Unsupervised Occupancy Learning from Sparse Point Cloud

Amine Ouasfi   Adnane Boukhayma

Inria, Univ. Rennes, CNRS, IRISA, M2S, France

Abstract

Implicit Neural Representations have gained prominence as a powerful framework for capturing complex data modalities, encompassing a wide range from 3D shapes to images and audio. Within the realm of 3D shape representation, Neural Signed Distance Functions (SDF) have demonstrated remarkable potential in faithfully encoding intricate shape geometry. However, learning SDFs from 3D point clouds in the absence of ground truth supervision remains a very challenging task. In this paper, we propose a method to infer occupancy fields instead of SDFs as they are easier to learn from sparse inputs. We leverage a margin-based uncertainty measure to differentiably sample from the decision boundary of the occupancy function and supervise the sampled boundary points using the input point cloud. We further stabilise the optimization process at the early stages of the training by biasing the occupancy function towards minimal entropy fields while maximizing its entropy at the input point cloud. Through extensive experiments and evaluations, we illustrate the efficacy of our proposed method, highlighting its capacity to improve implicit shape inference with respect to baselines and the state-of-the-art using synthetic and real data.

1. Introduction

Capturing full 3D shape from limited and corrupted data is a long standing problem. In this regard, one of the data type instances that has received increasing attention and investigation from computer vision, graphics and machine learning is point clouds. This interest emanates primarily from the ubiquity of this light, although topologically incomplete 3D representation, either as acquired *e.g.* through industrial and commodity depth sensors, or as an intermediate representation within computational photogrammetry [66, 67] pipelines for instance. While classical optimization methods such as Poisson Reconstruction [36] or Moving Least Squares [30] can be effective with dense, clean point sets and accurate normal pre-estimations, recent deep learning-based alternatives provide more robust predictions, particularly for noisy and sparse inputs, eliminating the need for

normal data in many cases.

Several methods rely on priors learned from large fully labeled data such as the synthetic dataset ShapeNet [14]. However, this strategy entails computationally expensive trainings, and the resulting models can still be prone to out-of-distribution generalization issues, as pointed in [15, 55], whether caused by change in the input size or domain shift. For instance, Table 2 shows that our unsupervised method outperforms supervised generalizable models when testing on data that is sparser and different in nature from their training corpus. Hence, it is important to design learning frameworks that can lead to robust reconstruction under such extreme constraints.

Neural implicit representaions (INR) were established recently as a powerful representation for 3D shape, generally taking the form of coordinate based MLPs. They have been applied to model various shape presentations, the most popular ones of which remain arguably signed distance functions (SDF) [58] and binary occupancy fields [52] for watertight shapes. For the task of point cloud reconstruction, there is a wide literature on learning SDFs from dense labels, as well as from point clouds solely (unsupervised), either in the generalizable setting or in instance specific optimization. However, apart from its applications in the supervised generalizable setting [8, 20, 60], it seems that occupancy learning from point sets has not been as widely considered in other scenarios. In particular, to the best of our knowledge, using occupancy fields to learn shapes from sparse noisy unoriented point clouds (*i.e.* unsupervisedly) has not been explored thoroughly. This is even the more intriguing as it seems to be easier to learn a binary classification task as opposed to regressing a continuous field that must additionally respect specific properties [28].

Concordantly, we introduce occupancy in this work for learning shape from a sparse point cloud. While other representations such as SDF can readily incorporate surface sample supervision directly (*e.g.* [28]), among other strategies, it is not entirely straightforward to define such supervision for occupancy fields. Hence, we propose a novel strategy for our training that combines a loss applied near the boundary decision, harnessing the input point samples, and an accompanying regularization. For the first loss, we encourage

the decision boundary of our occupancy to align with the point cloud. We achieve this by supervising an uncertainty sampling mechanism for our occupancy field with the point cloud samples. We approximate this sampling through root-finding of a margin function associated to the field. As regularization, we propose to minimize the uncertainty of our field, while maximizing it near the decision boundary, via the entropy of our occupancy.

Through extensive experiments under several real and synthetic benchmarks for object, non-rigid and scene level shape reconstruction, our results show that our method outperforms the existing state-of-the-art literature in reconstruction from sparse point cloud, using standard metrics and through superior visual results. Our ablation studies validate our design choices and showcase the importance of the components of the supervision scheme that we propose.

2. Related Work

Shape Representations in Deep Learning Shapes can be depicted in deep learning through either intrinsic or extrinsic representations. Intrinsic representations focus on discriminating the shape itself. When explicitly implemented, for instance, using tetrahedral or polygonal meshes [35, 79] or point clouds [25], the output topology is predefined, thereby restricting the range of shapes that can be generated. Among other intrinsic representations, 2D patches [23, 29, 83] may introduce discontinuities, while the simplicity of shape primitives like cuboids [78, 93], planes [43], and Gaussians [26] constrains their expressiveness. On the other hand, extrinsic shape representations model the entire space containing the scene or object of interest. Voxel grids [86, 87] are the most popular, serving as a direct extension of 2D pixels to the 3D domain. However, their capacity is constrained by the memory cost associated with cubic resolution. Sparse representations, such as octrees [64, 76, 81], can mitigate this issue to some extent.

Implicit Neural Shape Representations Implicit Neural Shape Representations have recently emerged as a significant approach for modeling extrinsic shape and radiance fields (*e.g.* [13, 33, 53, 80, 88]). These representations address many of the limitations associated with classical representations by showcasing the ability to represent shapes with arbitrary topologies at virtually infinite resolution. Typically, they are parameterized with MLPs, which map spatial locations or features to properties such as occupancy [52], signed [58], or unsigned [21, 91] distances relative to the target shape. The level-set derived from these MLPs can be visualized through techniques like ray marching [31] or tessellated into an explicit shape using methods like Marching Cubes [47]. Another notable line of research involves the development of hybrid implicit/explicit representations [19, 22, 57, 89], primarily based on differentiable space partitioning. To simultaneously represent

collections of shapes, implicit neural models necessitate conditioning mechanisms. These mechanisms encompass features and latent code concatenation, batch normalization, hypernetworks [17, 70, 72, 73, 82], and gradient-based meta-learning [54, 71]. Concatenation-based conditioning was initially implemented using single global latent codes [18, 52, 58] and has been further refined with the incorporation of local features [20, 24, 27, 34, 39, 60, 74, 77].

Reconstruction from Point Clouds Classical approaches to reconstruction include combinatorical methods where the shape is defined based on input point clouds through space partitioning, employing techniques such as alpha shapes [6], Voronoi diagrams [1], or triangulation [11, 45, 63]. Alternatively, the input samples can contribute to defining an implicit function, with its zero level set representing the target shape. This is achieved through global smoothing priors [40, 84, 85], such as radial basis functions [10] and Gaussian kernel fitting [65], or local smoothing priors like moving least squares [30, 37, 46, 51]. Another approach involves solving a boundary-conditioned Poisson equation [36]. Recent literature suggests parameterizing these implicit functions with deep neural networks and learning their parameters through gradient descent, either in a supervised or unsupervised manner.

Supervised Implicit Neural Reconstruction In supervised methods, there is an assumption of having labeled training data, typically in the form of dense samples containing ground truth shape information. Auto-decoding techniques [12, 34, 39, 58, 77] necessitate test-time optimization to adapt to a new point cloud. On the other hand, encoder-decoder methods allow for swift feed-forward inference. Initially introduced for this purpose, Pooling-based set encoders [18, 27, 52] like PointNet [62] have been found to underfit the context. State-of-the-art performance is achieved by convolutional encoders, utilizing local features defined either in explicit volumes and planes [20, 41, 60, 61] or solely at the input points [8, 56]. Peng *et al.* [61] proposed a differentiable Poisson solving layer that efficiently converts predicted normals into an indicator function grid. However, its applicability is limited to small scenes due to the cubic memory requirement in grid resolution. The work in [32, 85] suggests constructing generalizable models with implicit decoder functions as a kernel regression. Despite these advancements, many of the generalizable methods still face challenges related to generalization.

Unsupervised Implicit Neural Reconstruction For unsupervised approaches, a neural network is typically fitted to the input point cloud without additional information. Convergence improvements can be achieved through regularization techniques, such as the spatial gradient constraint based on the Eikonal equation introduced by Gropp et al. [28], a spatial divergence constraint as described in [5], and

Lipschitz regularization on the network [44]. Periodic activations were introduced in [72]. Lipman [42] learns a function that converges to occupancy, while its log transform converges to a distance function. Atzmon *et al.* [2] learn an SDF from unsigned distances, further supervising the spatial gradient of the function with normals [3]. Ma *et al.* [48] express the nearest point on the surface as a function of the neural signed distance and its gradient. They also utilize self-supervised local priors to handle very sparse inputs [49] and enhance generalization [50]. [38] guides the implicit field learning with an Octree based labelling. [9] predicts occupancy fields by learning whether a dropped needle goes across the surface or no. [15] learns a surface parameterization leveraged to provide additional coarse surface supervision to the shape network. In [84], infinitely wide shallow MLPs are learned as random feature kernels using points and their normals. However, most of the methods mentioned above encounter challenges when dealing with sparse and noisy input, primarily due to the lack of supervision. Differently from this literature, we propose here to apply occupancy to unsupervised sparse point cloud neural fitting for the first time to the best of our knowledge.

3. Method

Given a sparse noisy unoriented input point cloud $\mathcal{P} \subset \mathbb{R}^{3 \times N}$, our goal is to recover a 3D watertight shape surface $\mathcal{S}$ that best explains this observation, *i.e.* points from $\mathcal{P}$ approximating noisy samples from $\mathcal{S}$.

Based on the proven success of implicit neural representations, we propose to address this problem by fitting an implicit neural shape function to the scarce observation, which amounts to learning an MLP mapping coordinates to shape attributes. In this context, dubbed unsupervised, signed distance fields are the representation of choice in the community so far (*e.g.* [15, 16, 38, 48, 61, 84]). Oddly enough, binary occupancy fields were not explored extensively yet as a representation for this context to the best of our knowledge, even-though in theory, it seems *easier* to learn occupancy than SDFs. In fact, it is safe to assume that binary classification is an easier task than continuous regression in general. Besides, we can see intuitively that occupancy is only the *sign* part of the SDF information. Additionally, occupancy fields do not seem to stipulate any strong structural requirement, as it is the case for distance fields, whose gradient ought to satisfy the Eikonal constraint [28] for instance. Hence, we propose here to learn a binary occupancy field.

When labeled samples inside and outside the shape are available, such a neural occupancy field could be learned in this supervised setting via a standard cross-entropy classification loss (*e.g.* [52]). However when only surface points are available, as in our case ($\mathcal{P}$), it is not entirely clear how a successful learning can be achieved using such occupancy

decision boundary samples alone. We derive in the following a methodology to approach this task.

3.1. Learning Occupancy Through Margin Uncertainty Sampling

We are interested in learning the following posterior $P(y|\mathbf{x}, \theta)$, where $\mathbf{x}$ represents query points in $\mathbb{R}^3$, and label $y \in \{0, 1\}$ stands for binary shape occupancy. P is parameterized with a Softmax activated MLP θ similarly to [8]. Since we do not have points with their occupancy labels, all we can learn with is the input point cloud samples.

Let us consider the following margin function:

$$U_\theta(\mathbf{x}) = P(y = 1|\mathbf{x}, \theta) - P(y = 0|\mathbf{x}, \theta). \quad (1)$$

Borrowing from active learning terminology, finding the roots of this expression is akin to a form of uncertainty sampling from our occupancy field (margin sampling [68]). Intuitively, the smaller the margin of a point the more uncertain its shape prediction. In this particular case, the margin function zeroes correspond to the samples with upmost level of uncertainty.

Given that our input point cloud points are samples from the surface, *i.e.* the ground truth occupancy function decision boundary, we propose to learn our occupancy field by supervising uncertainty sampling with the point cloud, *i.e.* minimizing the distance between uncertain samples and their nearest point cloud samples. We define our uncertainty sampling as root finding of our margin function. We Initialize this root finding near the target surface. We hypothesis that training our occupancy field by making its uncertain samples coincide with the surface through minimal amount of steps of a root finding algorithm will encourage it to converge quickly and decisively.

We start by generating a pool of query points near $\mathcal{P}$, *i.e.* near the ground-truth surface, by sampling around the points, *i.e.* $\{\mathbf{q} \sim \mathcal{N}(\mathbf{p}, \sigma_{\mathbf{p}}\mathbf{I}_3)\}$. Here, $\sigma_{\mathbf{p}}$ is chosen as the maximal euclidean distance to the K nearest points to $\mathbf{q}$ in $\mathcal{P}$ (as in [28]). We subsequently recompute the nearest point $\mathbf{p}$ in $\mathcal{P}$ to each sample $\mathbf{q}$, thus forming the following set of training pairs:

$$\mathcal{Q} := \{(\mathbf{q}, \mathbf{p}), \mathbf{p} =_{\mathbf{v} \in \mathcal{P}} ||\mathbf{v} - \mathbf{q}||_2\}. \quad (2)$$

Given a pair $(\mathbf{q}, \mathbf{p})$ in $\mathcal{Q}$, we train by bounding one step of Newton-Raphson [90] root finding on our occupancy margin, initialized at $\mathbf{q}$, to result in an uncertain sample near $\mathbf{p}$ in mean squared error terms.

Let us recall that a Generalized Newton [4] (Theorem 1, with $m = 1$ and $n = 3$) iteration updates query point $\mathbf{q}$ accordingly: $\mathbf{q} - \nabla U_\theta(\mathbf{q})^\dagger U_\theta(\mathbf{q})$, where $\nabla U_\theta(\mathbf{q})^\dagger$ is the Moore-Penrose pseudoinverse of Jacobian $\nabla U_\theta(\mathbf{q})$. As U_θ is a scalar function ($\mathbb{R}^3 \to \mathbb{R}$), ∇U_θ is a row 3-vector (we can use it or its transpose interchangeably), and the pseudoinverse writes: $\nabla U_\theta(\mathbf{q})^\dagger = \frac{\nabla U_\theta(\mathbf{q})}{||\nabla U_\theta(\mathbf{q})||_2^2}$.

In conclusion, our margin uncertainty sampling loss can be expressed as follows:

$$\mathcal{L}_{\text{samp}}(\theta, \mathcal{Q}) = \mathop{\mathbb{E}}_{(\mathbf{q},\mathbf{p})\sim\mathcal{Q}} ||\mathbf{q} - U_\theta(\mathbf{q}) \cdot \frac{\nabla U_\theta(\mathbf{q})}{||\nabla U_\theta(\mathbf{q})||_2^2} - \mathbf{p}||_2^2, \tag{3}$$

where spatial gradient ∇U_θ can be computed efficiently through automatic differentiation (*e.g.* PyTorch [59]).

This loss offers two primary advantages. Firstly, it serves as a form of supervision for the occupancy of the query points. The sign of the margin function differs on opposite sides of the decision boundary. Consequently, based on this sign, a Newton-Raphson iteration will move points along or against the direction of its gradient resulting in an occupancy field whose decision boundary is aligned with the input point cloud. Additionally, this loss operates as a smoothness constraint as Newton-Raphson root finding relies on approximating the function at the roots by its first-order approximation on the initial points $\mathbf{q}$, hence smoothing the margin function around the input points.

3.2. Entropy Based Regularization

To ensure better convergence of our training, we use the MLP initialization in [2], while adapting it such that the last layer predicts binary occupancy for an r-radius sphere.

Furthermore, to ease and improve our learning, we suggest to urge our occupancy field to have minimal uncertainty almost everywhere in space. We use the entropy [69] as an uncertainty measure here. However, this is the opposite behavior that we expect from our network to display at the decision boundary *i.e.* near the ground-truth surface. Hence, we propose to maximize the uncertainty (*i.e.* entropy) conversely at the input point cloud $\mathcal{P}$ simultaneously.

Given a random variable X, Shannon entropy [69] is defined as follows:

$$\mathbb{H}(X) = -\mathbb{E} \log(P(X)). \tag{4}$$

We create a set $\Omega \subset \mathbb{R}^3$ of query points $\mathbf{x}$ by sampling space uniformly within a bounding box of input point cloud $\mathcal{P}$. As a measure of uncertainty we consider the entropy of the occupancy distribution conditioned on the observed spatial location:

$$\mathcal{L}_{\text{entr}}(\theta, \Omega, \mathcal{P}) = \mathbb{E}_{\mathbf{x}\sim\Omega}\mathbb{H}(y|\mathbf{x},\theta) - \mathbb{E}_{\mathbf{p}\sim\mathcal{P}}\mathbb{H}(y|\mathbf{p},\theta). \tag{5}$$

This entropy polarization loss acts both as an initialization of the field and a regularization mid-learning, as we empirically found that it performs best when its weight (λ in Equation 6) is reduced progressively in the final loss in the midst of training.

Finally we can train our model using the following combined empirical risk minimization:

$$\min_{\theta} \mathcal{L}_{\text{samp}}(\theta, \mathcal{Q}) + \lambda\mathcal{L}_{\text{entr}}(\theta, \Omega, \mathcal{P}). \tag{6}$$
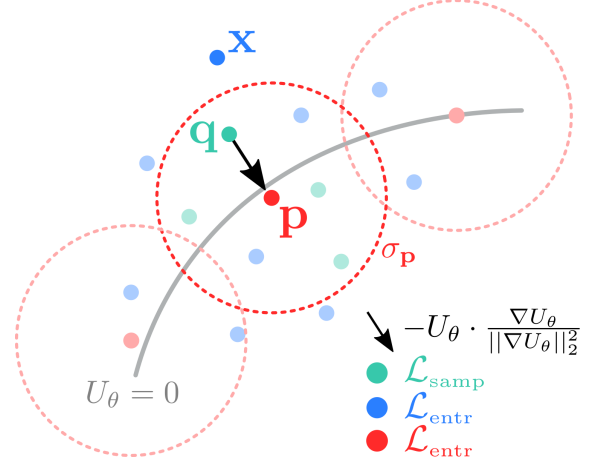

Figure 1. Illustration of our training. Our method learns a neural binary occupancy field without off-surface labels. It uses the combination of a margin uncertainty sampling loss near the surface (Green), maximizing entropy at the input point cloud samples (Red), and minimizing entropy everywhere else (Blue).

Algorithm 1 shows a summary of our training procedure. We note that once an implicit occupancy field is learned, an explicit triangle mesh $\hat{\mathcal{S}}$ can be obtained with it through the Marching Cubes algorithm [47].

Algorithm 1 The training procedure of our method.

Input: Point cloud $\mathcal{P}$, learning rate α, number of iterations N_{it}, number of near surface queries $N_{\mathcal{Q}}$, number of uniform queries N_{Ω}, number of batch input points $N_{\mathcal{P}}$, number of uniform queries N_{Ω}, reg. loss weight λ.
Output: Optimal weights θ^*.
 Compute local st. dev. $\{\sigma_{\mathbf{P}} = \max_{\mathbf{v}\in K_{\text{nn}}(\mathbf{p},\mathbf{P})} ||\mathbf{v} - \mathbf{p}||_2\}$.
 Generate sets $\mathcal{Q}$ (Sec.3.1) and Ω (Sec.3.2).
 Initialize λ (Sec.4).
 for N_{it} times **do**
  Sample a batch $\mathcal{Q}_b$ of size $N_{\mathcal{Q}}$ from $\mathcal{Q}$.
  Sample a batch Ω_b of size N_{Ω} from , Ω_b.
  Sample a batch $\mathcal{P}_b$ of size $N_{\mathcal{P}}$ from $\mathcal{P}$.
  Compute losses $\mathcal{L}_{\text{samp}}(\theta, \mathcal{Q}_b)$ (Equ.3).
  Compute losses $\mathcal{L}_{\text{entr}}(\theta, \Omega_b, \mathcal{P}_b)$ (Equ.5).
  $\theta \leftarrow \theta - \alpha\nabla_\theta(\mathcal{L}_{\text{samp}} + \lambda\mathcal{L}_{\text{entr}})$.
  Update λ (Sec.4).
 end for

---

4. Implementation Details

We set number of queries as $N_{\mathcal{Q}} = 1000k$ and $N_{\Omega} = 10k$. Our MLP follows the architecture in [48], and similarly to the latter we set $K = 51$. We train for $N_{it} = 40k$ iterations on a Nvidia RTX A6000 GPU using the Adam optimizer with learning rate $\alpha = 0.001$. Our training takes roughly 5 minutes for a 1024 sized input point cloud. The regularization parameter λ (Equation 6) follows an exponential decay

	CD1	CD2	NC	FS
SPSR [36]	2.34	0.224	0.74	0.50
OG-INR [38]	1.36	0.051	0.55	0.55
N-Pull [48]	1.16	0.074	0.84	0.75
G-Pull [16]	1.07	0.032	0.70	0.74
NTPS [15]	1.11	0.067	**0.88**	0.74
Ours	**0.76**	**0.020**	**0.88**	**0.83**

Table 1. ShapeNet [14] reconstructions from sparse noisy unoriented point clouds.

	CD1	CD2	NC	FS
POCO [8]	0.308	0.002	0.934	0.981
CONet [60]	1.26	0.028	0.829	0.599
On-Surf [49]	0.584	0.012	0.936	0.915
SPSR [36]	0.751	0.028	0.871	0.839
G-Pull [16]	0.495	0.005	0.887	0.945
NTPS [15]	0.737	0.015	0.943	0.844
Ours	**0.260**	**0.002**	**0.952**	**0.974**

Table 2. Faust [7] reconstructions from sparse noisy unoriented point clouds.

of the form $\exp -\kappa t$ where κ is set to 1.84×10^{-2}.

5. Results

To test the efficacy of our approach, we assess our capacity to infer implicit representations of shapes when presented with sparse and noisy point clouds. We employ datasets from standard reconstruction benchmarks, showcasing a diverse array of challenges associated with implicit shape function inference from sparse inputs. Consistently with existing literature, we evaluate the performance of our method by quantifying the accuracy of 3D explicit shape models obtained post-convergence from our MLPs. It is worth noting that our approach entails fitting an MLP independently for each point cloud, operating without reliance on pre-learned priors or additional training data. We compare quantitatively and qualitatively to the the state-of-the-art in our problem setting, *i.e.* unsupervised reconstruction from unoriented point, including deep learning methods N-Pull [48], NTPS [15], G-Pull [16], OG-INR [38], DiGs [5], NDrop [9], SAP [61]. We show results for NSpline [84] even-though it requires normals. We also compare to classical Poisson Reconstruction (SPSR [36]). We note that NTPS is the closest method to ours as it focuses specifically on the sparse input case. For comprehensive evaluation, we also include comparisons with feed-forward generalizable methods, namely POCO [8] and CONet [60], alongside the prior-based optimization method On-Surf [49] dedicated to sparse inputs. Unless stated differently, we use the publicly available official implementations of existing methods.

5.1. Metrics

Following seminal work, we evaluate our method and the competition w.r.t. the ground truth using standard metrics for the 3D reconstruction task. Namely, the L1 **Chamfer Distance (CD1)** ($\times 10^2$), L2 **Chamfer Distance (CD2)** ($\times 10^2$), the **Hausdorff distance (HD)** and the euclidean distance based **F-Score (FS)** when ground truth points are available, and finally **Normal Consistency (NC)** when ground truth normals are available. We detail the expressions of these metrics in the supplementary material.

5.2. Datasets and Input Definitions

ShapeNet [14] consists of various instances of 13 different synthetic 3D object classes. We follow the train/test splits defined in [84]. We generate noisy input point clouds by sampling 1024 points from the meshes and adding Gaussian noise of standard deviation 0.005 following the literature (*e.g.* [8, 60]). For brevity we show results on classes Tables, Chairs and Lamps. **Faust** [7] consists of real scans of 10 human body identities in 10 different poses. We sample sets of 1024 points from the scans as inputs. **3D Scene** [92] contains large scale complex real world scenes obtained with a handheld commodity range sensor. We follow [15, 34, 48] and sample our input point clouds with a sparse density of 100 per m^2, and we report performance similarly for scenes Burghers, Copyroom, Lounge, Stonewall and Totempole. **Surface Reconstruction Benchmark (SRB)** [83] consists of five object scans, each with different challenges such as complex topology, high level of detail, missing data and varying feature scales. We sample 1024 points from the scans for the sparse input experiment, and we also experiment using the dense inputs.

5.3. Object Level Reconstruction

We conduct the reconstruction of ShapeNet [14] objects using sparse and noisy point clouds. Table 1 and Figure 2 present a quantitative and qualitative comparison with other methods. Our approach consistently outperforms the competition across all metrics, as seen in the visual superiority of our reconstructions. We excel in capturing fine structures and details with greater fidelity. While NTPS demonstrates good overall coarse reconstructions, its thin plate spline smoothing prior appears to limit its expressivity. Additionally, OG-INR struggles to achieve satisfactory results in the sparse and noisy regime, despite its effective Octree-based sign field guidance in denser settings.

5.4. Real Articulated Shape Reconstruction

We undertake the reconstruction of Faust [7] human shapes using sparse and noisy point clouds. Table 2 and Figure 3 provide a numerical and qualitative comparison with other

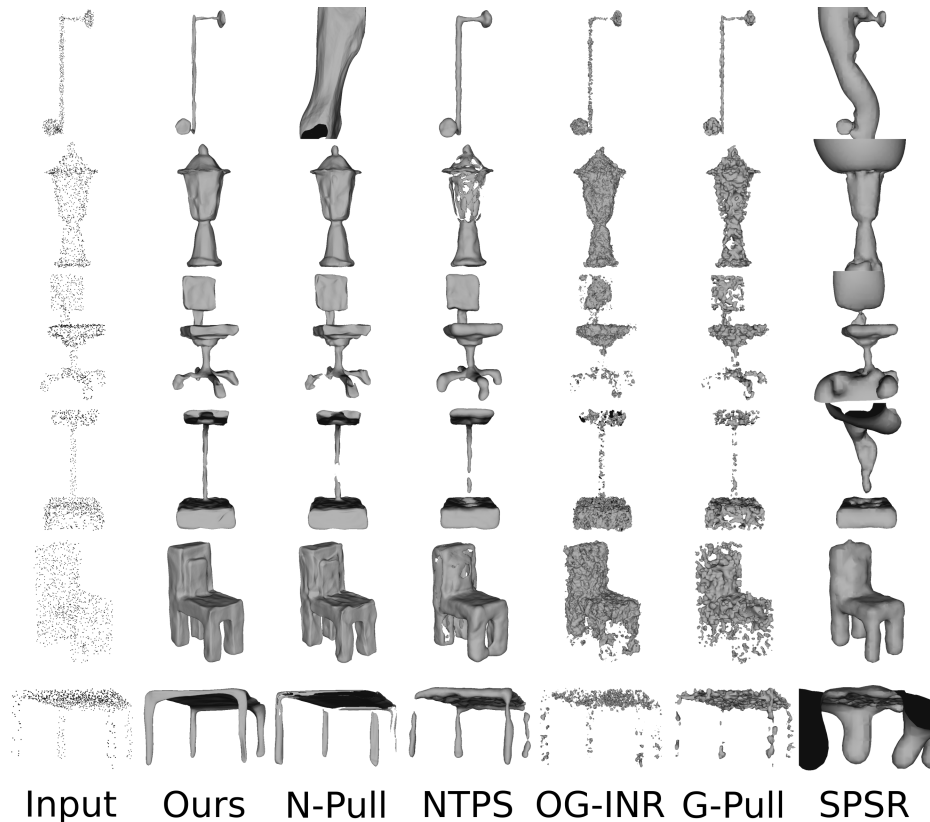| Input | Ours | N-Pull | NTPS | OG-INR | G-Pull | SPSR |

Figure 2. ShapeNet [14] reconstructions from sparse noisy unoriented point clouds.

methods. Our approach surpasses the alternatives across all metrics, with visually superior reconstructions, as especially noticeable at the extremities of the body. Analogous to the fine structures in the ShapeNet experiment, these areas pose challenges due to limited input point cloud samples, making shape prediction inherently difficult and ambiguous. NTPS reconstructions, similarly, exhibit coarser and less detailed results on this dataset. It is worth highlighting that our method outperforms not only feed-forward generalizable approaches POCO and CONet but also the prior-based optimization method On-Surf. These methods rely on priors trained on ShapeNet, thereby constraining their generalization ability.

5.5. Real Scene Level Reconstruction

In accordance with [15], we present the reconstruction outcomes on the 3D Scene [92] dataset derived from spatially sparse point clouds. Table 3 provides a summary of numerical results, incorporating results for methods NTPS, N-Pull, SAP, NDrop, and NSpline as reported in the state-of-the-art method NTPS. Our method excels in this benchmark, surpassing the competition in most metrics. Notably, our baseline N-Pull exhibits more conspicuous failures in this extensive sparse setup. Figure 5 offers qualitative compar-

isons against baselines N-Pull and SPSR.

6. Ablation Studies

Loss. Our approach centers around aligning the decision boundary of our occupancy function with the shape surface through a loss function based on uncertainty sampling, denoted as $\mathcal{L}_{\text{samp}}$ (Equation 3). To facilitate and enhance our learning process, we impose constraints on the network to minimize uncertainty almost everywhere in space while maximizing it at the decision boundary. Entropy serves as our chosen uncertainty measure. However, other measures can be considered. We justify this choice in Table 4. We compare our entropy-based regularization to margin-based regularization (Margin reg), using the absolute value of our margin function U_θ (Equation 3) as an uncertainty measure instead of entropy. Additionally, we compare it to the sign-agnostic occupancy supervision strategy introduced in [75] denoted as SA-Occ. Although this strategy fails to converge in the absence of priors, we demonstrate its beneficial use as a regularization in place of our entropy-based approach. It is noteworthy that while margin-based regularization (Margin reg) yields better results, it is still outperformed by our entropy-based regularization. Figure 4 shows an example of validation plots, displaying the role of our regularization

	Burghers			Copyroom			Lounge			Stonewall			Totemple			**Mean**		
	CD1	CD2	NC	CD1	CD2	NC	CD1	CD2	NC	CD1	CD2	NC	CD1	CD2	NC	CD1	CD2	NC
SPSR [36]	0.178	0.205	0.874	0.225	0.286	0.861	0.280	0.365	0.869	0.300	0.480	0.866	0.588	1.673	0.879	0.314	0.602	0.870
NDrop [9]	0.200	0.114	0.825	0.168	0.063	0.696	0.156	0.050	0.663	0.150	0.081	0.815	0.203	0.139	0.844	0.175	0.089	0.769
N-Pull [48]	0.064	0.008	0.898	0.049	0.005	0.828	0.133	0.038	0.847	0.060	0.005	0.910	0.178	0.024	0.908	0.097	0.016	0.878
SAP [61]	0.153	0.101	0.807	0.053	0.009	0.771	0.134	0.033	0.813	0.070	0.007	0.867	0.474	0.382	0.725	0.151	0.100	0.797
NSpline [84]	0.135	0.123	0.891	0.056	0.023	0.855	0.063	0.039	0.827	0.124	0.091	0.897	0.378	0.768	0.892	0.151	0.209	0.88
NTPS [15]	0.055	0.005	**0.909**	0.045	0.003	**0.892**	0.129	0.022	**0.872**	0.054	0.004	**0.939**	0.103	0.017	0.935	0.077	0.010	**0.897**
Ours	**0.022**	**0.001**	0.871	**0.041**	**0.012**	0.812	**0.021**	**0.001**	0.870	**0.028**	**0.003**	0.931	**0.026**	**0.001**	0.936	**0.027**	**0.003**	0.886

Table 3. 3D Scene [92] reconstructions from sparse point clouds.

Figure 4. CD1 distance to GT for reconstructions of shape Gargoyle of benchmark SRB [83] from a sparse unoriented point cloud.

	CD1	NC
N-Pull [48] (baseline)	1.10	0.85
$\mathcal{L}_{samp}$ (Equ.3)	1.02	0.86
$\mathcal{L}_{samp}$ + Margin reg	0.80	0.88
$\mathcal{L}_{samp}$ + SA-Occ [75] reg	0.95	0.88
Ours	**0.77**	**0.89**

Table 4. Ablation of our method on the Tables class of ShapeNet [14] .

	Sparse		Dense	
	CD1	HD	CD1	HD
SPSR [36]	2.27	21.1	1.25	22.59
DiGs [5]	0.68	6.05	**0.19**	**3.52**
OG-INR [38]	0.85	7.10	0.20	4.06
NTPS [15]	0.73	7.78	-	-
N-Pull [48]	0.58	8.90	0.23	4.46
Ours	**0.49**	**6.04**	0.20	3.94

Table 5. Ablation of point cloud density.

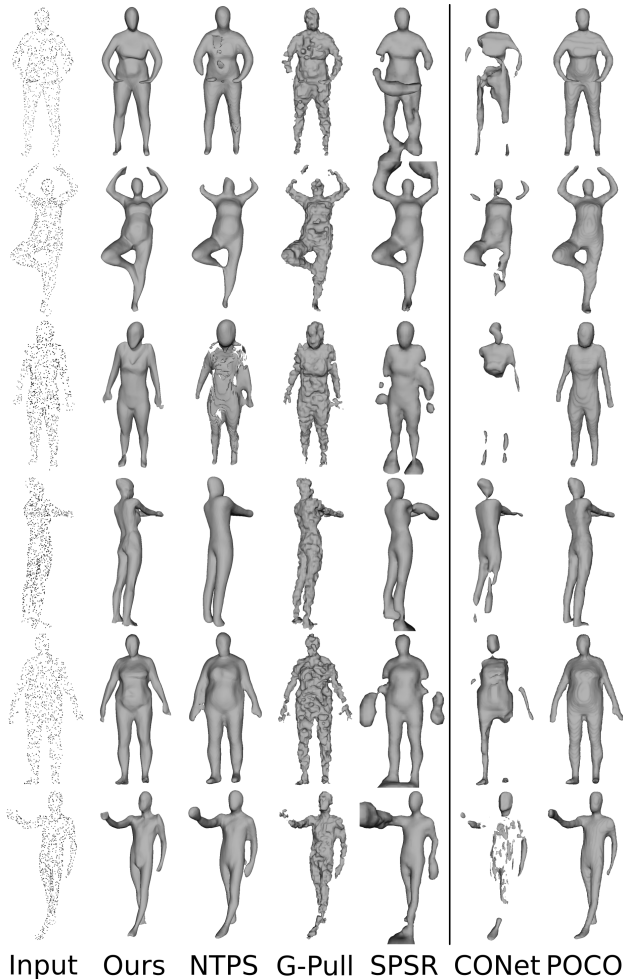Input Ours NTPS G-Pull SPSR CONet POCO

Figure 3. Faust [7] reconstructions from sparse unoriented point clouds.

in avoiding overfitting. The baseline in this example is SDF method N-Pull [48].

Point Cloud Density. We utilize the SRB [83] benchmark to evaluate the performance of our method across varying point cloud densities. Table 5 presents comparative results for sparse (1024) and dense input point clouds . We include results for the competition from OG-INR in the dense setting. Our method outperforms the competition in the sparse
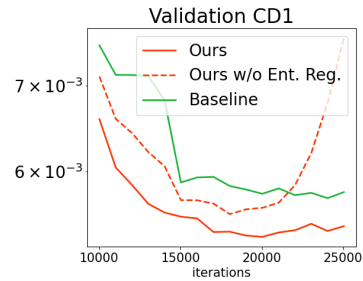
case and performs comparably to the state-of-the-art in the dense case. The improvement over our baseline is substantial for both sparse and dense inputs, as highlighted visually in Figure 6, where we showcase reconstructions for both sparse and dense cases. Notably, we achieve better topologies in the sparse case and enhanced, more accurate details in the dense case, as indicated by the red boxes. These results underscore the utility and advantages of our contribution, even in the dense setting.

Running Time and Performance against Number of Queries and Input Points. Fig.7 shows our best performance over time (minutes) for various input point cloud
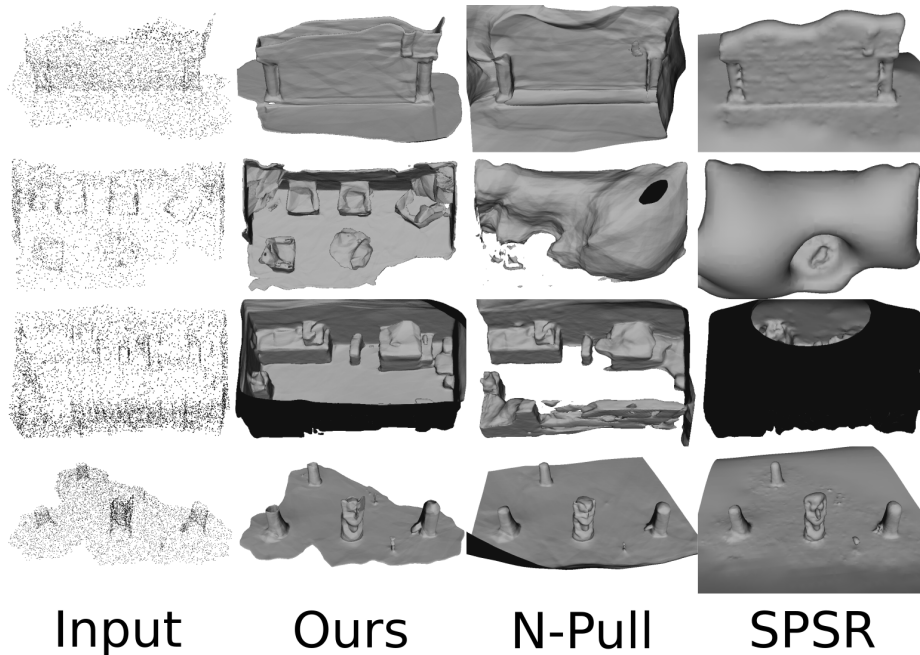
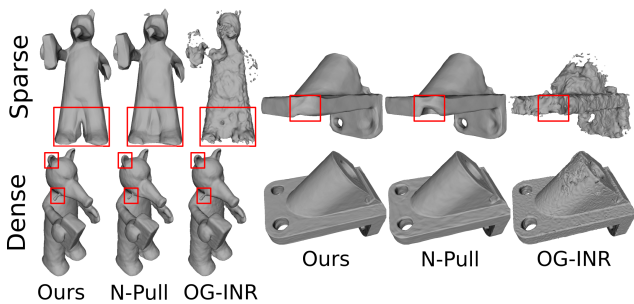Figure 5. 3D Scene [92] reconstructions from sparse unoriented point clouds.

Figure 6. SRB [83] reconstructions from sparse and dense unoriented inputs.

Figure 7. Performance over time (varying input size).
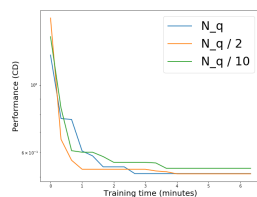
Figure 8. Performance over time (varying query set size).

noise std. dev.	0	0.005	0.025
CD1 $\downarrow$	0.56	0.77	2.16
NC $\uparrow$	0.93	0.89	0.68

Table 6. Noise ablation.

with various noise levels on class Table of ShapeNet. Performance remains reasonable even under heavily corrupted inputs, which confirms our qualitative superiority to competition in Fig.2, and our resilience to noise.

sizes. Models using larger inputs reach the baseline performance at earlier times. Fig.8 shows our best performance over time while decreasing the total number of query samples. Performance does not deteriorate drastically under less queries. These Experiments were performed using shape Gargoyle of benchmark SRB.

Robustness to Noise We evaluate here (Tab.6) our method

7. Limitations

We notice that excess regularization can cause reconstructions to be overly smooth in some cases, which can affect our NC numbers is some benchmarks. This can be improved by tuning the hyperparameters of the method per benchmark or per object/scene.

8. Conclusion

We presented a method for implicit shape reconstruction from sparse, noisy and unoriented point cloud. Our results demonstrate that occupancy fields offer and effective implicit shape representation for this task, and that they can be learned successfully even in the unsupervised setting. Under our carefully designed losses, the occupancy outperforms state-of-the-art methods based on SDF, as-well-as other occupancy based baselines that we show in ablation studies.
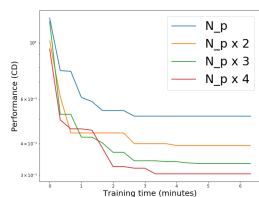
References

[1] Nina Amenta, Sunghee Choi, and Ravi Krishna Kolluri. The power crust, unions of balls, and the medial axis transform. *CG*, 2001. 2

[2] Matan Atzmon and Yaron Lipman. Sal: Sign agnostic learning of shapes from raw data. In *CVPR*, 2020. 3, 4

[3] Matan Atzmon and Yaron Lipman. Sald: Sign agnostic learning with derivatives. In *ICML*, 2020. 3

[4] Adi Ben-Israel. A newton-raphson method for the solution of systems of equations. In *Journal of Mathematical analysis and applications*, 1966. 3

[5] Yizhak Ben-Shabat, Chamin Hewa Koneputugodage, and Stephen Gould. Digs: Divergence guided shape implicit neural representation for unoriented point clouds. In *CVPR*, 2022. 2, 5, 7

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

[7] Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *CVPR*, 2014. 5, 7

[8] Alexandre Boulch and Renaud Marlet. Poco: Point convolution for surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6302–6314, 2022. 1, 2, 3, 5

[9] Alexandre Boulch, Pierre-Alain Langlois, Gilles Puy, and Renaud Marlet. Needrop: Self-supervised shape representation from sparse point clouds using needle dropping. In *2021 International Conference on 3D Vision (3DV)*, pages 940–950. IEEE, 2021. 3, 5, 7

[10] Jonathan C Carr, Richard K Beatson, Jon B Cherrie, Tim J Mitchell, W Richard Fright, Bruce C McCallum, and Tim R Evans. Reconstruction and representation of 3d objects with radial basis functions. In *SIGGRAPH*, 2001. 2

[11] Frédéric Cazals and Joachim Giesen. *Effective Computational Geometry for Curves and Surfaces*. 2006. 2

[12] Rohan Chabra, Jan E Lenssen, Eddy Ilg, Tanner Schmidt, Julian Straub, Steven Lovegrove, and Richard Newcombe. Deep local shapes: Learning local sdf priors for detailed 3d reconstruction. In *ECCV*, 2020. 2

[13] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. Efficient geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16123–16133, 2022. 2

[14] 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. 1, 5, 6, 7

[15] Chao Chen, Zhizhong Han, and Yu-Shen Liu. Unsupervised inference of signed distance functions from single sparse point clouds without learning priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1, 3, 5, 6, 7

[16] Chao Chen, Yu-Shen Liu, and Zhizhong Han. Gridpull: Towards scalability in learning implicit representations from 3d point clouds. In *Proceedings of the ieee/cvf international conference on computer vision*, pages 18322–18334, 2023. 3, 5

[17] Yinbo Chen and Xiaolong Wang. Transformers as meta-learners for implicit neural representations. In *European Conference on Computer Vision*, 2022. 2

[18] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *CVPR*, 2019. 2

[19] Zhiqin Chen, Andrea Tagliasacchi, and Hao Zhang. Bsp-net: Generating compact meshes via binary space partitioning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 2

[20] Julian Chibane and Gerard Pons-Moll. Implicit feature networks for texture completion from partial 3d data. In *European Conference on Computer Vision*, pages 717–725. Springer, 2020. 1, 2

[21] Julian Chibane, Aymen Mir, and Gerard Pons-Moll. Neural unsigned distance fields for implicit function learning. In *NeurIPS*, 2020. 2

[22] Boyang Deng, Kyle Genova, Soroosh Yazdani, Sofien Bouaziz, Geoffrey Hinton, and Andrea Tagliasacchi. Cvxnet: Learnable convex decomposition. In *CVPR*, 2020. 2

[23] Theo Deprelle, Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. Learning elementary structures for 3d shape generation and matching. In *NeurIPS*, 2019. 2

[24] Philipp Erler, Paul Guerrero, Stefan Ohrhallinger, Niloy J Mitra, and Michael Wimmer. Points2surf learning implicit surfaces from point clouds. In *ECCV*, 2020. 2

[25] Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *CVPR*, 2017. 2

[26] Kyle Genova, Forrester Cole, Daniel Vlasic, Aaron Sarna, William T Freeman, and Thomas Funkhouser. Learning shape templates with structured implicit functions. In *ICCV*, 2019. 2

[27] Kyle Genova, Forrester Cole, Avneesh Sud, Aaron Sarna, and Thomas Funkhouser. Local deep implicit functions for 3d shape. In *CVPR*, 2020. 2

[28] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *ICML*, 2020. 1, 2, 3

[29] Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. A papier-mâché approach to learning 3d surface generation. In *CVPR*, 2018. 2

[30] Gaël Guennebaud and Markus Gross. Algebraic point set surfaces. In *ACM siggraph 2007 papers*, pages 23–es. 2007. 1, 2

[31] John C Hart. Sphere tracing: A geometric method for the antialiased ray tracing of implicit surfaces. *The Visual Computer*, 1996. 2

[32] Jiahui Huang, Zan Gojcic, Matan Atzmon, Or Litany, Sanja Fidler, and Francis Williams. Neural kernel surface reconstruction. In *Proceedings of the IEEE/CVF Conference*

on Computer Vision and Pattern Recognition, pages 4369–4379, 2023. 2

[33] Ajay Jain, Ben Mildenhall, Jonathan T. Barron, Pieter Abbeel, and Ben Poole. Zero-shot text-guided object generation with dream fields. 2022. 2

[34] Chiyu Jiang, Avneesh Sud, Ameesh Makadia, Jingwei Huang, Matthias Nießner, Thomas Funkhouser, et al. Local implicit grid representations for 3d scenes. In *CVPR*, 2020. 2, 5

[35] Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. Neural 3d mesh renderer. In *CVPR*, 2018. 2

[36] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *TOG*, 2013. 1, 2, 5, 7

[37] Ravikrishna Kolluri. Provably good moving least squares. *TALG*, 2008. 2

[38] Chamin Hewa Koneputugodage, Yizhak Ben-Shabat, and Stephen Gould. Octree guided unoriented surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16717–16726, 2023. 3, 5, 7

[39] Tianyang Li, Xin Wen, Yu-Shen Liu, Hua Su, and Zhizhong Han. Learning deep implicit functions for 3d shapes with dynamic code clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12840–12850, 2022. 2

[40] Siyou Lin, Dong Xiao, Zuoqiang Shi, and Bin Wang. Surface reconstruction from point clouds without normals by parametrizing the gauss formula. *ACM Transactions on Graphics*, 42(2):1–19, 2022. 2

[41] Stefan Lionar, Daniil Emtsev, Dusan Svilarkovic, and Songyou Peng. Dynamic plane convolutional occupancy networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1829–1838, 2021. 2

[42] Yaron Lipman. Phase transitions, distance functions, and implicit neural representations. In *ICML*, 2021. 3

[43] Chen Liu, Jimei Yang, Duygu Ceylan, Ersin Yumer, and Yasutaka Furukawa. Planenet: Piece-wise planar reconstruction from a single rgb image. In *CVPR*, 2018. 2

[44] Hsueh-Ti Derek Liu, Francis Williams, Alec Jacobson, Sanja Fidler, and Or Litany. Learning smooth neural functions via lipschitz regularization. *arXiv preprint arXiv:2202.08345*, 2022. 3

[45] Minghua Liu, Xiaoshuai Zhang, and Hao Su. Meshing point clouds with predicted intrinsic-extrinsic ratio guidance. In *ECCV*, 2020. 2

[46] Shi-Lin Liu, Hao-Xiang Guo, Hao Pan, Peng-Shuai Wang, Xin Tong, and Yang Liu. Deep implicit moving least-squares functions for 3d reconstruction. In *CVPR*, 2021. 2

[47] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *SIGGRAPH*, 1987. 2, 4

[48] 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. In *ICML*, 2021. 3, 4, 5, 7

[49] 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, 5

[50] 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

[51] Corentin Mercier, Thibault Lescoat, Pierre Roussillon, Tamy Boubekeur, and Jean-Marc Thiery. Moving level-of-detail surfaces. *ACM Transactions on Graphics (TOG)*, 41(4):1–10, 2022. 2

[52] 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. 1, 2, 3

[53] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 2

[54] Amine Ouasfi and Adnane Boukhayma. Few 'zero level set'-shot learning of shape signed distance functions in feature space. In *ECCV*, 2022. 2

[55] Amine Ouasfi and Adnane Boukhayma. Robustifying generalizable implicit shape networks with a tunable nonparametric model. In *NeurIPS*, 2023. 1

[56] Amine Ouasfi and Adnane Boukhayma. Mixing-denoising generalizable occupancy networks. In *3DV*, 2024. 2

[57] David Palmer, Dmitriy Smirnov, Stephanie Wang, Albert Chern, and Justin Solomon. Deepcurrents: Learning implicit representations of shapes with boundaries. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18665–18675, 2022. 2

[58] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019. 1, 2

[59] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *NeurIPS*, 2019. 4

[60] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *European Conference on Computer Vision*, pages 523–540. Springer, 2020. 1, 2, 5

[61] 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. 2, 3, 5, 7

[62] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, 2017. 2

[63] Marie-Julie Rakotosaona, Noam Aigerman, Niloy Mitra, Maks Ovsjanikov, and Paul Guerrero. Differentiable surface triangulation. In *SIGGRAPH Asia*, 2021. 2

[64] Gernot Riegler, Ali Osman Ulusoy, and Andreas Geiger. Octnet: Learning deep 3d representations at high resolutions. In *CVPR*, 2017. 2

[65] Bernhard Schölkopf, Joachim Giesen, and Simon Spalinger. Kernel methods for implicit surface modeling. In *NeurIPS*, 2004. 2

[66] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[67] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016. 1

[68] Burr Settles. Active learning literature survey. 2009. 3

[69] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948. 4

[70] Vincent Sitzmann, Michael Zollhoefer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. In *NeurIPS*, 2019. 2

[71] Vincent Sitzmann, Eric R Chan, Richard Tucker, Noah Snavely, and Gordon Wetzstein. Metasdf: Meta-learning signed distance functions. In *NeurIPS*, 2020. 2

[72] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *NeurIPS*, 2020. 2, 3

[73] Vincent Sitzmann, Semon Rezchikov, William T Freeman, Joshua B Tenenbaum, and Fredo Durand. Light field networks: Neural scene representations with single-evaluation rendering. In *NeurIPS*, 2021. 2

[74] Towaki Takikawa, Joey Litalien, Kangxue Yin, Karsten Kreis, Charles Loop, Derek Nowrouzezahrai, Alec Jacobson, Morgan McGuire, and Sanja Fidler. Neural geometric level of detail: Real-time rendering with implicit 3d shapes. In *CVPR*, 2021. 2

[75] Jiapeng Tang, Jiabao Lei, Dan Xu, Feiying Ma, Kui Jia, and Lei Zhang. Sa-convonet: Sign-agnostic optimization of convolutional occupancy networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6504–6513, 2021. 6, 7

[76] Maxim Tatarchenko, Alexey Dosovitskiy, and Thomas Brox. Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs. In *ICCV*, 2017. 2

[77] Edgar Tretschk, Ayush Tewari, Vladislav Golyanik, Michael Zollhöfer, Carsten Stoll, and Christian Theobalt. Patchnets: Patch-based generalizable deep implicit 3d shape representations. In *ECCV*, 2020. 2

[78] Shubham Tulsiani, Hao Su, Leonidas J. Guibas, Alexei A. Efros, and Jitendra Malik. Learning shape abstractions by assembling volumetric primitives. In *CVPR*, 2017. 2

[79] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. In *ECCV*, 2018. 2

[80] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021. 2

[81] Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, and Xin Tong. O-cnn: Octree-based convolutional neural networks for 3d shape analysis. *TOG*, 2017. 2

[82] Shaofei Wang, Marko Mihajlovic, Qianli Ma, Andreas Geiger, and Siyu Tang. Metaavatar: Learning animatable clothed human models from few depth images. *Advances in Neural Information Processing Systems*, 34:2810–2822, 2021. 2

[83] Francis Williams, Teseo Schneider, Claudio Silva, Denis Zorin, Joan Bruna, and Daniele Panozzo. Deep geometric prior for surface reconstruction. In *CVPR*, 2019. 2, 5, 7, 8

[84] Francis Williams, Matthew Trager, Joan Bruna, and Denis Zorin. Neural splines: Fitting 3d surfaces with infinitely-wide neural networks. In *CVPR*, 2021. 2, 3, 5, 7

[85] Francis Williams, Zan Gojcic, Sameh Khamis, Denis Zorin, Joan Bruna, Sanja Fidler, and Or Litany. Neural fields as learnable kernels for 3d reconstruction. In *CVPR*, 2022. 2

[86] Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T Freeman, and Joshua B Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *NeurIPS*, 2016. 2

[87] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *CVPR*, 2015. 2

[88] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34:4805–4815, 2021. 2

[89] Mohsen Yavartanoo, Jaeyoung Chung, Reyhaneh Neshatavar, and Kyoung Mu Lee. 3dias: 3d shape reconstruction with implicit algebraic surfaces. In *ICCV*, 2021. 2

[90] Tjalling J Ypma. Historical development of the newton–raphson method. *SIAM review*, 37(4):531–551, 1995. 3

[91] Junsheng Zhou, Baorui Ma, Liu Yu-Shen, Fang Yi, and Han Zhizhong. Learning consistency-aware unsigned distance functions progressively from raw point clouds. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 2

[92] Qian-Yi Zhou and Vladlen Koltun. Dense scene reconstruction with points of interest. *ACM Transactions on Graphics (ToG)*, 32(4):1–8, 2013. 5, 6, 7, 8

[93] Chuhang Zou, Ersin Yumer, Jimei Yang, Duygu Ceylan, and Derek Hoiem. 3d-prnn: Generating shape primitives with recurrent neural networks. In *CVPR*, 2017. 2