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

Baorui Ma[1]*, Junsheng Zhou[1]*, Yu-Shen Liu[1]†, Zhizhong Han[2]

[1]School of Software, BNRist, Tsinghua University, Beijing, China
[2]Department of Computer Science, Wayne State University, Detroit, USA

{mbr18,zhoujs21}@mails.tsinghua.edu.cn, liuyushen@tsinghua.edu.cn, h312h@wayne.edu

Abstract

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. Code and data are available at https://github.com/mabaorui/TowardsBetterGradient.

1. Introduction

Signed distance functions (SDFs) have shown remarkable abilities in representing high fidelity 3D geometry [6, 14,16,28,32,36,38,41,42,45,51,52,54,55,62,63,67–69,76]. 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 [9, 31, 45, 51, 52] or signed distances inferred from 3D point clouds [1,2,11,22,38,60,77] or multi-view images [19, 24, 66–69, 73, 74, 76]. 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 [1,23,32,60,68] 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 [1,23,32] 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 [38] 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 fields. Our contributions are listed below.

i) We introduce a level set alignment loss to achieve better gradient consistency for inferring SDFs without signed distance ground truth.

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.

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.

2. Related Work

Neural implicit representations have shown prominent performance in representing 3D geometry with details [14, 25, 32, 38, 43, 45, 46, 49, 51, 52, 68–70, 76]. With signed distances and occupancy labels as supervision, we can learn neural implicit representations as a regression [52] or classification [43] problem. In the following, we focus on reviewing methods inferring supervision from 3D point clouds [1, 23, 32] and multi-view images [46].

Supervision from 3D Point Clouds. Some methods learn SDFs or occupancy with 3D point clouds as conditions. They require signed distances and occupancy labels as supervision to learn global priors [8,18,21,30,35,44,51,61,64] or local priors [8, 9, 12, 31, 34, 65, 71], which can be generalized to unseen cases. With the ground truth field, these methods get benefits including perfect scalar fields with consistent gradients, but struggle to generalize the learned priors to unseen cases with large geometry variations.

Some other methods infer SDFs without supervision by training neural networks to overfit to single point cloud. These methods require additional constraints [1, 2, 5, 22, 75, 77], specially designed operations [11, 38, 54] or normals [7,33,72] to estimate signed distances or occupancy using point clouds as a reference. Using similar idea, we can infer unsigned distances from point clouds [13, 78]. Using inferred signed distances, some methods use the inferred SDFs as priors, and then guide the SDF inference from a novel point cloud [39, 40].

Supervision from Multi-View Images. With multi-view images as supervision, classic multi-view stereo (MVS) [58, 59] methods use multi-view consistency to estimate depth maps. With differentiable renderers [26,32,66], we can render images from the learned SDFs, and refine the learned SDFs by minimizing errors between the rendered images and ground truth images. Similarly, DVR [48] and IDR [74] infer the radiance on surfaces, where IDR models view direction as a condition to reconstruct high frequency details. However, these methods focus on surfaces, which makes them require masks of objects during optimization. Hence, we can not use them to reveal structures for scenes, where no masks are available.

NeRF [46] and the following work [47, 53, 56, 57] were proposed for novel view synthesis, and render images from radiance field use volume rendering without requiring masks. By simultaneously modelling geometry and color, we can infer signed distance or occupancy fields by minimizing rendering errors. With samples on rays shooting from pixels into the field, unisurf [49] and NeuS [68] use a revised rendering procedure to render occupancy and signed distance fields with radiance into pixel colors. Following methods improve accuracy of implicit functions using multi-view consistency [10, 20, 27, 68, 69, 76] or additional priors including depth [3, 76, 80], normals [24, 67, 76].

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

3. Method

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

$$s = f_\theta(\boldsymbol{q}), \tag{1}$$

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

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

where we regard zero level set $\mathcal{S}_0$ as the surface of the scene. We extract the surface as a triangle mesh by running the marching cubes algorithm [37].

Inferring SDFs. Without signed distance ground truth, current methods infer SDFs by mining supervision from 3D point clouds with normals [1,23,60], 3D point clouds without normals [11,38,54], or multi-view images [19,24,66–69,73,74,76]. Although these methods use supervision in different modalities, all of them minimize a general form of loss function E to infer the SDF f_θ below,

$$\min_\theta E(T(f_\theta), \boldsymbol{G}), \tag{3}$$

where $\boldsymbol{G}$ 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 $\boldsymbol{G}$, and E is the metric function that evaluates the error between the representation transformed from f_θ and the ground truth supervision $\boldsymbol{G}$. More specifically, NeuralPull [38] uses 3D point clouds without normals as $\boldsymbol{G}$, then the function T projects a query $\boldsymbol{q}$ on $\boldsymbol{G}$ using its singed distance $f_\theta(\boldsymbol{q})$ and gradient $\nabla f_\theta(\boldsymbol{q})$, and the loss E is mean squared error (MSE) between query projections and ground truth $\boldsymbol{G}$. Siren [60] uses 3D point clouds with normals as $\boldsymbol{G}$, 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 [68] uses a set of multi-view images as $\boldsymbol{G}$, then uses volume rendering as the function T to render f_θ along with radiance into images, and compares the rendered images to $\boldsymbol{G}$ 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 $\boldsymbol{q}$ in Fig. 1. If gradients are consistent, as shown in Fig. 1 (c), the gradient at query $\boldsymbol{q}$ and the gradient at its projection on each level set $\mathcal{S}_l$ 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 $\boldsymbol{q}$, we use cosine distance between gradients at query and its projection on a level set $\mathcal{S}_l$,

$$c(\boldsymbol{q}, \mathcal{S}_l) = 1 - \frac{\nabla f_\theta(\boldsymbol{q}) \cdot \nabla f_\theta(\boldsymbol{p}^l)}{||\nabla f_\theta(\boldsymbol{q})||_2 \cdot ||\nabla f_\theta(\boldsymbol{p}^l)||_2}, \tag{4}$$

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

One issue here is that it costs extensive computation if we evaluate $c(\boldsymbol{q}, \mathcal{S}_l)$ on each level set $\mathcal{S}_l$. Our solution here is to use zero level set $\mathcal{S}_0$ as a reference and project all queries onto the reference to evaluate the gradient consistency. In this pairwise way, we associate all level sets $\mathcal{S}_l$ to the zero level set $\mathcal{S}_0$, which can propagate consistency to all level

sets through the projections on $\mathcal{S}_0$ 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 $\mathcal{Q}$,

$$\min_\theta \sum_{\boldsymbol{q} \in \mathcal{Q}} c(\boldsymbol{q}, \mathcal{S}_0). \tag{5}$$

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,

$$\min_\theta E_{\boldsymbol{q} \in \mathcal{Q}}(T(f_\theta(\boldsymbol{q})), \boldsymbol{G}) + \alpha \sum_{\boldsymbol{q} \in \mathcal{Q}} \beta_{\boldsymbol{q}} c(\boldsymbol{q}, \mathcal{S}_0). \tag{6}$$

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

$$\beta_{\boldsymbol{q}} = exp(-\delta * |f_\theta(\boldsymbol{q})|), \tag{7}$$

where we model $\beta_{\boldsymbol{q}}$ 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 $|f_\theta(\boldsymbol{q})|$ is to use the distance between $\boldsymbol{q}$ and its nearest point in point cloud representing a surface. However, finding the nearest point for each query $\boldsymbol{q}$ 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 [38] 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 [68] 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.
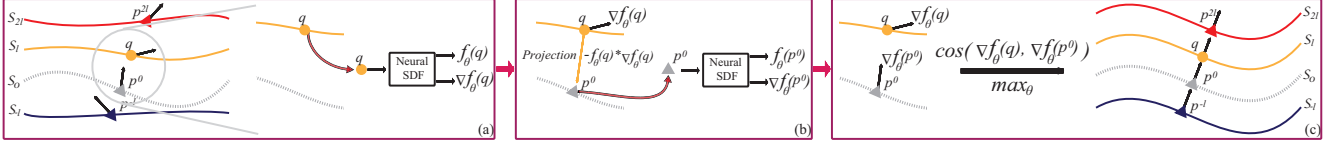
Figure 1. Overview of our level set alignment loss. We minimize our loss to pursue better gradient consistency in (c). The inconsistent gradient at a query q in (a) and its projections on zero level set in (b) are constrained to be consistent.
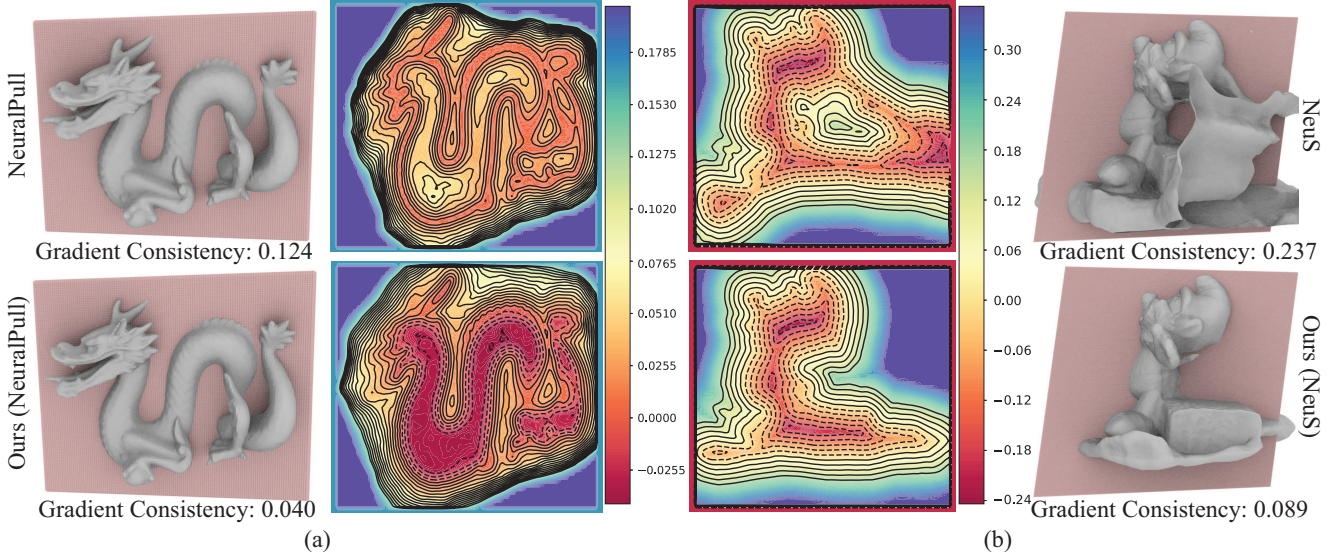

Figure 2. Visualization of level sets on a cross section. We pursue better gradient consistency in a field learned from 3D point clouds in (a) and multi-view images in (b). We minimize our level set alignment loss with NeuralPull in (a) and NeuS in (b), which leads to more accurate SDFs with better parallelism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

Projections on the Zero Level Set. We project a query q onto the zero level set $\mathcal{S}_0$, 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 [38], and use the predicted signed distance $f_\theta(q)$ and the gradient $\nabla f_\theta(q)$ to project q, as formulated by,

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

By replacing p^l in Eq. 4 into p^0, we obtain $c(q, \mathcal{S}_0)$ in Eq. 5 and Eq. 6 below,

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

4. Experiments

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 [37] 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.

4.1. Surface Reconstruction from 3D Point Clouds

Datasets. We evaluate our performance under three datasets including the one released by SIREN [60], Stanford Scanning [15] and 3D Scene [79]. 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 [22], SIREN [60], NeuralPull [38]. 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 normal 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 α 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 also report the comparison with normal supervision in our supplementary materials.

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 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.
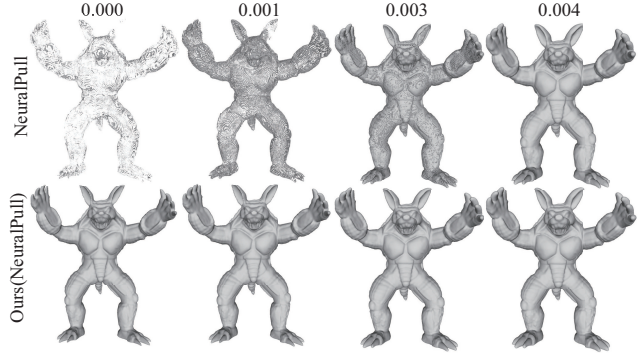

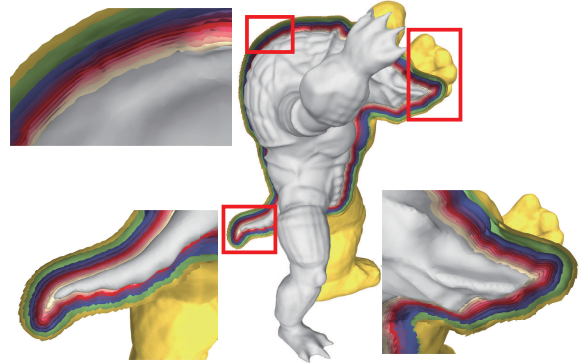Figure 3. Visual comparisons of level sets with NeuralPull.

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.

Metric	Thai		Room	
	SIREN	Ours(SIREN)	SIREN	Ours(SIREN)
CD	0.0043	**0.0011**	0.0189	**0.0023**

Table 1. Numerical comparison in SIREN dataset.

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

Table 2. Numerical comparison in Stanford scanning.

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 near the surface, reconstructing thinner and more compact surfaces, and sharpening surface edges. Moreover, we also report visual comparisons with methods using learned priors in our supplementary materials.

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 datatset [29]. Following previous methods [19, 49, 68, 69, 73, 74, 76], we

	Burghers		Lounge		Copyroom		Stonewall		Totempole	
	CD	NC	CD	NC	CD	NC	CD	NC	CD	NC
MPU [50]	0.456	0.720	0.206	0.817	0.062	0.832	0.428	0.800	0.671	0.763
ConvOcc [61]	0.077	0.865	0.042	0.857	0.045	0.848	0.066	0.866	0.016	0.925
LIG [31]	0.018	0.904	0.017	0.910	0.018	0.910	0.020	0.928	0.023	0.917
NP [38]	0.010	0.883	0.059	0.857	0.011	0.884	0.007	0.868	0.010	0.765
Ours (NP)	**0.008**	**0.947**	**0.020**	**0.936**	**0.009**	**0.941**	**0.006**	**0.972**	**0.008**	**0.968**
SIREN [60]	0.025	0.944	0.064	0.933	0.032	0.917	**0.026**	0.938	0.032	**0.952**
Ours (SIREN)	**0.016**	**0.948**	**0.021**	0.929	**0.026**	**0.922**	0.031	**0.940**	**0.028**	0.937

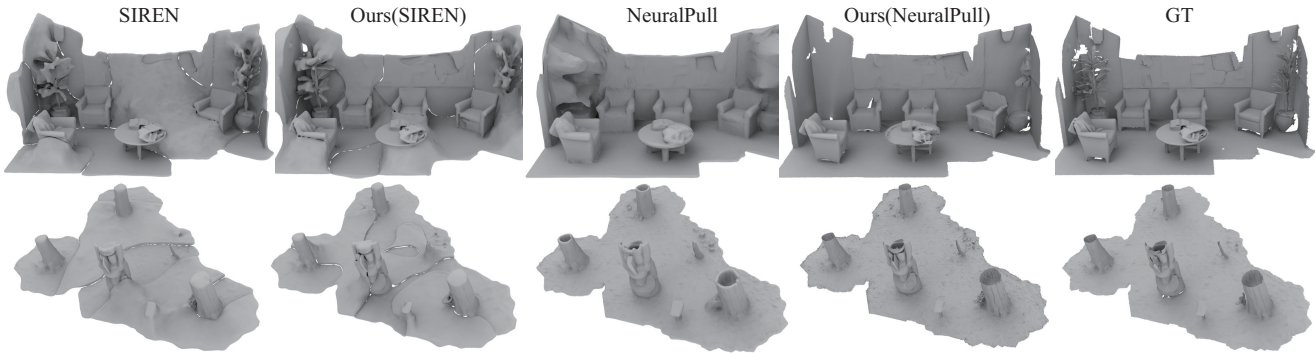Table 3. Numerical comparison with baselines in 3D scene dataset.

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

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 [17]. For fair comparison, we follow MonoSDF [76] 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 [19,49,68,69,73, 74,76], 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 [17], 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 [68] and MonoSDF [76] 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 α 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 evaluation protocol. Hence, we highlight our improvements in visual comparison in Fig. 7, where we show the reconstructed surfaces before the cleaning. More analysis and comparisons with NeuS can be found in supplementary materials.

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.

Method	24	37	40	55	63	65	69	83	97	105	106	110	114	118	122	Mean
COLMAP [59]	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
NeRF [46]	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
UNISURF [49]	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
VolSDF [73]	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
NeuS [68]	1.37	1.21	**0.73**	**0.40**	1.20	0.70	0.72	**1.01**	1.16	**0.82**	0.66	1.69	0.39	0.49	**0.51**	0.87
Ours(NeuS)	**0.88**	**0.90**	0.80	0.41	**1.13**	**0.63**	**0.58**	1.37	**1.157**	0.83	**0.51**	1.26	**0.33**	**0.48**	0.52	**0.78**

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

	COLMAP [59]	UNISURF [49]	NeuS [68]	VolSDF [73]	Manhattan-SDF [24]	MonoSDF [76]	Ours(MonoSDF)
CD	0.141	0.359	0.194	0.267	0.070	0.042	**0.041**
F-score	0.537	0.267	0.291	0.364	0.602	0.733	**0.750**

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

Weight α	0	0.001	0.01	0.1	1.0
CD×100	0.586	0.550	**0.394**	0.433	0.601
NC	0.955	0.955	**0.958**	0.957	0.940

Table 6. Effect of weight α.

Adaptive Weight β_q	0	1	10	100	Euclidean
CD×100	0.455	0.439	0.394	0.489	**0.391**
NC	0.955	0.957	**0.958**	0.956	**0.958**

Table 7. Effect of adaptive weight β_q.

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

Table 8. Ablation studies on the loss function.

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 α 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 adaptive weight β_q for each query in Tab. 7. We report the result without the weight β_q, and the result with β_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 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 δ to obtain β_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. Although 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.
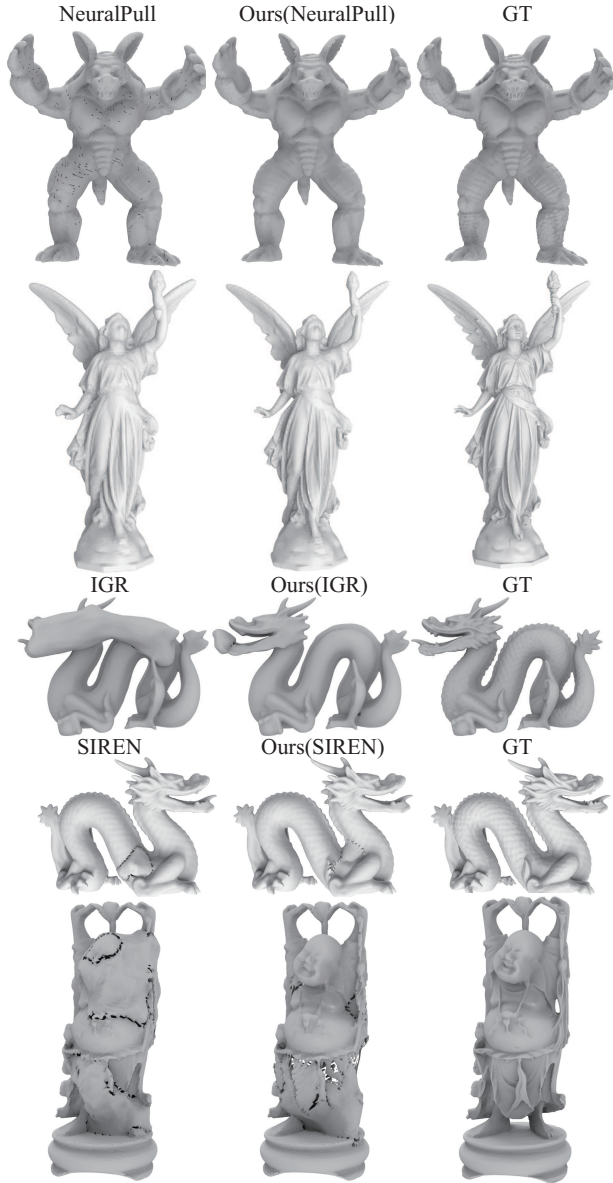
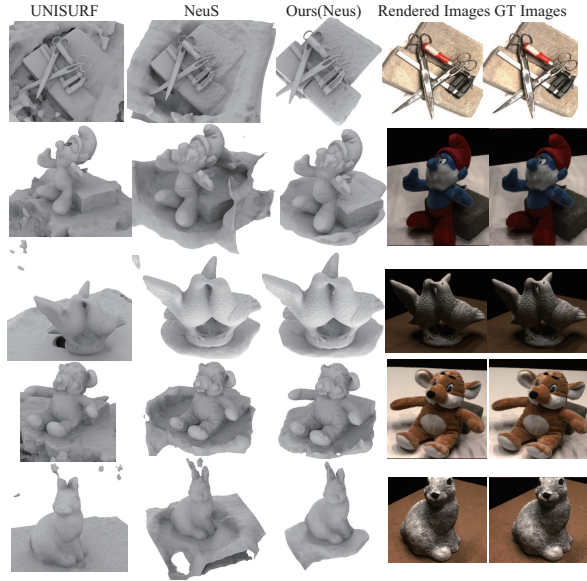Figure 5. Visual comparison with baselines in 3D scene dataset.

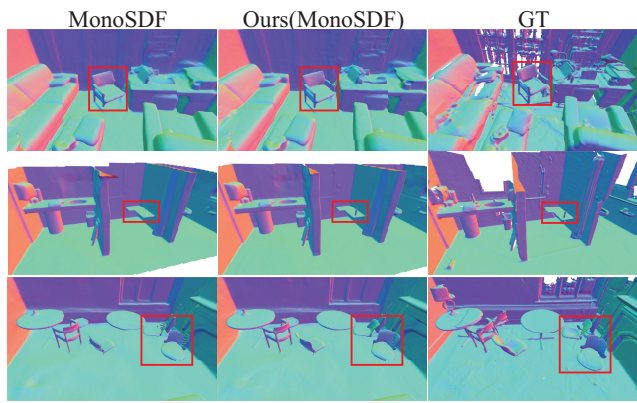

Figure 7. Visual comparison with baselines in DTU dataset.

Figure 8. Visual comparison with baselines in ScanNet.

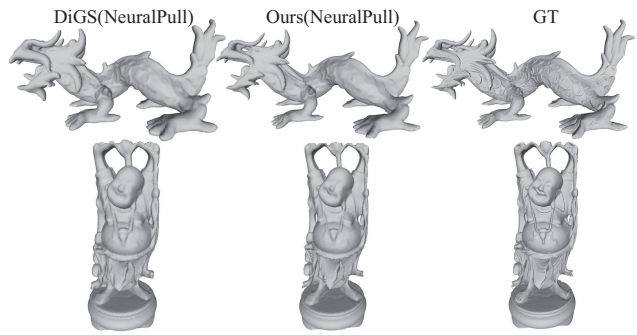Figure 9. Visual comparison with the constraint on second order derivatives in DiGS.

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.

References

[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, 3

[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. 7

[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] Wang Bing, Yu Zhengdi, Bo Yang, Qin Jie, Breckon Toby, Shao Ling, Niki Trigoni, and Andrew Markham. Rangeudf: Semantic surface reconstruction from 3d point clouds. *arXiv preprint arXiv:2204.09138*, 2022. 1

[7] 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

[8] 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

[9] 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

[10] Chao Chen, Zhizhong Han, Yu-Shen Liu, and Matthias Zwicker. Unsupervised learning of fine structure generation for 3D point clouds by 2D projection matching. In *IEEE International Conference on Computer Vision*, 2021. 2

[11] 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, 3

[12] Chao Chen, Yu-Shen Liu, and Zhizhong Han. Unsupervised inference of signed distance functions from single sparse point clouds without learning priors. In *Proceedings of the IEEE/CVF Conference on Computer Vsion and Pattern Recognition*, 2023. 2

[13] Julian Chibane, Aymen Mir, and Gerard Pons-Moll. Neural unsigned distance fields for implicit function learning. *arXiv*, 2010.13938, 2020. 2

[14] Gene Chou, Ilya Chugunov, and Felix Heide. GenSDF: Two-stage learning of generalizable signed distance functions. In *Proc. of Neural Information Processing Systems (NeurIPS)*, 2022. 1, 2

[15] 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

[16] Angela Dai and Matthias Nießner. Neural Poisson: Indicator functions for neural fields. *arXiv preprint arXiv:2211.14249*, 2022. 1

[17] 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

[18] 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

[19] Qiancheng Fu, Qingshan Xu, Yew-Soon Ong, and Wenbing Tao. Geo-Neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. 2022. 1, 3, 5, 6

[20] 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

[21] 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

[22] 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, 5

[23] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. *arXiv*, 2002.10099, 2020. 1, 2, 3

[24] 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, 3, 7

[25] Zhizhong Han, Chao Chen, Yu-Shen Liu, and Matthias Zwicker. ShapeCaptioner: Generative caption network for 3D shapes by learning a mapping from parts detected in multiple views to sentences. *ArXiv*, abs/1908.00120, 2019. 2

[26] Zhizhong Han, Chao Chen, Yu-Shen Liu, and Matthias Zwicker. DRWR: A differentiable renderer without rendering for unsupervised 3D structure learning from silhouette images. In *International Conference on Machine Learning*, 2020. 2

[27] Zhizhong Han, Baorui Ma, Yu-Shen Liu, and Matthias Zwicker. Reconstructing 3d shapes from multiple sketches using direct shape optimization. *IEEE Transactions on Image Processing*, 29:8721–8734, 2020. 2

[28] Jiahui Huang, Hao-Xiang Chen, and Shi-Min Hu. A neural galerkin solver for accurate surface reconstruction. *ACM Trans. Graph.*, 41(6), 2022. 1

[29] 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. 5

[30] Meng Jia and Matthew Kyan. Learning occupancy function from point clouds for surface reconstruction. *arXiv*, 2010.11378, 2020. 2

[31] 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

[32] 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

[33] Shujuan Li, Junsheng Zhou, Baorui Ma, Yu-Shen Liu, and Zhizhong Han. Neaf: Learning neural angle fields for point normal estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023. 2

[34] Tianyang Li, Xin Wen, Yu-Shen Liu, Hua Su, and Zhizhong Han. Learning deep implicit functions for 3D shapes with dynamic code clouds. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 12830–12840, 2022. 2

[35] Shi-Lin Liu, Hao-Xiang Guo, Hao Pan, Pengshuai Wang, Xin Tong, and Yang Liu. Deep implicit moving least-squares functions for 3D reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 2

[36] Xiaoxiao Long, Cheng Lin, Lingjie Liu, Yuan Liu, Peng Wang, Christian Theobalt, Taku Komura, and Wenping Wang. Neuraludf: Learning unsigned distance fields for multi-view reconstruction of surfaces with arbitrary topologies. *arXiv preprint arXiv:2211.14173*, 2022. 1

[37] 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

[38] 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, 5, 6

[39] 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

[40] 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

[41] Baorui Ma, Junsheng Zhou, Yu-Shen Liu, and Zhizhong Han. Towards better gradient consistency for neural signed distance functions via level set alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vsion and Pattern Recognition*, 2023. 1

[42] Julien N. P. Martel, David B. Lindell, Connor Z. Lin, Eric R. Chan, Marco Monteiro, and Gordon Wetzstein. ACORN: adaptive coordinate networks for neural scene representation. *CoRR*, abs/2105.02788, 2021. 1

[43] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3D reconstruction in function space. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 2

[44] Zhenxing Mi, Yiming Luo, and Wenbing Tao. SSRNet: Scalable 3D surface reconstruction network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 2

[45] Mateusz Michalkiewicz, Jhony K. Pontes, Dominic Jack, Mahsa Baktashmotlagh, and Anders P. Eriksson. Deep level sets: Implicit surface representations for 3D shape inference. *CoRR*, abs/1901.06802, 2019. 1, 2

[46] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision*, 2020. 2, 7

[47] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *arXiv:2201.05989*, 2022. 2

[48] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 2

[49] Michael Oechsle, Songyou Peng, and Andreas Geiger. UNISURF: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *International Conference on Computer Vision*, 2021. 2, 5, 6, 7

[50] Yutaka Ohtake, Alexander G. Belyaev, Marc Alexa, Greg Turk, and Hans-Peter Seidel. Multi-level partition of unity implicits. *ACM Transactions on Graphics*, 22(3):463–470, 2003. 6

[51] Amine Ouasfi and Adnane Boukhayma. Few 'zero level set'-shot learning of shape signed distance functions in feature space. In *European Conference on Computer Vision*, 2022. 1, 2

[52] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 1, 2

[53] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. *IEEE International Conference on Computer Vision*, 2021. 2

[54] Songyou Peng, Chiyu "Max" Jiang, Yiyi Liao, Michael Niemeyer, Marc Pollefeys, and Andreas Geiger. Shape as points: A differentiable poisson solver. In *Advances in Neural Information Processing Systems*, 2021. 1, 2, 3

[55] Albert Pumarola, Artsiom Sanakoyeu, Lior Yariv, Ali Thabet, and Yaron Lipman. Visco grids: Surface reconstruction with viscosity and coarea grids. In *Advances in Neural Information Processing Systems*, 2022. 1

[56] Darius Rückert, Linus Franke, and Marc Stamminger. Adop: Approximate differentiable one-pixel point rendering. *arXiv:2110.06635*, 2021. 2

[57] Sara Fridovich-Keil and Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 2

[58] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2

[59] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision*, 2016. 2, 7

[60] Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems*, 2020. 1, 3, 4, 5, 6

[61] Lars Mescheder Marc Pollefeys Andreas Geiger Songyou Peng, Michael Niemeyer. Convolutional occupancy networks. In *European Conference on Computer Vision*, 2020. 2, 6

[62] Towaki Takikawa, Joey Litalien, Kangxue Yin, Karsten Kreis, Charles Loop, Derek Nowrouzezahrai, Alec Jacobson, Morgan McGuire, and Sanja Fidler. Neural geometric level of detail: Real-time rendering with implicit 3D shapes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 1

[63] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS*, 2020. 1

[64] Jiapeng Tang, Jiabao Lei, Dan Xu, Feiying Ma, Kui Jia, and Lei Zhang. SA-ConvONet: Sign-agnostic optimization of convolutional occupancy networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. 2

[65] Edgar Tretschk, Ayush Tewari, Vladislav Golyanik, Michael Zollhöfer, Carsten Stoll, and Christian Theobalt. PatchNets: Patch-Based Generalizable Deep Implicit 3D Shape Representations. *European Conference on Computer Vision*, 2020. 2

[66] Delio Vicini, Sbastien Speierer, and Wenzel Jakob. Differentiable signed distance function rendering. *ACM Transactions on Graphics*, 41(4):125:1–125:18, 2022. 1, 2, 3

[67] Jiepeng Wang, Peng Wang, Xiaoxiao Long, Christian Theobalt, Taku Komura, Lingjie Liu, and Wenping Wang. NeuRIS: Neural reconstruction of indoor scenes using normal priors. In *European Conference on Computer Vision*, 2022. 1, 2, 3

[68] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Advances in Neural Information Processing Systems*, pages 27171–27183, 2021. 1, 2, 3, 5, 6, 7

[69] Yiqun Wang, Ivan Skorokhodov, and Peter Wonka. HF-NeuS: Improved surface reconstruction using high-frequency details. 2022. 1, 2, 3, 5, 6

[70] Xin Wen, Junsheng Zhou, Yu-Shen Liu, Hua Su, Zhen Dong, and Zhizhong Han. 3D shape reconstruction from 2D images with disentangled attribute flow. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 2

[71] Francis Williams, Teseo Schneider, Claudio Silva, Denis Zorin, Joan Bruna, and Daniele Panozzo. Deep geometric prior for surface reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 2

[72] Francis Williams, Matthew Trager, Joan Bruna, and Denis Zorin. Neural splines: Fitting 3D surfaces with infinitely-wide neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 9949–9958, 2021. 2

[73] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. In *Advances in Neural Information Processing Systems*, 2021. 1, 3, 5, 6, 7

[74] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems*, 33, 2020. 1, 2, 3, 5, 6

[75] Wang Yifan, Shihao Wu, Cengiz Oztireli, and Olga Sorkine-Hornung. Iso-Points: Optimizing neural implicit surfaces with hybrid representations. *CoRR*, abs/2012.06434, 2020. 2

[76] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. MonoSDF: Exploring monocular geometric cues for neural implicit surface reconstruction. *ArXiv*, abs/2022.00665, 2022. 1, 2, 3, 5, 6, 7

[77] Wenbin Zhao, Jiabao Lei, Yuxin Wen, Jianguo Zhang, and Kui Jia. Sign-agnostic implicit learning of surface self-similarities for shape modeling and reconstruction from raw point clouds. *CoRR*, abs/2012.07498, 2020. 1, 2

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

[79] Qian-Yi Zhou and Vladlen Koltun. Dense scene reconstruction with points of interest. *ACM Transactions on Graphics*, 32(4):112:1–112:8, 2013. 4

[80] Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R. Oswald, and Marc Pollefeys. Nice-slam: Neural implicit scalable encoding for slam. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 2