Sharpening Neural Implicit Functions with Frequency Consolidation Priors

Chao Chen[1], **Yu-Shen Liu**[1*], **Zhizhong Han**[2]

[1]School of Software, Tsinghua University, Beijing, China
[2]Department of Computer Science, Wayne State University, Detroit, USA
chenchao19@tsinghua.org.cn, liuyushen@tsinghua.edu.cn, h312h@wayne.edu

Abstract

Signed Distance Functions (SDFs) are vital implicit representations to represent high fidelity 3D surfaces. Current methods mainly leverage a neural network to learn an SDF from various supervisions including signed distances, 3D point clouds, or multi-view images. However, due to various reasons including the bias of neural network on low frequency content, 3D unaware sampling, sparsity in point clouds, or low resolutions of images, neural implicit representations still struggle to represent geometries with high frequency components like sharp structures, especially for the ones learned from images or point clouds. To overcome this challenge, we introduce a method to sharpen a low frequency SDF observation by recovering its high frequency components, pursuing a sharper and more complete surface. Our key idea is to learn a mapping from a low frequency observation to a full frequency coverage in a data-driven manner, leading to a prior knowledge of shape consolidation in the frequency domain, dubbed frequency consolidation priors. To better generalize a learned prior to unseen shapes, we introduce to represent frequency components as embeddings and disentangle the embedding of the low frequency component from the embedding of the full frequency component. This disentanglement allows the prior to generalize on an unseen low frequency observation by simply recovering its full frequency embedding through a test-time self-reconstruction. Our evaluations under widely used benchmarks or real scenes show that our method can recover high frequency component and produce more accurate surfaces than the latest methods.

Code — https://github.com/chenchao15/FCP

Introduction

Singed distance Functions (SDFs) can represent high fidelity 3D surfaces with arbitrary topology. An SDF is an implicit function that can predict signed distances at arbitrary 3D query locations. It describes a distance field in the 3D space hosting a surface, where we have iso-surfaces or level sets, each of which has the same signed distance values. One can extract the surface as the zero level set of the SDF using the marching cubes algorithm (Lorensen and Cline 1987).

*The corresponding author is Yu-Shen Liu.

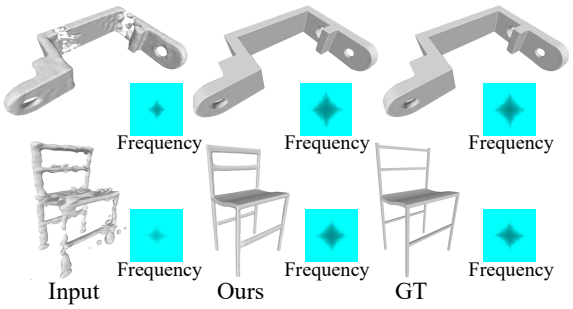

Figure 1: The concept of frequency consolidation priors. We also show averaged frequency weights across a band.

Recent methods (Mildenhall et al. 2020; Oechsle, Peng, and Geiger 2021; Takikawa et al. 2021) use neural networks to learn SDFs from 3D supervision (Jiang et al. 2020a; Park et al. 2019; Ouasfi and Boukhayma 2022; Takikawa et al. 2021; Liu et al. 2021), 3D point clouds (Genova et al. 2019), or multi-view images (Fu et al. 2022; Oechsle, Peng, and Geiger 2021; Wang et al. 2021; Jiang et al. 2020b; Wang et al. 2021; Vicini, Speierer, and Jakob 2022; Wang et al. 2022a), turning neural networks into neural implicit functions. However, due to issues like neural networks' bias towards low-frequency signals, 3D unaware sampling, sparsity in point clouds, and low-resolution images, neural SDFs still struggle with sharp or high-frequency geometries, particularly when inferred from point clouds or multi-view images. Techniques like positional encodings (Mildenhall et al. 2020) or feature grids (Sara Fridovich-Keil and Alex Yu et al. 2022; Chen, Liu, and Han 2023) attempt to recover high-frequency details, but often lead to unstable optimization (Jiang, Hua, and Han 2023) or discontinuous representations (Chen, Liu, and Han 2023), causing artifacts or noisy surfaces. Thus, recovering high-frequency components in neural implicit functions remains a challenge.

To address this challenge, we propose *frequency consolidation priors* to sharpen a neural SDF observation, as illustrated in Fig. 1. Since sharper features are usually represented by high frequency components, our key idea is to learn a mapping from a low frequency observation to its full frequency coverage in a data-driven manner. The prior knowledge learned by the mapping, dubbed frequency

consolidation priors, can produce sharper and more complete surfaces. To generalize a learned prior on unobserved low frequency SDFs better, we introduce to represent frequency components as embeddings, and disentangle the embedding of low frequency components from the one of its full frequency coverage. Our design enables the learned prior to recover full frequency embeddings by overfitting unseen low-frequency observations through a test-time self-reconstruction. We learn a frequency consolidation prior by establishing a dataset containing low and full frequency component pairs, where we produce low frequent components by removing high frequencies from full frequency coverage of a shape in the frequency domain. We demonstrate the effectiveness and the good generalization of our prior in shape and scene modeling. Benchmark comparisons show our method's superiority in accuracy and generalization over the latest methods. Our contributions are listed below.

- We present a novel method to sharpen neural SDFs for sharper and more complete surfaces in the frequency domain. Our frequency consolidation prior can recover full frequency coverage from a low frequency observation.
- We justify the idea of representing frequency components as embeddings. This design can prompt the generalization of learned priors by recovering the embedding of full frequency coverage via a test-time self-reconstruction on low frequency observations.
- We report the state-of-the-art results in shape or scene modeling by sharpening reconstructions from sparse point clouds or multi-view images.

Related Work

Neural Implicit Representations. Neural implicit representations have made huge progress in representing 3D geometry (Mescheder et al. 2019; Jiang et al. 2020b; Chen, Liu, and Han 2022; Hu and Han 2023; Jiang, Hua, and Han 2023; Chen, Liu, and Han 2024). One can learn neural implicit representations using coordinate-based MLP from supervision including 3D ground truth distances (Jiang et al. 2020a; Chabra et al. 2020; Songyou et al. 2020; Takikawa et al. 2021; Liu et al. 2021), 3D point clouds (Ma et al. 2021; Chen, Liu, and Han 2022, 2025; Noda et al. 2024; Zhou et al. 2024b,a), or multi-view images (Mildenhall et al. 2020; Fu et al. 2022; Oechsle, Peng, and Geiger 2021; Wang et al. 2021; Vicini, Speierer, and Jakob 2022; Wang et al. 2022a; Guo et al. 2022; Jiang, Hua, and Han 2023).

Differentiable renderers allow learning neural implicit representations by minimizing the error between 2D renderings and ground truth images. Surface rendering (Jiang et al. 2020b) and DVR (Niemeyer et al. 2020) estimate geometry in a radiance field, with IDR modeling view direction to reconstruct high-frequency details. These methods require masks to filter out the background as they focus on surface intersections.

NeRF (Mildenhall et al. 2020) and its variations (Sara Fridovich-Keil and Alex Yu et al. 2022; Zhang et al. 2024) simultaneously model geometry and color using volume rendering. They aim to generate novel views, and render images without masks. By deriving novel rendering equations, UNISURF (Oechsle, Peng, and Geiger 2021) and NeuS (Wang et al. 2021) are able to render occupancy and signed distance fields into RGB images, which measures the errors of implicit functions. Following methods improve accuracy of implicit functions using priors or losses including depth (Wang, Wang, and Agapito 2023; Hu and Han 2023), normals (Wang et al. 2022a; Guo et al. 2022), multi-view consistency (Fu et al. 2022), and segmentation priors (Haghighi et al. 2023).

Learning with Frequency. Learning neural implicit representations with multi-scale details improves interpretability, enabling progressive visualization at various scales (Takikawa et al. 2021). Curvature regulation (Ehret, Marí, and Facciolo 2022) can adjust surface details. A common approach is to use multiple frequency bands that cover the entire frequency range (Lindell et al. 2022; Grattarola and Vandergheynst 2022), creating multi-scale representations by reconstructing surfaces from different bands. Additionally, SAP learns an occupancy function by solving a Poisson equation in the frequency domain (Peng et al. 2021).

Recovering Sharp Structures. Several strategies recover sharp edges. Dual contouring (Chen et al. 2022) reconstructs sharper edges than marching cubes (Lorensen and Cline 1987) using gradients. Modeling displacements helps recover high-frequency details on surfaces (Yifan, Rahmann, and Sorkine-hornung 2021). Edges are crucial in CAD modeling, and NEF (Ye et al. 2023) learns an implicit function to represent edges from multi-view images. Some methods (Lambourne et al. 2022) sharpen edges directly on 3D shapes, though they work best on clean shapes. Consolidation (Metzer et al. 2021) sharpens point clouds, generating points with sharp features, removing noise and outliers.

Unlike previous methods, we aim to sharpen implicit field poorly recovered from point clouds or multi-views. We use a data-driven strategy to sharp a shape by learning priors in the frequency domain, which consolidates the frequency components and completes missing structures as well.

Method

Overview. Our method aims to sharpen a low frequency observation represented by an SDF f_L (or a point cloud), as illustrated in Fig. 2. With f_L, we intend to recover its full frequency coverage as another SDF f_F which represents a surface with sharper and more complete structures than the low frequency observation. Both f_L and f_F are learned by neural networks with parameters $\boldsymbol{\theta}_L$ and $\boldsymbol{\theta}_F$, respectively. At an arbitrary query q, f_L and f_F predict signed distances as $s_L = f_L(q, \boldsymbol{e}_L)$ and $s_F = f_F(q, \boldsymbol{e}_F)$, respectively, where $\boldsymbol{e}_L$ and $\boldsymbol{e}_F$ are learnable embeddings representing low frequency components and full frequency coverage (or shape identities), respectively, which are also conditions in f_L and f_F. We use $\boldsymbol{e}_L$ and $\boldsymbol{e}_F$ as bridges to connect f_L and f_F, where $\boldsymbol{e}_L$ is formed by $\boldsymbol{e}_C$ and $\boldsymbol{e}_F$.

We learn a frequency consolidation prior by learning f_L and f_F in a data-driven manner using supervision established from ground truth meshes. During testing, given an unseen SDF with low frequency components, we generalize the learned prior by conducting a test-time self-

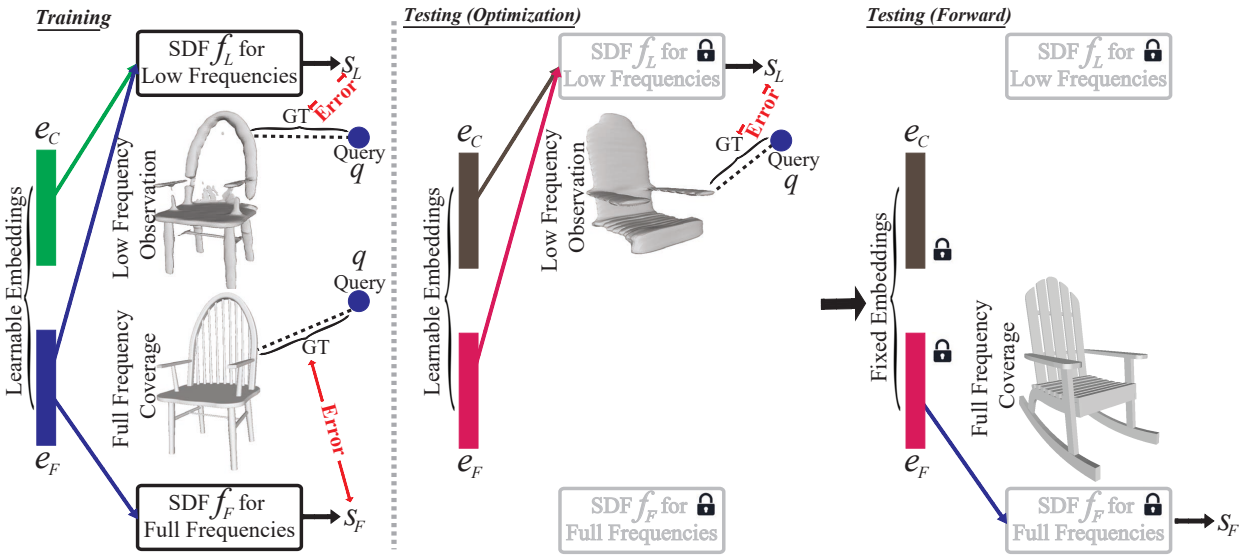

Figure 2: The overview of our method.

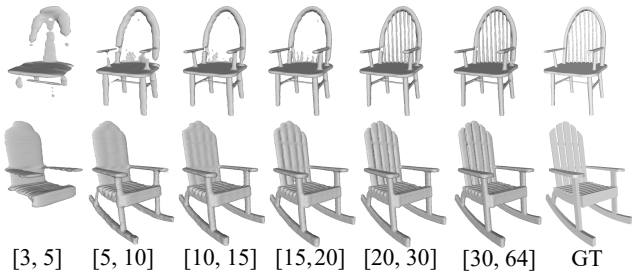

Figure 3: The illustration of low frequency observations and the full frequency coverage.

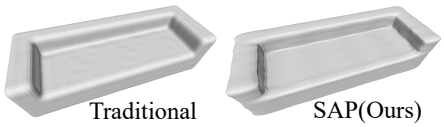

Figure 4: Over-smoothed surfaces.

reconstruction, which learns the embeddings $\boldsymbol{e}_L$ and $\boldsymbol{e}_F$ of the shape using f_L with the fixed parameters $\boldsymbol{\theta}_L$. Then, we further sharpen the shape by decoding the learned $\boldsymbol{e}_F$ using f_F with the fixed parameters $\boldsymbol{\theta}_F$.

Supervisions for Learning Priors. We establish supervisions from ground truth meshes. For a shape M, we produce its low frequency observations M_L by randomly removing its high frequency components from its full frequency coverage M_F. To decompose a 3D mesh into frequency domain, the traditional method like the spectral geometry theory (Zhang, van Kaick, and Dyer 2007) does eigen-decomposition of the discrete Laplace–Beltrami operator and regards the eigenvectors as frequency components. Some learning-based methods (Lindell et al. 2022; Takikawa et al. 2021) are also alternatives. However, eigen-decomposing a large matrix whose dimension is determined by the vertex number is usually limited due to the large space complexity, while learning based methods are too slow to get enough samples as supervisions. Instead, we introduce to manipulate frequency components in solving Poisson surface reconstruction from point clouds for efficiency. As a fast solving PDE strategy, spectral methods solve a Poisson surface reconstruction problem using Fast Fourier Transform (FFT) (Peng et al. 2021). The low frequency observations (SDF) established by our method with proper low frequency band can produce over smoothed surfaces, similar to the one produced by spectral geometry theory in Fig. 4.

We first randomly sample dense points on the mesh of M and estimate an occupancy field by solving a Poisson surface reconstruction equation, obtaining magnitudes for frequencies in a frequency band. We then reconstruct M using the marching cubes algorithm (Lorensen and Cline 1987) with the estimated occupancy function, resulting in the full frequency coverage M_F. Simultaneously, we produce low-frequency occupancy functions by removing some high-frequency components (setting their corresponding magnitudes to 0). Each manipulated occupancy function is used to reconstruct a mesh as a low-frequency observation M_L. To ensure low-frequency observations cover all frequencies in the band $[0, 64]$ (Fig.3), we randomly select one frequency from one of six subbands (e.g., [3,5]" or [5,10]") and remove all frequencies larger than the selected one. For example, if we select 4 in "[3,5]", we remove all frequencies from 5 to 64, producing a low-frequency observation M_L. Each low-frequency observation M_L is paired with the full frequency coverage M_F as a training sample. Fig.3 illustrates six pairs of low-frequency observations M_L and their corresponding full frequency coverage M_F (rightmost) on two shapes. Note that we do not use GT meshes directly as M_F to avoid non-watertight meshes during training.

Moreover, we produce more low frequency observations in the band $[3, 30]$, as most reconstructions in real applications contain very low-frequency components. These observations not only have smooth surfaces but also severe structural corruptions. Using them as training samples enables our prior to handle extremely poor reconstructions well.

Frequency Component Modeling. With low frequency observation M_L and its corresponding full frequency coverage M_F, we learn a frequency consolidation prior as a mapping from M_L to M_F in Fig. 2. We represent both M_L to M_F as SDFs f_L and f_F which are approximated by a two-branch network parameterized by $\boldsymbol{\theta_L}$ and $\boldsymbol{\theta_F}$. At each query q, one branch predicts a signed distance $s_L = f_L(q, \boldsymbol{e}_L)$ around M_L, the other predicts a signed distance $s_F = f_F(q, \boldsymbol{e}_F)$ around M_F. We use embeddings $\boldsymbol{e}_L$ and $\boldsymbol{e}_F$ to model M_L and M_F, which are also used as conditions to distinguish the low frequency band and shape identity when sharing the same neural network implementation.

For an embedding $\boldsymbol{e}_L$, we formulate it as a learnable 256-dimensional vector, and assign it to a low frequency observation M_L. Similarly, we formulate an embedding $\boldsymbol{e}_F$ as a learnable 128-dimensional vector, assign it to the full frequency coverage M_F, and more importantly, make it shareable to all low frequency observations $\{M_L\}$ of shape M.

We bridge the two neural SDFs f_L and f_F by disentangling $\boldsymbol{e}_F$ from $\boldsymbol{e}_L$. We formulate $\boldsymbol{e}_L$ as a concatenation of $\boldsymbol{e}_F$ representing a shape identity M and $\boldsymbol{e}_C$ representing a frequency corruption on the specific M_F below,

$$\boldsymbol{e}_L = [\boldsymbol{e}_F \;\; \boldsymbol{e}_C]. \tag{1}$$

This disentangling makes the frequency modeling interpretable, compacts the embedding space, synchronizes the learning of f_L and f_F, and more importantly, increases the generalization ability of the learned frequency consolidation prior which will show in experiments.

Learning Frequency Consolidation Priors. To learn the prior, we train the two-branch network to regress signed distances at query q. With a low frequency observation M_L and its target M_F, we sample queries q around M_F and record the ground truth signed distances s_L^{gt} and s_F^{gt}. We optimize parameters by minimizing the prediction error denoted by

$$\min_{\boldsymbol{\theta}_L, \boldsymbol{\theta}_F, \{\boldsymbol{e}_F\}, \{\boldsymbol{e}_C\}} ||s_L - s_L^{gt}||_2^2 + ||s_F - s_F^{gt}||_2^2, \tag{2}$$

where $s_L = f_L(q, [\boldsymbol{e}_F \;\; \boldsymbol{e}_C])$ and $s_F = f_F(q, \boldsymbol{e}_F)$ are signed distance predictions.

Generalizing Frequency Consolidation Priors. We generalize the learned prior to sharpen an unseen low frequency observation M_L'. M_L' can be represented as an SDF, a point cloud, or a mesh. To leverage the learned prior, we transform a point cloud or a mesh into an SDF using surface reconstruction methods like NeuralPull (Ma et al. 2021).

With our disentangling of $\boldsymbol{e}_F$ from $\boldsymbol{e}_L$, we can estimate the shape identity $\boldsymbol{e}_F$ through a test-time optimization in self-reconstruction on M_L' as auto-decoding (Park et al. 2019). To this end, we sample queries q around M_L' and record signed distances $s_L^{gt'}$ as supervision. We estimate $\boldsymbol{e}_L' = [\boldsymbol{e}_F' \;\; \boldsymbol{e}_C']$ with fixed parameters $\boldsymbol{\theta}_L$ by minimizing the reconstruction errors below,

$$\min_{\boldsymbol{e}_F', \boldsymbol{e}_C'} ||s_L - s_L^{gt'}||_2^2. \tag{3}$$

After optimization, we represent the SDF of the full frequency coverage M_F' as $f_F(q, \boldsymbol{e}_F')$, and reconstruct its surface using marching cubes (Lorensen and Cline 1987).

Implementation Details. We adopt two Gaussian functions centered at each point with standard deviations σ_1 and σ_2 to sample queries. Starting from meshes, we sample dense point clouds as surface points, and sample queries around each surface point. We set σ_1 to 8 for full-space sampling, allowing the network to perceive a large space and cover various shape variations. σ_2 is set to 0.2, enabling queries to be sampled close to the surface. These two types of queries are sampled with a one-to-one weighting ratio, dynamically sampling 16,384 queries in each iteration.

We learn $\boldsymbol{e}_L$ and $\boldsymbol{e}_F$ by 3 fully connected layers with 128 hidden units and a ReLU on each layer. We employ two SDF-decoder networks similar to DeepSDF (Park et al. 2019) to learn f_L and f_F. The Adam optimizer is used with an initial embedding learning rate of 0.0005 and an SDF-decoder learning rate of 0.001, both decreased by 0.5 every 500 epochs. We train our model in 2000 epochs. During test-time optimization, we overfit f_L on a low frequency observation in 800 iterations with a learning rate of 0.005.

Method	$CD_{L1} \times 10$	$CD_{L2} \times 100$	NC
DeepSDF	0.287	0.381	0.804
ConvOcc	0.306	0.451	0.805
LIG	0.292	0.430	0.809
IDF	0.287	0.390	0.815
NDC	0.269	0.358	0.768
POCO(pretrained)	0.259	0.374	0.812
POCO	0.217	0.284	0.858
ALTO(pretrained)	0.253	0.367	0.819
ALTO	0.213	0.285	0.861
Ours	**0.187**	**0.216**	**0.871**

Table 1: Reconstruction accuracy of 13 classes on ShapeNet in terms of CD_{L1}, CD_{L2} and NC. The accuracy for each class is provided in the supplement.

Experiments and Anaylysis

Datasets and Metrics. We evaluate our method on ShapeNet (Chang et al. 2015), ABC (Koch et al. 2019), and ScanNet (Dai et al. 2017) datasets. For ShapeNet, we report results on 13 classes from 3D-R2N2 (Choy et al. 2016) and 8 classes from NeuralTPS (Chen, Han, and Liu 2023). For ABC and ScanNet, we follow the train/test splits from Points2Surf (Erler et al. 2020) and Neural Part Priors (Bokhovkin and Dai 2022). All experiments use marching cubes (Lorensen and Cline 1987) on a 256^3 grid to reconstruct meshes.

For ShapeNet, we measure errors using L1 Chamfer Distance CD_{L1}, CD_{L2}, and normal consistency (NC), randomly sampling 100k points from both reconstructed and ground truth meshes. For ScanNet, we follow Neural Part Priors to report CD_{L1} between the reconstructed and ground truth meshes in ScanNet space, evaluating shapes based on

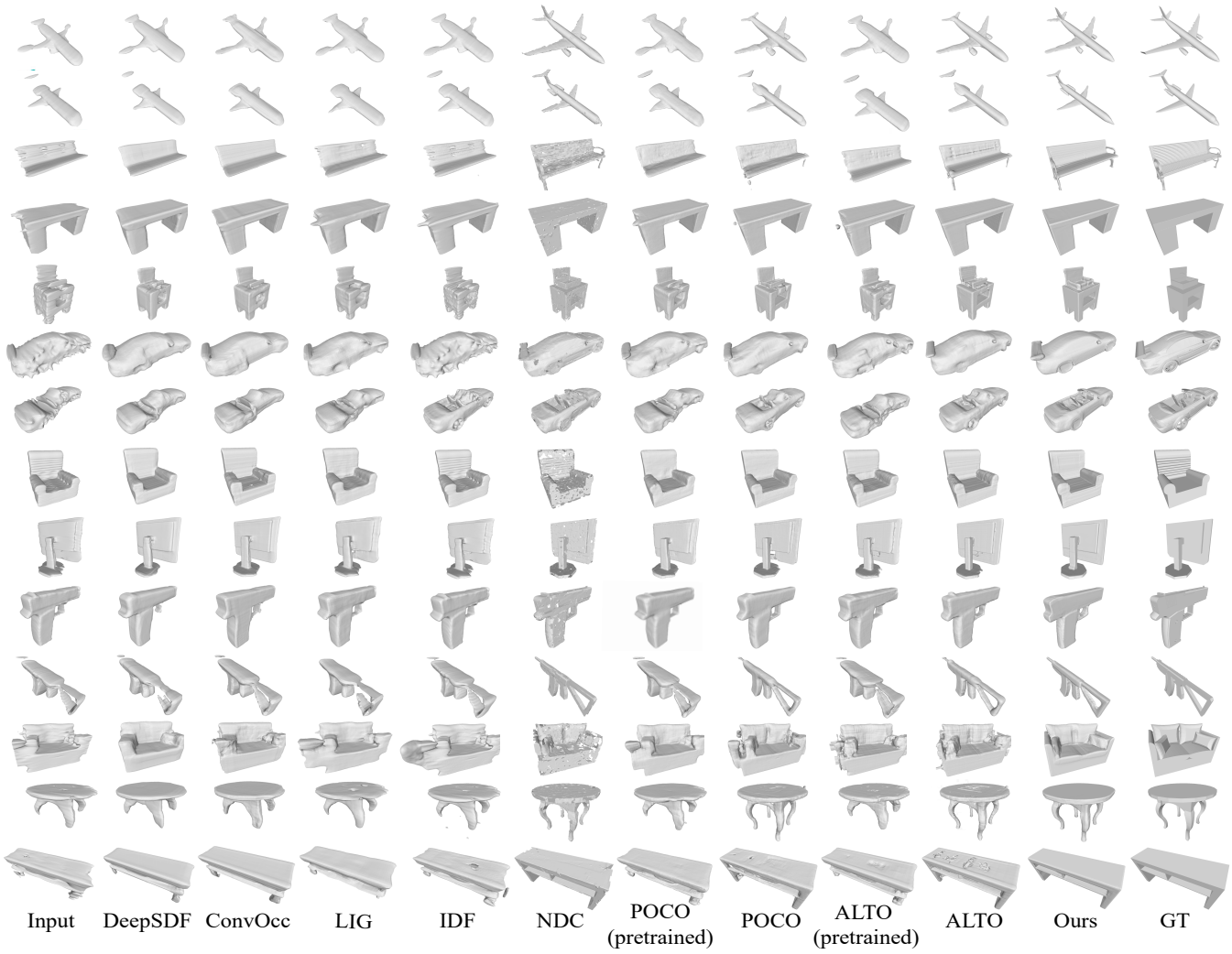

Figure 5: Visual comparison with the state-of-the-art on ShapeNet.

the union of CD_{L1} values of predicted and ground truth parts. We report two sets of results: one using annotated shapes from Scan2CAD (Avetisyan et al. 2019) and the other using shapes extracted from ScanNet segmentation masks.

Learning Frequency Consolidation Priors. For training, we use the first 5 low-frequency observations (Fig. 3) and full frequency coverage to learn the frequency consolidation prior, sampling queries around both and recording signed distances to both meshes as supervision.

Evaluations

Evaluation on ShapeNets. We evaluate using frequency consolidation priors learned from training samples. For each test shape, we generate low-frequency observations as described and use the worst observation to assess all methods. Tab. 1 reports average evaluations across all classes, showing our method achieves the best performance. Detailed per-class results are in the supplementary materials.

We evaluate DeepSdf (Park et al. 2019), POCO (Boulch and Marlet 2022), and ALTO (Wang et al. 2022b) using pre-trained or retrained parameters on our data to map point clouds into an SDF. For training, we sample low-frequency observations and record signed distances near their full-frequency coverages for POCO and ALTO. Our method significantly outperforms these approaches. Visual comparisons in Fig.5 show that DeepSDF, POCO, and ALTO struggle with low-frequency observations. NDC(Chen et al. 2022) generates sharper edges than marching cubes (Lorensen and Cline 1987), but fails to generalize on low-frequency shapes, producing noisy surfaces. IDF (Yifan, Rahmann, and Sorkine-hornung 2021) and ConvOcc (Songyou et al. 2020), though effective at generating high-frequency geometry from corrupted shape, fail to handle large geometry variations, as seen in Tab.1 and Fig.5.

Refining Reconstructions from Sparse Point clouds. We further evaluate the generalization ability of our learned prior. In the previous experiment, we produce test shapes from a known frequency band used in training. How good the performance of our prior is on unobserved frequency bands will be evaluated in this experiment. We use recon-

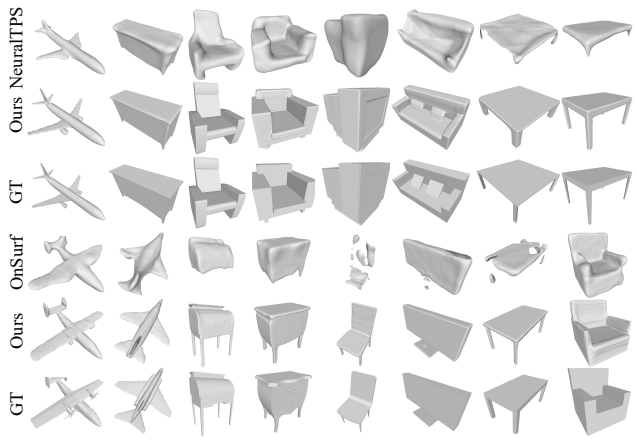

Figure 6: Reconstruction results for test-time optimization with sparse reconstructions from NeuralTPS and OnSurf as low-frequency observations, respectively.

structions from sparse point clouds by the latest methods as test shapes, which are barely with any geometry details as shown in Fig. 6 and have unobserved frequency components.

We use NeuralTPS (Chen, Han, and Liu 2023) and OnSurf Prior (Ma et al. 2022) for sparse point cloud reconstruction, generating test shapes from 300 points across 8 ShapeNet classes. Our learned prior recovers high-frequency components, improving reconstruction with sharper edges, flatter planes, and more complete surfaces, as shown in Tab.2, Fig.6, and Fig. 9 (first two rows). Detailed per-class evaluations are in the supplementary materials.

Evaluations on CAD Modeling. We learn a prior from shapes in the ABC dataset (Koch et al. 2019) using Points2Surf (Erler et al. 2020) to generate training samples with 5 low-frequency observations and full-frequency coverage from each shape. For each test shape, we generate 5 low-frequency observations and compute the mean and variance of these evaluations. Tab.3 shows our method outperforms Points2Surf(Erler et al. 2020) and SECAD-Net (Li et al. 2023) in reconstruction accuracy and stability. Visual comparisons in Fig. 7 demonstrate that our method produces sharper edges and more accurate structures, while Points2Surf struggles with surface variations, and SECAD-Net generates sharp edges but lacks generalization ability.

Reconstruction in Scenes. Our learned prior also applies to scene modeling. We train on ShapeNet classes present in the scenes used for evaluation, which are reconstructed from real scans. Using GT segmentation masks, we segment shapes into partial meshes that contain incomplete geometry, unobserved frequency bands, or severe corruption. We use pose and scale information from Scan2CAD (Avetisyan et al. 2019) to determine the layout in our visualizations.

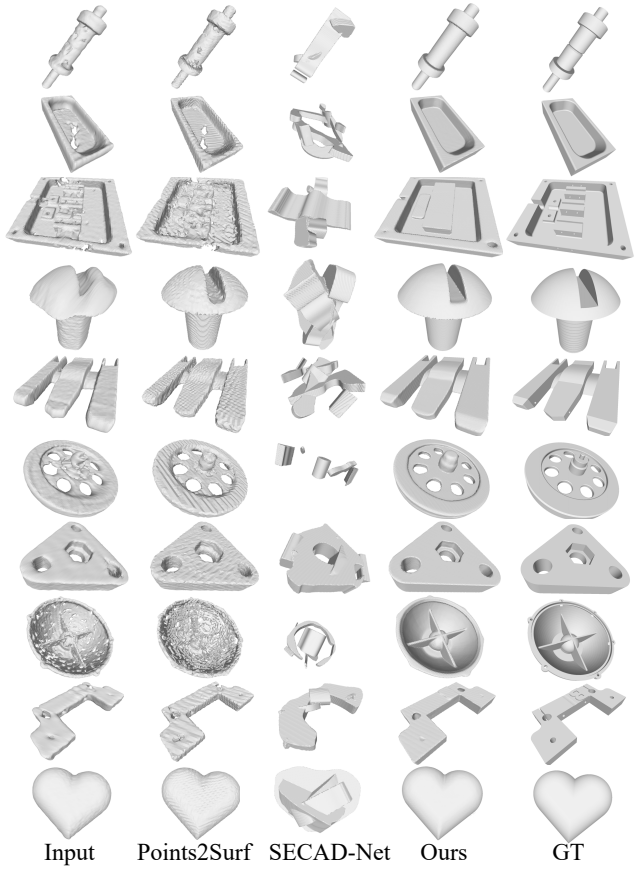

Figure 7: Visual comparisons on ABC.

In our evaluations (Tab. 1 and Tab. 2 in the supplement), we use two types of GT shapes. The first type comes from Scan2CAD, retrieved from ShapeNet. These shapes are complete but may differ significantly from the real scans. The second type is obtained directly from real scans using GT segmentation masks. These shapes are mostly incomplete but more closely resemble the real scenes.

We segment a shape from a scene and use NeuralPull to reconstruct a coarse, watertight mesh as a low-frequency observation. We then generalize the learned prior to recover its full frequency coverage (Rec). Using the low-frequency observation, we also produce two results by retrieving all low-frequency observations from our training set. Specifically, we render 30 images from viewpoints around each

Method	$CD_{L1} \times 10$	$CD_{L2} \times 100$	NC
Onsurf	0.214	0.223	0.845
Onsurf+Ours	**0.180**	**0.165**	**0.886**
NeuralTPS	0.141	0.093	50.899
NeuralTPS+Ours	**0.115**	**0.088**	**0.918**

Table 2: Numerical comparisons with sparse point cloud reconstruction on ShapeNet.

Method	CD_{L1}		NC	
	Mean	Variance	Mean	Variance
Points2Surf (Erler et al. 2020)	0.014	0.431	0.902	5.166
SECAD-Net (Li et al. 2023)	0.041	0.178	0.800	1.370
Ours	**0.011**	**0.015**	**0.962**	**1.076**

Table 3: Accuracy of reconstruction on ABC dataset in terms of CD_{L1} and NC. We multiply both variances by 10^4.

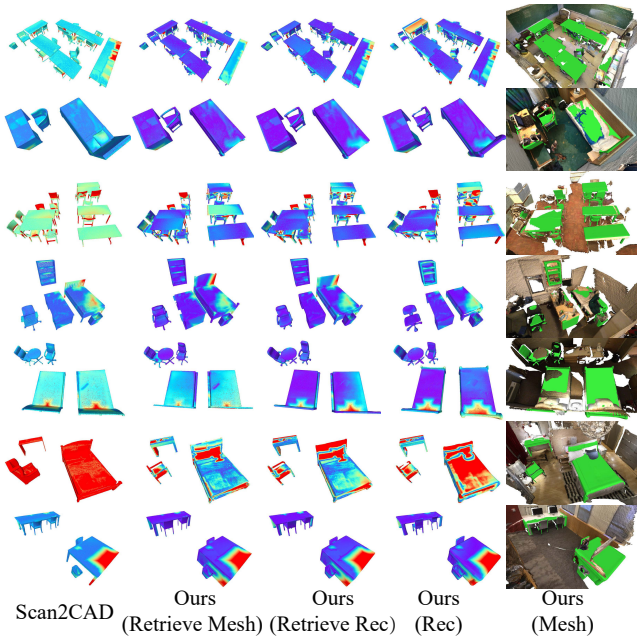

Figure 8: Visual comparisons on ScanNet dataset. The red in error maps indicates larger errors.

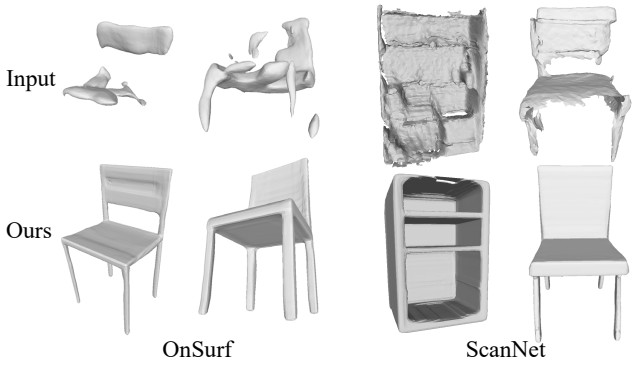

Figure 9: Reconstrcutions from severely corrupted cases.

shape, extract features using the CLIP image encoder (Radford et al. 2021), and evaluate the distance between two sets of images using a single-direction CD distance .

For each retrieved low frequency observation, we use the reconstructed high frequency coverage or their GT meshes from ShapeNet to report the results including (Retrieve Rec) and (Retrieval Mesh). We report the evaluations of these three results in both Tab. 1 and Tab. 2 in our supplement.

We outperform NeuralPartPriors (Bokhovkin and Dai 2022) and PartUnderstanding (Bokhovkin et al. 2021) using Scan2CAD as ground truth. Tab. 2 in our supplement also shows our superiority over Scan2CAD with segmented meshes. Fig.8 presents error maps of our results, while Fig.9 (last 2 columns) highlights plausible reconstructions of bad ScanNet scenes, demonstrating that shapes retrieved by Scan2CAD are not identical to real scans.

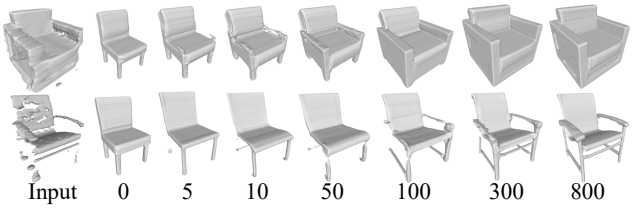

Figure 10: Visualization of the test-time optimization.

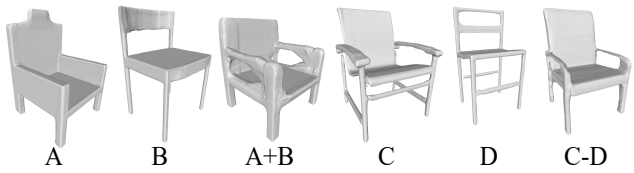

Figure 11: Embedding manipulations for shape generation.

Ablation Studies and Analysis

Semantic Latent Space. The latent space we learn is semantic, with the optimization process in self-reconstruction visualized in Fig.10, showing semantic shapes evolving along the optimization path. Embeddings can be semantically manipulated, as seen in Fig.11, and TSNE (van der Maaten and Hinton 2008) reveals semantic structures, with paths diverging towards different shapes (Fig.12).

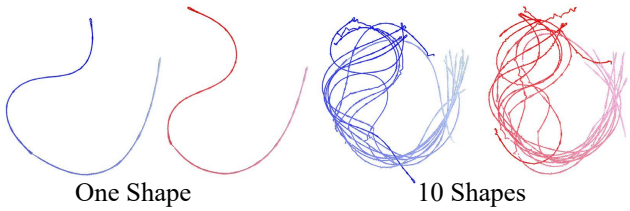

Figure 12: Embeddings for low and full frequency shapes.

Conclusion

We introduce frequency consolidation priors to sharpen neural implicit functions by learning from training pairs of low frequency components and full frequency coverage. These priors improve generalization, enabling the recovery of high-frequency components from low-frequency observations, which sharpens surfaces and completes missing structures. Our comparisons with state-of-the-art methods on widely used shape and scene datasets show that our approach recovers higher-frequency geometries, enhancing low-frequency SDF observations more effectively than existing methods.

Acknowledgments

This work was supported by National Key R&D Program of China (2022YFC3800600), the National Natural Science Foundation of China (62272263), and in part by Tsinghua-Kuaishou Institute of Future Media Data.

References

Avetisyan, A.; Dahnert, M.; Dai, A.; Savva, M.; Chang, A. X.; and Niessner, M. 2019. Scan2CAD: Learning CAD Model Alignment in RGB-D Scans. In *The IEEE Conference on Computer Vision and Pattern Recognition*.

Bokhovkin, A.; and Dai, A. 2022. Neural Part Priors: Learning to Optimize Part-Based Object Completion in RGB-D Scans.

Bokhovkin, A.; Ishimtsev, V.; Bogomolov, E.; Zorin, D.; Artemov, A.; Burnaev, E.; and Dai, A. 2021. Towards Part-Based Understanding of RGB-D Scans. In *In the IEEE Conference on Computer Vision and Pattern Recognition*.

Boulch, A.; and Marlet, R. 2022. POCO: Point Convolution for Surface Reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Chabra, R.; Lenssen, J. E.; Ilg, E.; Schmidt, T.; Straub, J.; Lovegrove, S.; and Newcombe, R. A. 2020. Deep Local Shapes: Learning Local SDF Priors for Detailed 3D Reconstruction. In *European Conference on Computer Vision*, volume 12374, 608–625.

Chang, A. X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; Xiao, J.; Yi, L.; and Yu, F. 2015. ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago.

Chen, C.; Han, Z.; and Liu, Y.-S. 2023. Unsupervised Inference of Signed Distance Functions from Single Sparse Point Clouds without Learning Priors. In *In the IEEE Conference on Computer Vision and Pattern Recognition*.

Chen, C.; Liu, Y.-S.; and Han, Z. 2022. Latent Partition Implicit with Surface Codes for 3D Representation. In *European Conference on Computer Vision*.

Chen, C.; Liu, Y.-S.; and Han, Z. 2023. GridPull: Towards Scalability in Learning Implicit Representations from 3D Point Clouds. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Chen, C.; Liu, Y.-S.; and Han, Z. 2024. Inferring Neural Signed Distance Functions by Overfitting on Single Noisy Point Clouds through Finetuning Data-Driven based Priors. In *Advances in Neural Information Processing Systems*.

Chen, C.; Liu, Y.-S.; and Han, Z. 2025. NeuralTPS: Learning Signed Distance Functions without Priors from Single Sparse Point Clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(1): 565–582.

Chen, Z.; Tagliasacchi, A.; Funkhouser, T.; and Zhang, H. 2022. Neural Dual Contouring. *ACM Transactions on Graphics (Special Issue of SIGGRAPH)*, 41(4).

Choy, C. B.; Xu, D.; Gwak, J.; Chen, K.; and Savarese, S. 2016. 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction. In *European Conference on Computer Vision*, 628–644.

Dai, A.; Chang, A. X.; Savva, M.; Halber, M.; Funkhouser, T.; and Nießner, M. 2017. ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes. In *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*.

Ehret, T.; Marí, R.; and Facciolo, G. 2022. Regularization of NeRFs using differential geometry. arXiv:2206.14938.

Erler, P.; Guerrero, P.; Ohrhallinger, S.; Mitra, N. J.; and Wimmer, M. 2020. Points2Surf: Learning Implicit Surfaces from Point Clouds. In *European Conference on Computer Vision*.

Fu, Q.; Xu, Q.; Ong, Y.-S.; and Tao, W. 2022. Geo-Neus: Geometry-Consistent Neural Implicit Surfaces Learning for Multi-view Reconstruction. In *Advances in Neural Information Processing Systems*.

Genova, K.; Cole, F.; Vlasic, D.; Sarna, A.; Freeman, W. T.; and Funkhouser, T. 2019. Learning Shape Templates with Structured Implicit Functions. In *International Conference on Computer Vision*.

Grattarola, D.; and Vandergheynst, P. 2022. Generalised implicit neural representations. *Advances in Neural Information Processing Systems*.

Guo, H.; Peng, S.; Lin, H.; Wang, Q.; Zhang, G.; Bao, H.; and Zhou, X. 2022. Neural 3D Scene Reconstruction with the Manhattan-world Assumption. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Haghighi, Y.; Kumar, S.; Thiran, J.-P.; and Gool, L. V. 2023. Neural Implicit Dense Semantic SLAM. arXiv:2304.14560.

Hu, P.; and Han, Z. 2023. Learning Neural Implicit through Volume Rendering with Attentive Depth Fusion Priors. In *Advances in Neural Information Processing Systems*.

Jiang, C.; Sud, A.; Makadia, A.; Huang, J.; Nießner, M.; and Funkhouser, T. 2020a. Local Implicit Grid Representations for 3D Scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Jiang, S.; Hua, J.; and Han, Z. 2023. Coordinate Quantized Neural Implicit Representations for Multi-view 3D Reconstruction. In *IEEE International Conference on Computer Vision*.

Jiang, Y.; Ji, D.; Han, Z.; and Zwicker, M. 2020b. SDFDiff: Differentiable Rendering of Signed Distance Fields for 3D Shape Optimization. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Koch, S.; Matveev, A.; Jiang, Z.; Williams, F.; Artemov, A.; Burnaev, E.; Alexa, M.; Zorin, D.; and Panozzo, D. 2019. ABC: A Big CAD Model Dataset For Geometric Deep Learning. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Lambourne, J. G.; Willis, K.; Jayaraman, P. K.; Zhang, L.; Sanghi, A.; and Malekshan, K. R. 2022. Reconstructing editable prismatic CAD from rounded voxel models. In *SIGGRAPH Asia 2022 Conference Papers*. ACM.

Li, P.; Guo, J.; Zhang, X.; and ming Yan, D. 2023. SECAD-Net: Self-Supervised CAD Reconstruction by Learning Sketch-Extrude Operations. In *In IEEE Conference on Computer Vision and Pattern Recognition*.

Lindell, D. B.; Van Veen, D.; Park, J. J.; and Wetzstein, G. 2022. BACON: Band-limited coordinate networks for multiscale scene representation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Liu, S.-L.; Guo, H.-X.; Pan, H.; Wang, P.; Tong, X.; and Liu, Y. 2021. Deep Implicit Moving Least-Squares Functions for 3D Reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Lorensen, W. E.; and Cline, H. E. 1987. Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics*, 21(4): 163–169.

Ma, B.; Han, Z.; Liu, Y.-S.; and Zwicker, M. 2021. Neural-Pull: Learning Signed Distance Functions from Point Clouds by Learning to Pull Space onto Surfaces. In *International Conference on Machine Learning*.

Ma, B.; Liu, Y.-S.; Zwicker, M.; and Han, Z. 2022. Reconstructing Surfaces for Sparse Point Clouds with On-Surface Priors. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Mescheder, L.; Oechsle, M.; Niemeyer, M.; Nowozin, S.; and Geiger, A. 2019. Occupancy Networks: Learning 3D Reconstruction in Function Space. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Metzer, G.; Hanocka, R.; Giryes, R.; and Cohen-Or, D. 2021. Self-Sampling for Neural Point Cloud Consolidation. 40(5).

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision*.

Niemeyer, M.; Mescheder, L.; Oechsle, M.; and Geiger, A. 2020. Differentiable Volumetric Rendering: Learning Implicit 3D Representations without 3D Supervision. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Noda, T.; Chen, C.; Zhang, W.; Liu, X.; Liu, Y.-S.; and Han, Z. 2024. MultiPull: Detailing Signed Distance Functions by Pulling Multi-Level Queries at Multi-Step. In *Advances in Neural Information Processing Systems*.

Oechsle, M.; Peng, S.; and Geiger, A. 2021. UNISURF: Unifying Neural Implicit Surfaces and Radiance Fields for Multi-View Reconstruction. In *International Conference on Computer Vision*.

Ouasfi, A.; and Boukhayma, A. 2022. Few 'Zero Level Set'-Shot Learning of Shape Signed Distance Functions in Feature Space. In *European Conference on Computer Vision*.

Park, J. J.; Florence, P.; Straub, J.; Newcombe, R.; and Lovegrove, S. 2019. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Peng, S.; Jiang, C. M.; Liao, Y.; Niemeyer, M.; Pollefeys, M.; and Geiger, A. 2021. Shape As Points: A Differentiable Poisson Solver. In *Advances in Neural Information Processing Systems*.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning Transferable Visual Models From Natural Language Supervision. *CoRR*, abs/2103.00020.

Sara Fridovich-Keil and Alex Yu; Tancik, M.; Chen, Q.; Recht, B.; and Kanazawa, A. 2022. Plenoxels: Radiance Fields without Neural Networks. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Songyou, P.; Michael, N.; Lars, P., Mescheder adn Marc; and Andreas, G. 2020. Convolutional Occupancy Networks. In *European Conference on Computer Vision*.

Takikawa, T.; Litalien, J.; Yin, K.; Kreis, K.; Loop, C.; Nowrouzezahrai, D.; Jacobson, A.; McGuire, M.; and Fidler, S. 2021. Neural Geometric Level of Detail: Real-time Rendering with Implicit 3D Shapes. In *IEEE Conference on Computer Vision and Pattern Recognition*.

van der Maaten, L.; and Hinton, G. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9: 2579–2605.

Vicini, D.; Speierer, S.; and Jakob, W. 2022. Differentiable Signed Distance Function Rendering. *ACM Transactions on Graphics*, 41(4): 125:1–125:18.

Wang, H.; Wang, J.; and Agapito, L. 2023. Co-SLAM: Joint Coordinate and Sparse Parametric Encodings for Neural Real-Time SLAM. *In the IEEE Conference on Computer Vision and Pattern Recognition*.

Wang, J.; Wang, P.; Long, X.; Theobalt, C.; Komura, T.; Liu, L.; and Wang, W. 2022a. NeuRIS: Neural Reconstruction of Indoor Scenes Using Normal Priors. In *European Conference on Computer Vision*.

Wang, P.; Liu, L.; Liu, Y.; Theobalt, C.; Komura, T.; and Wang, W. 2021. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. In *Advances in Neural Information Processing Systems*.

Wang, Z.; Zhou, S.; Park, J. J.; Paschalidou, D.; You, S.; Wetzstein, G.; Guibas, L.; and Kadambi, A. 2022b. ALTO: Alternating Latent Topologies for Implicit 3D Reconstruction. *arXiv preprint arXiv:2212.04096*.

Ye, Y.; Yi, R.; Gao, Z.; Zhu, C.; Cai, Z.; and Xu, K. 2023. NEF: Neural Edge Fields for 3D Parametric Curve Reconstruction From Multi-View Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8486–8495.

Yifan, W.; Rahmann, L.; and Sorkine-hornung, O. 2021. Geometry-Consistent Neural Shape Representation with Implicit Displacement Fields. In *International Conference on Learning Representations*.

Zhang, H.; van Kaick, O.; and Dyer, R. 2007. Spectral Methods for Mesh Processing and Analysis. In *Proc. of Eurographics State-of-the-art Report*, 1–22.

Zhang, W.; Shi, K.; Liu, Y.-S.; and Han, Z. 2024. Learning Unsigned Distance Functions from Multi-view Images with Volume Rendering Priors. In *European Conference on Computer Vision*.

Zhou, J.; Ma, B.; Li, S.; Liu, Y.-S.; Fang, Y.; and Han, Z. 2024a. CAP-UDF: Learning Unsigned Distance Functions Progressively from Raw Point Clouds with Consistency-Aware Field Optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 7475–7492.

Zhou, J.; Ma, B.; Liu, Y.-S.; and Han, Z. 2024b. Fast Learning of Signed Distance Functions from Noisy Point Clouds via Noise to Noise Mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 8936–8953.