Self-Supervised Implicit Attention Priors for Point Cloud Reconstruction

Kyle Fogarty
University of Cambridge

Chenyue Cai
Princeton University

Jing Yang
University of Cambridge

Zhilin Guo
University of Cambridge

Cengiz Öztireli
University of Cambridge

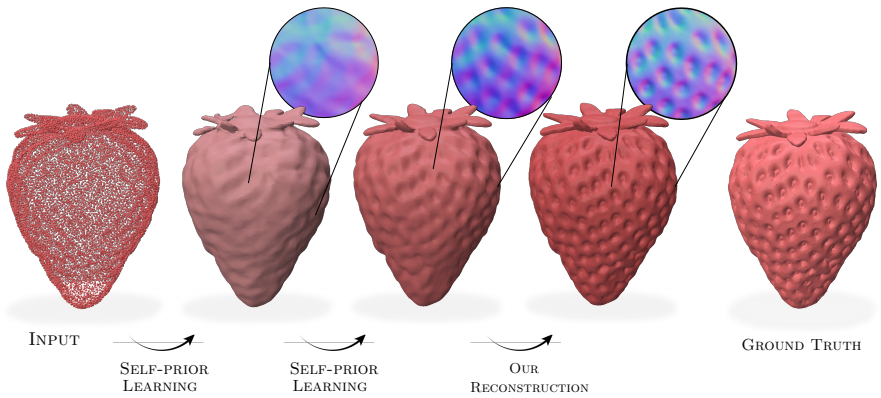

Figure 1. We propose a novel approach that incorporates a shape-specific self-prior for reconstructing high-fidelity surfaces from irregular point clouds. It *iteratively* leverages shape-specific priors via cross-attention with a compact, learnable dictionary, capturing repeating structures without external training data.

Abstract

Recovering high-quality surfaces from irregular point cloud is ill-posed unless strong geometric priors are available. We introduce an implicit self-prior approach that distills a shape-specific prior directly from the input point cloud itself and embeds it within an implicit neural representation. This is achieved by jointly training a small dictionary of learnable embeddings with an implicit distance field; at every query location, the field attends to the dictionary via cross-attention, enabling the network to capture and reuse repeating structures and long-range correlations inherent to the shape. Optimized solely with self-supervised point cloud reconstruction losses, our approach requires no external training data. To effectively integrate this learned prior while preserving input fidelity, the trained field is then sampled to extract densely distributed points and analytic normals via automatic differentiation. We integrate the resulting dense point cloud and corresponding normals into a robust implicit moving least squares (RIMLS) formulation. We show this hybrid strategy preserves fine geometric details in the input data, while leveraging the learned prior to regularize sparse regions. Experiments show that our method outperforms both classical and learning-based approaches in generating high-fidelity surfaces with superior detail preservation and robustness to common data degradations.

1. Introduction

Recovering continuous surface geometry from discrete, unstructured point clouds is a fundamental problem in computer graphics and 3D vision, crucial for numerous downstream applications. However, the inherent challenges of noise, sparsity, and outliers render this task ill-posed. Consequently, successful reconstruction algorithms must incorporate a *prior* to regularize the problem and guide the surface reconstruction process [4]. These priors represent assumptions about the underlying geometry, influencing the trade-offs between fidelity to the input data and the plausibility of the resulting surface.

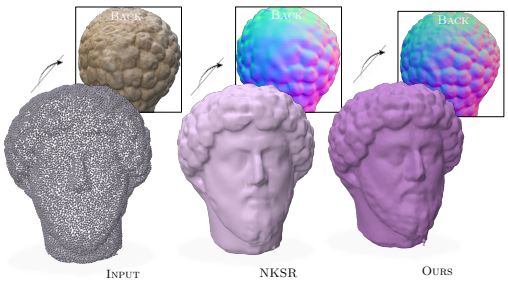

Figure 2. Comparison with NKSR [15] on the 'Bust of Marcus Aurelius' [41]. Though NKSR is trained on a large 3D dataset and configured for maximum detail preservation, our method better captures fine surface details, particularly at the back of the head.

A common strategy is to impose a global smoothness prior, as done in Poisson Surface Reconstruction (PSR) [17, 18], which solves a partial differential equation to produce globally smooth surfaces. Although effective for many objects and robust thanks to its global solution, approaches that rely solely on smoothness are limited: such priors often struggle to preserve sharp features and to leverage the intricate structural patterns and repetitive details common in real-world geometry. This motivates the development of priors capable of capturing richer geometric structure. The success of non-local methods in image processing [5, 7], which exploit self-similarity by connecting distant yet structurally alike patches, highlights the potential of extending this idea to geometry. Since real-world objects often exhibit strong internal repetition, self-similarity offers a compelling and intuitive geometric prior.

This naturally raises the question of how to effectively leverage self-similarity for 3D reconstruction. Point2Mesh [13] addressed this by learning a self-prior directly on an explicit mesh, sharing MeshCNN kernel weights across the surface to predict displacement vectors that refine the mesh to fit the point cloud. While this approach powerfully demonstrated the utility of learned self-similarity for capturing intricate details on a given mesh topology, its reliance on deforming an *explicit* surface representation fundamentally limits topological flexibility (see Figure 3), can blur sharp features, and is less suited for arbitrary shapes or reconstructing directly from unoriented point clouds. This motivates the use of alternative representations, such as implicit functions, which are continuous and inherently support more flexible surface topologies.

While implicit surface representations, such as Signed Distance Functions (SDFs), offer a powerful and flexible framework for surface reconstruction, incorporating a learned self-prior into these models remains challenging. Classical approaches like Implicit Moving Least Squares (IMLS) [33, 39] excel at producing surface reconstructions but cannot exploit structural cues beyond point neighborhoods. More recently, neural implicit methods have emerged as a promising alternative, representing surfaces as continuous fields parameterized by neural networks. Several works have adapted these models for unstructured point cloud reconstruction without ground-truth SDFs, leveraging geometric constraints such as the Eikonal loss [10], sign-agnostic supervision [2], or gradient-based point projection [30]. Others introduce local surface priors by operating on overlapping patches [9] or regularize the field using MLS-inspired smoothness terms [43]. While these approaches achieve impressive reconstructions and accommodate complex topologies, the priors they impose, whether based on smoothness, local patches, or global latent codes, are inherently limited in their ability to capture compact shape specific features (see Fig. 2).

To bridge this gap, we draw inspiration from signal processing, where *dictionaries* [8] represent complex signals as sparse combinations of shared *atoms*, effectively capturing structure by exploiting internal redundancy. This concept aligns powerfully with geometric self-similarity. We propose that a learnable dictionary can effectively encode such self-priors within a neural field. Attention mechanisms [42] provide a natural and flexible means to dynamically query and aggregate information from these dictionary elements to inform local surface predictions. Indeed, recent successes in generative 3D modeling, such as 3DShape2VecSet [46], have demonstrated the power of set-latent approaches using attention for high-fidelity 3D shape representation, suggesting their promise for reconstruction tasks as well.

Building on learned self-priors and attention mechanisms for implicit representations, we introduce a novel ap-

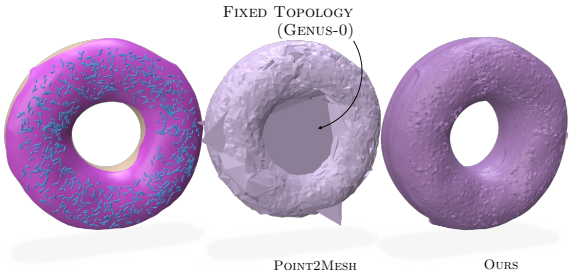

Figure 3. Deformation-based reconstruction methods, such as Point2Mesh, are limited by their reliance on a fixed input topology. In contrast, our *implicit* approach provides greater flexibility in representing complex geometries.

proach for surface reconstruction. Our method learns a neural field that implicitly captures a distance field corresponding to the input point cloud, trained using self-supervised loss functions. Crucially, we leverage a learned dictionary and cross-attention to enable the field to recognize and exploit non-local structure, effectively achieving a form of spatial weight sharing. This learned self-prior allows the resulting distance field to guide the densification of sparse input regions. We show that this field can be further refined through a robust moving-least-squares (MLS) surface approximation, enabling the extraction of high-fidelity surfaces with rich geometric detail and flexible topology. Our main contributions are: (1) a self-supervised implicit framework that learns a shape-specific geometric prior from the input point cloud via cross-attention to a learned dictionary, enabling non-local structure modeling; (2) demonstration that the learned field can be effectively refined using Robust Implicit MLS, yielding accurate, flexible surface reconstructions; and (3) state-of-the-art performance on self-similar shapes, highlighting the benefits of our attention-based self-prior.

2. Related Work

Classical Implicit Surface Reconstruction Reconstructing continuous surfaces from discrete point clouds is a foundational challenge in 3D vision and computer graphics [3]. Classical implicit methods typically represent surfaces as level sets of scalar functions, with Poisson Surface Reconstruction (PSR) being one of the most prominent examples [17, 18]. PSR is well-regarded for producing smooth, watertight meshes when provided with high-quality normal estimates; however, its performance deteriorates in the presence of noisy data and may excessively smooth out fine details. Another influential approach is Implicit Moving Least Squares (IMLS) [19, 21], which defines an implicit function by locally fitting polynomials to the point cloud. IMLS is capable of preserving geometric detail and exhibits robustness to non-uniform sampling, but it remains heavily dependent on the availability of accurate and consistently oriented normals, a significant constraint when working with raw point clouds.

Neural Implicit Representations In recent years, deep neural networks have revolutionized 3D surface reconstruction. Neural implicit functions, such as signed distance fields (SDFs) and occupancy fields parameterized by MLPs, were first introduced as data-driven shape representations. Notable examples include DeepSDF [34] and Occupancy Networks [31], which learn continuous volumetric functions by training on large shape datasets. While demonstrating the power of learned priors for complex topologies and shape interpolation, these early models typically rely on signed distances or occupancy labels for supervision, making them less robust to raw, unoriented, or noisy point clouds acquired in practice. Moreover, the learned shape spaces can struggle with out-of-distribution geometry, and the MLP representation itself can impose an implicit smoothness bias [40].

To eliminate the need for ground-truth signed distances, a subsequent wave of methods trains neural implicit fields directly on the input point cloud in a self-supervised manner. Implicit Geometric Regularization (IGR) fits an MLP by encouraging it to vanish on input points and have unit gradient norm (Eikonal loss) [10]. Sign Agnostic Learning (SAL) devises loss functions invariant to the sign of the distance, enabling distance learning from raw unoriented points [2]. NeuralPull [27] introduced an explicit point-to-surface pulling loss, significantly improving reconstruction accuracy. While these self-supervised approaches (e.g., [2, 10, 27]) avoid large training sets, they mainly exploit local surface priors or collapse the shape into a single global latent code. As highlighted by [13], such priors may oversimplify or miss fine recurring details, as they do not explicitly model non-local self-similarity where repeating geometry could be leveraged.

Learned Geometric Priors for Reconstruction Recent advances have therefore explored more expressive learned priors. Some methods leverage external datasets: for instance, this can be done by pre-training models and then specializing them to new instances by optimizing query points [29], or by specializing a decoder on local patches and projecting queries onto a learned surface via the decoder [28]. Neural Kernel surface reconstruction (NKSR) [15] learns a multi-scale prior for fast reconstruction of large scenes. In contrast, we aim to learn a *self-prior* derived from the input itself. Deep Geometric Prior (DGP) [45] fits an ensemble of small neural implicits to local patches, relying on the network's bias and consistency constraints, but lacks explicit global weight sharing. Point2Mesh [13] explicitly learns a self-prior by optimizing a MeshCNN [12] to deform an initial explicit mesh, effectively using shared convolutional kernels for self-similarity. Chu et al. [6] adopt the deep prior paradigm, completing shapes from a single instance using a CNN interpreted through the neural tangent kernel (NTK) framework. Their method captures global self-similarity implicitly via shared features and NTK-informed architectural choices. However, it does not operate within a neural implicit surface framework or explicitly model reusable geometric patterns, leaving the explicit handling of global self-similarity in implicit representations largely underexplored.

Attention-Driven Shape Representations Our approach employs a learned dictionary of geometric tokens, accessed

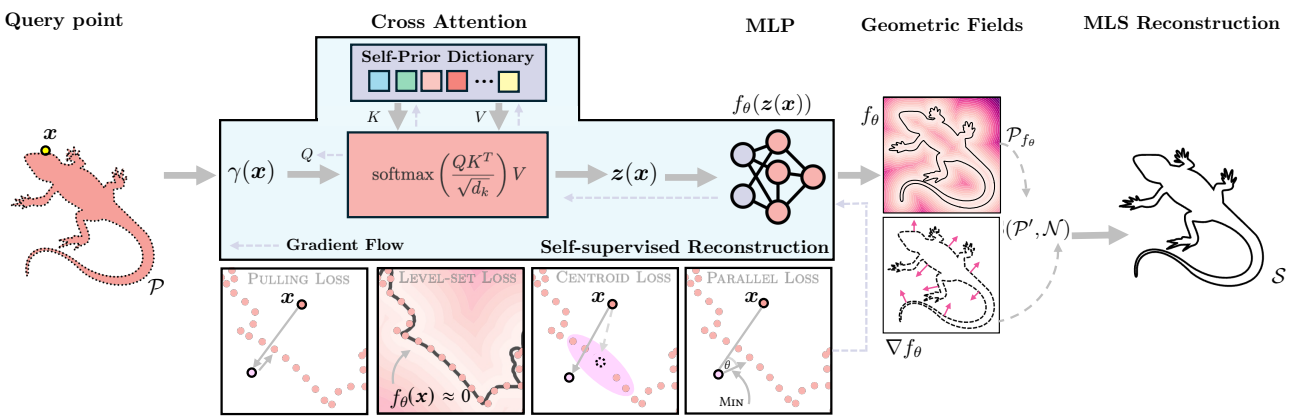

Figure 4. We present a self-supervised surface reconstruction method based on a neural field conditioned via cross-attention. We assume access to an unoriented point cloud $\mathcal{P}$. Each input query point x is encoded via positional encoding $\gamma(x)$ and interacts with a shared, learnable embedding dictionary to produce a latent representation $z(x)$ that captures shape-specific geometric priors. An MLP then predicts geometric field, which we train with a number of geometric losses to recover the surface from the point cloud. Finally, Moving Least Squares (MLS) reconstruction is applied to refine the implicit surface.

via an attention mechanisms, to implement a non-local self-prior. Attention has proven effective in enhancing the generalization of neural fields by conditioning them on sets of latent tokens [16, 36, 42]. Notably, set-latent representations leveraging attention, such as 3DShape2VecSet [46], have achieved high-fidelity 3D shape modeling by learning over collections of latent features. Inspired by these advancements, we introduce a framework where a neural field, conditioned by a learnable dictionary of geometric tokens through cross-attention, implicitly learn a non-local self-prior.

3. Our Approach

We address the problem of reconstructing high-fidelity 3D surfaces from point clouds through a two-stage pipeline (see in Figure 4). The core motivation behind our method is that self-similar patterns in local surface patches frequently recur across different regions of an object's surface. Our method is designed to learn inherent self-prior and is able to infer missing or inaccurate geometric details.

(Stage 1) : Learning a neural distance field with implicit self-prior: We train an MLP f_θ to approximate the neural SDF field of the target shape. The self-prior is encoded within a compact learnable dictionary. For each query point $\boldsymbol{x}$, we perform cross-attention between its encoded position and dictionary entries to produce a feature representation $\boldsymbol{z}(\boldsymbol{x})$. We input $\boldsymbol{z}(\boldsymbol{x})$ to f_θ to predict its SDF. The surface is defined as the zero-level set of the learned SDF field, while per-point normals follow from its spatial gradient.

(Stage 2) : Geometric projection. We discretize the learned geometric field and employ Robust Implicit Moving Least Squares (RIMLS) to define the final shape. This refinement step leverages the expressive capacity of f_θ alongside the feature-preserving properties of RIMLS, resulting in reconstructions that are both globally consistent and rich in detail.

3.1. Dictionary-Conditioned Neural Field

A neural field is a continuous function, typically parameterized by a Multi-Layer Perceptron (MLP), that maps input coordinates to some target property. In our case, we aim to represent a continuous 3D shape by the level sets of an implicit distance function. Neural fields typically condition their predictions solely on spatial coordinates, requiring the network to reconstruct the entire shape from strictly local information. This localized perspective makes it challenging to capture long-range symmetries, repeated structures, and other global regularities that are implicitly shared across shapes. To overcome this limitation, we augment each point query via cross-attention to a shape-specific dictionary of learned embeddings. Because the same dictionary is accessible to all coordinates, the model can exchange information across distant regions and exploit the object's structure.

Cross-Attention Dictionary: More specifically, we adopt a decoder-only architecture, with an overview shown in Fig. 4. We begin by orthogonally initializing the embedding dictionary $\mathbf{E} \in \mathbb{R}^{N_k \times d_e}$, via QR decomposition of a random matrix [37], to foster initial feature diversity and enhance learning stability. This dictionary consists of N_k latent feature vectors, which are optimized jointly with the rest of the model parameters during training.

For a query point $\boldsymbol{x} \in \mathbb{R}^3$, we apply sinusoidal positional [40] encoding $\gamma(\boldsymbol{x})$ yielding a query vector $\gamma(\boldsymbol{x}) \in \mathbb{R}^{d_q}$. We then linearly project the query position to yield $\mathbf{q} = W_\gamma \gamma(\boldsymbol{x})$, such that $\mathbf{q} \in \mathbb{R}^{d_e}$. To perform cross-attention between the query position $\boldsymbol{x}$ and the embedding dictionary $\mathbf{E}$, we apply *multi-headed attention* (MHA) with H heads. For each head $h = 1, \ldots, H$, the query, keys, and values are linearly projected via learned matrices $(\mathbf{W}_q^{(h)}, \mathbf{W}_k^{(h)}, \mathbf{W}_v^{(h)})$ such that:

$$\mathbf{Q}_h = \mathbf{q}\mathbf{W}_q^{(h)}, \quad \mathbf{K}_h = \mathbf{E}\mathbf{W}_k^{(h)}, \quad \mathbf{V}_h = \mathbf{E}\mathbf{W}_v^{(h)}. \quad (1)$$

Scaled dot-product attention is computed independently for each head:

$$\mathrm{Attn}(\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h) = \mathrm{softmax}\left(\frac{\mathbf{Q}_h \mathbf{K}_h^\top}{\sqrt{d_k}}\right)\mathbf{V}_h, \quad (2)$$

where d_k is the key dimensionality per head. The outputs from all heads are concatenated and projected through $\mathbf{W}_o$ to produce the final context vector $\boldsymbol{z}(\boldsymbol{x}) \in \mathbb{R}^{d_{\text{out}}}$:

$$\boldsymbol{z}(\boldsymbol{x}) = \mathrm{Concat}(\mathrm{Attn}_1, \ldots, \mathrm{Attn}_H)\mathbf{W}_o. \quad (3)$$

Signed-distance Prediction Head: To preserve fine-grained spatial information, we introduce a learned linear projection of the raw coordinates, defined as $\tilde{\boldsymbol{x}} = W_{\text{proj}}\boldsymbol{x}$, where $\tilde{\boldsymbol{x}} \in \mathbb{R}^{d_{\text{out}}}$. This projected signal is added to the context vector, yielding the final input $\bar{\boldsymbol{z}}(\boldsymbol{x}) = \boldsymbol{z}(\boldsymbol{x}) + \tilde{\boldsymbol{x}}$. We consider a multilayer perceptron (MLP) $f_\theta : \mathbb{R}^{d_{\text{out}}} \to \mathbb{R}$ with L hidden layers of width d_{hid} and ReLU activations. To ease optimization in deeper networks, we follow [23] and add a single skip connection by concatenating the original input to the intermediate representation after the $\lfloor L/2 \rfloor$-th layer. Following [1, 10], we adopt geometric initialization to encourage signed distance function (SDF) behavior. Specifically, (i) weights in the hidden layers are initialized from a normal distribution $\mathcal{N}(0,\, 2/d_{\text{hid}})$; and (ii) the final layer is initialized with zero bias and small weights, ensuring outputs are near zero at initialization and promoting stable training.

3.2. Training the Neural Field

We denote the full attentive neural field $f_\theta(\bar{\boldsymbol{z}}(\boldsymbol{x}))$ by $g_\phi(\boldsymbol{x})$. Following [23, 27], training uses two complementary supervision sets derived from the input point cloud $\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^N$: off-surface points $\mathcal{Q}$, obtained by adding Gaussian noise $\boldsymbol{\delta} \sim \mathcal{D}$ to uniformly sampled $\mathbf{p} \in \mathcal{P}$, and on-surface points $\mathcal{G}$, directly subsampled from $\mathcal{P}$ to anchor the zero-level set. The complete training set is $\mathcal{T} = \mathcal{Q} \cup \mathcal{G}$. We estimate the unit normal:

$$\boldsymbol{\nu}(\boldsymbol{x}) = \frac{\nabla_{\boldsymbol{x}} g_\phi(\boldsymbol{x})}{\|\nabla_{\boldsymbol{x}} g_\phi(\boldsymbol{x})\|},$$

and define the projection operator:

$$\mathscr{P}(\boldsymbol{x}) = \boldsymbol{x} - g_\phi(\boldsymbol{x}) \cdot \boldsymbol{\nu}(\boldsymbol{x}),$$

which moves points toward the surface; $\mathscr{P}_m$ denotes m successive applications. We employ established geometric losses from prior neural distance function works [23, 27, 48]:

Global Surface Loss: Enforces that projected points lie close to surface samples:

$$\mathcal{L}_\alpha = \mathbb{E}_{\mathbf{q}\sim\mathcal{Q}}\left[\|\mathscr{P}_2(\mathbf{q}) - \hat{\mathbf{q}}\|^2\right] + \mathbb{E}_{\mathbf{g}\sim\mathcal{G}}\left[\|\mathscr{P}_2(\mathbf{g}) - \mathbf{g}\|^2\right],$$

where $\hat{\mathbf{q}}$ is the nearest neighbor of $\mathbf{q}$ in $\mathcal{P}$. This pulls off-surface points toward the data and anchors on-surface points.

Level-set Loss: Encourages $g_\phi \approx 0$ for on surface points and also after one applications of the projection operator:

$$\mathcal{L}_\beta = \mathbb{E}_{\mathbf{g}\sim\mathcal{G}}\left[g_\phi(\mathbf{g})^2\right] + \mathbb{E}_{\mathbf{t}\sim\mathcal{T}}\left[g_\phi(\mathscr{P}(\mathbf{t}))^2\right].$$

The first term constrains known surface samples and the second regularises refined points.

Local Displacement Loss: Aligns predicted displacements with local geometric estimates across scales s:

$$\mathcal{V}_s(\mathbf{q}) = \mathbf{q} - \frac{1}{K_s}\sum_{k=1}^{K_s}\mathbf{p}_k.$$

Here, $\mathcal{V}_s(\mathbf{q})$ denotes the displacement from $\mathbf{q}$ to the centroid of its K_s nearest neighbours in $\mathcal{P}$. The local displacement loss is:

$$\mathcal{L}_\gamma = \sum_s \mathbb{E}_{\mathbf{q}\sim\mathcal{Q}}\left[\|(\mathbf{q} - \mathscr{P}_1(\mathbf{q})) - \mathcal{V}_s(\mathbf{q})\|^2\right],$$

which promotes consistency across varying point densities.

Normal Consistency Loss: Encourages stable surface normals during refinement:

$$\mathcal{L}_\delta = \mathbb{E}_{\mathbf{x}\sim\mathcal{T}}\left[w(\mathbf{x}) \cdot \left(1 - d_{\cos}(\boldsymbol{\nu}(\mathbf{x}), \boldsymbol{\nu}(\mathscr{P}(\mathbf{x})))\right)\right],$$

with $w(\mathbf{x}) = \exp(-\rho|g_\phi(\mathbf{x})|)$. The loss enforces minimal change in normals as points move closer to the surface.

Total Loss The training objective is a weighted sum of all terms:

$$\mathcal{L} = \alpha\,\mathcal{L}_\alpha + \beta\,\mathcal{L}_\beta + \gamma\,\mathcal{L}_\gamma + \delta\,\mathcal{L}_\delta,$$

where $\alpha, \beta, \gamma, \delta$ control the relative influence of global alignment, level-set consistency, local displacement, and normal smoothness, respectively.

3.3. Geometric Quantity Estimation

We use the learnt implicit field g_ϕ both to lightly inpaint sparse regions and to obtain normals; this allows the self-prior to guide the surface reconsturction as in Fig. 5.

Inpainting We extract the zero level set of g_ϕ with Marching Cubes [26] and uniformly sample it to obtain a dense auxiliary set $\tilde{\mathcal{P}}$. For each $\tilde{\boldsymbol{p}}_j \in \tilde{\mathcal{P}}$, let $d(\tilde{\boldsymbol{p}}_j, \mathcal{P}) = \min_{\boldsymbol{p}\in\mathcal{P}} \|\tilde{\boldsymbol{p}}_j - \boldsymbol{p}\|_2$ be its nearest-neighbour distance to the input cloud $\mathcal{P}$ (i.e., a one-sided Chamfer term). With σ_d the standard deviation of $\{d(\tilde{\boldsymbol{p}}_j, \mathcal{P})\}$, we keep only far points $\mathcal{P}_{\text{fill}} = \{\tilde{\boldsymbol{p}}_j \mid d(\tilde{\boldsymbol{p}}_j, \mathcal{P}) \geq 3\sigma_d\}$. The augmented set is then given by $\mathcal{P}' = \mathcal{P} \cup \mathcal{P}_{\text{fill}}$.

Normals For each $\boldsymbol{p}_i \in \mathcal{P}'$, we estimate a normal by the normalized gradient of the field, $\boldsymbol{n}_i = \nabla g_\phi(\boldsymbol{p}_i)/\|\nabla g_\phi(\boldsymbol{p}_i)\|_2$, and denote the full set of normals as $\mathcal{N} = \{\boldsymbol{n}_i\}$.

MLS Refinement To move beyond the stability–detail trade-off inherent in direct reconstruction from sparse, uneven point clouds, we refine the surface using Robust Implicit Moving Least Squares (RIMLS) [33]. The field g_ϕ first inpaints gaps and provides coherent normals, yielding $(\mathcal{P}', \mathcal{N})$ as an enhanced point set. RIMLS then reconstructs the surface while preserving sharp features and fine detail, guided by these refinements. Finally, we evaluate the implicit function on a grid and extract the mesh using Marching Cubes [26].

4. Experiments

We evaluate of our method through a series of qualitative and quantitative experiments involving shapes with low density regions, noise, and different topologies. We provide implementation details in section 7 of the supplementary materials.

4.1. Experimental Setup

Datasets: We evaluate our method on four datasets. First, we use the Surface Reconstruction Benchmark (SRB) [3], which contains five range-scan models and is a standard dataset for surface reconstruction. Second, we curate a set of objects with strong self-similarity to assess performance on inputs with repeated structure. Third, we test robustness using a subset of Thingi10K [49], specifically a variant containing noise from [9], which include Gaussian noise to simulate sensor imperfections. Finally, we evaluate on the full set of models provided in the public release of Point2Mesh [13].

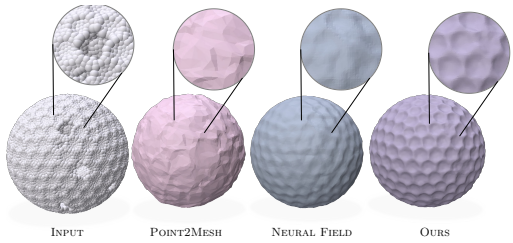

Figure 5. Our method combines a learned self-prior with explicit point cloud control to preserve surface detail, outperforming approaches that rely solely on learned priors (Point2Mesh) or neural fields without attention.

Comparison: We evaluate our method against a broad spectrum of surface reconstruction techniques, including analytical, optimization-based, and learning-driven approaches. Analytical baselines include Screened Poisson Surface Reconstruction (SPSR) [17], which produces smooth, complete surfaces under clean input conditions, but is sensitive to noise and requires oriented normals, and Diffusing Winding Gradients (DWG) [25], a recent non-learning method that reconstructs from unoriented point clouds via diffusion of generalized winding number gradients, offering strong scalability but lacking learned priors. Optimization-based methods such as Shape-as-Points (SAP) [35] formulate classical objectives as differentiable losses, minimizing Chamfer distance in a Poisson-inspired setting. Learning-based approaches include Neural Kernel Surface Reconstruction (NKSR) [15], which learns transferable shape priors; Point2Mesh (P2M) [13], which learns per-shape self-priors through mesh deformation with fixed connectivity; and other techniques like Deep Geometric Prior (DGP) [45], Neural-IMLS (NIMLS) [43], and Predictive Context Prior (PCP) [29], which differ in supervision and prior modeling strategies. We also compare against recent neural field–based methods, including PG-SDF [20] and Neural Singular Hessian [44], which define implicit surfaces using continuous neural fields trained directly on point clouds.

4.2. Experimental Results

SRB Dataset: Table 1 reports the reconstruction metrics of our method on the SRB dataset. Our approach achieves state-of-the-art performance across multiple evaluation criteria. Notably, it yields the lowest Chamfer Distance and Hausdorff Distance, indicating superior geometric accuracy and surface completeness. Qualitative reconstruction results are provided in Sec. 10 of the appendix.

Self-similar Dataset: Table 2 further demonstrates the effectiveness of our method in reconstructing shapes characterized by strong self-similarity. Our approach achieves the best performance across Chamfer Distance, Normal Consistency, and F-score, highlighting its robustness and accuracy in challenging reconstruction scenarios. Qualitative comparisons are shown in Fig. 6.

Thingi10K Dataset: Table 3 highlights the robustness of

Table 1. Surface reconstruction metrics on the SRB dataset, proposed in [3]. Lower is better for CD and HD; higher is better for NC and F-score.

METHOD	CD (↓)	HD (↓)	NC (↑)	FS (↑)
SPSR	0.413	1.498	0.919	71.63
DGP	0.022	0.701	0.951	75.67
P2M	0.177	0.902	0.857	24.47
PCP	0.283	2.039	0.900	49.39
NIMLS	0.283	1.992	0.913	54.62
NKSR	0.019	0.614	0.949	75.98
SAP	0.024	0.682	0.936	75.49
Ours	0.016	0.484	0.956	75.54

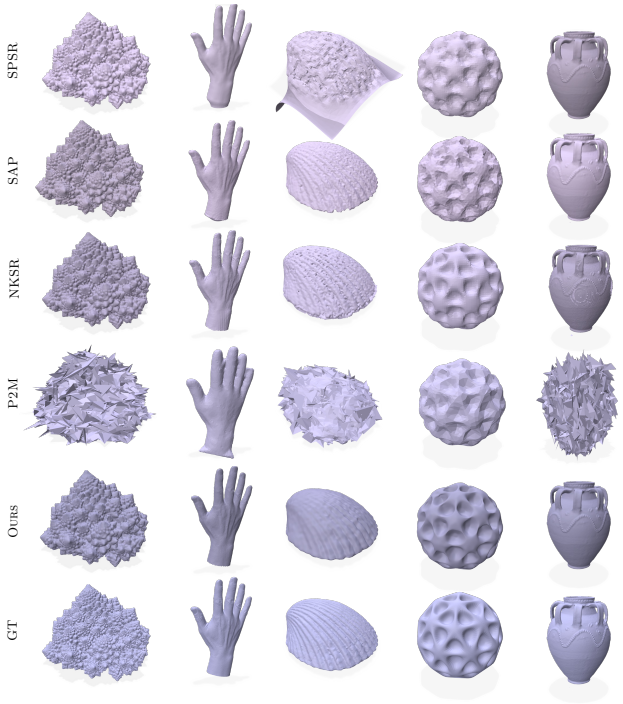

Figure 6. We present qualitative comparison between our method and other leading reconstruction methods on shapes with high amounts of self-similarity; Our approach excels at capturing global shape properties while retaining local shape details.

Table 2. Surface reconstruction metrics on our dataset consisting of objects with large self-similarity. Lower is better for CD and HD; higher is better for NC and F-score.

METHOD	CD (↓)	HD (↓)	NC (↑)	FS (↑)
SPSR	0.248	1.475	0.866	61.89
SAP	0.021	0.690	0.906	71.21
NKSR	0.019	0.512	0.897	68.56
P2M	0.239	1.384	0.695	19.24
NSH	0.043	0.971	0.859	61.95
WDG	0.038	1.246	0.785	48.39
PG-SDF	0.019	0.397	0.872	66.38
Ours	0.017	0.438	0.907	72.44

Table 3. Surface reconstruction metrics over samples from the Thingi10K dataset. Lower is better for CD and HD; higher is better for NC and F-score.

METHOD	CD (↓)	HD (↓)	NC (↑)	FS (↑)
SPSR	0.032	0.620	0.899	67.31
SAP	0.022	0.385	0.734	62.77
NKSR	0.019	0.398	0.939	70.50
PG-SDF	0.151	0.596	0.860	1.72
Ours	0.021	0.458	0.931	63.71

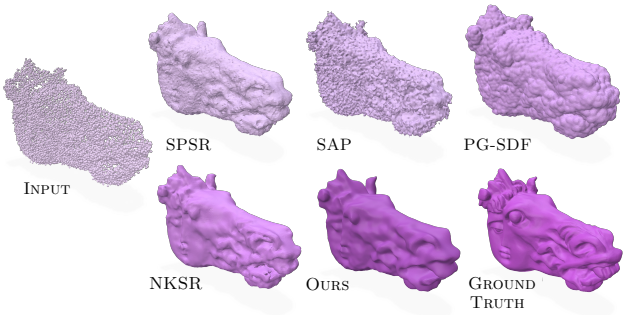

Figure 7. Comparison on a Thinki10K sample with added Gaussian noise. Even with noise, our method preserves fine-grained details and similar patterns better than competing approaches.

our method on the noised Thingi10K dataset. While NKSR achieves the top performance due to explicit noise-aware training, our approach ranks second in Chamfer Distance and Normal Consistency, demonstrating strong resilience to noise and sparsity. As shown in Fig. 7, our method effectively suppresses input noise while preserving fine geometric details.

Point2Mesh Dataset: Since Point2Mesh is the most directly comparable method in terms of learning a self-prior for surface reconstruction, we further evaluate our approach on the publicly available dataset provided by its authors. The results, shown in Fig. 11 of the appendix, indicate that while both methods capture the underlying geometry, our approach produces smoother surfaces and better preserves sharp features.

Interpretability: To better understand the behavior of our attention mechanism, we visualize the attention weight similarity across the surface relative to a selected query point. We train our model on the strawberry point cloud with a dictionary size of 16. In Fig. 8, the similarity is measured via the dot product of attention weights, where

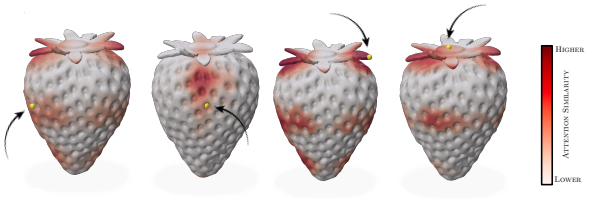

Figure 8. Attention weight similarity across the surface, relative to a yellow query point. Similarity (white: low, red: high) is computed via dot product of attention weights.

warmer colors (red) indicate higher similarity and cooler colors (white) denote lower similarity. We observe that the model learns to couple non-local regions; for example, when the query is located on or near the leaf structure, the attention shifts to highlight other leaf regions, despite their spatial separation. This behavior indicates that the model learns a meaningful self-prior, effectively linking similar but spatially distant regions of the shape.

Ablation studies: We evaluate the contributions of the RIMLS refinement and the attentive dictionary through ablation studies on the self-similar dataset. To this end, we compare three configurations: (i) the full model, (ii) a variant without RIMLS refinement, and (iii) a variant without both RIMLS and attention (No MLS + No Attn); for the latter we increase the parameter count of the neural field to approximately match the full model's parameter count. Quantitative results for all configurations are reported in Table 4. The base model without attention or MLS refinement has the weakest performance across all metrics. Introducing attention substantially improves reconstruction quality, while the addition of RIMLS refinement yields further gains, demonstrating the complementary benefits of both components.

Table 4. Ablations on the self-similar dataset. Lower is better for CD and HD; higher is better for NC and F-score.

Attn	MLS	Params	**CD (↓)**	**HD (↓)**	**NC (↑)**	**FS (↑)**
✗	✗	1.18 M	0.021	0.534	0.876	66.62
✓	✗	1.16 M	0.019	0.494	0.903	67.52
✓	✓	1.16 M	**0.017**	**0.438**	**0.907**	**72.44**

Figure 9 illustrates the effect of the MLS refinement. In this example, the raw zero-level set of the neural field fails to capture a thin pipe structure. By applying the refinement, the structure is successfully recovered, demonstrating its ability to preserve fine geometric details.

4.3. Normal Estimation Experiments

While designed for surface reconstruction, we evaluate our method on surface normal estimation using the PCPNet dataset, following the protocol of [23]. The results highlight the robustness of our approach across different noise levels and point cloud densities. Detailed comparisons and additional results are provided in Sec. 6.1 of the appendix.

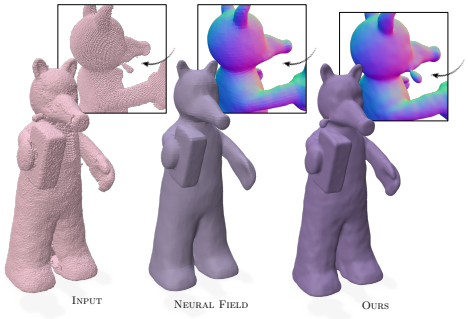

Figure 9. The zero level set of the attentive neural field may miss fine geometric details (e.g., the missing tube structure), but its gradient field still captures meaningful surface normals. Left: input point cloud. Middle: zero level set. Right: our hybrid method uses gradients for more complete reconstruction.

5. Conclusion

We introduced a self-supervised approach for high-fidelity point cloud reconstruction, leveraging an implicit attention prior. The method learns a shape-specific prior directly from the input by training an implicit neural field conditioned on a learnable dictionary of geometric tokens via cross-attention. This enables the network to capture non-local self-similarities and repeating structural patterns without external training data, guiding both sparse-region densification and high-quality analytic normal estimation. These features are integrated into a robust implicit moving least squares (RIMLS) framework, combining the global structural awareness of the learned prior with the local accuracy of classical reconstruction. Experiments suggest that our self-prior demonstrates competitive performance, showing strengths in detail preservation, topological adaptability, and robustness to noise and sparsity compared to both classical and learning-based methods. By learning complex, shape-specific priors from input alone, our approach overcomes key limitations of traditional methods and provides a flexible foundation for challenging scenarios. Future directions include extensions to dynamic or large-scale scenes, transfer learning between shapes, and generative modeling of novel shapes that inherit the structural traits of a reference.

References

[1] Matan Atzmon and Yaron Lipman. Sald: Sign agnostic learning with derivatives. *arXiv preprint arXiv:2006.05400*, 2020. 5, 2

[2] Matan Atzmon, Amos Gropp, and Yaron Lipman. SAL: Sign agnostic learning of shapes from raw point clouds. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 21499–21509, 2020. 2, 3

[3] Matthew Berger, Joshua A Levine, Luis Gustavo Nonato, Gabriel Taubin, and Claudio T Silva. A benchmark for surface reconstruction. *ACM Transactions on Graphics (TOG)*, 32(2):1–17, 2013. 3, 6, 7

[4] Matthew Berger, Andrea Tagliasacchi, Lee M Seversky, Pierre Alliez, Joshua A Levine, Andrei Sharf, and Claudio T Silva. State of the art in surface reconstruction from point clouds. In *35th Annual Conference of the European Association for Computer Graphics, Eurographics 2014-State of the Art Reports*. The Eurographics Association, 2014. 1

[5] 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)*, pages 60–65. Ieee, 2005. 2

[6] Lei Chu, Hao Pan, and Wenping Wang. Unsupervised shape completion via deep prior in the neural tangent kernel perspective. *ACM Transactions on Graphics (TOG)*, 40(3):1–17, 2021. 3

[7] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 2

[8] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006. 2

[9] Philipp Erler, Paul Guerrero, Niloy J Mitra, and Leonidas Guibas. Points2Surf: Learning implicit surfaces from point clouds. In *European Conference on Computer Vision (ECCV)*, pages 414–431. Springer, 2020. 2, 6

[10] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *International Conference on Machine Learning (ICML)*, pages 3789–3798. PMLR, 2020. 2, 3, 5

[11] Paul Guerrero, Yanir Kleiman, Maks Ovsjanikov, and Niloy J Mitra. Pcpnet learning local shape properties from raw point clouds. In *Computer graphics forum*, pages 75–85. Wiley Online Library, 2018. 1

[12] 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

[13] Rana Hanocka, Amir Hertz, Philipp Trettner, Amit Bermano, and Daniel Cohen-Or. Point2Mesh: A self-prior for deformable meshes. *ACM Transactions on Graphics (TOG)*, 39(4):1–14, 2020. 2, 3, 6, 4

[14] 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. 1

[15] Jiacheng Huang, Ruizhen Zhang, Zhiqin Jiang, Haozhi Liu, Leonidas J. Guibas, and Min Zhang. Neural kernel surface reconstruction. In *European Conference on Computer Vision (ECCV)*, pages 345–361. Springer, 2022. 2, 3, 6

[16] Wei Jiang, Eduard Trulls, Jan Hosang, Andrea Tagliasacchi, and Kwang Moo Yi. Cotr: Correspondence transformer for matching across images. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6207–6217, 2021. 4

[17] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (TOG)*, 32(3):1–13, 2013. 2, 3, 6

[18] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, pages 61–70. ACM, 2006. 2, 3

[19] Ravikrishna Kolluri. Provably good moving least squares. *ACM Transactions on Algorithms (TALG)*, 4(2):1–25, 2008. 3

[20] Chamin Hewa Koneputugodage, Yizhak Ben-Shabat, Dylan Campbell, and Stephen Gould. Small steps and level sets: Fitting neural surface models with point guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21456–21465, 2024. 6

[21] David Levin. The approximation power of moving least-squares. *Mathematics of computation*, 67(224):1517–1531, 1998. 3

[22] Qing Li, Yu-Shen Liu, Jin-San Cheng, Cheng Wang, Yi Fang, and Zhizhong Han. Hsurf-net: Normal estimation for 3d point clouds by learning hyper surfaces. *Advances in Neural Information Processing Systems*, 35:4218–4230, 2022. 1

[23] Qing Li, Huifang Feng, Kanle Shi, Yue Gao, Yi Fang, Yu-Shen Liu, and Zhizhong Han. Neuralgf: Unsupervised point normal estimation by learning neural gradient function. *Advances in Neural Information Processing Systems*, 36:66006–66019, 2023. 5, 8, 1, 2

[24] Qing Li, Huifang Feng, Kanle Shi, Yue Gao, Yi Fang, Yu-Shen Liu, and Zhizhong Han. Shs-net: Learning signed hyper surfaces for oriented normal estimation of point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13591–13600, 2023. 1

[25] Weizhou Liu, Jiaze Li, Xuhui Chen, Fei Hou, Shiqing Xin, Xingce Wang, Zhongke Wu, Chen Qian, and Ying He. Diffusing winding gradients (dwg): A parallel and scalable method for 3d reconstruction from unoriented point clouds. *ACM Transactions on Graphics*, 44(2):1–18, 2025. 6

[26] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Seminal Graphics: Pioneering Efforts that Shaped the Field*, pages 347–353. ACM Press, New York, NY, USA, 1998. 6

[27] Baorui Ma, Zhizhong Han, Yu-Shen Liu, and Matthias Zwicker. Neural-pull: Learning signed distance functions from point clouds by learning to pull space onto surfaces. *arXiv preprint arXiv:2011.13495*, 2020. 3, 5

[28] Baorui Ma, Yu-Shen Liu, and Zhizhong Han. Reconstructing surfaces for sparse point clouds with on-surface priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6315–6325, 2022. 3

[29] Baorui Ma, Yu-Shen Liu, Matthias Zwicker, and Zhizhong Han. Surface reconstruction from point clouds by learning predictive context priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6326–6337, 2022. 3, 6

[30] Zixiong Ma, Linyu Liu, Zexi Luo, Bohan Han, Zhidong Wei, Jianjiang Wang, Jianwei Wu, and Lei Zhang. NeuralPull: Learning signed distance functions from point clouds by learning to pull space onto the surface. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 15602–15614, 2021. 2

[31] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4460–4470, 2019. 3

[32] Gal Metzer, Rana Hanocka, Denis Zorin, Raja Giryes, Daniele Panozzo, and Daniel Cohen-Or. Orienting point clouds with dipole propagation. *ACM Transactions on Graphics (TOG)*, 40(4):1–14, 2021. 1

[33] A Cengiz Öztireli, Gael Guennebaud, and Markus Gross. Feature preserving point set surfaces based on non-linear kernel regression. In *Computer graphics forum*, pages 493–501. Wiley Online Library, 2009. 2, 6

[34] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019. 3

[35] Songyou Peng, Chiyu Jiang, Yiyi Liao, Michael Niemeyer, Marc Pollefeys, and Andreas Geiger. Shape as points: A differentiable poisson solver. *Advances in Neural Information Processing Systems*, 34:13032–13044, 2021. 6

[36] Mehdi SM Sajjadi, Henning Meyer, Etienne Pot, Urs Bergmann, Klaus Greff, Noha Radwan, Suhani Vora, Mario Lučić, Daniel Duckworth, Alexey Dosovitskiy, et al. Scene representation transformer: Geometry-free novel view synthesis through set-latent scene representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6229–6238, 2022. 4

[37] Andrew M Saxe, James L McClelland, and Surya Ganguli. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *arXiv preprint arXiv:1312.6120*, 2013. 4

[38] Nico Schertler, Bogdan Savchynskyy, and Stefan Gumhold. Towards globally optimal normal orientations for large point clouds. In *Computer Graphics Forum*, pages 197–208. Wiley Online Library, 2017. 1

[39] Chen Shen, James F. O'Brien, and Jonathan R. Shewchuk. Interpolating and approximating implicit surfaces from polygon soup. In *ACM SIGGRAPH 2004 Papers*, pages 896–904. ACM, New York, NY, USA, 2004. 2

[40] 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. 3, 5

[41] The Fitzwilliam Museum. Head. Web page, 2025. Accessed: 2025-05-19 14:07:15. 2

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 2, 4

[43] Zixiong Wang, Pengfei Wang, Qiujie Dong, Junjie Gao, Shuangmin Chen, Shiqing Xin, and Changhe Tu. Neural-imls: learning implicit moving least-squares for surface reconstruction from unoriented point clouds. *arXiv preprint arXiv:2109.04398*, 1(2):3, 2021. 2, 6

[44] Zixiong Wang, Yunxiao Zhang, Rui Xu, Fan Zhang, Peng-Shuai Wang, Shuangmin Chen, Shiqing Xin, Wenping Wang, and Changhe Tu. Neural-singular-hessian: Implicit neural representation of unoriented point clouds by enforcing singular hessian. *ACM Transactions on Graphics (TOG)*, 42 (6):1–14, 2023. 6

[45] Francis Williams, Christoph H. Rönz, Milo Roux, Ian Reid, Gerhard Neumann, Mauricio A. Valenzuela-Escárcega, and Lourdes Agapito Vázquez. Deep geometric prior for surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5470–5479, 2021. 3, 6

[46] Biao Zhang, Jiapeng Tang, Matthias Niessner, and Peter Wonka. 3dshape2vecset: A 3d shape representation for neural fields and generative diffusion models. *ACM Transactions On Graphics (TOG)*, 42(4):1–16, 2023. 2, 4

[47] Jie Zhang, Junjie Cao, Xiuping Liu, Jun Wang, Jian Liu, and Xiquan Shi. Point cloud normal estimation via low-rank subspace clustering. *Computers & Graphics*, 37(6):697–706, 2013. 1

[48] Junsheng Zhou, Baorui Ma, Shujuan Li, Yu-Shen Liu, and Zhizhong Han. Learning a more continuous zero level set in unsigned distance fields through level set projection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3181–3192, 2023. 5

[49] Qingnan Zhou and Alec Jacobson. Thingi10k: A dataset of 10,000 3d-printing models. *arXiv preprint arXiv:1605.04797*, 2016. 6

[50] 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 *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6118–6127, 2021. 1