Few-Shot Unsupervised Implicit Neural Shape Representation Learning with Spatial Adversaries

Amine Ouasfi [1] **Adnane Boukhayma** [1]

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 sparse 3D point clouds in the absence of ground truth supervision remains a very challenging task. While recent methods rely on smoothness priors to regularize the learning, our method introduces a regularization term that leverages adversarial samples around the shape to improve the learned SDFs. Through extensive experiments and evaluations, we illustrate the efficacy of our proposed method, highlighting its capacity to improve SDF learning with respect to baselines and the state-of-the-art using synthetic and real data.

1. Introduction

Obtaining faith-full and intelligible neural representations of the 3D world from limited and corrupted point clouds is a challenge of paramount importance, that finds applications in countless downstream computer vision and graphics tasks. While many methods rely on data priors learned from large fully labeled datasets, these priors can fail to generalize to unseen shapes especially under very sparse unoriented inputs (Chen et al., 2023a; Ouasfi & Boukhayma, 2024b). Hence, it is important to design learning frameworks that can lead to efficient and robust learning of implicit shape representations under such extreme constraints.

In this context, the learning strategy introduced by (Ma et al., 2021) (dubbed NeuralPull) have shown to be one of the most successful ones in learning implicit shapes from point cloud unsupervisedly. However, upon observing the behavior of the training and validation errors of this method under sparse and dense input point clouds (Figure 1), we notice that the validation error starts increasing quite early on in the training in the sparse input case, whilst the training loss keeps on decreasing. This suggests an overfitting problem evidently intensifying in the sparse setting. Qualitatively, this increase in the validation error is usually synonymous to deterioration in the extracted shape with symptoms varying between shape instances, including shape hallucinations, missing shape parts, and shape becoming progressively wavy, bumpy and noisy. In extreme cases, shapes can also break into separate components or clusters around input points. When the input is additionally noisy, these phenomena are further exacerbated.

Recent work in the field relies on various smoothness priors (*e.g.* (Chen et al., 2023a; Gropp et al., 2020; Ben-Shabat et al., 2022; Ouasfi & Boukhayma, 2024c)) to regularize the implicit shape functions, and hence reduce overfitting. One side of the problem that remains underexplored however is how training data is sampled during learning, and understanding to which extent this sampling could affect performance. This is even the more an important question in our situation. In fact, while standard supervised learning uses typically data/label sample pairs, fitting implicit representations entails mapping spatial coordinates to labels or pseudo labels, where these spatial queries can be sampled uniformly or normally around the input point cloud. In the case of our baseline NeuralPull, the nearest point could sample to a spatial query is a pseudo-label approximating the unavailable nearest groundtruth surface point in the training. Hence, both inherent input point cloud noise and its sparsity represent a *noise* (*i.e.* displacement) on the perfect surface labels. This composite noise can affect both the SDF function and gradient orientation. In practice, we notice the network first produces a very smooth shape. When it tries to refine it, it tends to overfit to the noise present in the supervision signal. At this stage, further fitting on easy samples (predominant samples) means overfitting on this noise. The samples that can benefit the implicit representation can be drowned within easy samples.
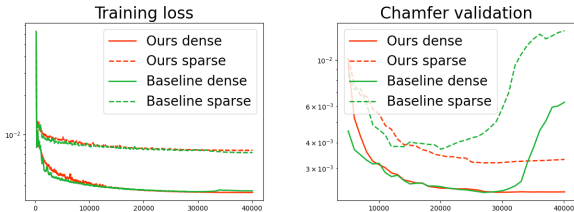
[1]Inria, Univ. Rennes, CNRS, IRISA, M2S, France. Correspondence to: Adnane Boukhayma <adnane.boukhayma@gmail.com>.

Figure 1. While the training loss (left) is decreasing for both our baseline (Ma et al., 2021) and our method, the Chamfer distance of reconstructions *w.r.t.* the GT starts increasing quite early on especially in the sparse input point cloud case for the baseline. This undesirable behaviour is remedied by our adversarial query mining. We report here metrics for unit box normalized meshes, using shape Gargoyle of dataset SRB (Williams et al., 2019).

Among literature interested in such a problem, active learning advocates sampling based on informativeness and diversity (Huang et al., 2010). New samples are queried from a pool of unlabeled data given a measure of these criteria. Informative samples are usually defined as samples that can reduce the uncertainty of a statistical model. However, several heuristics coexist as it is impossible to obtain a universal active learning strategy that is effective for any given task (Dasgupta, 2005). In our setting it is not clear what samples are the most informative for our implicit shape function and how to sample from the uncertainty regions of the implicit field. Recent work on distributionally robust optimization (DRO) (Volpi et al., 2018; Rahimian & Mehrotra, 2019) provides a mathematical framework to model uncertainty. In this framework, the loss is minimized over the worst case distribution in a neighborhood of the observed training data distribution. As a special case, adversarial training (Madry et al., 2017) uses pointwise adversaries rather than adversarial joint perturbations of the entire training set.

Inspired by this literature, we propose to use adversarial samples to regularize the learning of implicit shape representations from sparse point clouds. We build on SDF projection minimization error loss in training (See figure 2). Typically query points are pre-sampled around the input point cloud to train such a method. We augment these queries with adversarial samples during training. To ensure the diversity of these additional samples, we generate them in the vicinity of original queries within locally adapted radii. These radii modulate the adversarial samples density in concordance with the input point cloud density, thus allowing us to adapt to the local specificities of the input during the neural fitting. Our adversarial training strategy, focuses on samples that can still benefit the network, which prevents the aforementioned overfitting while refining the implicit representation.

To test our idea, we devise experiments on real and synthetic reconstruction benchmarks, including objects, articulated

shapes and large scenes. Our method outperforms the baseline as well as the most related competition both quantitatively and qualitatively. We notice that our adversarial loss helps our model most in places where shape prediction is the hardest and most ambiguous, such as fine and detailed structures and body extremities. Experiments on a dense reconstruction setting show that our method can be useful in this setup as well. Finally, thanks to our method, and as illustrated in Figure 1, validation stabilizes and plateaus at convergence unlike our baseline, which makes it easier for us to decide the evaluation model epoch, given that evaluation measurements are normally unavailable in unsupervised settings.

2. Related work

Classical shape modelling from point cloud includes combinatorial methods where the shape is defined through an input point cloud based space partitioning, through *e.g.* alpha shapes (Bernardini et al., 1999) Voronoi diagrams (Amenta et al., 2001) or triangulation (Cazals & Giesen, 2006; Liu et al., 2020; Rakotosaona et al., 2021). Differently, the input samples can be used to define an implicit function whose zero level set represents the target shape, using global smoothing priors (Williams et al., 2022; Lin et al., 2022; Williams et al., 2021) *e.g.* radial basis function (Carr et al., 2001) and Gaussian kernel fitting (Schölkopf et al., 2004), local smoothing priors such as moving least squares (Mercier et al., 2022; Guennebaud & Gross, 2007; Kolluri, 2008; Liu et al., 2021), or by solving a boundary conditioned Poisson equation (Kazhdan & Hoppe, 2013). The recent literature proposes to parameterise these implicit functions with deep neural networks and learn their parameters with gradient descent, either in a supervised (*e.g.* (Ouasfi & Boukhayma, 2024a; 2022; Boulch & Marlet, 2022; Peng et al., 2020; Lionar et al., 2021; Peng et al., 2021)) or unsupervised manner. These implicit neural representations (Mescheder et al., 2019; Park et al., 2019) overcome many of the limitations of explicit ones (*e.g.* meshes (Wang et al., 2018; Kato et al., 2018; Jena et al., 2022) and point clouds (Fan et al., 2017; Aliev et al., 2020; Kerbl et al., 2023)) in modelling shape, radiance and light fields (*e.g.* (Mildenhall et al., 2020; Yariv et al., 2021; Wang et al., 2021; Jain et al., 2022; Chan et al., 2022; Li et al., 2023a;b)), as they allow to represent functions with arbitrary topologies at virtually infinite resolution.

We are interested in unsupervised implicit neural shape learning. In this scenario, an MLP is typically fitted to the input point cloud without extra priors or information. Regularizations can compensate for the lack of supervision. For instance, (Gropp et al., 2020) introduced a spatial gradient constraint based on the Eikonal equation. (Ben-Shabat et al., 2022) introduces a spatial divergence constraint. (Liu et al.,

2022) propose a Lipschitz regularization on the network. (Ma et al., 2021; Ouasfi & Boukhayma, 2024c) expresses the nearest point on the surface as a function of the neural signed distance and its gradient. (Peng et al., 2021) proposed a differentiable Poisson solving layer that converts predicted normals into an indicator function grid efficiently. (Koneputugodage et al., 2023) guides the implicit field learning with an Octree based labelling. (Boulch et al., 2021) predicts occupancy fields by learning whether a dropped needle goes across the surface or not. (Chen et al., 2023a) learns a surface parametrization leveraged to provide additional coarse surface supervision to the shape network. Most methods can benefit from normals if available. (Atzmon & Lipman, 2020) proposed to supervise the gradient of the implicit function with normals, while (Williams et al., 2021) uses the inductive bias of kernel ridge regression. In the absence of normals and learning-based priors, and under input scarcity, most methods still display considerable failures. Differently from all existing work, we explore the use of adversarial samples in training implicit neural shape representations.

In the adversarial training literature, a trade-off between accuracy and robustness has been observed empirically in different datasets (Raghunathan et al., 2019; Madry et al., 2018). This has led prior work to claim that this tradeoff may be inevitable for many classification tasks (Tsipras et al., 2018; Zhang et al., 2019). However, several recent papers challenged this claim. (Yang et al., 2020) showed theoretically that this tradeoff is not inherent. It is rather a consequence of current robustness methods. These findings are corroborated empirically in recent work (Stutz et al., 2019; Xie et al., 2020; Herrmann et al., 2021). Our baseline relies on a pseudo-labeling strategy that introduces noise as the input gets sparser. Our method robustifies the learning against this noise, providing regularization and additional informative samples. The regularization helps prevent overfitting and enhances generalization, *i.e.* ensuring the loss behavior on the "training" query points is generalized in the 3D space, while informative samples aid in refining the shape function during advanced training stages.

3. Method

Given a noisy, sparse unoriented point cloud $\mathbf{P} \subset \mathbb{R}^{3 \times N_p}$, our objective is to obtain a corresponding 3D shape reconstruction, *i.e.* the shape surface $\mathcal{S}$ that best explains the observation $\mathbf{P}$. In other terms, the input point cloud elements should approximate noised samples from $\mathcal{S}$.

In order to achieve this goal, we learn a shape function f parameterised with an MLP f_θ. The function represents the implicit signed distance field relative to the target shape $\mathcal{S}$. That is, for a query euclidean space location $q \in \mathbb{R}^3$, $f(q) := s \cdot \min_{v \in \mathcal{S}} ||v - q||_2$, where $s := 1$ if q is inside
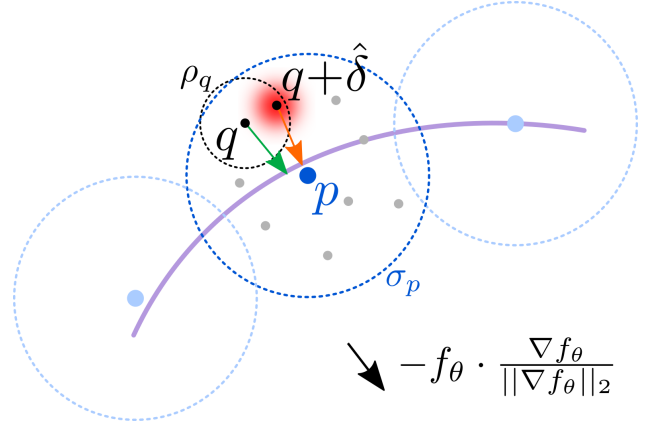

Figure 2. We learn an implicit shape representation f_θ from a point cloud (blue points) by minimizing the error between projection (through f_θ) of spatial queries q (gray points) onto the level set of the field (purple) and the nearest input point p. We introduce adversarial queries $q + \hat{\delta}$ to the optimization. They are defined as samples maximizing the loss in the vicinity of original queries.

$\mathcal{S}$, and $s := -1$ otherwise. The inferred shape $\hat{\mathcal{S}}$ can be obtained as the zero level set of the SDF (signed distance function) f_θ at convergence:

$$\hat{\mathcal{S}} = \{q \in \mathbb{R}^3 \mid f_\theta(q) = 0\}. \tag{1}$$

Practically, an explicit triangle mesh for $\hat{\mathcal{S}}$ can be obtained through the Marching Cubes algorithm (Lorensen & Cline, 1987), while querying neural network f_θ. We note also that $\hat{\mathcal{S}}$ can be rendered through ray marching (Hart, 1996) through the SDF field inferred by f_θ.

3.1. Background: Learning an SDF by pulling queries onto the surface.

Several state-of-the-art reconstruction from point cloud methods (*e.g.* (Chen et al., 2023a; Ma et al., 2022b;a; Chen et al., 2022; Ma et al., 2021)), including the state-of-the-art unsupervised reconstruction from sparse point cloud method (Chen et al., 2023a), build on the neural SDF training procedure introduced in (Ma et al., 2021) named NeuralPull. The latter is inspired by the observation that the distance field guided projection operator $q \mapsto q - f(q) \cdot \nabla f(q)$ ((Chibane et al., 2020; Perry & Frisken, 2001; Wolter, 1993; Zhao et al., 2021)) yields the nearest surface point when applied near the surface, where ∇f is the spatial gradient of f.

In practice, query points $q \in Q$ are sampled around the input point cloud $\mathbf{P}$, specifically from normal distributions centered at input samples $\{p\}$, with locally defined standard deviations $\{\sigma_p\}$:

$$Q := \bigcup_{p \in \mathbf{P}} \{q \sim \mathcal{N}(p, \sigma_p \mathbf{I}_3)\}, \tag{2}$$

where σ_p is chosen as the maximum euclidean distance to the K nearest points to p in $\mathbf{P}$. For each query q, the nearest point p in $\mathbf{P}$ is computed subsequently, and the following objective is optimized in (Ma et al., 2021) yielding a neural signed distance field f_θ whose zero level set concurs with the samples in $\mathbf{P}$:

$$\mathcal{L}(\theta, q) = ||q - f_\theta(q) \cdot \frac{\nabla f_\theta(q)}{||\nabla f_\theta(q)||_2} - p||_2^2,$$
$$\text{where} \quad p = \arg\min_{t \in \mathbf{P}} ||t - q||_2. \tag{3}$$

In the case of (Ma et al., 2021), the training procedure is empirical risk minimization (ERM), namely:

$$\min_\theta \mathop{\mathbb{E}}_{q \sim Q} \mathcal{L}(\theta, q), \tag{4}$$

where Q constitutes the empirical distribution over the training samples.

3.2. Local adversarial queries for the few shot setting

As introduced in the first section, input point cloud noise and sparsity are akin to noisy labels for query points q. Hence, training through standard ERM under sparse input point clouds $\mathbf{P}$ leads to an overfitting on this noise (Figure 1), *i.e.* useful information carried out by query samples decreasing during training thus leading to poor convergence. Hence, differently from existing work in the field of learning based reconstruction from point cloud, we propose to focus on the manner in which query points q are sampled at training, as we hypothesise that there could be a different sampling strategy from the one proposed in NeuralPull (*i.e.* Equation 2) that can lead to better results. This hypothesis is backed by literature showing that hard sample mining can lead to improved generalization and reduced over-fitting (Xie et al., 2020; Chawla, 2010; Fernández-Delgado et al., 2014; Krawczyk, 2016; Shrivastava et al., 2016). Intuitively, exposing the network to the worst cases in training is likely to make it more robust and less specialized on the more common easy queries.

We explore a different procedure from standard ERM (Equation 4). Ideally, we wish to optimize θ under the worst distribution Q' of query points $\{q\}$ in terms of our objective function, meaning:

$$\min_\theta \max_{Q'} \mathop{\mathbb{E}}_{q \sim Q'} \mathcal{L}(\theta, q). \tag{5}$$

Such a training procedure is akin to a distributionally robust optimization (Sagawa* et al., 2020; Rahimian & Mehrotra, 2019) which is hard to achieve in essence. It was shown however that a special more attainable case of the latter consists in harnessing hard samples locally (Staib & Jegelka, 2017), that is looking for hard samples in the vicinity of the

original empirical samples:

$$\min_\theta \mathop{\mathbb{E}}_{q \sim Q} \max_{\delta, ||\delta||_2 \le \rho} \mathcal{L}(\theta, q + \delta). \tag{6}$$

Let us consider optimum perturbations $\hat{\delta}$. Using a first order Taylor expansion on loss $\mathcal{L}(\theta, q + \delta)$, we can write:

$$\begin{aligned} \hat{\delta} &= \arg\max_{||\delta||_2 \le \rho} \mathcal{L}(\theta, q + \delta), \\ &\approx \arg\max_{||\delta||_2 \le \rho} \mathcal{L}(\theta, q) + \delta^\top \nabla_q \mathcal{L}(\theta, q), \\ &\approx \arg\max_{||\delta||_2 \le \rho} \delta^\top \nabla_q \mathcal{L}(\theta, q). \end{aligned} \tag{7}$$

Leveraging this approximation, we can obtain the optimum value $\hat{\delta}$ as the solution to a classical dual norm problem, by using the equality case of Cauchy–Schwarz inequality applied to the scalar product $\delta^\top \nabla_q \mathcal{L}(\theta, q)$ for δ in a closed ball of radius ρ and center 0:

$$\hat{\delta} = \rho \frac{\nabla_q \mathcal{L}(\theta, q)}{||\nabla_q \mathcal{L}(\theta, q)||_2}, \tag{8}$$

where gradients $\nabla_q \mathcal{L}$ can be computed efficiently through automatic differentiation in a deep-learning framework (*e.g.* PyTorch (Paszke et al., 2019)).

Algorithm 1 The training procedure of our method.

Input: Point cloud $\mathbf{P}$, learning rate α, number of iterations N_{it}, batch size N_b.
Output: Optimal parameters θ^*.
  Compute local st. devs. $\{\sigma_p\}$ ($\sigma_p = \max_{t \in K\text{nn}(p, \mathbf{P})} ||t - p||_2$).
  $Q \leftarrow \text{sample}(\mathbf{P}, \{\sigma_p\})$ (Equ. 2)
  Compute nearest points in $\mathbf{P}$ for all samples in Q.
  Compute local radii $\{\rho_q\}$ ($\rho_q = \sigma_p \times 10^{-2}$, $p := \text{nn}(q, \mathbf{P})$).
  Initialize $\lambda_1 = \lambda_2 = 1$.
  for N_{it} times **do**
    Sample N_b query points $\{q, q \sim Q\}$.
    Compute losses $\{\mathcal{L}(\theta, q)\}$. (Equ. 3)
    Compute loss gradients $\{\nabla_q \mathcal{L}(\theta, q)\}$ with autodiff.
    Compute 3D offsets $\{\hat{\delta}\}$. (Equ. 8, using radii $\{\rho_q\}$)
    Compute adversarial losses $\{\mathcal{L}(\theta, q + \hat{\delta})\}$. (Equ. 3)
    Compute combined losses $\{\mathfrak{L}(\theta, q)\}$. (Equ. 9)
    $(\theta, \lambda_1, \lambda_2) \leftarrow (\theta, \lambda_1, \lambda_2) - \alpha \nabla_{\theta, \lambda_1, \lambda_2} \Sigma_q \mathfrak{L}(\theta, q)$
  end for

We found empirically that using local radii $\{\rho_q\}$ in our context improves over using a single global radius ρ and we provide and ablation later on of this design choice. We recall that each query point q has a nearest counterpart p in $\mathbf{P}$. As we want our adversarial sample $q + \hat{\delta}$ to still remain relatively close to p, we define $\{\rho_q\}$ as a fraction of local standard deviation σ_p of the nearest point p (*e.g.* $\rho_q = \sigma_p \times 10^{-2}$). Linking ρ_q to σ_p also allows us to adjust the local radii to the local input point cloud density. We remind that σ_p was used for sampling query points q around input point p, as explained in the previous background section.

To ensure the stability of our learning, we train our neural network by backpropagating a hybrid loss combining the original objective and the adversarial one, using the strategy in (Liebel & Körner, 2018) for multi-task learning:

$$\mathcal{L}(\theta, q) = \frac{1}{2\lambda_1}\mathcal{L}(\theta, q) + \frac{1}{2\lambda_2}\mathcal{L}(\theta, q + \hat{\delta}) \\ + \ln(1 + \lambda_1) + \ln(1 + \lambda_2), \quad (9)$$

where λ_1 and λ_2 are learnable weights. A summary of our training procedure is shown in Algorithm 1. A visual illustration of our training can be seen in Figure 2.

4. Results

To evaluate our method, we assess our ability to learn implicit shape representations given sparse and noisy point clouds. We use datasets from standard reconstruction benchmarks. These datasets highlight a variety of challenges of fitting coordinate based MLPs to sparse data as well as reconstruction more broadly. Following the literature, we evaluate our method by measuring the accuracy of 3D explicit shape models extracted after convergence from our MLPs. We compare quantitatively and qualitatively to the the state-of-the-art in our problem setting, *i.e.* unsupervised reconstruction from unoriented point cloud, including methods designed for generic point cloud densities and methods dedicated to the sparse setting. For the former, we compare to fully implicit deep learning methods such as NP (Ma et al., 2021), SAP (Peng et al., 2021), DIGS (Ben-Shabat et al., 2022), in addition to hybrid methods combining implicit and grid based representations such as OG-INR (Koneputugodage et al., 2023) and GP (GridPull) (Chen et al., 2023b). When it comes to methods dedicated to the sparse setting we compare to NTPS (Chen et al., 2023a) which is the closest method to ours as it focuses specifically on the sparse input case. We additionally compare to NDrop (Boulch et al., 2021). We show results for NSpline (Williams et al., 2021) even though it requires normals. We also compare to classical Poisson Reconstruction SPSR (Kazhdan & Hoppe, 2013). We note also that comparisons to NP (NeuralPull, our baseline) also serves additionally as an ablation of our adversarial loss through out our experiments. For comprehensive evaluation, we also include comparisons to supervised methods including state-of-the-art feed-forward generalizable methods, namely POCO (Boulch & Marlet, 2022), CONet (Peng et al., 2020) and NKSR (Huang et al., 2023), alongside the finetuning method SAC (Tang et al., 2021) and the prior-based optimization method dedicated to sparse inputs On-Surf (Ma et al., 2022a). Unless stated differently, we use the publicly available official implementations of existing methods. For sparse inputs, we experimented with point clouds of size $N_p = 1024$.

4.1. 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 CD_1** ($\times 10^2$), L2 **Chamfer Distance CD_2** ($\times 10^2$), 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 appendix.

4.2. Datasets and input definitions

ShapeNet (Chang et al., 2015) consists of various instances of 13 different synthetic 3D object classes. We follow the train/test splits defined in (Williams et al., 2021). 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.* (Boulch & Marlet, 2022; Peng et al., 2020)). For brevity we show results on classes Tables, Chairs and Lamps.

Faust (Bogo et al., 2014) 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 (Zhou & Koltun, 2013) contains large scale complex real world scenes obtained with a handheld commodity range sensor. We follow (Chen et al., 2023a; Jiang et al., 2020; Ma et al., 2021) 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) (Williams et al., 2019) 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.

4.3. Implementation details

Our MLP (f_θ) follows the architecture in NP (Ma et al., 2021). We train for $N_{it} = 40000$ iterations using the Adam optimizer. We use batches of size $N_b = 5000$. Following NP, we set $K = 51$ for estimating local standard deviations σ_p. We train on a NVIDIA RTX A6000 GPU. Our method takes 8 minutes in average to converge for a 1024 sized input point cloud. In the interest of practicality and fairness in our comparisons, we decide the evaluation epoch for all the methods for which we generated results (including our main baseline) in the same way: we chose the best epoch for each method in terms of chamfer distance between the reconstruction and the input point cloud.
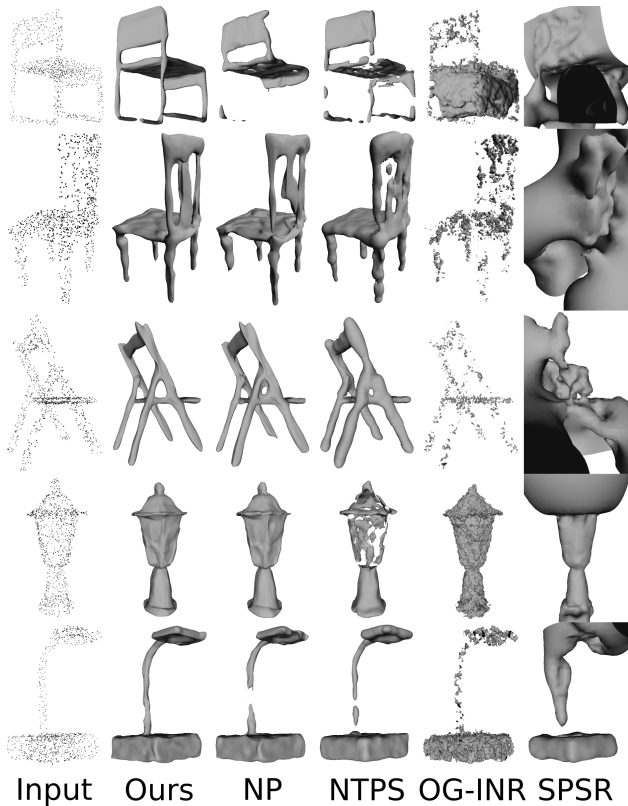
Figure 3. ShapeNet (Chang et al., 2015) reconstructions from sparse noisy unoriented point clouds.

	CD1	CD2	NC	FS
SPSR (Kazhdan & Hoppe, 2013)	2.34	0.224	0.74	0.50
OG-INR (Koneputugodage et al., 2023)	1.36	0.051	0.55	0.55
NP (Ma et al., 2021)	1.16	0.074	0.84	0.75
GP (Chen et al., 2023b)	1.07	0.032	0.70	0.74
NTPS (Chen et al., 2023a)	1.11	0.067	**0.88**	0.74
Ours	**0.76**	**0.020**	0.87	**0.83**

Table 1. ShapeNet (Chang et al., 2015) reconstructions from sparse noisy unoriented point clouds.

	CD1	CD2	NC	FS
POCO (Boulch & Marlet, 2022)	0.308	0.002	0.934	0.981
CONet (Peng et al., 2020)	1.260	0.048	0.829	0.599
On-Surf (Ma et al., 2022a)	0.584	0.012	0.936	0.915
SAC (Tang et al., 2021)	0.261	0.002	0.935	0.975
NKSR (Huang et al., 2023)	0.274	0.002	0.945	0.981
SPSR (Kazhdan & Hoppe, 2013)	0.751	0.028	0.871	0.839
GP (Chen et al., 2023b)	0.495	0.005	0.887	0.945
NTPS (Chen et al., 2023a)	0.737	0.015	0.943	0.844
NP (Ma et al., 2021)	0.269	0.003	0.951	0.973
Ours	**0.220**	**0.001**	**0.956**	**0.981**

Table 2. Faust (Bogo et al., 2014) reconstructions from sparse noisy unoriented point clouds.
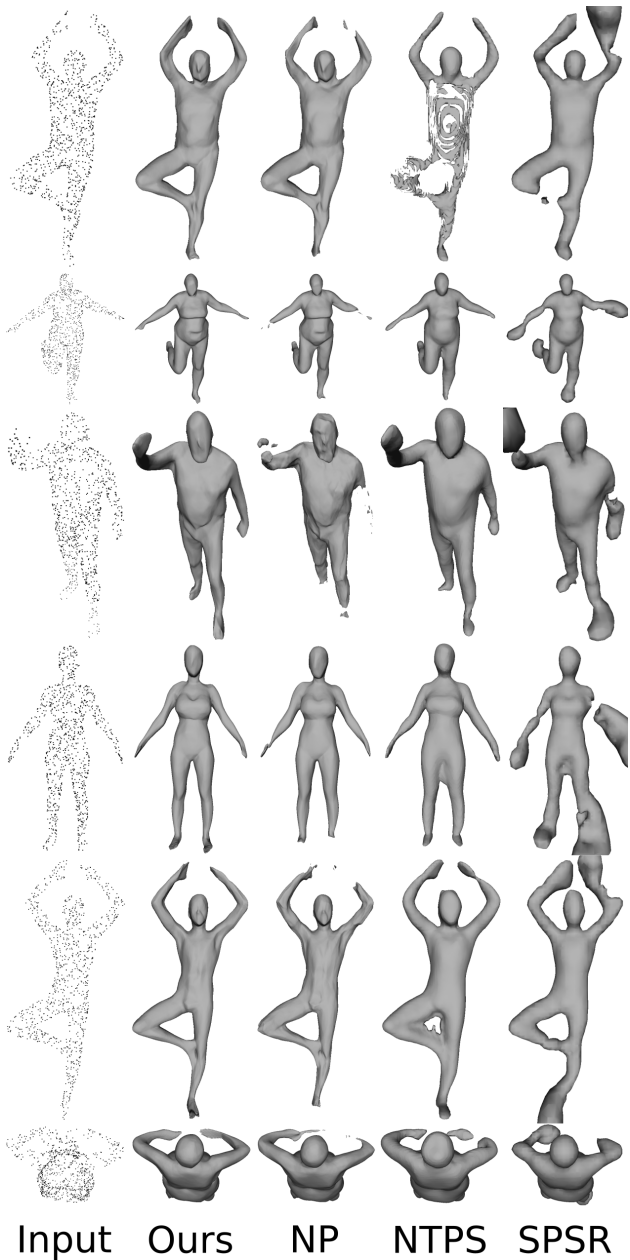

Figure 4. Faust (Bogo et al., 2014) reconstructions from sparse noisy unoriented point clouds.

4.4. Object level reconstruction

We perform reconstruction of ShapeNet (Chang et al., 2015) objects from sparse and noisy point clouds. Table 1 and Figure 3 show respectively a numerical and qualitative comparison to the competition. We outperform the competition across all metrics, as witnessed by the visual superiority of our reconstructions. We manage to recover fine structures and details with more fidelity. Although it obtains overall good coarse reconstructions, the thin plate spline smoothing

	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 (Kazhdan & Hoppe, 2013)	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 (Boulch et al., 2021)	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
NP (Ma et al., 2021)	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 (Peng et al., 2021)	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 (Williams et al., 2021)	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 (Chen et al., 2023a)	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	**0.051**	**0.006**	0.881	**0.037**	**0.002**	0.833	**0.044**	**0.011**	0.862	**0.035**	**0.003**	0.912	**0.042**	**0.002**	0.925	**0.041**	**0.004**	0.881

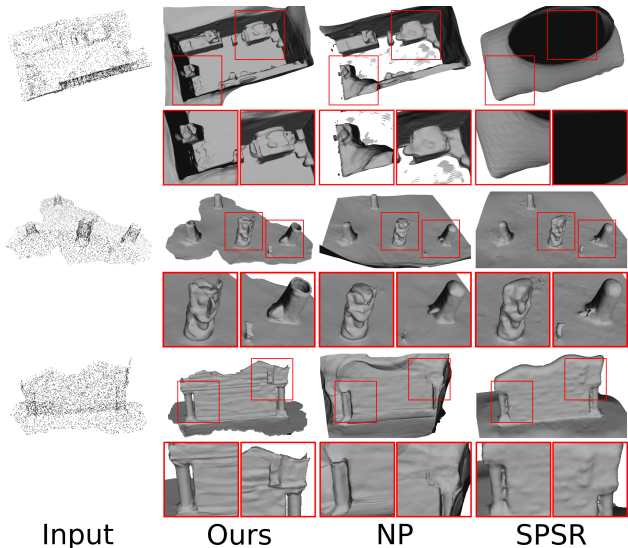*Table 3.* 3D Scene ((Zhou & Koltun, 2013)) reconstructions from sparse point clouds.

Figure 5. 3D Scene (Zhou & Koltun, 2013) reconstructions from sparse unoriented point clouds.

prior of NTPS seems to be hindering its expressivity. We noticed OG-INR fails to converge to satisfactory results under the sparse and noisy regime despite its effective Octree based sign field guidance in dense settings.

4.5. Real articulated shape reconstruction

We perform reconstruction of Faust ((Chang et al., 2015)) human shapes from sparse and noisy point clouds. Table 2 and Figure 4 show respectively a numerical and qualitative comparison to the competition. We outperform the other methods across all metrics. Visually, our reconstructions are particularly better at body extremities. Similarly to fine structures in the ShapeNet experiment, these are areas where the input point cloud samples are scarce and shape prediction is hard and ambiguous. NTPS reconstructions tend to be coarser and less detailed on this data as well. Notice that ShapeNet trained Generalizable methods (Top half of the table) do not all necessarily generalize well in this experiment.

4.6. Real scene level reconstruction

Following (Chen et al., 2023a), we report reconstruction results on the 3D Scene ((Zhou & Koltun, 2013)) data from spatially sparse point clouds. Table 3 summarizes numerical results. We compiled results for methods NTPS, NP, SAP, NDrop and NSpline as reported in state-of-the-art method NTPS. We outperform the competition is this benchmark thanks to our loss, as our baseline NP displays more blatant failures in this large scale sparse setup. Figure 5 shows qualitative comparisons to our baseline NP and SPSR. Red boxes highlight areas where our method displays particularly better details and fidelity in the reconstruction.

4.7. Varying the point cloud density

We use the SRB (Williams et al., 2019) benchmark to assess the behavior of our method across different point cloud densities. Table 4 indicates comparative results under both 1024 sized and dense input point clouds. We compiled results for the competition from OG-INR in the dense setting. We outperform our competition in the sparse case, and we perform on par with the state-of-the-art in the dense case. Our improvement *w.r.t.* our baseline (NP) is substantial for both sparse and dense inputs. This can be seen visually in Figure 6, where we show reconstructions for both sparse and dense cases. Notice how we recover better topologies in the sparse case and improved and more accurate details in the dense case, as pinpointed by the red boxes. These results showcase the utility and benefit of our contribution even in the dense setting. We note that SPSR becomes a very viable contestant qualitatively in the dense setting.

5. Ablation studies

The ablation of our main contribution is present throughout all Tables and Figures. In fact while we use the combined loss in Equation 9, our baseline (*i.e.* NP) uses solely the query projection loss in Equation 3. The improvement brought by our additional loss is visible across real/synthetic, scenes/objects, sparse/dense point clouds.

Perturbation radii We perform an ablation of using local *vs.* global radii ρ (Equation 8) and the choice of value of local radii ρ_q in Table 5. Results show that using local radii (ρ_q) is a superior strategy as it intuitively allows for
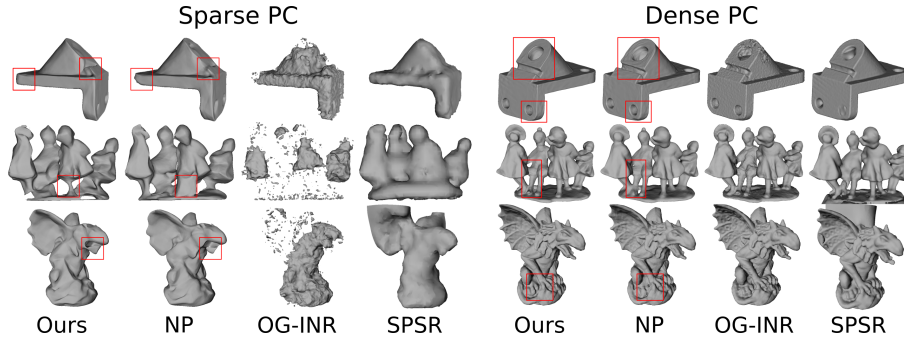
Figure 6. SRB (Williams et al., 2019) reconstructions from unoriented sparse and dense inputs.

	Sparse	Dense
SPSR (Kazhdan & Hoppe, 2013)	2.27	1.25
DIGS (Ben-Shabat et al., 2022)	0.68	**0.19**
OG-INR (Koneputugodage et al., 2023)	0.85	0.20
NTPS (Chen et al., 2023a)	0.73	-
NP (Ma et al., 2021)	0.58	0.23
Ours	**0.49**	**0.19**

Table 4. Ablation of point cloud density

	CD1	NC
NP (Ma et al., 2021) (baseline)	1.10	0.85
Ours (local) $\rho_q = \sigma_p/10$	0.92	0.86
Ours (local) $\rho_q = \sigma_p/100$	**0.75**	0.86
Ours (local) $\rho_q = \sigma_p/500$	0.88	0.86
Ours (local) $\rho_q = \sigma_p/1000$	1.02	0.84
Ours (global) $\rho = \sigma/100$	0.98	0.86

Table 5. Ablation of hard sample search radii.

Figure 7. Training time for shape Gargoyle of benchmark SRB (Williams et al., 2019).

a spatially adaptive search of hard samples (global radius is $\sigma \times 10^{-2}$, where σ is the average σ_p). We note that our baseline NP constitutes the special case $\rho_q = 0$. A value of $\sigma_p \times 10^{-2}$ achieves empirically satisfactory results (p being the nearest point to the query q in the input point cloud). Decreasing ρ_q leads expectedly to worse results as less hard queries are available for sampling. However, we also note that very large values of ρ_q can lead to spurious pseudo supervision, as adversarial samples $q + \delta$ run the risk of no longer having the same nearest point in the input point cloud as their original sample q. We performed this ablation on class Table of ShapeNet.

Multitask loss To guarantee the stability of our learning process, we employ a hybrid loss that combines the original objective with an adversarial one. This approach becomes crucial when a shape-specific trade-off between adversarial regularization and the original loss is required for the convergence of the shape function. In practical terms, this strategy outperformed using the adversarial loss alone, leading to an improvement in CD1 from 0.78 to 0.75 in class Table of
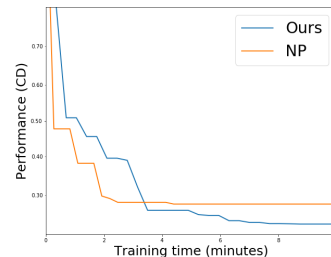
ShapeNet.

Increasing the number of query points We increase the number of NP original query samples to equal the total number of samples in our method (*i.e.* original queries + adversarial queries). We find that the performance of NP with extra queries only leads occasionally to marginal improvement (On average Chamfer distance goes from 0.581 to 0.576 in SRB).

Training time We provide in Figure 7 a plot showing the best performance attained by our method and the baseline (NP) *w.r.t.* the training time (RTX A6000). As it can be seen in this figure, within less than 4 minutes of training, we already surpass the best baseline performance. NP's performance stagnates from that point onwards due to overfitting, while our performance keeps on improving. We reach our optimal result after 8 minutes of training. The experiment was conducted on the challenging shape Gargoyle of benchmark SRB using a 1024 sized input point cloud.

6. Limitations

As can be seen in *e.g.* Figure 3, even though we improve on our baseline, we still face difficulties in learning very thin and convoluted shape details, such as chair legs and slats. Although our few-shot problem is inherently challenging

and can be under-constrained in many cases, we still believe there is room for improvement for our proposed strategy in this setting. For instance, one element that could be a source of hindrance to our accuracy is the difficulty of balancing empirical risk minimization and adversarial learning. In this work, we used an off-the-shelf self-trained loss weighting strategy, and we would like to address this challenge further as part of our future work.

7. Conclusion

We explored in this work a novel idea for regularizing implicit shape representation learning from sparse unoriented point clouds. We showed that harnessing adversarial samples locally in training can lead to numerous desirable outcomes, including superior results, reduced over fitting and easier evaluation model selection.

Impact statement

This paper presents work whose goal is to advance the fields of Machine Learning and 3D Computer Vision, specifically implicit neural shape representation learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Aliev, K.-A., Sevastopolsky, A., Kolos, M., Ulyanov, D., and Lempitsky, V. Neural point-based graphics. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16*, pp. 696–712. Springer, 2020.

Amenta, N., Choi, S., and Kolluri, R. K. The power crust, unions of balls, and the medial axis transform. *CG*, 2001.

Atzmon, M. and Lipman, Y. Sal: Sign agnostic learning of shapes from raw data. In *CVPR*, 2020.

Ben-Shabat, Y., Koneputugodage, C. H., and Gould, S. Digs: Divergence guided shape implicit neural representation for unoriented point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19323–19332, 2022.

Bernardini, F., Mittleman, J., Rushmeier, H., Silva, C., and Taubin, G. The ball-pivoting algorithm for surface reconstruction. *TVCG*, 1999.

Bogo, F., Romero, J., Loper, M., and Black, M. J. FAUST: Dataset and evaluation for 3D mesh registration. In *CVPR*, 2014.

Boulch, A. and Marlet, R. Poco: Point convolution for surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6302–6314, 2022.

Boulch, A., Langlois, P.-A., Puy, G., and Marlet, R. Needrop: Self-supervised shape representation from sparse point clouds using needle dropping. In *2021 International Conference on 3D Vision (3DV)*, pp. 940–950. IEEE, 2021.

Carr, J. C., Beatson, R. K., Cherrie, J. B., Mitchell, T. J., Fright, W. R., McCallum, B. C., and Evans, T. R. Reconstruction and representation of 3d objects with radial basis functions. In *SIGGRAPH*, 2001.

Cazals, F. and Giesen, J. *Effective Computational Geometry for Curves and Surfaces*. 2006.

Chan, E. R., Lin, C. Z., Chan, M. A., Nagano, K., Pan, B., De Mello, S., Gallo, O., Guibas, L. J., Tremblay, J., Khamis, S., et al. Efficient geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16123–16133, 2022.

Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

Chawla, N. V. Data mining for imbalanced datasets: An overview. *Data mining and knowledge discovery handbook*, pp. 875–886, 2010.

Chen, C., Liu, Y.-S., and Han, Z. Latent partition implicit with surface codes for 3d representation. In *European Conference on Computer Vision (ECCV)*, 2022.

Chen, C., Han, Z., and Liu, Y.-S. 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)*, 2023a.

Chen, C., Liu, Y.-S., and Han, Z. Gridpull: Towards scalability in learning implicit representations from 3d point clouds. In *Proceedings of the ieee/cvf international conference on computer vision*, pp. 18322–18334, 2023b.

Chibane, J., Mir, A., and Pons-Moll, G. Neural unsigned distance fields for implicit function learning. In *NeurIPS*, 2020.

Dasgupta, S. Coarse sample complexity bounds for active learning. *Advances in neural information processing systems*, 18, 2005.

Fan, H., Su, H., and Guibas, L. J. A point set generation network for 3d object reconstruction from a single image. In *CVPR*, 2017.

Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research*, 15(1):3133–3181, 2014.

Gropp, A., Yariv, L., Haim, N., Atzmon, M., and Lipman, Y. Implicit geometric regularization for learning shapes. In *ICML*, 2020.

Guennebaud, G. and Gross, M. Algebraic point set surfaces. In *ACM siggraph 2007 papers*, pp. 23–es. 2007.

Hart, J. C. Sphere tracing: A geometric method for the antialiased ray tracing of implicit surfaces. *The Visual Computer*, 1996.

Herrmann, C., Sargent, K., Jiang, L., Zabih, R., Chang, H., Liu, C., Krishnan, D., and Sun, D. Pyramid adversarial training improves vit performance. *arXiv preprint arXiv:2111.15121*, 2021.

Huang, J., Gojcic, Z., Atzmon, M., Litany, O., Fidler, S., and Williams, F. Neural kernel surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4369–4379, June 2023.

Huang, S.-J., Jin, R., and Zhou, Z.-H. Active learning by querying informative and representative examples. *Advances in neural information processing systems*, 23, 2010.

Jain, A., Mildenhall, B., Barron, J. T., Abbeel, P., and Poole, B. Zero-shot text-guided object generation with dream fields. 2022.

Jena, S., Multon, F., and Boukhayma, A. Neural mesh-based graphics. In *European Conference on Computer Vision*, pp. 739–757. Springer, 2022.

Jiang, C., Sud, A., Makadia, A., Huang, J., Nießner, M., Funkhouser, T., et al. Local implicit grid representations for 3d scenes. In *CVPR*, 2020.

Kato, H., Ushiku, Y., and Harada, T. Neural 3d mesh renderer. In *CVPR*, 2018.

Kazhdan, M. and Hoppe, H. Screened poisson surface reconstruction. *TOG*, 2013.

Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023.

Kolluri, R. Provably good moving least squares. *TALG*, 2008.

Koneputugodage, C. H., Ben-Shabat, Y., and Gould, S. Octree guided unoriented surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16717–16726, 2023.

Krawczyk, B. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4):221–232, 2016.

Li, Q., Multon, F., and Boukhayma, A. Learning generalizable light field networks from few images. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023a.

Li, Q., Multon, F., and Boukhayma, A. Regularizing neural radiance fields from sparse rgb-d inputs. In *2023 IEEE International Conference on Image Processing (ICIP)*, pp. 2320–2324. IEEE, 2023b.

Liebel, L. and Körner, M. Auxiliary tasks in multi-task learning. *arXiv preprint arXiv:1805.06334*, 2018.

Lin, S., Xiao, D., Shi, Z., and Wang, B. Surface reconstruction from point clouds without normals by parametrizing the gauss formula. *ACM Transactions on Graphics*, 42(2):1–19, 2022.

Lionar, S., Emtsev, D., Svilarkovic, D., and Peng, S. Dynamic plane convolutional occupancy networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1829–1838, 2021.

Liu, H.-T. D., Williams, F., Jacobson, A., Fidler, S., and Litany, O. Learning smooth neural functions via lipschitz regularization. *arXiv preprint arXiv:2202.08345*, 2022.

Liu, M., Zhang, X., and Su, H. Meshing point clouds with predicted intrinsic-extrinsic ratio guidance. In *ECCV*, 2020.

Liu, S.-L., Guo, H.-X., Pan, H., Wang, P.-S., Tong, X., and Liu, Y. Deep implicit moving least-squares functions for 3d reconstruction. In *CVPR*, 2021.

Lorensen, W. E. and Cline, H. E. Marching cubes: A high resolution 3d surface construction algorithm. In *SIGGRAPH*, 1987.

Ma, B., Han, Z., Liu, Y.-S., and Zwicker, M. Neural-pull: Learning signed distance functions from point clouds by learning to pull space onto surfaces. In *ICML*, 2021.

Ma, B., Liu, Y.-S., and Han, Z. Reconstructing surfaces for sparse point clouds with on-surface priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6315–6325, 2022a.

Ma, B., Liu, Y.-S., Zwicker, M., and Han, Z. Surface reconstruction from point clouds by learning predictive context priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6326–6337, 2022b.

Madry, A., Makelov, A., Schmidt, L., Tsipras, D., and Vladu, A. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.

Madry, A., Makelov, A., Schmidt, L., Tsipras, D., and Vladu, A. Towards deep learning models resistant to adversarial attacks. In *International Conference on Learning Representations*, 2018.

Mercier, C., Lescoat, T., Roussillon, P., Boubekeur, T., and Thiery, J.-M. Moving level-of-detail surfaces. *ACM Transactions on Graphics (TOG)*, 41(4):1–10, 2022.

Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., and Geiger, A. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4460–4470, 2019.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

Ouasfi, A. and Boukhayma, A. Few'zero level set'-shot learning of shape signed distance functions in feature space. In *ECCV*, 2022.

Ouasfi, A. and Boukhayma, A. Mixing-denoising generalizable occupancy networks. In *3DV*, 2024a.

Ouasfi, A. and Boukhayma, A. Robustifying generalizable implicit shape networks with a tunable non-parametric model. *Advances in Neural Information Processing Systems*, 36, 2024b.

Ouasfi, A. and Boukhayma, A. Unsupervised occupancy learning from sparse point cloud. *CVPR*, 2024c.

Park, J. J., Florence, P., Straub, J., Newcombe, R., and Lovegrove, S. Deepsdf: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. Pytorch: An imperative style, high-performance deep learning library. *NeurIPS*, 2019.

Peng, S., Niemeyer, M., Mescheder, L., Pollefeys, M., and Geiger, A. Convolutional occupancy networks. In *European Conference on Computer Vision*, pp. 523–540. Springer, 2020.

Peng, S., Jiang, C., Liao, Y., Niemeyer, M., Pollefeys, M., and Geiger, A. Shape as points: A differentiable poisson solver. *Advances in Neural Information Processing Systems*, 34:13032–13044, 2021.

Perry, R. N. and Frisken, S. F. Kizamu: A system for sculpting digital characters. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 47–56, 2001.

Raghunathan, A., Xie, S. M., Yang, F., Duchi, J. C., and Liang, P. Adversarial training can hurt generalization. *arXiv preprint arXiv:1906.06032*, 2019.

Rahimian, H. and Mehrotra, S. Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*, 2019.

Rakotosaona, M.-J., Aigerman, N., Mitra, N., Ovsjanikov, M., and Guerrero, P. Differentiable surface triangulation. In *SIGGRAPH Asia*, 2021.

Sagawa*, S., Koh*, P. W., Hashimoto, T. B., and Liang, P. Distributionally robust neural networks. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=ryxGuJrFvS.

Schölkopf, B., Giesen, J., and Spalinger, S. Kernel methods for implicit surface modeling. In *NeurIPS*, 2004.

Shrivastava, A., Gupta, A., and Girshick, R. Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 761–769, 2016.

Staib, M. and Jegelka, S. Distributionally robust deep learning as a generalization of adversarial training. In *NIPS workshop on Machine Learning and Computer Security*, volume 3, pp. 4, 2017.

Stutz, D., Hein, M., and Schiele, B. Disentangling adversarial robustness and generalization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Tang, J., Lei, J., Xu, D., Ma, F., Jia, K., and Zhang, L. Sa-convonet: Sign-agnostic optimization of convolutional occupancy networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6504–6513, 2021.

Tsipras, D., Santurkar, S., Engstrom, L., Turner, A., and Madry, A. Robustness may be at odds with accuracy. *arXiv.org*, 2018. abs/1805.12152.

Volpi, R., Namkoong, H., Sener, O., Duchi, J. C., Murino, V., and Savarese, S. Generalizing to unseen domains via adversarial data augmentation. *Advances in neural information processing systems*, 31, 2018.

Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W., and Jiang, Y.-G. Pixel2mesh: Generating 3d mesh models from single rgb images. In *ECCV*, 2018.

Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., and Wang, W. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021.

Williams, F., Schneider, T., Silva, C., Zorin, D., Bruna, J., and Panozzo, D. Deep geometric prior for surface reconstruction. In *CVPR*, 2019.

Williams, F., Trager, M., Bruna, J., and Zorin, D. Neural splines: Fitting 3d surfaces with infinitely-wide neural networks. In *CVPR*, 2021.

Williams, F., Gojcic, Z., Khamis, S., Zorin, D., Bruna, J., Fidler, S., and Litany, O. Neural fields as learnable kernels for 3d reconstruction. In *CVPR*, 2022.

Wolter, F.-E. Cut locus and medial axis in global shape interrogation and representation. 1993.

Xie, C., Tan, M., Gong, B., Wang, J., Yuille, A. L., and Le, Q. V. Adversarial examples improve image recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 819–828, 2020.

Yang, Y.-Y., Rashtchian, C., Zhang, H., Salakhutdinov, R., and Chaudhuri, K. A closer look at accuracy vs. robustness. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Yariv, L., Gu, J., Kasten, Y., and Lipman, Y. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34:4805–4815, 2021.

Zhang, H., Yu, Y., Jiao, J., Xing, E., Ghaoui, L. E., and Jordan, M. Theoretically principled trade-off between robustness and accuracy. *arXiv.org*, 2019. abs/1901.08573.

Zhao, F., Wang, W., Liao, S., and Shao, L. Learning anchored unsigned distance functions with gradient direction alignment for single-view garment reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12674–12683, 2021.

Zhou, Q.-Y. and Koltun, V. Dense scene reconstruction with points of interest. *ACM Transactions on Graphics (ToG)*, 32(4):1–8, 2013.

A. Additional visualizations

Figures 8,9,10 show more multi-view qualitative comparisons to our baseline NP.
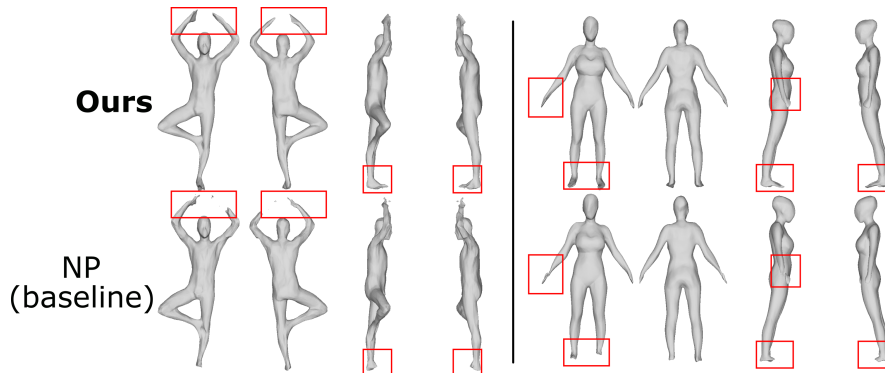

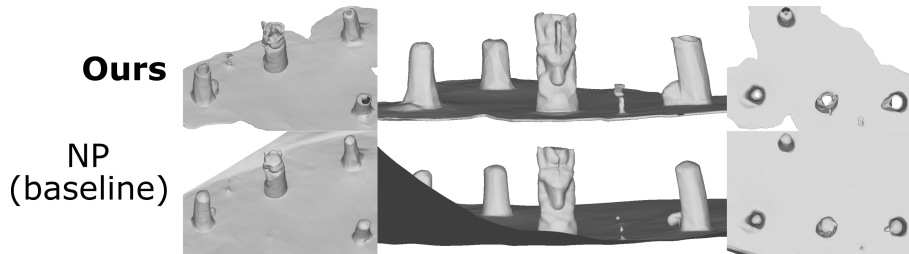*Figure 8.* Faust (Bogo et al., 2014) reconstructions from sparse noisy unoriented point clouds (1024 pts).

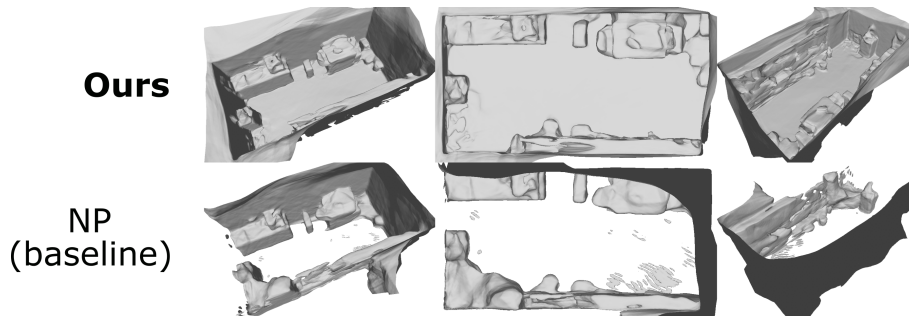*Figure 9.* 3D Scene (Zhou & Koltun, 2013) reconstructions from sparse unoriented point clouds (100 pts per m^2).

Figure 10. 3D Scene (Zhou & Koltun, 2013) reconstructions from sparse unoriented point clouds (100 pts per m^2).

B. Metrics

Following the definitions from (Boulch & Marlet, 2022) and (Williams et al., 2019), we present here the formal definitions for the metrics that we use for evaluation in the main submission. We denote by $\mathcal{S}$ and $\hat{\mathcal{S}}$ the ground truth and predicted mesh respectively. All metrics are approximated with 100k samples from the groundtruth mesh $\mathcal{S}$ and reconstruction $\hat{\mathcal{S}}$.

Chamfer Distance (CD$_1$) The L$_1$ Chamfer distance is based on the two-ways nearest neighbor distance:

$$\text{CD}_1 = \frac{1}{2|\mathcal{S}|} \sum_{v \in \mathcal{S}} \min_{\hat{v} \in \hat{\mathcal{S}}} \|v - \hat{v}\|_2 + \frac{1}{2|\hat{\mathcal{S}}|} \sum_{\hat{v} \in \hat{\mathcal{S}}} \min_{v \in \mathcal{S}} \|\hat{v} - v\|_2.$$

Chamfer Distance (CD$_2$)   The L$_2$ Chamfer distance is based on the two-ways nearest neighbor squared distance:

$$\text{CD}_2 = \frac{1}{2|\mathcal{S}|} \sum_{v \in \mathcal{S}} \min_{\hat{v} \in \hat{\mathcal{S}}} \|v - \hat{v}\|_2^2 + \frac{1}{2|\hat{\mathcal{S}}|} \sum_{\hat{v} \in \hat{\mathcal{S}}} \min_{v \in \mathcal{S}} \|\hat{v} - v\|_2^2.$$

F-Score (FS)   For a given threshold τ, the F-score between the meshes $\mathcal{S}$ and $\hat{\mathcal{S}}$ is defined as:

$$\text{FS}\left(\tau, \mathcal{S}, \hat{\mathcal{S}}\right) = \frac{2 \, \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}},$$

where

$$\text{Recall}\left(\tau, \mathcal{S}, \hat{\mathcal{S}}\right) = \mid \left\{ v \in \mathcal{S}, \text{ s.t. } \min_{\hat{v} \in \hat{\mathcal{S}}} \|v - \hat{v}\|_2 \langle \tau \right\} \mid,$$
$$\text{Precision}\left(\tau, \mathcal{S}, \hat{\mathcal{S}}\right) = \mid \left\{ \hat{v} \in \hat{\mathcal{S}}, \text{ s.t. } \min_{v \in \mathcal{S}} \|v - \hat{v}\|_2 \langle \tau \right\} \mid.$$

Following (Mescheder et al., 2019) and (Peng et al., 2020), we set τ to 0.01.

Normal consistency (NC)   We denote here by n_v the normal at a point v in $\mathcal{S}$. The normal consistency between two meshes $\mathcal{S}$ and $\hat{\mathcal{S}}$ is defined as:

$$\text{NC} = \frac{1}{2|\mathcal{S}|} \sum_{v \in \mathcal{S}} n_v \cdot n_{\text{closest}(v, \hat{\mathcal{S}})} + \frac{1}{2|\hat{\mathcal{S}}|} \sum_{\hat{v} \in \hat{\mathcal{S}}} n_{\hat{v}} \cdot n_{\text{closest}(\hat{v}, \mathcal{S})},$$

where

$$\text{closest}(v, \hat{\mathcal{S}}) = \text{argmin}_{\hat{v} \in \hat{\mathcal{S}}} \|v - \hat{v}\|_2.$$