	0.715	0.925	0.583	1.205	0.879
POCO [9]	0.867	0.845	0.943	0.510	1.721	0.911
N-P [5]	0.397	1.359	0.945	0.257	2.027	0.901
SAL [2]	0.767	1.823	0.937	0.328	8.467	0.880
SALD [3]	0.724	1.209	0.926	0.255	3.472	0.919
IGR [17]	0.714	7.316	0.918	<u>0.697</u>	3.480	0.889
DiGS [6]	0.877	0.868	<u>0.951</u>	0.544	1.273	0.717
Ours	0.917	0.567	0.962	0.685	0.994	0.957

Table 3. Quantitative comparisons on the data with different levels of noise and different density variation on PCPNet [18].

DS	SS	LRS	PE	F-score \uparrow	CD- L_1 ↓ ($\times 100$)	NC \uparrow
✓	✓	✓	✓	0.943	0.520	0.960
✗	✓	✓	✓	0.940 (-0.004)	0.524 (+0.004)	0.959 (-0.001)
✓	✗	✓	✓	0.881 (-0.062)	0.626 (+0.106)	0.952 (-0.008)
✓	✓	✗	✓	0.883 (-0.060)	0.831 (+0.311)	0.947 (-0.013)
✓	✓	✓	✗	0.877 (-0.066)	0.808 (+0.288)	0.912 (-0.048)

Table 4. Ablation study conducted on the Thingi10K [44] dataset. SS, DS, LRS, PE represent “Signed supervision”, “Derivative supervision”, “Loss-based per-region sampling”, and “Positional encoding”, respectively.

tion, it still has difficulty in fitting the fine structures, since the points in these regions have not been adaptively sampled as many times as ours due to its uniform sampling strategy. See the abdomen of the “dinosaur” shown in Fig. 6. Though SAP [37] could successfully reconstruct an accurate surface, it occasionally creates artifacts for the complicated region.

Varying-density data To evaluate the robustness of our method to varying sampling density, we conduct experiments on the varying-density dataset released by PCPNet [18]. As shown in Tab. 3, our method performs the best on all three metrics. As shown Fig. 7, our method produces more complete surfaces compared with the baseline methods. Note that POCO [9] has difficulty in obtaining the accurate surface

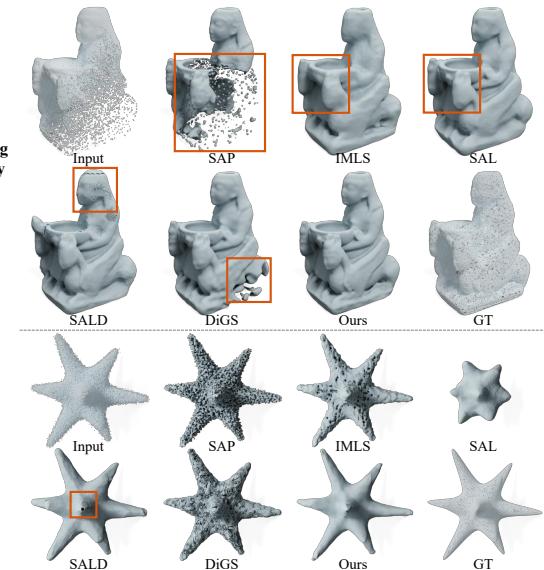


Figure 7. Visual comparisons on varying-density data from [18].

since the test shapes’ sampling density is not similar to their training-set shapes. More results can be found in Supp.

Noisy data To test the robustness of our method to noise, we conduct experiments on the noisy test data from PCPNet [18]. This dataset includes 57 point cloud shapes of different levels of Gaussian noise. Our method significantly outperforms others on the CD- L_1 and NC metrics; see Tab. 3. The visual comparisons in Fig. 7 also show that our method reconstructs more accurate and smooth surfaces for the noisy point clouds. More results can be found in Supp.

4.4. Ablation study

Signed-supervision & derivative supervision. We conduct experiments to show the effectiveness of signed-supervision (SS) and derivative supervision (DS) in Tab. 4 (1st & 2nd rows). For the ablation on SS, we compare the result optimized by our proposed SS with the result by the existing unsigned supervision on the full space. As expected, removing either SS or DS leads to inaccurate fitting results.

Loss-based per-region sampling & positional encoding. We conduct ablation on loss-based per-region sampling (LRS) and positional encoding (PE). For LRS, we compare the result optimized by our LRS with the result optimized by the traditional Gaussian sampling (GS) strategy, which is widely used in the existing methods [2, 3, 5, 17]. As shown in Tab. 4 (3rd), removing LRS in our method results in significant performance drop. Since the sampled points by GS are mostly near to the surface, this may lead to inefficient sampling in the outside region for adding signed supervision. Besides, removing PE decreases the overall performance as shown in Tab. 4 (4th).

Method	Signed Super.	F-score \uparrow	CD- $L_1 \downarrow (\times 100)$	NC \uparrow
IGR [17]		0.308	6.471	0.631
IGR [17]	✓	0.636	1.676	0.729

Table 5. Quantitative comparisons of results optimized by IGR and IGR equipped with our signed supervision on Thingi10k [44].

Method	Signed Super.	F-score \uparrow	CD- $L_1 \downarrow (\times 100)$	NC \uparrow
DiGS [6]		0.411	4.018	0.881
DiGS [6]	✓	0.500	2.218	0.920

Table 6. Quantitative comparison of results optimized by DiGS and DiGS equipped with our signed supervision on the ABC [26] hard case subset.

Method	Norm. Info.	F-score \uparrow	CD- $L_1 \downarrow (\times 100)$	NC \uparrow
IGR	<u>Oriented</u> (GT)	0.917	0.785	0.94
IGR	None	0.308	6.471	0.631
IGR	<u>Oriented</u> (Calc.)	0.748	2.889	0.933
IGR	<u>Unoriented</u>	0.584	4.353	0.858
Ours	<u>Unoriented</u>	0.943	0.520	0.960

Table 7. Quantitative comparison of results optimized from oriented and unoriented point clouds. "Calc." means "calculated".

4.5. Further analysis

Is the proposed signed supervision applicable to other SOTA methods? To answer this question, we verify the effectiveness of our signed supervision on two baselines, IGR [17] and DiGS [14]. For IGR, we conduct experiments on Thingi10k [44], since we observe that IGR had difficulty in fitting coarse shapes for several objects in Thingi10k. For DiGS, we choose the top 10 hard cases in the ABC subset [15] based on the CD- L_1 values obtained experimentally. We compare the results optimized with an extra term, *i.e.*, our signed supervision, to the results from the original baselines (IGR and DiGS). As shown in Tab. 5 and Tab. 6, our signed supervision can significantly improve the performance of both baseline methods (especially on the CD- L_1 metric).

Difficulty of optimization from unoriented point clouds. We conduct experiments to explore the difficulty of optimization from unoriented point clouds in Tab. 7. IGR [17] reconstructs relatively accurate shapes, if GT oriented normals are given (1st row in Tab. 7). However, the accurate oriented normals are usually difficult to obtain. When there is no normal supervision, the performance decreases significantly (2nd row in Tab. 7). One solution to this problem is that we can adopt traditional methods (*e.g.*, the minimum spanning tree [45]) to estimate the normals' orientations. However, the normal orientation task is notoriously difficult in regions with complicated structures, and the inaccurate normal orientations limit the subsequent reconstruction quality (3rd row in Tab. 7). Another possibility is that we only use the unsigned derivative supervision based on unoriented

Method	Ours	DiGS [6]	IGR [17]	SAL [2]	SALD [3]	NP [5]	
Opti. Time	Prep. ~1	10K ~15	10K ~15	20K ~27	20K ~20	20K ~22	40K ~15

Table 8. The optimization epochs & approximate time (in minutes) for optimization-based neural methods. "Opti." means "Optimization" and "Prep." means "Preprocessing".

σ	F-score \uparrow			CD- $L_1 \downarrow (\times 100)$			NC \uparrow		
	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
	0.939	0.943	0.941	0.532	0.520	0.538	0.959	0.960	0.960

Table 9. Thingi10K results using 3 voxel sizes (scaled by σ).

normals for the optimization. However, optimization from unsigned supervision occasionally suffers from the difficulty in finding coarse shapes, leading to large fitting errors for complex shapes, as shown in Tab. 7. In contrast, our SSP achieves a significant improvement with solely unoriented normal information.

Optimization timing comparison. We list the optimization epochs and average time for the optimization-based methods in Tab. 8, our method takes the similar time with DiGS [6] and converges faster than other optimization-based baselines (*e.g.*, IGR [17], SAL [2], *etc.*).

Influence of voxel size. We conduct experiments using different voxel sizes to study the influence of different voxel sizes in Tab. 9. The results are stable within a reasonable range around the voxel size determined by our automatic algorithm (details in Supp).

5. Conclusion and Limitation

We present a novel semi-signed prioritized fitting-based method (SSP) for neural surface reconstruction from unoriented point clouds. First, we propose to utilize the extra coarse signed supervision to help the neural network quickly learn coarse shapes for target (complicated) objects. To obtain more details, we propose a loss-based region importance sampling strategy and progressive PE to prioritize the optimization. SSP method has achieved improved performance on multiple datasets. One limitation of our method is that it can not handle the open surfaces or the scene-level input data with large missing parts and we leave this as future work.

6. Acknowledgement

This work was supported by the InnoHK of the Government of the Hong Kong Special Administrative Region via the Hong Kong Centre for Logistics Robotics and RGC TRS project (T45-401/22-N). We thank Yuan Liu for his constructive comments.

References

[1] Matan Atzmon, Niv Haim, Lior Yariv, Ofer Israelov, Haggai Maron, and Yaron Lipman. Controlling neural level sets. In *NeurIPS*, 2019. 2, 3

[2] Matan Atzmon and Yaron Lipman. SAL: Sign agnostic learning of shapes from raw data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2565–2574, 2020. 1, 2, 3, 4, 6, 7, 8

[3] Matan Atzmon and Yaron Lipman. Sald: Sign agnostic learning with derivatives. In *International Conference on Learning Representations*, 2020. 2, 3, 4, 6, 7, 8

[4] Abhishek Badki, Orazio Gallo, Jan Kautz, and Pradeep Sen. Meshlet priors for 3D mesh reconstruction. In *CVPR*, 2020. 3

[5] Ma Baorui, Han Zhizhong, Liu Yu-shen, and Zwicker Matthias. Neural-Pull: Learning signed distance functions from point clouds by learning to pull space onto surfaces. 2021. 2, 3, 6, 7, 8

[6] Yizhak Ben-Shabat, Chamin Hewa Koneputugodage, and Stephen Gould. DiGS: Divergence guided shape implicit neural representation for unoriented point clouds. In *CVPR*, 2022. 2, 3, 6, 7, 8

[7] Matthew Berger, Andrea Tagliasacchi, Lee M. Seversky, Pierre Alliez, Gael Guennebaud, Joshua A Levine, Andrei Sharf, and Claudio T. Silva. A survey of surface reconstruction from point clouds. In *Comput. Graph. Forum*, 2017. 2

[8] Fausto Bernardini, Joshua Mittleman, Holly Rushmeier, Cláudio Silva, and Gabriel Taubin. The ball-pivoting algorithm for surface reconstruction. *IEEE TVCG*, 1999. 2

[9] Alexandre Boulch and Renaud Marlet. Poco: Point convolution for surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6302–6314, 2022. 6, 7

[10] Jonathan C. Carr, Richard K. Beatson, Jon B. Cherrie, Tim J. Mitchell, W. Richard Fright, Bruce C. McCallum, and Tim R. Evans. Reconstruction and representation of 3D objects with radial basis functions. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001. 1, 2

[11] Angel X. Chang, Thomas A. Funkhouser, Leonidas J. 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. *arXiv preprint arXiv:1512.03012*, 2015. 3

[12] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *CVPR*, 2019. 3

[13] Michael G. Crandall and Pierre-Louis Lions. Viscosity solutions of Hamilton-Jacobi equations. *Transactions of the American mathematical society*, 1983. 3

[14] Angela Dai, Christian Diller, and Matthias Nießner. SG-NN: Sparse generative neural networks for self-supervised scene completion of RGB-D scans. In *CVPR*, 2020. 1, 2, 3, 8

[15] Philipp Erler, Paul Guerrero, Stefan Ohrhallinger, Niloy J. Mitra, and Michael Wimmer. Points2Surf: Learning implicit surfaces from point clouds. In *ECCV*, 2020. 2, 3, 5, 6, 8

[16] Kyle Genova, Forrester Cole, Avneesh Sud, Aaron Sarna, and Thomas Funkhouser. Local deep implicit functions for 3D shape. In *CVPR*, 2020. 3

[17] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. 2020. 2, 3, 4, 6, 7, 8

[18] Paul Guerrero, Yanir Kleiman, Maks Ovsjanikov, and Niloy J. Mitra. PCPNet: Learning local shape properties from raw point clouds. In *Comput. Graph. Forum*. Wiley Online Library, 2018. 2, 5, 7

[19] Amir Hertz, Or Perel, Raja Giryes, Olga Sorkine-Hornung, and Daniel Cohen-Or. SAPE: Spatially-adaptive progressive encoding for neural optimization. In *NeurIPS*, 2021. 5

[20] Hugues Hoppe, Tony DeRose, Tom Duchamp, John McDonald, and Werner Stuetzle. Surface reconstruction from unorganized points. In *Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, 1992. 6

[21] Fei Hou, Chiyu Wang, Wencheng Wang, Hong Qin, Chen Qian, and Ying He. Iterative poisson surface reconstruction (ipsr) for unoriented points. *arXiv preprint arXiv:2209.09510*, 2022. 2

[22] Chiyu Jiang, Avneesh Sud, Ameesh Makadia, Jingwei Huang, Matthias Nießner, Thomas Funkhouser, et al. Local implicit grid representations for 3D scenes. In *CVPR*, 2020. 3

[23] Michael Kazhdan. Reconstruction of solid models from oriented point sets. 2005. 2, 3

[24] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. 2006. 2, 6

[25] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. 2013. 1, 2, 6, 7

[26] Sebastian Koch, Albert Matveev, Zhongshi Jiang, Francis Williams, Alexey Artemov, Evgeny Burnaev, Marc Alexa, Denis Zorin, and Daniele Panozzo. ABC: A big CAD model dataset for geometric deep learning. In *CVPR*, 2019. 6, 8

[27] Guo Li, Ligang Liu, Hanlin Zheng, and Niloy J. Mitra. Analysis, reconstruction and manipulation using arterial snakes. 2010. 2

[28] Yiyi Liao, Simon Donne, and Andreas Geiger. Deep marching cubes: Learning explicit surface representations. In *CVPR*, 2018. 3

[29] Shi-Lin Liu, Hao-Xiang Guo, Hao Pan, Peng-Shuai Wang, Xin Tong, and Yang Liu. Deep implicit moving least-squares functions for 3D reconstruction. In *CVPR*, 2021. 3, 6, 7

[30] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 1987. 6

[31] Josiah Manson, Guergana Petrova, and Scott Schaefer. Streaming surface reconstruction using wavelets. In *Comput. Graph. Forum*, 2008. 2, 3

[32] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3D reconstruction in function space. In *CVPR*, 2019. 3

[33] Zhenxing Mi, Yiming Luo, and Wenbing Tao. Ssnet: Scalable 3D surface reconstruction network. In *CVPR*, 2020. 3

[34] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF:

Representing scenes as neural radiance fields for view synthesis. *ECCV*, 2020. 5

- [35] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019. 3
- [36] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B. Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *ICCV*, 2021. 5
- [37] Songyou Peng, Chiyu Jiang, Yiyi Liao, Michael Niemeyer, Marc Pollefeys, and Andreas Geiger. Shape as points: A differentiable poisson solver. In *NeurIPS*, 2021. 2, 6, 7
- [38] Carlos Eduardo Scheidegger, Shachar Fleishman, and Cláudio T. Silva. Triangulating point set surfaces with bounded error. 2005. 2
- [39] Andrei Sharf, Thomas Lewiner, Ariel Shamir, Leif Kobbelt, and Daniel Cohen-Or. Competing fronts for coarse-to-fine surface reconstruction. In *Comput. Graph. Forum*, 2006. 2
- [40] Chen Shen, James F. O'Brien, and Jonathan R. Shewchuk. Interpolating and approximating implicit surfaces from polygon soup. In *ACM SIGGRAPH 2004 Papers*. 2004. 1, 2, 3
- [41] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33:7462–7473, 2020. 6
- [42] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33:7537–7547, 2020. 2
- [43] Yiqun Wang, Ivan Skorokhodov, and Peter Wonka. Improved surface reconstruction using high-frequency details. *arXiv preprint arXiv:2206.07850*, 2022. 5
- [44] Qingnan Zhou and Alec Jacobson. Thingi10k: A dataset of 10,000 3D-printing models. *arXiv preprint arXiv:1605.04797*, 2016. 2, 3, 5, 6, 7, 8
- [45] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *arXiv preprint arXiv:1801.09847*, 2018. 6, 8
- [46] Runsong Zhu, Yuan Liu, Zhen Dong, Yuan Wang, Tengping Jiang, Wenping Wang, and Bisheng Yang. AdaFit: Rethinking learning-based normal estimation on point clouds. In *ICCV*, 2021. 6