

000  
001  
002054  
055  
056

# Towards Better Gradient Consistency for Neural Signed Distance Functions via Level Set Alignment

057  
058  
059003  
004  
005  
006  
007060  
061  
062  
063

Anonymous CVPR submission

008  
009  
010  
011  
012064  
065  
066  
067

Paper ID 4834

013  
014  
015068  
069  
070

## Abstract

016  
017  
018  
019  
020071  
072  
073  
074  
075

Neural signed distance functions (SDFs) have shown remarkable capability in representing geometry with details. However, without signed distance supervision, it is still a challenge to infer SDFs from point clouds or multi-view images using neural networks. In this paper, we claim that gradient consistency in the field, indicated by the parallelism of level sets, is the key factor affecting the inference accuracy. Hence, we propose a level set alignment loss to evaluate the parallelism of level sets, which can be minimized to achieve better gradient consistency. Our novelty lies in that we can align all level sets to the zero level set by constraining gradients at queries and their projections on the zero level set in an adaptive way. Our insight is to propagate the zero level set to everywhere in the field through consistent gradients to eliminate uncertainty in the field that is caused by the discreteness of 3D point clouds or the lack of observations from multi-view images. Our proposed loss is a general term which can be used upon different methods to infer SDFs from 3D point clouds and multi-view images. Our numerical and visual comparisons demonstrate that our loss can significantly improve the accuracy of SDFs inferred from point clouds or multi-view images under various benchmarks.

039  
040  
041  
042076  
077  
078  
079

## 1. Introduction

043  
044  
045  
046  
047080  
081  
082  
083  
084

Signed distance functions (SDFs) have shown remarkable abilities in representing high fidelity 3D geometry [23, 26, 29, 32, 38, 39, 48, 49, 54, 55, 61]. Current methods mainly use neural networks to learn SDFs as a mapping from 3D coordinates to signed distances. Using gradient descent, we can train neural networks by adjusting parameters to minimize errors to either signed distance ground truth [8, 22, 32, 38, 39] or signed distances inferred from 3D point clouds [1, 2, 9, 17, 26, 46, 62] or multi-view images [14, 19, 52–55, 58, 59, 61]. However, factors like the discreteness in point clouds and the lack of observations in

multi-view images result in 3D ambiguity, which makes inferring SDFs without ground truth signed distances remain a challenge.

Recent solutions [1, 18, 23, 46, 54] impose additional constraints on gradients with respect to input coordinates. The gradients determine the rate of change of signed distances in a field, which is vital for the accuracy of SDFs. Specifically, Eikonal term [1, 18, 23] is widely used to learn SDFs, which constrains the norm of gradients to be one at any location in the field. This regularization ensures the networks to predict valid signed distances. NeuralPull [26] constrains the directions of gradients to pull arbitrary queries onto the surface. One issue here is that these methods merely constrain gradients at single locations, without considering gradient consistency to their corresponding projections on different level sets. This results in inconsistent gradients in the field, indicated by level sets with poor parallelism, which drastically decreases the accuracy of inferred SDFs.

To resolve this issue, we introduce a level set alignment loss to pursue better gradient consistency for SDF inference without ground truth signed distances. Our loss is a general term which can be used to train different networks for learning SDFs from either 3D point clouds or multi-view images. Our key idea is to constrain gradients at corresponding locations on different level sets of the inferred SDF to be consistent. We achieve this by minimizing the cosine distance between the gradient of a query and the gradient of its projection on the zero level set. Minimize our loss is equivalent to aligning all level sets onto a reference, i.e. the zero level set, in a pairwise way. This enables us to propagate the zero level set to everywhere in the field, which eliminates uncertainty in the field that is caused by the discreteness of 3D point clouds or the lack of observations from multi-view images. Moreover, we introduce an adaptive weight to focus more on the gradient consistency nearer to the zero level set. We evaluate our loss upon the latest methods in surface reconstruction and multi-view 3D reconstruction under the widely used benchmarks. Our improvements over baselines justify not only our effectiveness but also the importance of gradient consistency to the inference of signed distance

108 fields. Our contributions are listed below.  
 109

- 110 i) We introduce a level set alignment loss to achieve better gradient consistency for inferring SDFs without signed distance ground truth.
- 111 ii) We justify the importance of gradient consistency to the accuracy of SDFs inferred from 3D point clouds and multi-view images, and show that aligning level sets together is an effective way of learning more consistent gradients for eliminating 3D ambiguity.
- 112 iii) We show our superiority over the state-of-the-art methods in surface reconstruction from point clouds and multi-view 3D reconstruction under the widely used benchmarks.

## 124 2. Related Work

125 Neural implicit representations have shown prominent  
 126 performance in representing 3D geometry with details [23, 26, 30, 32, 33, 36, 38, 39, 54, 55, 61]. With signed distances and occupancy labels as supervision, we can learn  
 127 neural implicit representations as a regression [39] or classification [30] problem. In the following, we focus on reviewing  
 128 methods inferring supervision from 3D point clouds [1, 18, 23] and multi-view images [33].

129 **Supervision from 3D Point Clouds.** Some methods learn  
 130 SDFs or occupancy with 3D point clouds as conditions.  
 131 They require signed distances and occupancy labels as supervision to learn global priors [7, 13, 16, 21, 24, 31, 38, 47, 50]  
 132 or local priors [7, 8, 22, 51, 56], which can be generalized to  
 133 unseen cases. With the ground truth field, these methods  
 134 get benefits including perfect scalar fields with consistent  
 135 gradients, but struggle to generalize the learned priors to  
 136 unseen cases with large geometry variations.

137 Some other methods infer SDFs without supervision by  
 138 training neural networks to overfit to single point cloud.  
 139 These methods require additional constraints [1, 2, 5, 17,  
 140 60, 62], specially designed operations [9, 26, 41] or normals  
 141 [6, 57] to estimate signed distances or occupancy using  
 142 point clouds as a reference. Using similar idea, we can infer  
 143 unsigned distances from point clouds [10, 63]. Using  
 144 inferred signed distances, some methods use the inferred  
 145 SDFs as priors, and then guide the SDF inference from a  
 146 novel point cloud [27, 28].

147 **Supervision from Multi-View Images.** With multi-view  
 148 images as supervision, classic multi-view stereo (MVS) [44,  
 149 45] methods use multi-view consistency to estimate depth  
 150 maps. With differentiable renderers [23, 52], we can render  
 151 images from the learned SDFs, and refine the learned SDFs  
 152 by minimizing errors between the rendered images and  
 153 ground truth images. Similarly, DVR [35] and IDR [59]  
 154 infer the radiance on surfaces, where IDR models view  
 155 direction as a condition to reconstruct high frequency details.

156 However, these methods focus on surfaces, which makes  
 157 them require masks of objects during optimization. Hence,  
 158 we can not use them to reveal structures for scenes, where  
 159 no masks are available.

160 NeRF [33] and the following work [34, 40, 42, 43] were  
 161 proposed for novel view synthesis, and render images  
 162 from radiance field use volume rendering without requiring  
 163 masks. By simultaneously modelling geometry and  
 164 color, we can infer signed distance or occupancy fields by  
 165 minimizing rendering errors. With samples on rays shooting  
 166 from pixels into the field, unisurf [36] and NeuS [54]  
 167 use a revised rendering procedure to render occupancy and  
 168 signed distance fields with radiance into pixel colors. Following  
 169 methods improve accuracy of implicit functions using  
 170 multi-view consistency [15, 54, 55, 61] or additional priors  
 171 including depth [3, 61, 65], normals [19, 53, 61].

172 The SDFs inferred these methods are not accurate, due  
 173 to the poor gradient consistency in signed distance fields,  
 174 indicated by level sets with poor parallelism. This is the  
 175 key factor that impacts on the accuracy of inferred SDFs  
 176 through neural rendering in a multi-view context or reasoning  
 177 on point clouds. We improve gradient consistency by  
 178 aligning level sets on the zero level set via minimizing a  
 179 level set alignment loss. Our loss is a general term that can  
 180 be used upon different methods.

## 184 3. Method

185 **Neural SDFs and Level Sets.** We focus on inferring an  
 186 SDF  $f$  from a 3D point cloud or a set of multi-view images  
 187 which does not provide ground truth signed distances.  $f$   
 188 predicts a signed distance  $s \in \mathbb{R}$  for an arbitrary query point  
 189  $q \in \mathbb{R}^3$ , as formulated by,

$$s = f_\theta(q), \quad (1)$$

190 where we use a neural network parameterized by  $\theta$  to learn  
 191 the SDF  $f$ . Level sets of  $f_\theta$  are denoted as  $\{\mathcal{S}_l\}$ , each of  
 192 which is a set of points where  $f_\theta$  takes on a given constant  
 193 value  $l$ ,

$$\mathcal{S}_l = \{q | f_\theta(q) = l\}, \quad (2)$$

194 where we regard zero level set  $\mathcal{S}_0$  as the surface of the scene.  
 195 We extract the surface as a triangle mesh by running the  
 196 marching cubes algorithm [25].

197 **Infering SDFs.** Without signed distance ground truth, current  
 198 methods infer SDFs by mining supervision from 3D  
 199 point clouds with normals [1, 18, 46], 3D point clouds without  
 200 normals [9, 26, 41], or multi-view images [14, 19, 52–  
 201 55, 58, 59, 61]. Although these methods use supervision in  
 202 different modalities, all of them minimize a general form of  
 203 loss function  $E$  to infer the SDF  $f_\theta$  below,

$$\min_{\theta} E(T(f_\theta), \mathbf{G}), \quad (3)$$

where  $G$  is the supervision including 3D point clouds with or without normals or multi-view images,  $T$  is a transformation function that transforms signed distances into a representation in the same modality of  $G$ , and  $E$  is the metric function that evaluates the error between the representation transformed from  $f_\theta$  and the ground truth supervision  $G$ . More specifically, NeuralPull [26] uses 3D point clouds without normals as  $G$ , then the function  $T$  projects a query  $q$  on  $G$  using its signed distance  $f_\theta(q)$  and gradient  $\nabla f_\theta(q)$ , and the loss  $E$  is mean squared error (MSE) between query projections and ground truth  $G$ . Siren [46] uses 3D point clouds with normals as  $G$ , the function  $T$  is the Eikonal term regulating gradients and MSE over signed distances of surface points, the loss  $E$  is an energy based metric. NeuS [54] uses a set of multi-view images as  $G$ , then uses volume rendering as the function  $T$  to render  $f_\theta$  along with radiance into images, and compares the rendered images to  $G$  using a MSE  $E$ .

**Gradient Consistency.** Our main contribution lies in pursuing better gradient consistency. We illustrate gradient consistency in the field using one query  $q$  in Fig. 1. If gradients are consistent, as shown in Fig. 1 (c), the gradient at query  $q$  and the gradient at its projection on each level set  $S_l$  should point to the same direction, which leads to level sets with great parallelism, while inconsistent gradients shown in Fig. 1 (a) result in level sets with poor parallelism. To evaluate gradient consistency at a query  $q$ , we use cosine distance between gradients at query and its projection on a level set  $S_l$ ,

$$c(q, S_l) = 1 - \frac{\nabla f_\theta(q) \cdot \nabla f_\theta(p^l)}{\|\nabla f_\theta(q)\|_2 \cdot \|\nabla f_\theta(p^l)\|_2}, \quad (4)$$

where  $p^l$  is the projection of  $q$  on the level set  $S_l$ .  $c(q, S_l)$  is in a range of  $[0, 2]$ , where 0 indicates  $f_\theta(q)$  and  $f_\theta(p^l)$  are pointing to the same direction, which are the most consistent.

One issue here is that it costs extensive computation if we evaluate  $c(q, S_l)$  on each level set  $S_l$ . Our solution here is to use zero level set  $S_0$  as a reference and project all queries onto the reference to evaluate the gradient consistency. In this pairwise way, we associate all level sets  $S_l$  to the zero level set  $S_0$ , which can propagate consistency to all level sets through the projections on  $S_0$  since we randomly sample a large amount of queries in each iteration during optimization. Hence, we pursue a better gradient consistency by minimizing Eq. 4 over all sampled queries  $\mathcal{Q}$ ,

$$\min_{\theta} \sum_{q \in \mathcal{Q}} c(q, S_0). \quad (5)$$

Minimizing the loss in Eq. 5 is equivalent to align all level sets to the zero level set, which we named it as level set alignment loss, as illustrated in Fig. 1 (c).

**Loss Function.** We use our level set alignment loss upon methods for inferring neural SDF without signed distance ground truth. We formulate our loss function by combining Eq. 3 and Eq. 5 below,

$$\min_{\theta} E_{q \in \mathcal{Q}}(T(f_\theta(q)), G) + \alpha \sum_{q \in \mathcal{Q}} \beta_q c(q, S_0). \quad (6)$$

where  $\alpha$  is the balance weight for our level set alignment loss, and it scales the per point weight  $\beta_q$  which is an adaptive weight indicating the importance of each query  $q$ , as defined below,

$$\beta_q = \exp(-\delta * |f_\theta(q)|), \quad (7)$$

where we model  $\beta_q$  according to the predicted signed distance, which aims to encourage the optimization to focus more on the area near the surface. Another option to replace  $|f_\theta(q)|$  is to use the distance between  $q$  and its nearest point in point cloud representing a surface. However, finding the nearest point for each query  $q$  increases computational burden. Moreover, we can not use it in scenarios without point clouds such as multi-view images.

By optimizing the objective function in Eq. 6, we can achieve more consistent gradients in the field, as illustrated in Fig. 2. By optimizing with our level set alignment loss, we reformulate the loss function of NeuralPull [26] into Eq. 6, which improves the gradient consistency in the field learned from 3D point clouds without normals. Fig. 2 (a) shows that we improve the parallelism of level sets, especially near the surface and inside of the dragon, where we visualize the signed distance field on a cross section of the reconstructed surface. This enables us to eliminate the swollen effect on the reconstructed surface of NeuralPull, which achieves a more compact surface with sharper edges. Similarly, we reformulate the loss function of NeuS [54] into Eq. 6 by adding our level set alignment loss, which improves the gradient consistency in the field learned from multi-view images. Fig. 2 (b) shows that the better gradient consistency leads to level sets with better parallelism, which propagates the zero level sets to everywhere in the field. This is a key factor to eliminate the artifacts in the empty space.

**Projections on the Zero Level Set.** We project a query  $q$  onto the zero level set  $S_0$ , and use the projection  $p^0$  to evaluate the gradient consistency defined in Eq. 4. As illustrated in Fig. 1(b), we follow the differentiable pulling operation in [26], and use the predicted signed distance  $f_\theta(q)$  and the gradient  $\nabla f_\theta(q)$  to project  $q$ , as formulated by,

$$p^0 = q - |f_\theta(q)| \frac{\nabla f_\theta(q)}{\|\nabla f_\theta(q)\|_2}. \quad (8)$$

By replacing  $p^l$  in Eq. 4 into  $p^0$ , we obtain  $c(q, S_0)$  in Eq. 5 and Eq. 6 below,

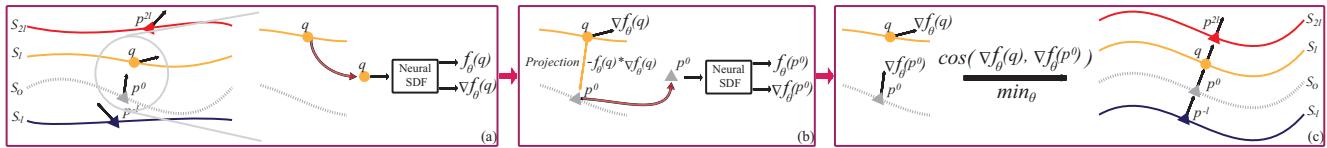
324  
325  
326  
327  
328

Figure 1. Overview of our level set alignment loss. We minimize our loss to pursue better gradient consistency in (c). The inconsistent gradient at a query  $\mathbf{q}$  in (a) and its projections on zero level set in (b) are constrained to be consistent.

331

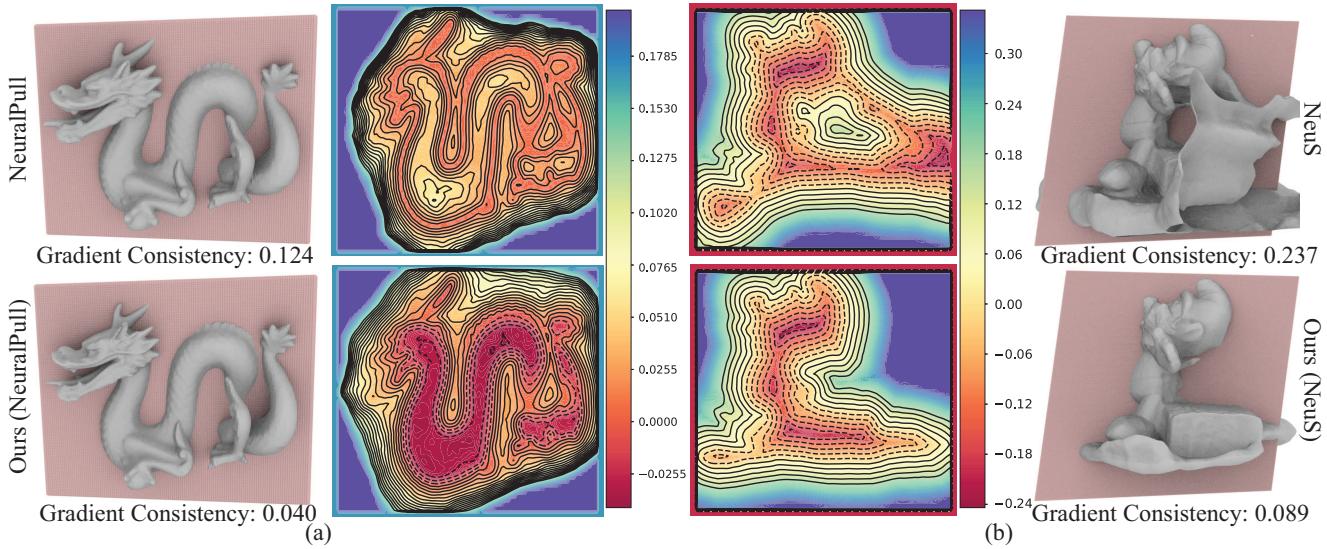


Figure 2. Visualization of level sets on a cross section. We pursue better gradient consistency in a field learned from 3D point clouds in (a) and multi-view images in (b). We minimize our level set alignment loss with NeuralPull in (a) and NeuS in (b), which leads to more accurate SDFs with better parallelism of level sets and less artifacts in empty space.

$$c(\mathbf{q}, \mathcal{S}_0) = 1 - \frac{\nabla f_\theta(\mathbf{q}) \cdot \nabla f_\theta(\mathbf{p}^0)}{\|\nabla f_\theta(\mathbf{q})\|_2 \cdot \|\nabla f_\theta(\mathbf{p}^0)\|_2}. \quad (9)$$

356

## 4. Experiments

357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369

We conduct experiments to evaluate our method in learning neural signed distance functions for 3D reconstruction from 3D point clouds and multi-view images. We use our level set alignment loss upon different methods to improve the performance by encouraging more consistent gradients in the field. We extract the zero level set of the learned signed distance functions using the marching cubes algorithm [25] as a surface. Note that we do not evaluate our performance with methods learning from signed distance ground truth, since the supervision provides perfect gradient consistency in the field, which dose not highlight our inference capability.

370

### 4.1. Surface Reconstruction from 3D Point Clouds

371  
372  
373  
374  
375  
376  
377

**Datasets.** We evaluate our performance under three datasets including the one released by SIREN [46], Stanford Scanning [11] and 3D Scene [64]. These datasets contains challenging cases including single objects and scenes with arbitrary topology and complex geometry. We use the point clouds in the dataset released by SIREN, which each scene

contains millions points, and we sample 2 million points for each shape or scene in Stanford Scanning and 3D Scene.

**Metrics.** We evaluate the accuracy of the learned SDFs using the error between the reconstructed meshes and ground truth. We use L1 Chamfer distance (CD) and normal consistency (NC) to measure the error. We sample 100k points on the reconstructed mesh and ground truth to calculate CD in Stanford Scanning dataset, and sample 1 million points to calculate CD in SIREN dataset and 3D Scene. We also use the normals estimated on the reconstructed meshes for the calculation of NC.

**Baselines.** We report our performance with the latest methods learning SDFs from 3D point clouds including I-GR [17], SIREN [46], NeuralPull [26]. These methods infer SDFs by training neural networks to overfit single 3D point cloud, with learning priors from large scale dataset. Specifically, IGR and SIREN adopt similar strategy to infer SDFs. They use Eikonal term to constrain the length of gradients to be one at everywhere, employ additional point normals to constrain gradients at points on surface, and set signed distances on surface to be zero. While NeuralPull constrains the gradients and signed distances together via pulling a query onto the surface without normals.

**Details.** To highlight our capability of inferring consistent gradients, we do not use the ground truth normal to produce our results with IGR and SIREN, since the ground truth nor-

378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431

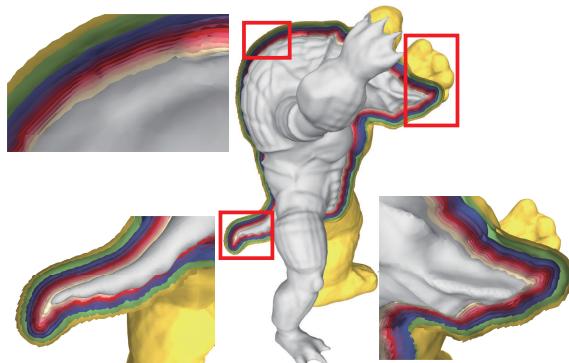


Figure 4. Visualization of level sets learned with our loss. We visualize the internal and external level sets, where the red surface represents the zero level set.

mal is a direct supervision for gradients. We report our result upon the baselines using their official code. We use the loss function of the baseline to replace the first term in Eq. 6, which is combined with our level set alignment loss into a loss function we use to report our results. We set the weight  $\alpha$  to make our loss contribute equally as the loss of the baseline.

**Comparison.** We report numerical evaluations in SIREN dataset in Tab. 1. In these point clouds with high frequency details, SIREN performs not well without using normals as supervision. Since ground truth normals determines the distance field near point clouds, which is the key to reconstruct accurate surface. But, without normals, the other loss terms in SIREN, such as the Eikonal term, can not infer accurate signed distances. While minimizing our loss can achieve more accurate signed distance field, which reveals surfaces with details even without using normal supervision.

We report our evaluation in Stanford scanning dataset in Tab. 2. Without using normals as supervision, SIREN and IGR reconstruct surfaces with artifacts. With our loss, we improve the gradient consistency in the field, especially near the surfaces. As shown in Fig. 5, we can eliminate the artifacts in the empty area, and obtain more completed and smoother surface. Since NeuralPull can not infer the zero level set very accurately, its reconstructed surfaces look a little bit “fat”. The swollen effect is mainly caused by inference uncertainty near surfaces. Training with our loss can make NeuralPull infer more accurate distance fields with much less uncertainty. This leads to more compact surfaces with more details, as shown in Fig. 5.

We further visualize the level sets learned with our loss in Fig. 3 and Fig. 4. The comparisons of level sets shown in Fig. 3 indicate that better gradient consistency can achieve more completed level sets near the surface. More visualization of level sets can be found in Fig. 4.

We further evaluate our method in 3D scene dataset in Tab. 3. Our level set alignment loss significantly improves the performance of baselines. Visual comparisons in Fig. 6 illustrate that we improve the field by removing artifacts n-

Metric	Thai		Room	
	SIREN	Ours(SIREN)	SIREN	Ours(SIREN)
CD	0.0043	<b>0.0011</b>	0.0189	<b>0.0023</b>
NC	0.942	<b>0.948</b>	0.946	<b>0.947</b>

Table 1. Numerical comparison in SIREN dataset.

	SIREN	Ours(SIREN)	IGR	Ours(IGR)	NP	Ours(NP)
CD	0.0130	<b>0.0129</b>	0.020	<b>0.011</b>	0.006	<b>0.004</b>
NC	0.942	<b>0.948</b>	0.946	<b>0.947</b>	0.955	<b>0.958</b>

Table 2. Numerical comparison in Stanford scanning.

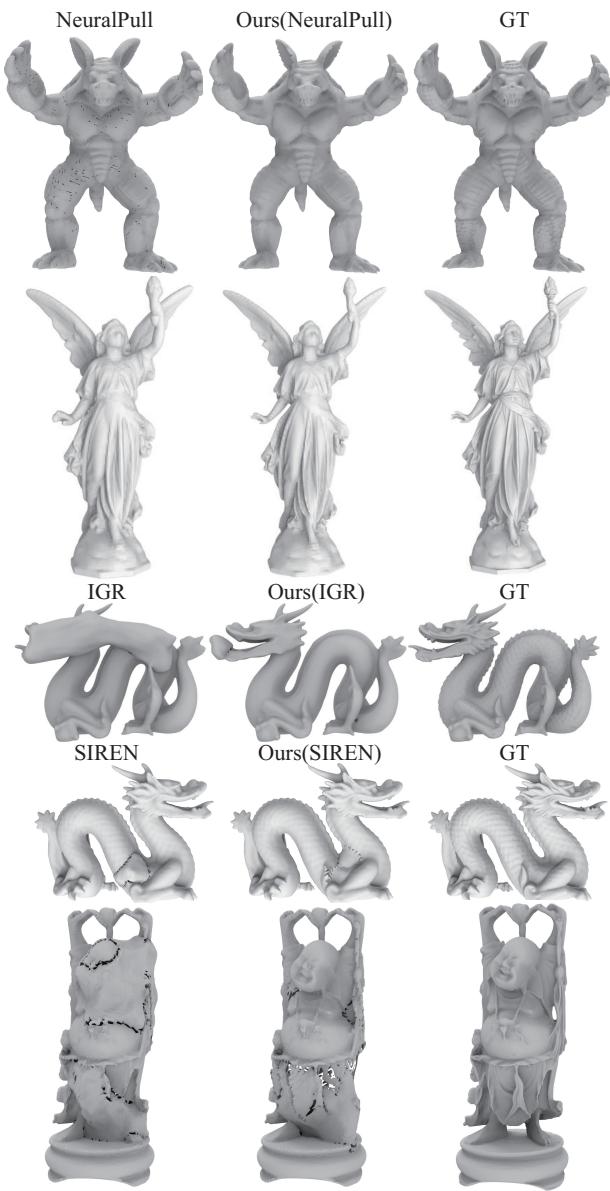


Figure 5. Visual comparison with baselines in 3D scene dataset.

ear the surface, reconstructing thinner and more compact surfaces, and sharpening surface edges.

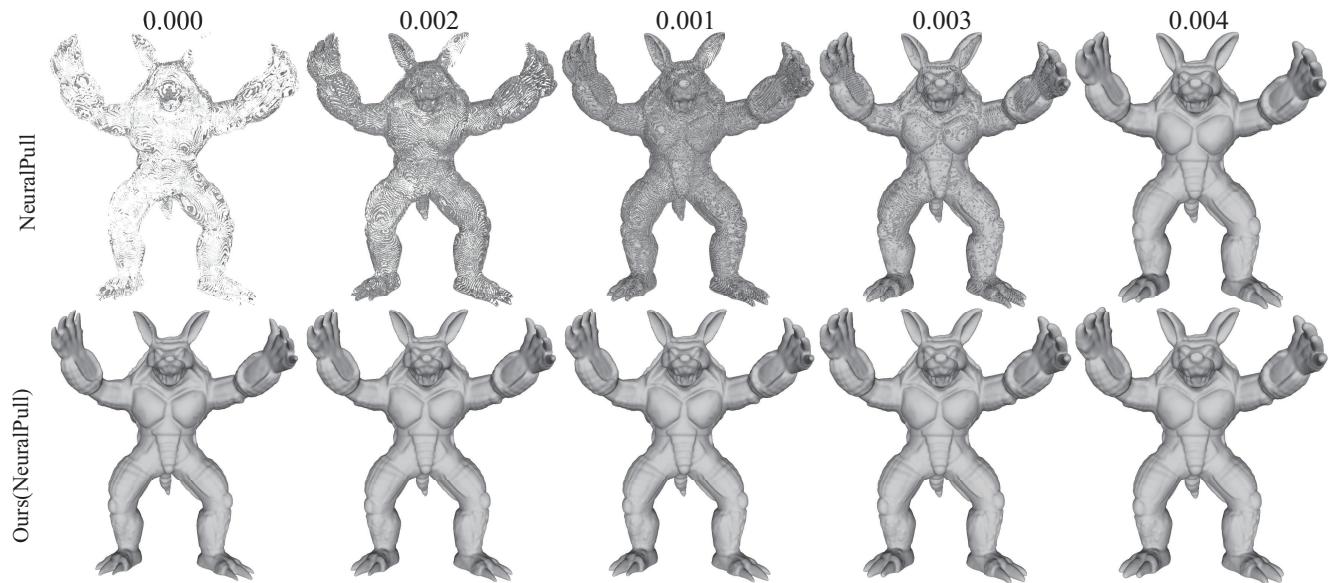


Figure 3. Visual comparisons of level sets with NeuralPull.

	Burghers		Lounge		Copyroom		Stonewall		Totempole	
	CD	NC								
MPU [37]	0.456	0.720	0.206	0.817	0.062	0.832	0.428	0.800	0.671	0.763
ConvOcc [47]	0.077	0.865	0.042	0.857	0.045	0.848	0.066	0.866	0.016	0.925
LIG [22]	0.018	0.904	0.017	0.910	0.018	0.910	0.020	0.928	0.023	0.917
NP [26]	0.010	0.883	0.059	0.857	0.011	0.884	0.007	0.868	0.010	0.765
Ours (NP)	<b>0.008</b>	<b>0.947</b>	<b>0.020</b>	<b>0.936</b>	<b>0.009</b>	<b>0.941</b>	<b>0.006</b>	<b>0.972</b>	<b>0.008</b>	<b>0.968</b>
SIREN [46]	0.025	0.944	0.064	0.933	0.032	0.917	<b>0.026</b>	0.938	0.032	<b>0.952</b>
Ours (SIREN)	<b>0.016</b>	<b>0.948</b>	<b>0.021</b>	<b>0.929</b>	<b>0.026</b>	<b>0.922</b>	0.031	<b>0.940</b>	<b>0.028</b>	0.937

Table 3. Numerical comparison with baselines in 3D scene dataset.

## 4.2. 3D Reconstruction from Multi-view Images

**Dataset.** We further evaluate our loss in reconstructing 3D shapes from multi-view images in the DTU dataset [20]. Following previous methods [14, 36, 54, 55, 58, 59, 61], we report our results on the widely used 15 scenes, each of which shows single object with background in 49 to 64 images with different shape appearances. For larger scale scenes, we report our results under ScanNet [12]. For fair comparison, we follow MonoSDF [61] to conduct evaluations using the same scenes.

**Metrics.** For evaluations under DTU dataset, we use L1 Chamfer distance to evaluate the error of points randomly sampled on the reconstructed surfaces compared to the ground truth. Following previous methods [14, 36, 54, 55, 58, 59, 61], we clean the reconstructed meshes using the respective masks. We use the official evaluation code released by the DTU dataset to measure our accuracy. For evaluations under ScanNet [12], we use the same metrics as MonoSDF, which includes Chamfer distance, F-Scores with a threshold of 5cm, and normal consistency to measure the error between the reconstructed surface and the ground truth surface in ScanNet.

**Baselines.** We add our loss on the latest methods for learning SDFs from multi-view images. we use NeuS [54] and MonoSDF [61] as baselines. NeuS does not use priors, and infer an SDF using multi-view consistency through volume rendering. MonoSDF adopts the same strategy and learns SDFs with depth and normal priors on images.

**Details.** We use the official code released by NeuS and MonoSDF to produce our results with our level set alignment loss. We use the loss function of the baseline to replace the first term in Eq. 6, which is combined with our level set alignment loss into a loss function we use to report our results. We set the weight  $\alpha$  to make our loss contribute equally as the loss of the baseline.

**Comparison.** We report numerical evaluations in DTU Tab. 4. We achieve better performance in 10 out of 15 scenes, and get comparable results in the other 5 scenes. In terms of the Chamfer distance, our improvements over NeuS are subtle. The reason is that the advantages of better gradient consistency lie in the ability of improving the smoothness of surfaces and removing artifacts in empty space. However, the smoothness does not significantly improve the numerical results, and artifacts in empty space has been cleaned using respective masks following the evalua-

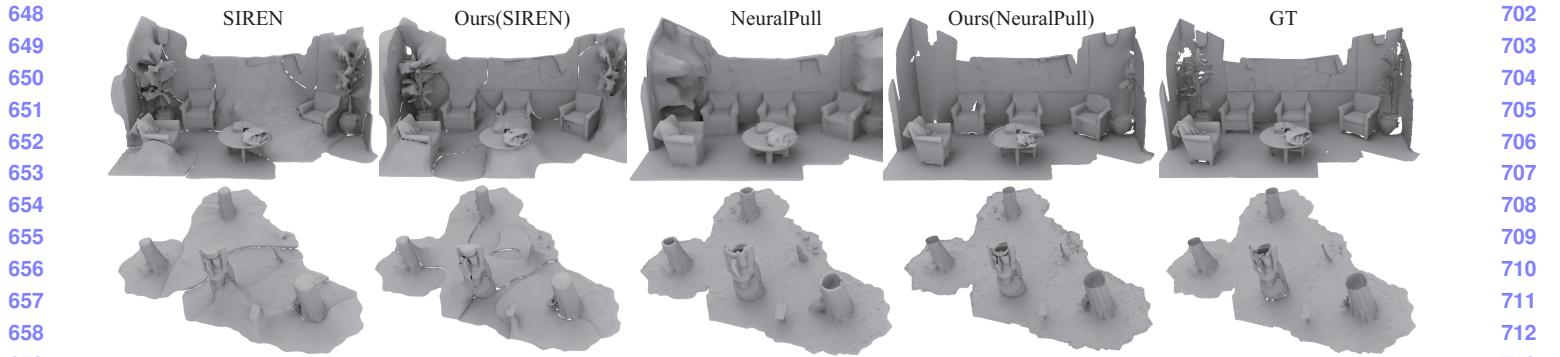


Figure 6. Visual comparison with baselines in 3D scene dataset.

Weight $\alpha$	0	0.001	0.01	0.1	1.0
CD $\times 100$	0.586	0.550	<b>0.394</b>	0.433	0.601
NC	0.955	0.955	<b>0.958</b>	0.957	0.940

Table 6. Effect of weight  $\alpha$ .

tion protocol. Hence, we highlight our improvements in visual comparison in Fig. 7, where we show the reconstructed surfaces before the cleaning.

As we can see, unisurf and NeuS learn neural implicit fields with lots of uncertainty, which is caused by the lack of multi-view consistency constraints or the ambiguity with the textureless background. This uncertainty results in artifacts especially in empty space. By minimizing our level set alignment loss, we can propagate the zero level set to all other level sets everywhere in the field through consistent gradients, which eliminates the uncertainty that can not get inferred from multi-view images. Hence, our results produce much less artifacts even in the area that few images can cover. We also show our rendered images as reference.

We further evaluate our method in ScanNet. We report average performance over each scene in Tab. 5. The numerical comparison show that we achieve the best performance among the state-of-the-art methods. Visual comparisons in Fig. 8 show that better gradient consistency reveals more geometry details.

### 4.3. Ablation Studies

We conduct ablation studies to justify the effectiveness of modules in our method. We use NeuralPull as a baseline and train it using our loss as one term in the loss function. We report our ablation studies under 3D scene dataset.

**Weights.** We explore the effect of our loss by adjusting the weight  $\alpha$  in Eq. 6. We report our results with different candidates  $\{0, 0.001, 0.01, 0.1, 1.0\}$ . The comparison in Tab. 6 shows that our level set alignment loss can improve the accuracy of inferred SDFs, it may affect the optimization to converge if we weight it too much.

**Adaptive per Point Weights.** We show the effect of the

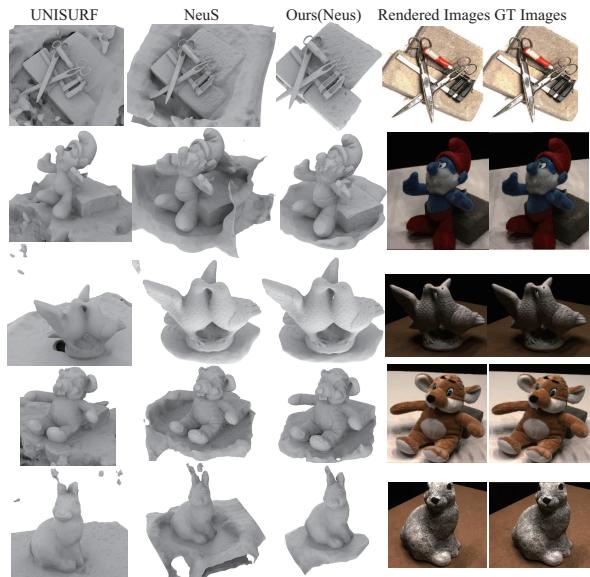


Figure 7. Visual comparison with baselines in DTU dataset.

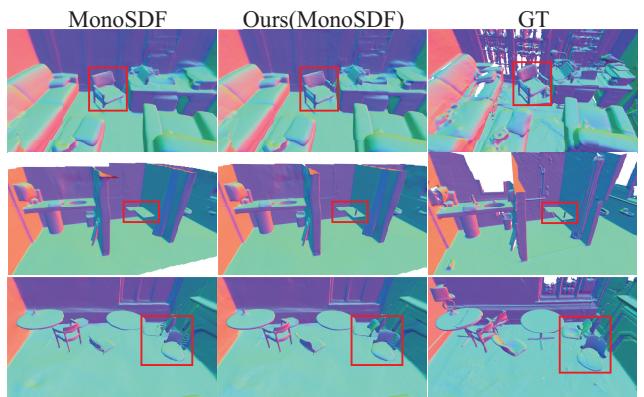


Figure 8. Visual comparison with baselines in ScanNet.

adaptive weight  $\beta_q$  for each query in Tab. 7. We report the result without the weight  $\beta_q$ , and the result with  $\beta_q$  that is obtained using the nearest distance to the point cloud rather than the predicted signed distance. The result of “0” indicates that weighting queries nearer to the surface more

756	Method	24	37	40	55	63	65	69	83	97	105	106	110	114	118	122	Mean	810
757	COLMAP [45]	0.81	2.05	0.73	1.22	1.79	1.58	1.02	3.05	1.40	2.05	1.00	1.32	0.49	0.78	1.17	1.36	811
758	NeRF [33]	1.90	1.60	1.85	0.58	2.28	1.27	1.47	1.67	2.05	1.07	0.88	2.53	1.06	1.15	0.96	1.49	812
759	UNISURF [36]	1.32	1.36	1.72	0.44	1.35	0.79	0.80	1.49	1.37	0.89	0.59	1.47	0.46	0.59	0.62	1.02	813
760	VolSDF [58]	1.14	1.26	0.81	0.49	1.25	0.70	0.72	1.29	1.18	0.70	0.66	1.08	0.42	0.61	0.55	0.86	814
761	NeuS [54]	1.37	1.21	<b>0.73</b>	<b>0.40</b>	1.20	0.70	0.72	<b>1.01</b>	1.16	<b>0.82</b>	0.66	1.69	0.39	0.49	<b>0.51</b>	0.87	815
762	Ours(NeuS)	<b>0.88</b>	<b>0.90</b>	0.80	0.41	<b>1.13</b>	<b>0.63</b>	<b>0.58</b>	1.37	<b>1.157</b>	0.83	<b>0.51</b>	<b>1.26</b>	<b>0.33</b>	<b>0.48</b>	0.52	<b>0.78</b>	816

Table 4. Numerical comparison with baselines in DTU dataset. The bars above numbers indicate the best.

	COLMAP [45]	UNISURF [36]	NeuS [54]	VolSDF [58]	Manhattan-SDF [19]	MonoSDF [61]	Ours(MonoSDF)
CD	0.141	0.359	0.194	0.267	0.070	0.042	<b>0.041</b>
F-score	0.537	0.267	0.291	0.364	0.602	0.733	<b>0.750</b>

Table 5. Numerical comparison with the state-of-the-art in ScanNet. We show normals as the color map on the surface.

Adaptive Weight $\beta_q$	0	1	10	100	Euclidean
CD×100	0.455	0.439	0.394	0.489	<b>0.391</b>
NC	0.955	0.957	<b>0.958</b>	0.956	<b>0.958</b>

Table 7. Effect of adaptive weight  $\beta_q$ .

Loss	Fixed	MSE-Nor	MSE	Cosine	DiGS	DiGS+Cosine
CD×100	1.384	0.486	0.494	<b>0.394</b>	0.601	0.412
NC	0.941	0.953	0.951	<b>0.958</b>	0.938	0.951

Table 8. Ablation studies on the loss function.

is important for the level set alignment, since all level sets are aligned to the zero level set. The result of “Euclidean” indicates that using inferred SDF achieves comparable results (the results of “10”) with using its nearest distance to the point cloud, but finding the nearest point for each query may increase the computational burden in large scale point clouds. We also compare the decay parameters  $\delta$  to obtain  $\beta_q$  in Eq. 7, and  $\delta = 10$  performs the best.

**Consistency with Surface Points.** We further justify how we compute gradient consistency. We report results with maximizing consistency between gradients at queries and gradients at their nearest points on the surface, rather than their projections on the zero level set. Compared to the projections on the zero level set which is optimized in different iterations, the nearest point on surface is fixed. In Tab. 8, the results of “Fixed” degenerate from the results of “Cosine”. The reason is that, during the early stage of optimization, the surface may not be the zero level set of the learned SDF, which brings lots of ambiguity and conflict if we use the nearest point as a reference. Hence, using projections on the zero level set in current iteration produces better results.

**Cosine Distance.** We show the advantages of cosine distance in Eq. 4. We replace cosine distance using a mean squared error with normalized gradients or with gradients without normalization. The results of “MSE-Norm” and “MSE” in Tab. 8 show that cosine distance performs better than MSE in SDF inference.

**Constraint on Second Order Derivatives.** We compare our loss with the constraints on second order derivatives in [4] which aims to smooth the change of gradients. Al-

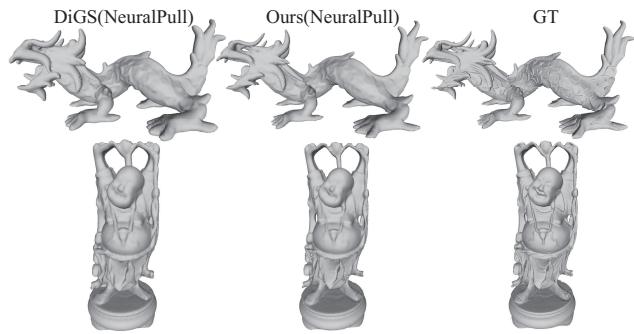


Figure 9. Visual comparison with the constraint on second order derivatives in DiGS.

though our loss also involves second order derivatives during gradient descent, we do not explicitly add constraints on the second order derivatives, which may result in unstable optimization. The comparison with the results of “DiGS” and the results of “DiGS+Cosine” indicate that our loss can reveal more accurate SDFs than the constraint on second order derivatives. The visual comparison with [4] in Fig. 9 shows that the constraint on second order derivatives can not achieve more compact and sharper surfaces as ours.

## 5. Conclusion

We improve the learning of SDFs without signed distance supervision by pursuing better gradient consistency. Our analysis shows that consistent gradients in the field are the key factor affecting the accuracy of inferred SDFs. To evaluate the gradient consistency, we introduce a level set alignment loss. By minimizing our loss, we successfully align all level sets onto the zero level set, which propagates the zero level set to eliminate 3D ambiguity through better gradient consistency. Our loss can be applied upon different methods a general term in loss function to improve the gradient consistency in the SDFs inferred from 3D point clouds or multi-view images. The visual and numerical comparisons with the state-of-the-art methods justify our effectiveness and show our superiority over the latest methods in SDF inference.

864

## References

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

- [1] Matan Atzmon and Yaron Lipman. Sal: Sign agnostic learning of shapes from raw data. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. [1](#), [2](#)
- [2] Matan Atzmon and yaron Lipman. SALD: sign agnostic learning with derivatives. In *International Conference on Learning Representations*, 2021. [1](#), [2](#)
- [3] Dejan Azinović, Ricardo Martin-Brualla, Dan B Goldman, Matthias Nießner, and Justus Thies. Neural rgb-d surface reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6290–6301, 2022. [2](#)
- [4] Yizhak Ben-Shabat, Chamin Hewa Koneputugodage, and Stephen Gould. Digs: Divergence guided shape implicit neural representation for unoriented point clouds. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. [8](#)
- [5] Yizhak Ben-Shabat, Chamin Hewa Koneputugodage, and Stephen Gould. Digs : Divergence guided shape implicit neural representation for unoriented point clouds. *CoRR*, abs/2106.10811, 2021. [2](#)
- [6] Alexandre Boulch, Pierre-Alain Langlois, Gilles Puy, and Renaud Marlet. Needrop: Self-supervised shape representation from sparse point clouds using needle dropping. In *International Conference on 3D Vision*, 2021. [2](#)
- [7] Alexandre Boulch and Renaud Marlet. Poco: Point convolution for surface reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6302–6314, 2022. [2](#)
- [8] Rohan Chabra, Jan Eric Lenssen, Eddy Ilg, Tanner Schmidt, Julian Straub, Steven Lovegrove, and Richard A. Newcombe. Deep local shapes: Learning local SDF priors for detailed 3D reconstruction. In *European Conference on Computer Vision*, volume 12374, pages 608–625, 2020. [1](#), [2](#)
- [9] Chao Chen, Yu-Shen Liu, and Zhizhong Han. Latent partition implicit with surface codes for 3d representation. In *European Conference on Computer Vision*, 2022. [1](#), [2](#)
- [10] Julian Chibane, Aymen Mir, and Gerard Pons-Moll. Neural unsigned distance fields for implicit function learning. *arXiv*, 2010.13938, 2020. [2](#)
- [11] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 1996. [4](#)
- [12] Angela Dai, Matthias Nießner, Michael Zollöfer, Shahram Izadi, and Christian Theobalt. Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface re-integration. *ACM Transactions on Graphics*, 2017. [6](#)
- [13] Philipp Erler, Paul Guerrero, Stefan Ohrhallinger, Niloy J. Mitra, and Michael Wimmer. Points2Surf: Learning implicit surfaces from point clouds. In *European Conference on Computer Vision*, 2020. [2](#)
- [14] Qiancheng Fu, Qingshan Xu, Yew-Soon Ong, and Wenbing Tao. Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. 2022. [1](#), [2](#), [6](#)
- [15] Qiancheng Fu, Qingshan Xu, Yew-Soon Ong, and Wenbing Tao. Geo-neus: Geometry-consistent neural implicit surfaces

- learning for multi-view reconstruction. In *Advances in Neural Information Processing Systems*, 2022. [2](#)
- [16] Kyle Genova, Forrester Cole, Daniel Vlasic, Aaron Sarna, William T. Freeman, and Thomas Funkhouser. Learning shape templates with structured implicit functions. In *International Conference on Computer Vision*, 2019. [2](#)
- [17] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3789–3799, 2020. [1](#), [2](#), [4](#)
- [18] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. *arXiv*, 2002.10099, 2020. [1](#), [2](#)
- [19] Haoyu Guo, Sida Peng, Haotong Lin, Qianqian Wang, Guofeng Zhang, Hujun Bao, and Xiaowei Zhou. Neural 3d scene reconstruction with the manhattan-world assumption. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. [1](#), [2](#), [8](#)
- [20] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engil Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 406–413, 2014. [6](#)
- [21] Meng Jia and Matthew Kyan. Learning occupancy function from point clouds for surface reconstruction. *arXiv*, 2010.11378, 2020. [2](#)
- [22] Chiyu Jiang, Avneesh Sud, Ameesh Makadia, Jingwei Huang, Matthias Nießner, and Thomas Funkhouser. Local implicit grid representations for 3D scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. [1](#), [2](#), [6](#)
- [23] Yue Jiang, Dantong Ji, Zhizhong Han, and Matthias Zwicker. SDFDiff: Differentiable rendering of signed distance fields for 3D shape optimization. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. [1](#), [2](#)
- [24] Shi-Lin Liu, Hao-Xiang Guo, Hao Pan, Pengshuai Wang, X-in Tong, and Yang Liu. Deep implicit moving least-squares functions for 3D reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. [2](#)
- [25] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics*, 21(4):163–169, 1987. [2](#), [4](#)
- [26] 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 *International Conference on Machine Learning*, 2021. [1](#), [2](#), [3](#), [4](#), [6](#)
- [27] Baorui Ma, Yu-Shen Liu, and Zhizhong Han. Reconstructing surfaces for sparse point clouds with on-surface priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6305–6315, 2022. [2](#)
- [28] Baorui Ma, Yu-Shen Liu, Matthias Zwicker, and Zhizhong Han. Surface reconstruction from point clouds by learning predictive context priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6316–6327, 2022. [2](#)
- [29] Julien N. P. Martel, David B. Lindell, Connor Z. Lin, Eric R. Chan, Marco Monteiro, and Gordon Wetzstein. ACORN:

- 972 adaptive coordinate networks for neural scene representation. *CoRR*, abs/2105.02788, 2021. 1 1026
- 973 [30] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: 1027 Learning 3D reconstruction in function space. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 2 1028
- 974 [31] Zhenxing Mi, Yiming Luo, and Wenbing Tao. Ssrnet: Scalable 3D surface reconstruction network. In *IEEE Conference 1029 on Computer Vision and Pattern Recognition*, 2020. 2 1030
- 975 [32] Mateusz Michalkiewicz, Jhony K. Pontes, Dominic Jack, 1031 Mahsa Baktashmotagh, and Anders P. Eriksson. Deep level 1032 sets: Implicit surface representations for 3D shape inference. *CoRR*, abs/1901.06802, 2019. 1, 2 1033
- 976 [33] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, 1034 Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: 1035 Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision*, 2020. 1, 2, 3, 4, 6 1036
- 977 [34] Thomas Müller, Alex Evans, Christoph Schied, and Alexander 1037 Keller. Instant neural graphics primitives with a multiresolution 1038 hash encoding. *arXiv:2201.05989*, 2022. 2 1039
- 978 [35] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and 1040 Andreas Geiger. Differentiable volumetric rendering: Learning 1041 implicit 3d representations without 3d supervision. In *IEEE 1042 Conference on Computer Vision and Pattern Recognition*, 2020. 2, 6 1043
- 979 [36] Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields 1044 for multi-view reconstruction. In *International Conference 1045 on Computer Vision*, 2021. 2, 6, 8 1046
- 980 [37] Yutaka Ohtake, Alexander G. Belyaev, Marc Alexa, Greg 1047 Turk, and Hans-Peter Seidel. Multi-level partition of unity 1048 implicits. *ACM Transactions on Graphics*, 22(3):463–470, 2003. 6 1049
- 981 [38] Amine Ouasfi and Adnane Boukhayma. Few ‘zero level set’- 1050 shot learning of shape signed distance functions in feature 1051 space. In *European Conference on Computer Vision*, 2022. 1, 2 1052
- 982 [39] Jeong Joon Park, Peter Florence, Julian Straub, Richard 1053 Newcombe, and Steven Lovegrove. DeepSDF: Learning 1054 continuous signed distance functions for shape representation. 1055 In *IEEE Conference on Computer Vision and Pattern 1056 Recognition*, 2019. 1, 2 1057
- 983 [40] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien 1058 Bouaziz, Dan B Goldman, Steven M. Seitz, and Ricardo 1059 Martin-Brualla. Nerfies: Deformable neural radiance fields. 1060 *IEEE International Conference on Computer Vision*, 2021. 2 1061
- 984 [41] Songyou Peng, Chiyu “Max” Jiang, Yiyi Liao, Michael 1062 Niemeyer, Marc Pollefeys, and Andreas Geiger. Shape as 1063 points: A differentiable poisson solver. In *Advances in Neural 1064 Information Processing Systems*, 2021. 2 1065
- 985 [42] Darius Rückert, Linus Franke, and Marc Stamminger. Adop: 1066 Approximate differentiable one-pixel point rendering. *arXiv:2110.06635*, 2021. 2 1067
- 986 [43] Sara Fridovich-Keil and Alex Yu, Matthew Tancik, Qinhong 1068 Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: 1069 Radiance fields without neural networks. In *IEEE Conference 1070 on Computer Vision and Pattern Recognition*, 2022. 2 1071
- 987 [44] Johannes Lutz Schönberger and Jan-Michael Frahm. 1072 Structure-from-motion revisited. In *IEEE Conference on 1073 Computer Vision and Pattern Recognition*, 2016. 2 1074
- 988 [45] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, 1075 and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on 1076 Computer Vision*, 2016. 2, 8 1077
- 989 [46] Vincent Sitzmann, Julien N.P. Martel, Alexander W. 1078 Bergman, David B. Lindell, and Gordon Wetzstein. Implicit 1079 neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems*, 2020. 1, 2, 3, 4, 6
- 990 [47] Lars Mescheder Marc Pollefeys Andreas Geiger Songyou 1080 Peng, Michael Niemeyer. Convolutional occupancy 1081 networks. In *European Conference on Computer Vision*, 2020. 2, 6 1082
- 991 [48] Towaki Takikawa, Joey Litalien, Kangxue Yin, Karsten 1083 Kreis, Charles Loop, Derek Nowrouzezahrai, Alec Jacobson, 1084 Morgan McGuire, and Sanja Fidler. Neural geometric level 1085 of detail: Real-time rendering with implicit 3D shapes. In *IEEE 1086 Conference on Computer Vision and Pattern Recognition*, 2021. 1 1087
- 992 [49] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara 1088 Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ra- 1089 mamoorthi, Jonathan T. Barron, and Ren Ng. Fourier 1090 features let networks learn high frequency functions in low 1091 dimensional domains. *NeurIPS*, 2020. 1 1092
- 993 [50] Jiapeng Tang, Jiabao Lei, Dan Xu, Feiying Ma, Kui Jia, 1093 and Lei Zhang. Sa-convolnet: Sign-agnostic optimization of 1094 convolutional occupancy networks. In *Proceedings of the 1095 IEEE/CVF International Conference on Computer Vision*, 2021. 2 1096
- 994 [51] Edgar Tretschk, Ayush Tewari, Vladislav Golyanik, Michael 1097 Zollhöfer, Carsten Stoll, and Christian Theobalt. PatchNets: 1098 Patch-Based Generalizable Deep Implicit 3D Shape 1099 Representations. *European Conference on Computer Vision*, 2020. 2 1100
- 995 [52] Delio Vicini, Sbastien Speierer, and Wenzel Jakob. Differentiable 1101 signed distance function rendering. *ACM Transactions 1102 on Graphics*, 41(4):125:1–125:18, 2022. 1, 2 1103
- 996 [53] Jiepeng Wang, Peng Wang, Xiaoxiao Long, Christian 1104 Theobalt, Taku Komura, Lingjie Liu, and Wenping Wang. 1105 Neuris: Neural reconstruction of indoor scenes using normal 1106 priors. In *European Conference on Computer Vision*, 2022. 1, 2 1107
- 997 [54] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku 1108 Komura, and Wenping Wang. Neus: Learning neural implicit 1109 surfaces by volume rendering for multi-view reconstruction. 1110 In *Advances in Neural Information Processing Systems*, 1111 pages 27171–27183, 2021. 1, 2, 3, 6, 8 1112
- 998 [55] Yiqun Wang, Ivan Skorokhodov, and Peter Wonka. Hf-neus: 1113 Improved surface reconstruction using high-frequency 1114 details. 2022. 1, 2, 6 1115
- 999 [56] Francis Williams, Teseo Schneider, Claudio Silva, Denis 1116 Zorin, Joan Bruna, and Daniele Panozzo. Deep geomet- 1117
- 1000

- 1080       ric prior for surface reconstruction. In *IEEE Conference on* 1134  
1081       *Computer Vision and Pattern Recognition*, 2019. 2 1135  
1082 [57] Francis Williams, Matthew Trager, Joan Bruna, and Denis 1136  
1083 Zorin. Neural splines: Fitting 3D surfaces with infinitely- 1137  
1084 wide neural networks. In *IEEE Conference on Computer* 1138  
1085 *Vision and Pattern Recognition*, pages 9949–9958, 2021. 2 1139  
1086 [58] Lior Yariv, Jitao Gu, Yoni Kasten, and Yaron Lipman. Vol- 1140  
1087 ume rendering of neural implicit surfaces. In *Advances in* 1141  
1088 *Neural Information Processing Systems*, 2021. 1, 2, 6, 8 1142  
1089 [59] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan 1143  
1090 Atzmon, Basri Ronen, and Yaron Lipman. Multiview neu- 1144  
1091 ral surface reconstruction by disentangling geometry and ap- 1145  
1092 pearance. *Advances in Neural Information Processing Sys-* 1146  
1093 *tems*, 33, 2020. 1, 2, 6 1147  
1094 [60] Wang Yifan, Shihao Wu, Cengiz Oztireli, and Olga Sorkine- 1148  
1095 Hornung. Iso-points: Optimizing neural implicit surfaces 1149  
1096 with hybrid representations. *CoRR*, abs/2012.06434, 2020. 2 1150  
1097 [61] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sat- 1151  
1098 tler, and Andreas Geiger. Monosdf: Exploring monocular 1152  
1099 geometric cues for neural implicit surface reconstruction. 1153  
1100 *ArXiv*, abs/2022.00665, 2022. 1, 2, 6, 8 1154  
1101 [62] Wenbin Zhao, Jiabao Lei, Yuxin Wen, Jianguo Zhang, and 1155  
1102 Kui Jia. Sign-agnostic implicit learning of surface self- 1156  
1103 similarities for shape modeling and reconstruction from raw 1157  
1104 point clouds. *CoRR*, abs/2012.07498, 2020. 1, 2 1158  
1105 [63] Junsheng Zhou, Baorui Ma, Yu-Shen Liu, Yi Fang, and 1159  
1106 Zhizhong Han. Learning consistency-aware unsigned 1160  
1107 distance functions progressively from raw point clouds. In *Ad-* 1161  
1108 *vances in Neural Information Processing Systems (NeurIPS)*, 1162  
1109 2022. 2 1163  
1110 [64] Qian-Yi Zhou and Vladlen Koltun. Dense scene reconstruc- 1164  
1111 tion with points of interest. *ACM Transactions on Graphics*, 1165  
32(4):112:1–112:8, 2013. 4 1166  
1112 [65] Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hu- 1167  
1113 jun Bao, Zhaopeng Cui, Martin R. Oswald, and Marc Polle- 1168  
1114 feys. Nice-slam: Neural implicit scalable encoding for slam. 1169  
1115 In *IEEE Conference on Computer Vision and Pattern Recog-* 1170  
1116 *nition*, 2022. 2 1171  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133