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

Wenbin Zhao[1*], Jiabao Lei[1*], Yuxin Wen[1], Jianguo Zhang[4], Kui Jia[123†]
[1]South China University of Technology. [2]Pazhou Lab. [3]Peng Cheng Lab.
[4]Department of Computer Science and Engineering, Southern University of Science and Technology.

{eemszhaowb,eejblei,wen.yuxin}@mail.scut.edu.cn,
zhangjg@sustech.edu.cn, kuijia@scut.edu.cn

Abstract

Shape modeling and reconstruction from raw point clouds of objects stand as a fundamental challenge in vision and graphics research. Classical methods consider analytic shape priors; however, their performance is degraded when the scanned points deviate from the ideal conditions of cleanness and completeness. Important progress has been recently made by data-driven approaches, which learn global and/or local models of implicit surface representations from auxiliary sets of training shapes. Motivated from a universal phenomenon that self-similar shape patterns of local surface patches repeat across the entire surface of an object, we aim to push forward the data-driven strategies and propose to learn a local implicit surface network for a shared, adaptive modeling of the entire surface for a direct surface reconstruction from raw point cloud; we also enhance the leveraging of surface self-similarities by improving correlations among the optimized latent codes of individual surface patches. Given that orientations of raw points could be unavailable or noisy, we extend sign-agnostic learning into our local implicit model, which enables our recovery of signed implicit fields of local surfaces from the unsigned inputs. We term our framework as Sign-Agnostic Implicit Learning of Surface Self-Similarities (SAIL-S3). With a global post-optimization of local sign flipping, SAIL-S3 is able to directly model raw, un-oriented point clouds and reconstruct high-quality object surfaces. Experiments show its superiority over existing methods.

1. Introduction

Surface reconstruction from point clouds is of significance during the course of digitally representing the world around us, especially when we have witnessed the development of scanning devices that makes it easier to acquire

---

[*]Equal contribution
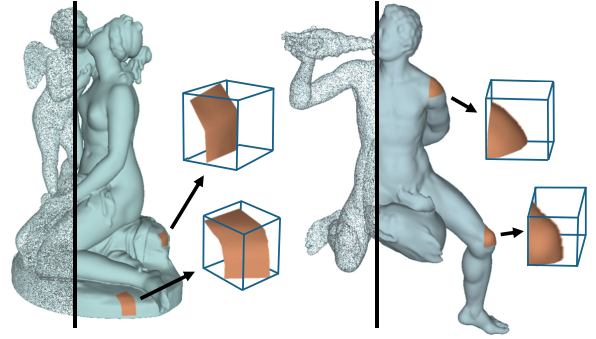[†]Correspondence to Kui Jia <kuijia@scut.edu.cn>

Figure 1: 3D reconstructions from our proposed Sign-Agnostic Implicit Learning of Surface Self-Similarities. For each sculpture, we visualize the raw, un-oriented point cloud on the left, and the reconstructed surface on the right. The surface is reconstructed by interpolation from the learned local implicit subfields, where we isolate some of their zero-level sets for better understanding.

point cloud data. This problem is severely ill-posed [5], since there could be infinite solutions of the continuous surface given the discrete approximation of point clouds, especially when the points are noisy, irregularly distributed, and/or incomplete. As such, proper priors of geometric regularity are necessary to tackle this problem. Classical methods adopt analytic priors such as local linearity and smoothness [26, 7, 2]. However, performance of these methods is degraded when encountering sensing imperfection, or unavailable of surface normals for the observed points.

More recently, deep neural networks are introduced to learn geometric priors from auxiliary shapes [20, 33] in a data-driven manner, which have shown their superiority over classical methods. The pipeline starts from approaches [33, 11, 31] that globally embed a shape into a latent shape space based on auto-encoders. Encoding global shape priors might simplify the problem, which, however, is limited in generalizing the learned priors to unseen shapes. To improve generalization, there has been some attempts

[10, 18, 25, 8, 34] that learn local shape priors and model a global shape as configuration of local shape parts. Indeed, for surface shapes of a certain object category, decomposing the global modeling into local ones prevents learning priors that are mostly concerned with the category-level shape configuration. These methods rely on learning priors from auxiliary training sets. This always risks their generalization in cases that testing samples are out of the distributions of the training ones. In this work, we aim to close the generalization gap by learning the shape priors directly from the input data themselves.

Our idea is motivated from an arguably universal phenomenon that self-similar shape patterns of local surface patches repeat across an entire object surface. Figure 1 gives an illustration. Such a phenomenon is similar to the self-similarities of local patches in a 2D image, which has motivated a plethora of methods in the literature of image modeling and restoration [6, 14, 13].

To implement this phenomenon for modeling and reconstructing a surface from raw observed points, a challenge remains due to the possibly unreliable surface normals associated with the observed individual points. One may compute approximate ones, which might not be precise enough to support a fine surface recovery especially when points are noisy or scanner information is absent. Learning to predict the surface normals [16, 21] is not applicable as well, since we may only have the observed points at hand. To this end, we propose in this paper a novel method, termed *Sign-Agnostic Implicit Learning of Surface Self-Similarities (SAIL-S3)*, for modeling and reconstruction of a continuous surface directly from a raw, un-oriented point cloud. We note that the property of surface self-similarities is also used in [23] to deform an initial mesh, where they implicitly leverage the property by training the mesh deformation network. In contrast, our proposed SAIL-S3 is a completely different local framework for sign-agnostic implicit surface modeling.

Specifically, SAIL-S3 is by design a local model that partitions a global implicit surface field into an *adaptive* set of overlapped, local subfields, each of which is expected to cover a surface patch. We leverage the property of surface self-similarities by incorporating the following designs into SAIL-S3: (1) we use a shared implicit model to learn these subfields, while allowing the individual latent representations of local subfields to be freely optimized, and (2) we use a learning objective that promotes correlated latent representations when their corresponding surface patches are of similar shape (cf. Section 4.1). We extend sign-agnostic learning [3] into our local framework, and propose provably model initialization that outputs a *signed* solution of implicit field function given the *unsigned* learning objective (cf. Section 4.2). The signed solutions of local implicit subfields may not be consistent globally. With a global post-

optimization of local sign flipping, SAIL-S3 is able to directly model raw, un-oriented point clouds and reconstructs high-quality object surfaces (cf. Section 4.3). We conduct thorough experiments on the objects from ShapeNet [9] and Threedscans [1] datasets. They include object instances with natural and complex topologies. Experiments show that given no auxiliary training set, our proposed SAIL-S3 outperforms existing methods in terms of reconstructing smooth and sharp surfaces, even though the comparative learning based methods use auxiliary training shapes. Robustness tests with noisy inputs again confirm the efficacy of our proposed method. We finally summarize our technical contributions as follows.

- We propose a novel method of *SAIL-S3* for surface modeling and reconstruction from raw, un-oriented point clouds. The method learns self-adaptive shape priors by implementing a universal phenomenon that an object surface contains self-similar shape patterns of local surface patches.

- SAIL-S3 uses adaptively learned local implicit functions to model the global implicit surface field. We extend sign-agnostic learning into the local SAIL-S3 framework, by proposing provably model initializations that can be optimized to produce *signed* solutions of local implicit function from the *unsigned* learning objective.

- With a global post-optimization of local sign flipping, SAIL-S3 is able to directly model raw, un-oriented point clouds and reconstructs high-quality surfaces of objects. Experiments demonstrate its superiority over existing methods.

2. Related works

In this section, we briefly review existing methods for surface modeling and reconstruction from raw point clouds. We focus on those works closely related to the present one.

Classical Methods Using Analytic Shape Priors – There have been a number of analytic priors proposed in the literature. Representative ones include Screened Poisson Surface Reconstruction (SPSR) [26], Radius Basis Functions (RBF) [7], and Moving Least Squares (MLS) [29]. SPSR a kind of method based on global surface smoothness priors. It casts the reconstruction as a spatial Poisson problem and solves it in the frequency domain. However, it relies on oriented normals of surface points. Likewise, RBF is also based on global surface smoothness priors. It produces reconstruction through a linear combination of radially symmetric basis functions. MLS [29] directly approximates the input points as spatially-varying low-degree polynomials, which adopts local surface smoothness as its priors.

Neural Priors for Explicit Surface Modeling – Given observed points, deep neural networks are recently proposed

to encode the points as a latent representation, and then decode it explicitly as a surface mesh. Among these methods, AtlasNet [20] defines the surface as a set of atlas charts, and trains a network to deform their vertices to form a complete surface mesh. Subsequent methods [39, 22, 37, 38] extend AtlasNet by deforming a single, initial mesh. Note that the property of self-similarities is also used in [23] to deform an initial mesh, where they assume these self-similarities are implicitly leveraged by training the network for surface deformation. However, mesh deformation cannot change surface topologies, and it is difficult for such method to generate surface of complex topologies. As a remedy, topology modification is proposed in [32] by pruning edges and faces during the deformation process. In general, such methods of explicit mesh deformation perform worse than those reconstructing a surface by learning deep implicit fields.

Neural Priors for Implicit Surface Modeling – More recently, learning deep networks as implicit surface fields is found to be an effective approach for modeling continuous surface [33, 11, 31], which should be extracted via [30, 28]. They typically learn a global surface field of Signed Distance Function (SDF) [33] or occupancy [31]. Subsequent methods [10, 18, 25, 8, 34] extend them as local implicit models for modeling local surface patches. For example, BAE-NET [10] adopts branched decoders for adaptively modeling surface parts; Deep Local Shape [8] and Convolutional OccNet [34] utilize a 3D grid of voxels to represent an SDF or occupancy field, which is memory-expensive; the method [25] avoids voxel-based SDF representations, however, it still requires auxiliary shapes for model training.

Sign-Agnostic Surface Modeling – Practically scanned raw points are usually short of oriented normals. Analytic computation can only give approximate results. It is thus appealing to model the raw points in a sign-agnostic manner [3, 4]. Atzmon and Lipman study this problem of Sign-Agnostic Learning (SAL). They propose unsigned objectives, which, given proper initialization of network weights, can produce signed solutions of implicit functions. Original SAL works with global shape modeling. In this work, we extend SAL into our local framework, and propose the corresponding network initialization.

3. Problem Statement

Given a set of observed points $\mathcal{P} = \{\boldsymbol{p}_i \in \mathbb{R}^3\}_{i=1}^n$ that represents a discrete sampling of an underlying object surface $\mathcal{S}$, we study a fundamental problem of modeling $\mathcal{S}$ and reconstruct it from the observed $\mathcal{P}$ [5]. The problem is severely ill-posed, since there could be infinitely many solutions of the continuous $\mathcal{S}$ given the discrete approximation $\mathcal{P}$; it becomes even more difficult considering that $\mathcal{P}$ may be obtained from practical scanning, and due to imperfection of sensing, the scanned points may be noisy, irregularly distributed, and/or incomplete. As such, proper
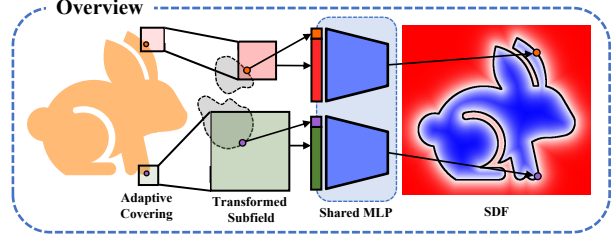

Figure 2: An illustration of our proposed Sign-Agnostic Implicit Learning of Surface Self-Similarities (SAIL-S3).

priors of geometric regularity is to be imposed in order to recover meaningful solutions. Classical methods adopt analytic priors such as local linearity and smoothness [26, 7, 2]; however, their performance degrades with the increased levels of sensing imperfection. These methods usually require the availability of surface normals for the observed points, which, however, are either unavailable or cannot be computed accurately. It is recently shown that learned neural priors from an auxiliary set of training shapes provide strong regularization on the recovery of continuous surface shapes, particularly those based implicit models, *e.g.*, SDF [33] or occupancy field [31, 11], and those extending these models as local ones [16, 25]. In this work, we also consider modeling the observed $\mathcal{P}$ with implicit SDFs. We improve over existing methods by learning directly from $\mathcal{P}$ itself, without relying on the auxiliary training set.

4. The Proposed Method

Our proposed method is primarily motivated from a universal phenomenon that self-similar shape patterns of local surface patches repeat across an entire object surface. Figure 1 gives an illustration. Such a phenomenon is similar to the self-similarities of local patches in a 2D image, which has motivated a plethora of methods in the literature of image modeling and restoration, including the representative non-local means [6], self-adaptive dictionary learning [14], and BM3D [13]; they have successfully shown that clean images can be restored from distorted images themselves. Motivated from the conceptually similar surface self-similarities, we aim to address the problem stated in Section 3, and propose a novel method termed *Sign-Agnostic Implicit Learning of Surface Self-Similarities (SAIL-S3)*. An illustration of it is shown in Figure 2. Details are presented as follows.

4.1. Local Implicit Modeling of Surface Self-Similarities

Denote $\mathcal{G} \subset \mathbb{R}^3$ as an implicit field of SDF, whose zero-level set represents the underlying surface $\mathcal{S}$ from which the observed points in $\mathcal{P}$ are sampled. Let $\hat{f} : \mathbb{R}^3 \times \mathbb{R}^d \to \mathbb{R}$ be the implicit model of $\mathcal{G}$; it takes as input a sampled point $\boldsymbol{q} \in \mathbb{R}^3$ in the 3D space and a latent representation

$\hat{z} \in \mathbb{R}^d$ encoding the surface, and outputs a value of signed distance between q and $\mathcal{S}$. Instead of obtaining $\hat{z}$ from any learned encoder, we follow [33] and resort to latent code optimization to fit with the observed $\mathcal{P}$.

The pre-assumed property of surface self-similarities suggests that some local surface patches of $\mathcal{S}$ are similar in terms of shape pattern. Our design for leveraging this property has the following two ingredients.

- Instead of directly learning $\hat{f}$ for the global field $\mathcal{G}$, we consider a number N of overlapped, local implicit subfields $\{\mathcal{F}_i\}_{i=1}^N$, each of which is responsible for a surface patch $\mathcal{S}_{\mathcal{F}_i}$; we use a shared implicit model $f_{\boldsymbol{\theta}} : \mathbb{R}^3 \times \mathbb{R}^d \to \mathbb{R}$, parameterized by $\boldsymbol{\theta}$, to learn these subfields, while allowing the individual latent representations $\{z_i \in \mathbb{R}^d\}_{i=1}^N$ to be freely optimized.
- We use a learning objective that promotes correlated latent representations when their corresponding surface patches are of similar shape; for example, a pair of correlated z_i and z_j outputs, through $f_{\boldsymbol{\theta}}$, implicit subfields $\mathcal{F}_i$ and $\mathcal{F}_j$ whose zero-level sets represent similar shapes.

While the latter ingredient is an explicit design to promote surface self-similarities, the former one implicitly does so by decoding the implicit subfield with the shared model $f_{\boldsymbol{\theta}}$. In this work, we implement $f_{\boldsymbol{\theta}}$ as a network of multi-layer perceptron (MLP). Each subfield $\mathcal{F}_i$, $i \in \{1, \dots, N\}$, is centered at $c_i \in \mathbb{R}^3$ and covers a local volume of size $a_i \times a_i \times a_i$. Both $\mathcal{C} = \{c_i\}_{i=1}^N$ and $\mathcal{A} = \{a_i\}_{i=1}^N$ are learnable parameters, which determine how $\{\mathcal{F}_i\}_{i=1}^N$ distribute in the global $\mathcal{G}$; we expect their optimizations to make each of $\{\mathcal{F}_i\}_{i=1}^N$ cover a patch of the surface $\mathcal{S}$ (cf. Section 4.2.2 for the details). Let $\mathcal{Z} = \{z_i \in \mathbb{R}^d\}_{i=1}^N$. SAIL-S3 is formally to learn the implicit SDF $f_{\boldsymbol{\theta}} : \mathbb{R}^3 \times \mathbb{R}^d \to \mathbb{R}$, parameterized by $(\boldsymbol{\theta}, \mathcal{Z}, \mathcal{C}, \mathcal{A})$, by fitting to the observed $\mathcal{P}$. We note that defining implicits over cubes practically supports more convenient space partitioning (cf. Section 4.2.2) and smoothing of the results in overlapped regions (cf. Section 4.4). We present the sign-agnostic implicit learning of SAIL-S3 as follows.

4.2. Sign-Agnostic Local Implicit Learning

An important challenge for shape modeling of raw point clouds is the possibly unreliable surface normals associated with the observed points. Approximate ones may be estimated from $\mathcal{P}$ via either covariance analysis [24] or learning-based methods [16, 21]; however, their performances usually degrade especially when inputs are noisy or camera information is absent, which is hard to support accurate surface recovery. SAL [3] is a promising solution to cope with the issue; it is proposed to learn a global implicit model to reconstruct an entire object surface, with no requirement on the availability of surface normals. However, it remains absent for how to use it to model a surface as a collection of local implicit subfields. In this work, we extend the SAL technique [3] into our local framework of SAIL-S3, as follows.

We first present the extension in a local implicit subfield $\mathcal{F}$. Before that, for any point q sampled in the global $\mathcal{G}$, we compute its *unsigned* distance $s(q) \in \mathbb{R}^+$ to the surface $\mathcal{S}$ of interest approximately as

$$s(q) = \|q - p\|_2 \text{ s.t. } p = \arg\min_{p' \in \mathcal{P}} \|q - p'\|_2. \quad (1)$$

The distance (1) approaches the true one when the number n of points in $\mathcal{P}$ goes to infinity. In practice, it would provide us an approximate supervision signal for learning the implicit model. Its unsigned nature relaxes the requirement on the knowledge of local surface orientations.

Assume that $\mathcal{F}$ contains a set $\mathcal{P}_{\mathcal{F}}$ of observed points in a local neighborhood of $\mathcal{P}$. We normalize any $p \in \mathcal{P}_{\mathcal{F}} \subset \mathcal{G}$ as $\bar{p} = (p - c)/a$ in a local coordinate system of the subfield $\mathcal{F}$, where c and a are the learnable center and side length. We applies the same to any sampled point $q \in \mathcal{F} \subset \mathcal{G}$, resulting in $\bar{q}$ after coordinate offset and scaling. We thus have the unsigned distance $s(\bar{q}) = s(q)/a$ inside $\mathcal{F}$. Given the supervision from $\{\bar{q} \in \mathcal{F}\}$, SAL [3] aims to find a *signed* solution of the implicit function $f_{\boldsymbol{\theta}}$, by solving the following *unsigned*, bi-level optimization problem

$$\min_{\boldsymbol{\theta}, z} \sum_{\bar{q} \in \mathcal{F}} ||f_{\boldsymbol{\theta}}(\bar{q}, z)| - s(\bar{q})| \text{ s.t. } s(\bar{q}) = s(q)/a, \quad (2)$$

where we have temporarily assumed that a and c for the subfield $\mathcal{F}$ are fixed, and $\{\bar{q} \in \mathcal{F}\}$ (equivalently, $\{q \in \mathcal{F}\}$) are usually sampled around individual $p \in \mathcal{P}_{\mathcal{F}}$ with densities inversely proportional to the distances. A signed solution of (2) means that for an optimal z^*, we have $f_{\boldsymbol{\theta}^*}(\bar{q}, z^*) > 0$ when $\bar{q}$ lies by one side of the local surface $\mathcal{S}_{\mathcal{F}}$, and $f_{\boldsymbol{\theta}^*}(\bar{q}, z^*) < 0$ otherwise; in other words, it produces signed distances to the surface even though the supervision is unsigned. However, this is not always guaranteed given that a flipped sign of $f_{\boldsymbol{\theta}}(\bar{q}, z)$ does not change its absolute value, and consequently the loss (2). The problem becomes even more involved when coupled with the simultaneous optimization of z.

Fortunately, it is suggested in [3] that a proper setting of network weights $\boldsymbol{\theta}^0$ would initialize a latent code-free function $f_{\boldsymbol{\theta}^0}(\cdot)$ as a signed function, and it is also empirically observed that optimization from such an initialization is stably in the signed local minima, without going across loss barriers to the unsigned solutions. In this work, we extend the weight initialization scheme in [3] for learning a signed solution of local implicit function from the loss (2), as presented shortly.

A Proper Model Initialization for Signed Solutions For the observed points $\{\bar{p} \in \mathcal{P}_{\mathcal{F}}\}$ in $\mathcal{F}$, we first use least squares to fit them with a surface patch on a radius-$\bar{r}$ sphere centered at $\bar{t}$ (cf. the supplementary material for more details). The following corollary shows a scheme of initializ-
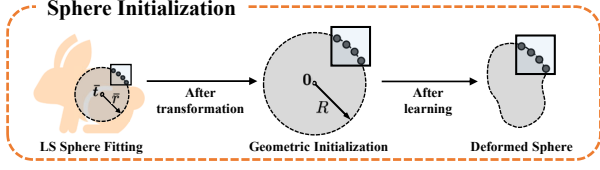
Figure 3: An illustration on our model initialization.

ing $\boldsymbol{\theta}^0$ and $\boldsymbol{z}^0$ that guarantees $f_{\boldsymbol{\theta}^0}(\bar{\boldsymbol{q}}, \boldsymbol{z}^0) \approx \|\bar{\boldsymbol{q}} - \bar{\boldsymbol{t}}\| - \bar{r}$; in other words, $f_{\boldsymbol{\theta}^0}(\cdot, \boldsymbol{z}^0)$ is initialized as a signed distance function whose zero-level set is approximately a $\bar{r}$-radius sphere centered at $\bar{\boldsymbol{t}}$, and the sphere approximately covers the local point set $\mathcal{P}_{\mathcal{F}}$. We experimentally find that our results are relatively stable w.r.t. LS sphere initialization. Figure 3 gives an illustration.

Corollary 4.1. *Let* $f : \mathbb{R}^{3+d} \to \mathbb{R}$ *be an* l-*layer MLP with ReLU activation* ν. *That is,* $f(\boldsymbol{p}, \boldsymbol{z}) = \boldsymbol{w}^T \nu(\boldsymbol{W}^l(\cdots \nu(\boldsymbol{W}_{\boldsymbol{p}}^1 \boldsymbol{p} + \boldsymbol{W}_{\boldsymbol{z}}^1 \boldsymbol{z} + \boldsymbol{b}^1)) + \boldsymbol{b}^l) + c$, *where* $\boldsymbol{W}_{\boldsymbol{p}}^1 \in \mathbb{R}^{d_{out}^1 \times 3}$ *and* $\boldsymbol{W}_{\boldsymbol{z}}^1 \in \mathbb{R}^{d_{out}^1 \times d}$ *denote the weight matrices of the first layer, and* $\boldsymbol{b}^1 \in \mathbb{R}^{d_{out}^1}$ *denotes the bias;* $\boldsymbol{W}^i \in \mathbb{R}^{d_{out}^i \times d_{out}^{i-1}}$ *and* $\boldsymbol{b}^i \in \mathbb{R}^{d_{out}^i}$ *denote parameters of the* i^{th} *layer;* $\boldsymbol{w} \in \mathbb{R}^{d_{out}^l}$ *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* $\boldsymbol{w} = \sqrt{\frac{\pi}{d_{out}^l}} \mathbf{1}$, $c = -\bar{r}$, $\bar{r} > 0$, *let all entries of* $\boldsymbol{W}^i$ $(2 \leq i \leq l)$ *follow i.i.d. normal* $\mathcal{N}(0, \frac{2}{d_{out}^i})$, *let entries of* $\boldsymbol{W}_{\boldsymbol{p}}^1$ *follow i.i.d. normal* $\mathcal{N}(0, \frac{2}{d_{out}^1})$, *and let* $\boldsymbol{b}^i = \mathbf{0}$ $(2 \leq i \leq l)$. *If* $\boldsymbol{W}_{\boldsymbol{z}}^1 = \boldsymbol{W}_{\boldsymbol{p}}^1 [\boldsymbol{I} \in \mathbb{R}^{3 \times 3}, \mathbf{0} \in \mathbb{R}^{3 \times (d-3)}]$ *and* $\boldsymbol{b}^1 = -\boldsymbol{W}_{\boldsymbol{p}}^1 \bar{\boldsymbol{t}}$, *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* $\bar{\boldsymbol{t}}$.

The proof to Corollary 4.1 is provided in the supplementary material.

4.2.1 Learning over a Collection of Local Implicit Subfields

We have so far presented how to learn an implicit function individually for a local subfield. For learning over the collection $\{\mathcal{F}_i\}_{i=1}^N$, we remind that SAIL-S3 shares the function $f_{\boldsymbol{\theta}}$ for these subfields, where $\boldsymbol{\theta}$ is initialized and optimized for all the N subfields, while the latent codes $\{\boldsymbol{z}_i\}_{i=1}^N$ are adaptively optimized. This brings an inconsistency when initializing the implicit function separately for each subfield based on Corollary 4.1. In practice, we circumvent this inconsistency by first initializing $\boldsymbol{\theta}^0$ as suggested by Corollary 4.1, and $\{\boldsymbol{z}_i^0\}_{i=1}^N$ as the samples drawn from a Gaussian distribution with small standard deviation, which essentially defines the zero-level set of $f_{\boldsymbol{\theta}^0}(\cdot, \boldsymbol{z}_i^0)$ as a radius-R sphere centered at the origin, where R is a hyperparameter; for the ith subfield, we then transform its sampled point $\bar{\boldsymbol{q}}^i \in \mathcal{F}_i$ as

$$\tilde{\boldsymbol{q}}^i = R(\bar{\boldsymbol{q}}^i - \bar{\boldsymbol{t}}_i)/\bar{r}_i, \tag{3}$$

where we use superscript i in $\bar{\boldsymbol{q}}^i$ (and $\tilde{\boldsymbol{q}}^i$) to indicate that it is the transformed coordinates of $\boldsymbol{q}$ in the ith subfield (note that a sampled $\boldsymbol{q}$ may appear in different subfields), and $\bar{\boldsymbol{t}}_i$ and $\bar{r}_i$ associated with the subfield $\mathcal{F}_i$ are the center and radius of a sphere obtained by solving a least square fitting problem. Note that by (3), any observed point $\bar{\boldsymbol{p}} \in \mathcal{P}_{\mathcal{F}_i}$ is transformed in the same way as $\tilde{\boldsymbol{p}}$; this means geometrically that each subfield is transformed such that its contained point observations approximately fit with a surface patch on the initialized zero-level sphere of radius R. Figure 3 gives the illustration.

Given $\{\tilde{\boldsymbol{q}}^i \in \mathcal{F}_i\}$ sampled from each local subfield, we have the following loss function for modeling the underlying surface $\mathcal{S}$ over the collection $\{\mathcal{F}_i\}_{i=1}^N$

$$\mathcal{L}^{\text{Modeling}}(\boldsymbol{\theta}, \mathcal{Z}, \mathcal{C}, \mathcal{A}) = \sum_{i=1}^N \sum_{\tilde{\boldsymbol{q}}^i \in \mathcal{F}_i} \left| |f_{\boldsymbol{\theta}}(\tilde{\boldsymbol{q}}^i(\boldsymbol{c}_i, a_i), \boldsymbol{z}_i)| - s(\tilde{\boldsymbol{q}}^i) \right| + \lambda \|\boldsymbol{Z}\|_*, \tag{4}$$

where $s(\tilde{\boldsymbol{q}}^i) = s(\boldsymbol{q})R/(\bar{r}_i a_i)$, and $\boldsymbol{Z} = [\boldsymbol{z}_1/\|\boldsymbol{z}_1\|_2; \cdots; \boldsymbol{z}_N/\|\boldsymbol{z}_N\|_2]$ is collection of all the normalized latent codes. We use nuclear norm penalty to improve correlations among latent codes, which ensures the self-similarities learned from the input itself. Besides, the above modeling loss can be further improved by leveraging the first derivative of function $f_{\boldsymbol{\theta}}(\cdot, \boldsymbol{z}_i)$ as detailed in [4].

4.2.2 An Adaptive Field Covering

Our motivation for modeling $\mathcal{S}$ with local self-similarities expects that each individual subfield $\mathcal{F}$ covers a roughly similar volume of the entire $\mathcal{G}$, and the covering would be evenly distributed along the surface $\mathcal{S}$. To this end, we initialize parameters $\{(\boldsymbol{c}_i, a_i)\}_{i=1}^N$ of local implicit subfields as follows. Given the observed $\mathcal{P}$, we use farthest point sampling [15] of $\mathcal{P}$ to initialize the subfield centers $\{\boldsymbol{c}_i\}_{i=1}^N$; we then initialize the covering size of subfield as $a_i = \alpha \min_{j \in \{1,\ldots,N\}/\{i\}} \|\boldsymbol{c}_i - \boldsymbol{c}_j\|_2$, where we set $\alpha \geq 1$ such that each observed point in $\mathcal{P}$ is covered by at least one subfield and these subfields have a certain amount of overlapping. This initialization can roughly meet our expectation; however, directly solving the objective (4) may update $\{(\boldsymbol{c}_i, a_i)\}_{i=1}^N$ such that a very few of the subfields are enlarged to cover large portions of the surface, while the remaining ones are moved to cover duplicate surface patches. To avoid these undesired solutions, we propose the following loss terms to constrain the optimization

Volume loss – We prevent undesirable enlarging of individual subfields by penalizing the volume of each subfield

$$\mathcal{L}^{\text{Volume}}(\mathcal{A}) = \sum_{i=1}^N \max\{a_i, 0\}. \tag{5}$$

Placing loss – Given the constraint from (5), we further en-

courage an even distribution of the subfields $\{\mathcal{F}_i\}_{i=1}^N$ on the surface by penalizing the Chamfer Distance [17] between each observed $\boldsymbol{p} \in \mathcal{P}$ and its closest subfield center, i.e.,

$$\mathcal{L}^{\text{Placing}}(\mathcal{C}) = \sum_{\boldsymbol{p} \in \mathcal{P}} \min_{\boldsymbol{c}_i \in \mathcal{C}} \|\boldsymbol{p} - \boldsymbol{c}_i\|_2^2 + \sum_{\boldsymbol{c}_i \in \mathcal{C}} \min_{\boldsymbol{p} \in \mathcal{P}} \|\boldsymbol{p} - \boldsymbol{c}_i\|_2^2. \quad (6)$$

Covering loss – To cover all the observed points in $\mathcal{P}$, we use the exterior signed distance field to penalize those uncovered point $\boldsymbol{p}$, i.e.

$$\mathcal{L}^{\text{Covering}} = \sum_{\substack{\boldsymbol{p} \in \mathcal{P} \\ \boldsymbol{p} \notin \mathcal{G}}} \sqrt{\min_{i \in \{1,\dots,N\}} \sum_{j=1}^3 \max\{|\pi_j(\boldsymbol{p} - \boldsymbol{c}_i)| - a_i, 0\}^2}, \quad (7)$$

where π_j is an operator that selects the jth element from a vector.

4.2.3 The Combined Learning objective

Given the observed $\mathcal{P}$, we use the following combined objective to reconstruct its underlying surface $\mathcal{S}$ via sign-agnostic implicit learning of surface self-similarities

$$\mathcal{L}^{\text{SAIL-S3}} = \mathcal{L}^{\text{Modeling}} + \lambda_1 \mathcal{L}^{\text{Volume}} + \lambda_2 \mathcal{L}^{\text{Placing}} + \lambda_3 \mathcal{L}^{\text{Covering}} \quad (8)$$

where λ_1, λ_2, and λ_3 are penalty parameters. Starting from initializations, as presented above, the objective (8) can be optimized simply via stochastic gradient descent. Penalty parameters λ_1, λ_2, and λ_3 can be found via grid search, and they are practically insensitive to all kinds of experimental settings.

4.3. A Global Post-Optimization of Local Sign Flipping

Solving the objective (8) produces N local implicit functions $f_{\boldsymbol{\theta}}(\cdot, \boldsymbol{z}_i)$, $i = 1, \dots, N$, which may be used for extraction of their zero-level iso-surfaces. For any $\boldsymbol{q} \in \mathcal{G}$ covered by a pair of neighboring i^{th} and j^{th} functions, however, their function evaluations may not be consistent in their signs because each subfield is optimized individually. Motivated by the procedure of reorienting point cloud in [24], we propose a post-optimization of local sign flipping to address this issue. More specifically, we treat each local implicit subfield $\mathcal{F}$ as a vertex, and construct a connected, undirected graph over the vertices as $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, \dots, v_N\}$ and an edge $e_{i,j} \in \mathcal{E}$ is connected once the subfields $\mathcal{F}_i$ and $\mathcal{F}_j$ have an overlapped region in the field $\mathcal{G}$. We associate each edge $e_{i,j}$ with a pair of weights defined as

$$w_{i,j}^1(e_{i,j}) = \sum_{\boldsymbol{q} \in \mathcal{F}_i \cap \mathcal{F}_j} \left| f_{\boldsymbol{\theta}}(\tilde{\boldsymbol{q}}^i, \boldsymbol{z}_i) - f_{\boldsymbol{\theta}}(\tilde{\boldsymbol{q}}^j, \boldsymbol{z}_j) \right|,$$

$$w_{i,j}^0(e_{i,j}) = \sum_{\boldsymbol{q} \in \mathcal{F}_i \cap \mathcal{F}_j} \left| f_{\boldsymbol{\theta}}(\tilde{\boldsymbol{q}}^i, \boldsymbol{z}_i) + f_{\boldsymbol{\theta}}(\tilde{\boldsymbol{q}}^j, \boldsymbol{z}_j) \right|.$$

This gives two weight sets $\mathcal{W}^1$ and $\mathcal{W}^0$ with $|\mathcal{W}^0| = |\mathcal{W}^1| = |\mathcal{E}|$, and we write $\mathcal{W} = \mathcal{W}^0 \cup \mathcal{W}^1$. We endow

each vertex $v \in \mathcal{V}$ with a sign variable $h(v) \in \{1, -1\}$, and use minimum spanning tree (MST) to determine $\{h(v_i)\}_{i=1}^N$ (details are given in the supplementary material).

We finally flip local implicit functions as $h(v_i)f_{\boldsymbol{\theta}}(\cdot, \boldsymbol{z}_i)$ (or equivalently, $h(\mathcal{F}_i)f_{\boldsymbol{\theta}}(\cdot, \boldsymbol{z}_i)$), $i = 1, \dots, N$. Our MST is based on the Prim's algorithm [35], which is guaranteed to find a solution of minimal cost; we empirically observe that it works well in practice.

4.4. Inference via Interpolation of Local Fields

After post-optimizing the local sign as described in Section 4.3, the predictions for some point $\boldsymbol{q} \in \mathcal{G}$ are not expected to be consistent since each prediction may have slight error. We extend trilinear interpolation to the case of arbitrary number of overlapping regions. The weights for interpolating are calculated by the following. For any $\boldsymbol{q} \in \mathcal{G}$ falling in a number M of overlapped subfields $\{\mathcal{F}_j\}_{j=1}^M$, we evaluate its signed distance to the underlying surface $\mathcal{S}$ as the following averaged one

$$\hat{f}(\boldsymbol{q}) = \sum_{j=1}^M \omega_j(\boldsymbol{q}) \cdot h(\mathcal{F}_j) \cdot \frac{\bar{r}_j a_j}{R} f_{\boldsymbol{\theta}}(\tilde{\boldsymbol{q}}^j, \boldsymbol{z}_j) \quad (9)$$

$$\text{s.t.} \omega_j(\boldsymbol{q}) = \frac{\left| \max_{k \in \{1,2,3\}} |\pi_k(\boldsymbol{q} - \boldsymbol{c}_j)| - a_j \right|}{\sum_{j'=1}^M \left| \max_{k \in \{1,2,3\}} |\pi_k(\boldsymbol{q} - \boldsymbol{c}_{j'})| - a_{j'} \right|},$$

Instead of directly using the inverse distance weighted average method mentioned in [36], the proposed interpolation ensures truly smooth transitions between different overlapped regions without discontinuity of the first kind when switching the neighbors. Such an inference via interpolation has the benefit of smoothing out the less consistent local subfields inferred from individual functions. We practically observe that training procedure will slow down if applying (9) during training, compared to simply adopting (4). Experiments confirmed the efficacy of our choice to use (9) as a post-processing method during inference.

Given the signed distance evaluation (9), we finally use marching cubes [30] to extract the zero-level set that reconstructs the surface $\mathcal{S}$.

5. Experiments

In this section, we present setups, comparative results, and robustness test to verify the efficacy of our proposed SAIL-S3, by comparing with the state-of-the-art methods for surface reconstruction from raw point clouds. ***Ablation studies are given in the supplementary material.***

Datasets – We conduct experiments on the ShapeNet [9] and ThreeDScans [1] datasets that respectively contain synthetic objects and objects of real scans. For ShapeNet, we randomly select 100 objects of *chair*; for ThreeDScans, we randomly select 30 sculptures. These objects are selected

due to their complex shape topologies. In addition, we evaluate the robustness of our method by adding point-wise Gaussian noise to sculptures from ThreeDScans. We pre-process these object surfaces by centering their origins and scaling them uniformly to fit within the unit sphere. We then sample points at densities of $50,000$ and $100,000$ respectively from each instance of ShapeNet and ThreeDScans as the raw inputs. For comparative methods that require oriented surface normals, we compute the normals via tangent plane estimation [24] and reorient the directions via minimal spanning tree, which follows [24].

Implementation Details – We adopt a 6-layer MLP as our local implicit model, and initialize it according to Section 4.2. We initialize elements in the latent z by sampling from $\mathcal{N}(0, (1\times10^{-3})^2)$. During learning, we optimize the objective (8) for $40,000$ iterations using Adam [27], with initial learning rates of 1×10^{-3} for θ and z, and 3×10^{-4} for subfield center c and scale a. The learning rates decay by 0.2 at $20,000$, $30,000$, $35,000$ and $38,000$ iterations. We set $\lambda_1 = 3 \times 10^{-4}$, $\lambda_2 = 1.0$, $\lambda_3 = 1.0$ in the objective (8). It takes around 70ms per iteration, and better results require more iterations.

Evaluation Metrics – We randomly sample $100,000$ points respectively on the ground truth and reconstructed mesh, and use the metrics of Chamfer Distance (CD), Normal Consistency (NC) and F-score (F) to quantitatively evaluate different methods, where F is evaluated under the threshold of 0.005.

Comparative Methods – We compare our method with three categories of existing methods, including the Screened Poisson Surface Reconstruction (SPSR) [26], global fitting methods such as Implicit Geometric Regularization (IGR) [19] (with normal data) and Sign Agnostic Learning (SAL) [3] (without normal data), and locally learned methods such as Local implicit Grid (LIG) [25], Convolutional Occupancy Networks (CON) [34], and Points2Surf (P2S) [16]. We implement SPSR in MeshLab [12] with the default hyper-parameters. For IGR and SAL, we directly fit the training points without learning from auxiliary data; for LIG, CON and P2S, we use the provided pre-trained models and the default settings from the original papers. Note that global fitting methods require no auxiliary data, which is the same as our method. We summarize working conditions of different methods in Table 1.

5.1. Comparative Results

To demonstrate the efficacy of our method to reconstruct high-fidelity surfaces from raw point clouds, we conduct surface reconstruction experiments on *chair* objects in ShapeNet [9] and sculptures in ThreeDScans [1]. Qualitative results are shown in Figure 4. For the methods relying on global representations, such as IGR [19] and SAL [3], they fail to generalize to the complex topologies. Conv. OccNet (CON) [34] and Point2Surf (P2S) [16] perform much

Methods	No requirement on surface normals	No requirement on auxiliary data	Local model
SPSR [26]	×	✓	✓
IGR [19]	×	✓	×
SAL [3]	✓	✓	×
LIG [25]	×	×	✓
CON [34]	∗	×	✓
P2S [16]	∗	×	✓
SAIL-S3	✓	✓	✓

Table 1: Working condition summary of different methods. Note that ∗ indicates that the method does not require surface normals during inference but does require during training. Our proposed SAIL-S3 requires neither surface normals nor auxiliary training data.

better, almost recovering the topologies of different chairs and sculptures; however, their details are either missing or rugged. Results from those methods requiring accurate surface normals, such as SPSR [26], IGR [19], and LIG [25], have unpredictable errors including non-watertight or wrongly folded meshes, since estimated normals might not be accurate enough. In contrast, our proposed SAIL-S3 recovers both the correct topologies and surface details from the raw, un-oriented input points, without requiring any auxiliary data. Quantitative results in Table 2 further confirm the superiority of our method over existing ones.

Datasets	ShapeNet[9]			ThreeDScans [1]		
Methods	CD↓	NC↑	F↑	CD↓	NC↑	F↑
SPSR [26]	0.009	0.966	0.832	**0.003**	0.968	0.864
IGR [19]	0.011	0.955	0.774	0.007	0.942	0.773
SAL [3]	0.015	0.897	0.360	0.009	0.899	0.332
LIG [25]	0.006	0.940	0.756	0.005	0.920	0.727
CON [34]	0.011	0.876	0.269	0.010	0.852	0.258
P2S [16]	**0.003**	0.928	0.798	0.005	0.869	0.669
SAIL-S3	**0.003**	**0.981**	**0.884**	**0.003**	**0.972**	**0.871**

Table 2: Quantitative results of *chair* instances in ShapeNet [9] and sculptures in ThreeDScans [1]. For CD, the smaller, the better; for NC and F, the larger, the better.

5.2. Robustness Evaluation

We further evaluate the robustness of different methods against noisy inputs. We use sculptures from ThreeDScans [1], and add point-wise Gaussian noise of varying levels. As shown in Figure 5, adding noise to the input points indeed degrades the performance of different methods, particularly for those methods that require estimation of surface normals. Our proposed SAIL-S3 stays more robust against the noise. Quantitative comparisons with different noise levels are given in the supplementary material.
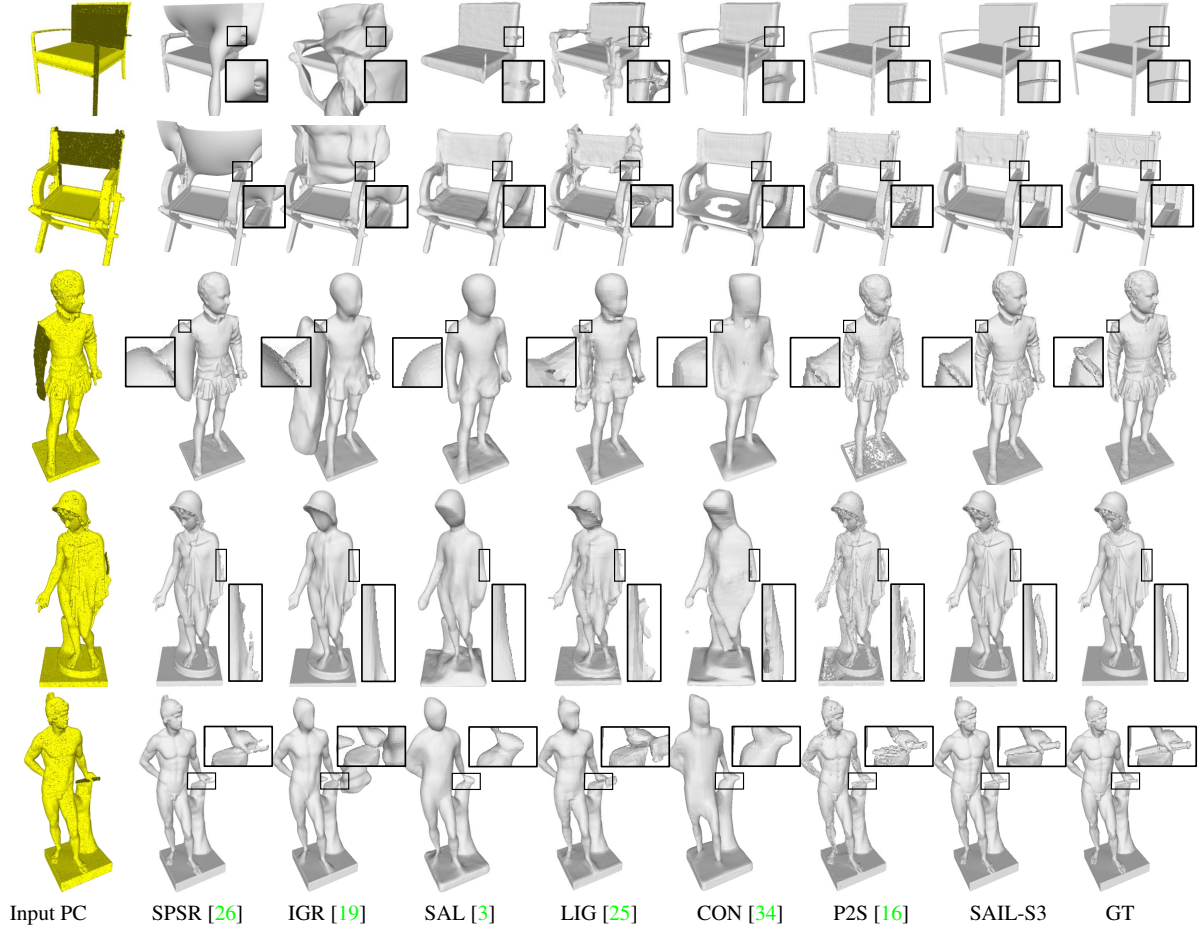
6. Acknowledgement

Figure 4: Qualitative results of different methods on *chair* instances in ShapeNet [9] (top two rows) and sculptures in ThreeDScans [1] (bottom three rows). Black points on the five 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. Refer to the supplementary material for more qualitative results.
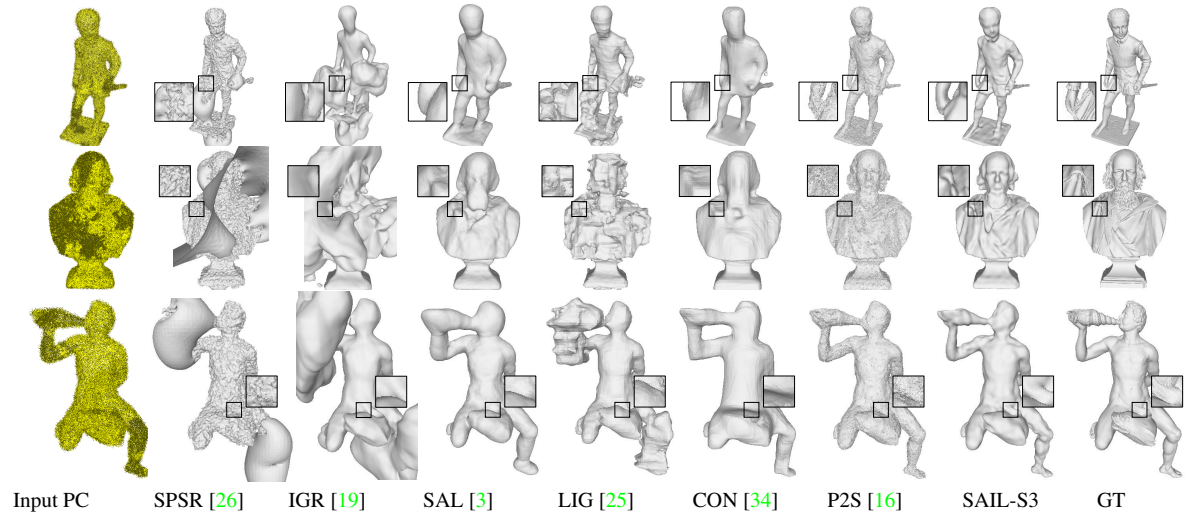

Figure 5: Qualitative results of different methods when adding point-wise Gaussian noise of standard deviation 0.01 to input points of sculptures in ThreeDScans [1]. Black points on the three inputs denote incorrect estimations of normal orientations. Note that IGR and SAL belong to global fitting methods, and LIG, CON, and P2S belong to locally learned methods.

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, 6, 7, 8

[2] Marc Alexa, Johannes Behr, Daniel Cohen-Or, Shachar Fleishman, David Levin, and Claudio T. Silva. Computing and rendering point set surfaces. *IEEE Transactions on visualization and computer graphics*, 9(1):3–15, 2003. 1, 3

[3] 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. 2, 3, 4, 7, 8

[4] Matan Atzmon and Yaron Lipman. Sald: Sign agnostic learning with derivatives, 2020. 3, 5

[5] 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 *Computer Graphics Forum*, volume 36, pages 301–329. Wiley Online Library, 2017. 1, 3

[6] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005. 2, 3

[7] 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*, pages 67–76, 2001. 1, 2, 3

[8] Rohan Chabra, Jan Eric Lenssen, Eddy Ilg, Tanner Schmidt, Julian Straub, Steven Lovegrove, and Richard Newcombe. Deep local shapes: Learning local sdf priors for detailed 3d reconstruction. In *Proceedings of the European Conference on Computer Vision*, 2020. 2, 3

[9] 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. 2, 6, 7, 8

[10] Zhiqin Chen, Kangxue Yin, Matthew Fisher, Siddhartha Chaudhuri, and Hao Zhang. Bae-net: Branched autoencoder for shape co-segmentation. *Proceedings of International Conference on Computer Vision (ICCV)*, 2019. 2, 3

[11] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5939–5948, 2019. 1, 3

[12] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. Meshlab: an open-source mesh processing tool. In *Eurographics Italian chapter conference*, volume 2008, pages 129–136. Salerno, 2008. 7

[13] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising with block-matching and 3d filtering. In *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, volume 6064, page 606414. International Society for Optics and Photonics, 2006. 2, 3

[14] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. 2, 3

[15] Yuval Eldar, Michael Lindenbaum, Moshe Porat, and Yehoshua Y Zeevi. The farthest point strategy for progressive image sampling. *IEEE Transactions on Image Processing*, 6(9):1305–1315, 1997. 5

[16] 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, 7, 8

[17] Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 605–613, 2017. 6

[18] Kyle Genova, Forrester Cole, Daniel Vlasic, Aaron Sarna, William T. Freeman, and Thomas Funkhouser. Learning shape templates with structured implicit functions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 2, 3

[19] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. *arXiv preprint arXiv:2002.10099*, 2020. 7, 8

[20] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. AtlasNet: A Papier-Mâché Approach to Learning 3D Surface Generation. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1, 3

[21] Paul Guerrero, Yanir Kleiman, Maks Ovsjanikov, and Niloy J Mitra. Pcpnet learning local shape properties from raw point clouds. In *Computer Graphics Forum*, volume 37, pages 75–85. Wiley Online Library, 2018. 2, 4

[22] Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. Meshcnn: a network with an edge. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019. 3

[23] Rana Hanocka, Gal Metzer, Raja Giryes, and Daniel Cohen-Or. Point2mesh: A self-prior for deformable meshes. *arXiv preprint arXiv:2005.11084*, 2020. 2, 3

[24] 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*, pages 71–78, 1992. 4, 6, 7

[25] 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, 7, 8

[26] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013. 1, 2, 3, 7, 8

[27] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[28] Jiabao Lei and Kui Jia. Analytic marching: An analytic meshing solution from deep implicit surface networks. In *International Conference on Machine Learning 2020 ICML-20*, 7 2020. 3

[29] David Levin. Mesh-independent surface interpolation. In *Geometric modeling for scientific visualization*, pages 37–49. Springer, 2004. 2

[30] 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. 3, 6

[31] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 3

[32] Junyi Pan, Xiaoguang Han, Weikai Chen, Jiapeng Tang, and Kui Jia. Deep mesh reconstruction from single rgb images via topology modification networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9964–9973, 2019. 3

[33] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 3, 4

[34] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *European Conference on Computer Vision*, 2020. 2, 3, 7, 8

[35] Robert Clay Prim. Shortest connection networks and some generalizations. *The Bell System Technical Journal*, 36(6):1389–1401, 1957. 6

[36] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, pages 5099–5108, 2017. 6

[37] Jiapeng Tang, Xiaoguang Han, Junyi Pan, Kui Jia, and Xin Tong. A skeleton-bridged deep learning approach for generating meshes of complex topologies from single rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4541–4550, 2019. 3

[38] Jiapeng Tang, Xiaoguang Han, Mingkui Tan, Xin Tong, and Kui Jia. Skeletonnet: A topology-preserving solution for learning mesh reconstruction of object surfaces from rgb images. *arXiv preprint arXiv:2008.05742*, 2020. 3

[39] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 52–67, 2018. 3