Sharpening Neural Implicit Functions with Frequency Consolidation Priors

Anonymous submission

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.

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).

Recent methods (Mildenhall et al. 2020; Oechsle, Peng, and Geiger 2021; Takikawa et al. 2021; Martel et al. 2021; Rematas, Martin-Brualla, and Ferrari 2021; Jun and Nichol 2023; Takikawa et al. 2021; Müller et al. 2022; Li et al. 2023b) use a neural network to learn an SDF from 3D supervision (Jiang et al. 2020a; Michalkiewicz et al. 2019; Park et al. 2019; Ouasfi and Boukhayma 2022; Chen and Zhang 2019; Takikawa et al. 2021; Liu et al. 2021), 3D point clouds (Williams et al. 2019; Liu, Zhang, and Su 2020; Mi, Luo, and Tao 2020; Genova et al. 2019; Atzmon and Lipman 2020; Zhao et al. 2020; Atzmon and Lipman 2021), or multi-view images (Fu et al. 2022b; Oechsle, Peng, and Geiger 2021; Wang et al. 2021; Yu et al. 2022; Sitzmann, Zollhöfer, and Wetzstein 2019; Liu et al. 2020; Jiang et al. 2020b; Zakharov et al. 2020; Liu et al. 2019; Wu and Sun 2020; Niemeyer et al. 2020a; Lin, Wang, and Lucey 2020; Yariv et al. 2021; Fu et al. 2022a; Wang et al. 2021; Yu et al. 2022; Wang, Skorokhodov, and Wonka 2022; Vicini, Speierer, and Jakob 2022; Wang et al. 2022a), which seamlessly turns a neural network into a neural implicit function. However, due to various reasons like neural networks' bias on low frequency signals, 3D unaware sampling, sparsity in point clouds, or low resolutions of images, neural SDFs still struggle to represent geometries with high frequency components like sharp structures, especially for the ones inferred from point clouds or multi-view images. Although positional encodings (Mildenhall et al. 2020; Park et al. 2021) or feature grids (Sara Fridovich-Keil and Alex Yu et al. 2022; Chen, Liu, and Han 2023) were proposed to recover high frequency components during the inference, their downsides cause either unstable optimization or discontinuous representations, resulting in either artifacts in empty space, or noisy surfaces, holes. Thus, how to recover high frequency components in neural implicit functions is still a challenge.

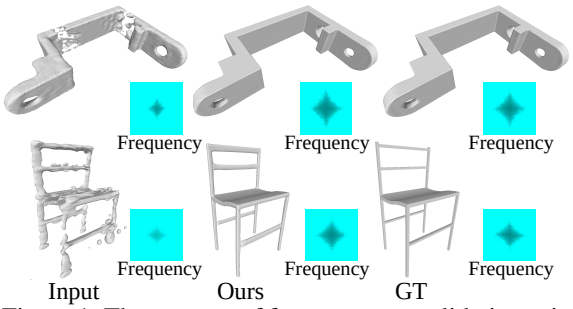

Figure 1: The concept of frequency consolidation priors. We also show averaged frequency weights across a band.

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 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 prior's effectiveness and good generalization in shape and scene modeling. Benchmark comparisons show our method's superiority in reconstruction 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 recovers full frequency coverage from a low frequency observation.
- We justify the idea of representing frequency components as embeddings. This design prompts the generalization of learned priors due to the simplicity of recovering the embedding of full frequency coverage by conducting 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 (Michalkiewicz et al. 2019; Mescheder et al. 2019; Chen and Zhang 2019; Jiang et al. 2020b; Chen, Liu, and Han 2022). 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 Peng 2020; Martel et al. 2021; Takikawa et al. 2021; Liu et al. 2021; Tang et al. 2021), 3D point clouds (Zhou et al. 2022; Ma et al. 2021; Gropp et al. 2020; Atzmon and Lipman 2020; Zhao et al. 2020; Atzmon and Lipman 2021; Chen, Liu, and Han 2022), or multi-view images (Mildenhall et al. 2020; Fu et al. 2022b; Oechsle, Peng, and Geiger 2021; Wang et al. 2021; Yu et al. 2022; Wang, Skorokhodov, and Wonka 2022; Vicini, Speierer, and Jakob 2022; Wang et al. 2022a; Guo et al. 2022; Rosu and Behnke 2023; Li et al. 2023b; Jiang, Hua, and Han 2023). With differentiable renderers, neural implicit representations can be learnt by minimizing errors between their 2D renderings and ground truth images. Using surface rendering (Jiang et al. 2020b), DVR (Niemeyer et al. 2020b) and IDR (Yariv et al. 2020) estimate geometry in a radiance field. IDR also models view direction as a condition to reconstruct high frequency details. Since these methods focus on intersections on surfaces, they need masks to filter out the background.

NeRF (Mildenhall et al. 2020) and its variations (Park et al. 2021; Rückert, Franke, and Stamminger 2021; Sara Fridovich-Keil and Alex Yu et al. 2022; Müller et al. 2022) simultaneously model geometry and color using volume rendering. These methods 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 additional priors or losses including depth (Yu et al. 2022; Azinović et al. 2022; Wang, Wang, and Agapito 2023; Zhu et al. 2022a,b; Hu and Han 2023), normals (Yu et al. 2022; Wang et al. 2022a; Guo et al. 2022), multi-view consistency (Fu et al. 2022b), and segmentation priors (Kong et al. 2023; Haghighi et al. 2023)

Learning with Frequency. Learning neural implicit representations with multi-scale details enhances interpretability. It allows progressively detailed visualization at different scales (Saragadam et al. 2022; Takikawa et al. 2021). Controlling curvature regulation (Ehret, Marí, and Facciolo 2022) can add or remove surface details. A common approach is learning neural implicits with several frequency bands that cover a whole frequency scope (Lindell et al. 2022; Yang et al. 2022; Shayan et al. 2022; Cho et al. 2022; Grattarola and Vandergheynst 2022; Benbarka et al. 2022). This yields a multi-scale representation by reconstructing surfaces or signals from different frequency bands. Moreover, SAP learns an occupancy function by solving a Poisson equation in the frequency domain (Peng et al. 2021a).

Recovering Sharp Structures. There are several different strategies to recover sharp edges. Dual contouring (Chen et al. 2022) can reconstruct sharper edges than the marching cubes (Lorensen and Cline 1987) with the help of gradients. By modeling displacements, more high frequency details can get recovered on surfaces (Yifan, Rahmann, and Sorkine-hornung 2021). Edges are also an important structure to recover, especially in CAD modeling. NEF (Ye et al. 2023) was proposed to learn an implicit function to represent edges from multi-view images. Some methods (Matveev et al. 2022; Cherenkova et al. 2023; Lambourne et al. 2022; Feng et al. 2023) focus on sharpening edges directly on 3D shapes. However, these methods merely work on relatively clean and plausible shapes. Consolidation is also a way to sharpen shapes, especially for point clouds (Metzer et al. 2021), which can generate points with sharp features or in sparse regions, and also remove noises and outliers.

Unlike previous methods, our approach aims to sharpen implicit functions 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 of a shape 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

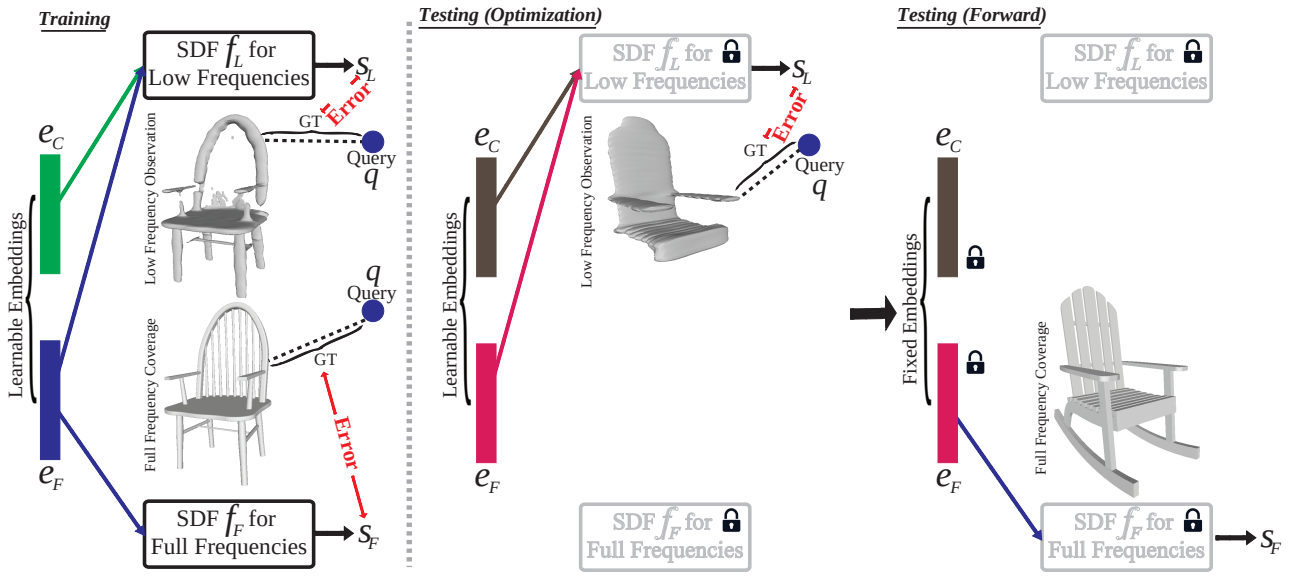

Figure 2: The overview of our method.

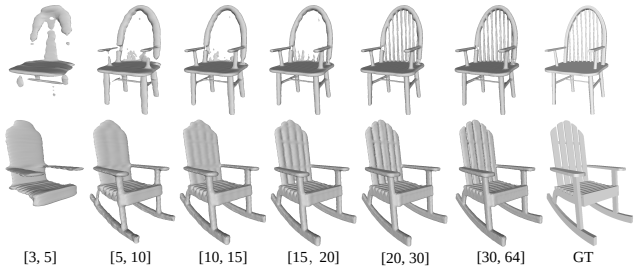

Figure 3: The illustration of low frequency observations and the full frequency coverage.

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-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 (Vallet and Levy 2008; Zhang, van Kaick, and Dyer 2007) does eigen-decomposition of the discrete Laplace–Beltrami operator and regards the eigenvectors as frequency components. Some recent learning-based methods (Lindell et al. 2022; Takikawa et al. 2021; Saragadam et al. 2022) 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. 2021b). The low frequency observations (SDF) established by our method with proper low frequency band could produce over smoothed surfaces which are very similar to the one produced by spectral geometry theory one meshes in Fig. 4.

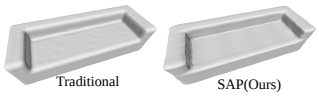

Figure 4: Over-smoothed surfaces.

Specifically, we first randomly sample dense points on the mesh of M, and then estimate an occupancy field by solving a Poisson surface reconstruction equation, where we obtain magnitudes for frequencies in a frequency band. We reconstruct M by running the marching cubes algorithms (Lorensen and Cline 1987) with the estimated occupancy function, and use the reconstruction as the full frequency coverage M_F. At the same time, we produce low frequency occupancy functions by removing some parts of high frequency components (setting their corresponding magnitudes to 0), and use each one of the manipulated occupancy functions to reconstruct a mesh as a low frequency observation M_L. To make low frequency observations cover all frequencies over the frequency band $[0, 64]$ illustrated

in Fig. 3, we randomly select one frequency from one of six subbands, such as "[3,5]" or "[5,10]", and remove all frequencies larger than the sampled one. For instance, we select 4 in "[3,5]", then we will get a low frequency observation M_L by removing all frequencies from 5 to 64. Each low frequency observation produced is paired with the full frequency coverage M_F as a training sample. Fig. 3 illustrates 6 pairs of low frequency observations $\{M_L\}$ and their corresponding full frequency coverage M_F (rightmost). Note that we do not simply use the ground truth meshes as a full frequency coverage M_F to avoid non-watertight meshes during training.

Moreover, we produce more low frequency observations in the frequency band $[3, 30]$, as most inputs during inference 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 their ground truth signed distances $s_L^{gt\prime}$. We estimate $\boldsymbol{e}'_L = [\boldsymbol{e}'_F \;\; \boldsymbol{e}'_C]$ with fixed parameters $\boldsymbol{\theta}_L$ by minimizing the reconstruction errors below,

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

After the optimization, we represent the SDF of a full frequency coverage M'_F as $f_F(q, \boldsymbol{e}'_F)$. We can reconstruct the surface of M'_F by running the marching cubes (Lorensen and Cline 1987) with $f_F(q, \boldsymbol{e}'_F)$.

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 (Park et al. 2019)	0.287	0.381	0.804
ConvOcc (Songyou Peng 2020)	0.306	0.451	0.805
LIG (Jiang et al. 2020a)	0.292	0.430	0.809
IDF (Yifan, Rahmann, and Sorkine-hornung 2021)	0.287	0.390	0.815
NDC (Chen et al. 2022)	0.269	0.358	0.768
POCO(pretrained)	0.259	0.374	0.812
POCO (Boulch and Marlet 2022)	0.217	0.284	0.858
ALTO(pretrained)	0.253	0.367	0.819
ALTO (Wang et al. 2022b)	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 by numerical and visual comparisons with the latest methods on the ShapeNet (Chang et al. 2015), ABC (Koch et al. 2019), and ScanNet datasets (Dai et al. 2017). For ShapeNet, we report evaluations under 13 classes with the train/test split from 3D-R2N2 (Choy et al. 2016) and also on 8 classes under the test split from NeuralTPS (Chen, Han, and Liu 2023) for fair comparisons. For ABC, we follow Points2Surf (Erler et al. 2020), and use its train/test splitting for evaluation. For ScanNet, we follow Neural Part Priors (Bokhovkin and Dai

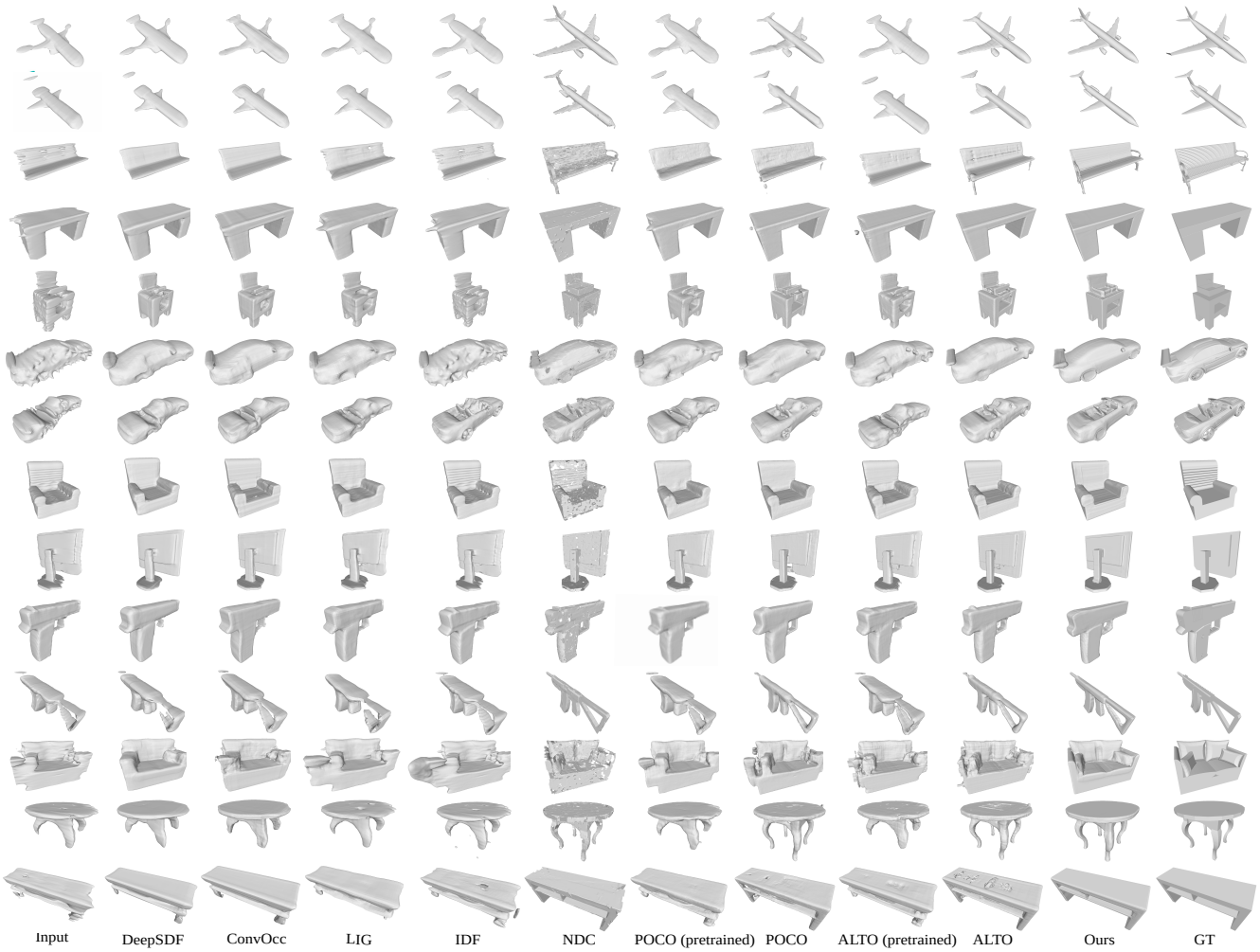

Figure 5: Visual comparison with the state-of-the-art on ShapeNet.

2022) and use the same 6 classes and test split. All experiments leverage marching cubes (Lorensen and Cline 1987) on a 256^3 grid to reconstruct meshes.

For the ShapeNet dataset, we measure errors using L1 Chamfer Distance (CD_{L1}), L2 Chamfer Distance (CD_{L2}), and normal consistency (NC). Following NeuralTPS (Chen, Han, and Liu 2023), we randomly sample 100k points on the reconstructed and ground truth meshes. Under the ScanNet dataset, we follow Neural Part Priors to report L1 Chamfer Distance (CD_{L1}) between the reconstructed meshes and the ground truth meshes which are transformed to the ScanNet coordinate space. Notice that in Neural Parts Prior, each shape is evaluated over the union of the CD_{L1} of predicted and ground truth parts, which is equivalent to the CD_{L1} between global predictions and global ground truths. We report two results using annotated shapes from Scan2CAD (Avetisyan et al. 2019) dataset and extracted shapes using ScanNet segmentation masks as ground truth, respectively.

Learning Frequency Consolidation Priors. For each shape used for training, we use the first 5 low frequency observations (as shown in Fig. 3) and the full frequency coverage to learn the frequency consolidation prior. We sample queries around both the low frequency obseravtions and the full frequency coverage, and record the signed distances to both of the low and full frequency meshes as supervision.

Evaluations with The Latest Methods

Evaluation on ShapeNets. We use our frequency consolidation priors learned from training samples for evaluation. For each shape in the testing dataset, we produce low-frequency observations using the method previously described and use the worst observation from each shape to evaluate all compared methods. We report the average evaluations over all classes in Tab. 1, demonstrating that our method achieves the best performance. Please refer to our supplementary materials for more detailed evaluation in each class.

We use pre-trained parameters or parameters retrained using our data from the latest methods. To train the latest methods like DeepSdf (Park et al. 2019), POCO (Boulch and Marlet 2022), and ALTO (Wang et al. 2022b) which map a point cloud into an SDF, we sample the low frequency observation as a point cloud, record the signed distances at queries near their full frequency coverages, forming training samples for both POCO and ALTO. Our method significantly outperforms these methods. Visual comparisons in Fig. 5 shows that DeepSDF, POCO and ALTO struggle to generalize their prior knowledge on various low-frequency observations. With pre-trained parameters, NDC (Chen et al. 2022) produces sharp edges with dual contouring, preserving more and sharper structures than marching cubes algorithms (Lorensen and Cline 1987). But NDC does not gener-

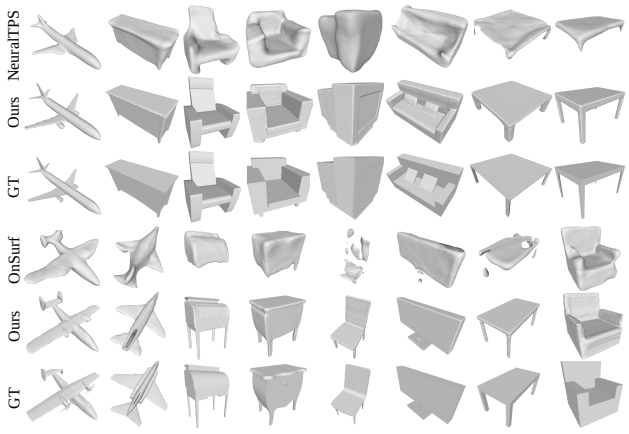

Figure 6: Reconstruction results for test-time optimization with sparse reconstructions from NeuralTPS and OnSurf as low-frequency observations, respectively.

alize well on low frequency shapes and produces broke and noisy surfaces, as shown in Fig. 5. IDF (Yifan, Rahmann, and Sorkine-hornung 2021) and ConvOcc (Songyou Peng 2020) pretrained on corrupted shapes also generate high frequency geometry on surfaces from corrupted point clouds in the spatial space. Comparisons in Tab. 1 and Fig. 5 show that it can not handle large geometry variations.

Refining Reconstruction 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. What the performance of our prior is on unobserved frequency band will be answered in this experiment. We use reconstruction from sparse point clouds by the latest method 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), the latest methods for reconstruction from sparse point clouds, to produce our test shapes. They provide shapes reconstructed from 300 points of each shape in 8 classes from ShapeNet. Thus, we use our learned prior to recover the high frequency components. Numerical and visual comparisons in Tab. 2, Fig. 6, and Fig. 9 (first two rows) indicate that our method can generalize the prior well on unknown frequency bands and improve the reconstruction accuracy by recovering sharper edges, flatter planes, and more complete surfaces. Please refer to our supplementary materials for more evaluation in each class.

Evaluations on CAD Modeling. CAD shapes usually contain many sharper edges. We learn a prior from shapes in the training dataset collected from ABC dataset (Koch et al. 2019) by Points2Surf (Erler et al. 2020). We use the same way to produce training samples using 5 low frequency and a full frequency coverage from each shape for training. For each testing shape, we also produce 5 low frequency observations, and use each one as a testing sample. We calculate the mean and variance of the 5 evaluations for each testing shape. Tab. 3 reports the average values over all testing shapes, comparing our method with Points2Surf (Erler et al. 2020) and SECAD-Net (Li et al. 2023a). Our method achieves the best reconstruction accuracy and stability. Visual comparisons in Fig. 7 show that our method can generate much sharper edges and more accurate structures. Points2Surf struggles with surface variations, and SECAD-Net generates sharper edges but does not generalize well.

Method	$CD_{L1} \times 10$	$CD_{L2} \times 100$	NC
Onsurf (Ma et al. 2022)	0.214	0.223	0.845
Onsurf+Ours	**0.180**	**0.165**	**0.886**
NeuralTPS (Chen, Han, and Liu 2023)	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.

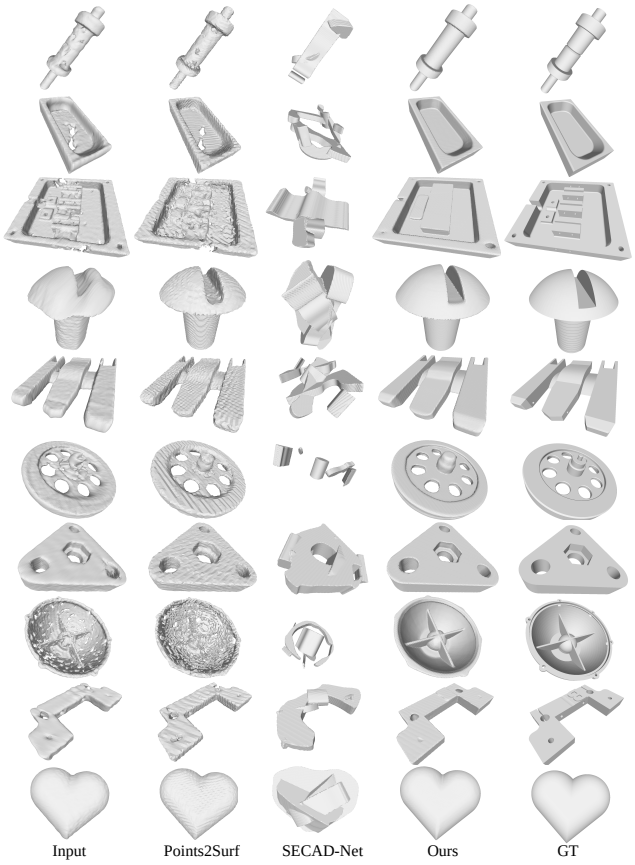

Figure 7: Visual comparisons on ABC.

Reconstruction in Scenes. Our learned prior also works with objects in scene modeling. We learn our priors on ShapeNet classes that appear in the scenes used for evaluation. These scenes are reconstructed from real scans. We use the GT segmentation masks to segment shapes as partial meshes. These partial meshes are also with artifacts, few geometry, unobserved frequency bands or severe corruption. We use the poses and scale information from Scan2CAD (Avetisyan et al. 2019) to determine the layout in our visualizations.

We use two kinds of GT shapes in evaluations in Tab. 1 and Tab. 2 in our supplement, respectively. One is the shapes provided by Scan2CAD, which are retrieved from ShapeNet. These shapes are complete but may drift away a lot from the real scans. The other kind is the shapes obtained directly from real scans with GT segmentation masks. These shapes are mostly incomplete but more identical to the real scenes.

With a shape segmented from a scene, we use NeuralPull to reconstruct a coare but watertight mesh, serving as a low

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. 2023a)	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.

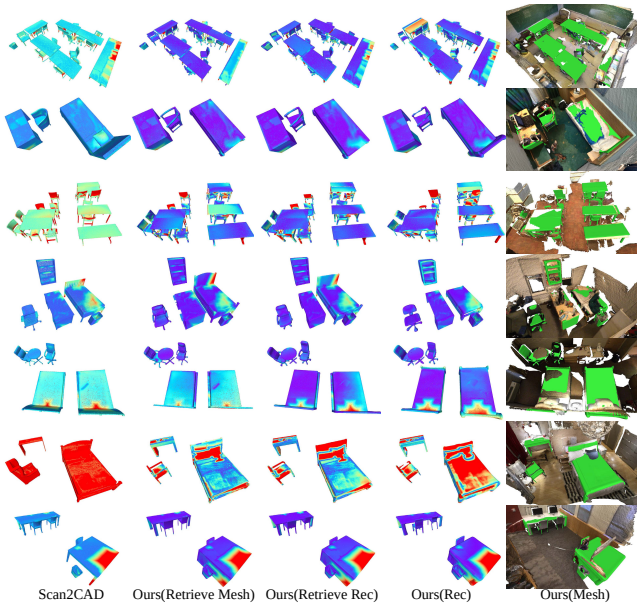

Figure 8: Visual comparisons on ScanNet dataset. The red in error maps indicates larger errors.

frequency observation, We then generalize the learned prior to recover its full frequency coverage (Rec). Using the low frequency observation, we can also produce two results by retrieving all low frequency observations in our training set. Specifically, we render 30 images from viewpoints around each shape. We use clip image encoder (Radford et al. 2021) to extract features of each image. Then, we use a single direction CD distance as a retrieval metric to evaluate the distance between two sets of images representing two shapes. 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.

Our reconstruction results produce more accurate reconstruction than NeuralPartPriors (Bokhovkin and Dai 2022) and PartUnderstanding (Bokhovkin et al. 2021) using Scan2CAD as GT shapes. Based on that, our retrieved results can produce even better results. Similarly, comparisons in Tab. 2 in our supplement show our superiority over Scan2CAD with segmented meshes as ground truth shapes. We detail our results with error maps in Fig. 8, and show plausible results on bad reconstructions from scenes in ScanNet in Fig. 9. We can see that shapes retrieved by Scan2CAD are not every identical to the real scans.

Figure 9: Reconstrcutions from severely corrupted cases.

Ablation Studies and Analysis

Semantic Latent Space. The latent space that we learn is semantic. We visualize the reconstructed full frequency coverage optimization process in the self-reconstruction during testing. The transformation from one latent code to another in Fig. 10 shows semantic shapes on the optimization path. Moreover, we can also manipulate embeddings in a semantic way like the plus and minus of embeddings for full frequency components in Fig. 11. We reduce the dimensions of all embeddings learned for low and full frequency components in each iteration using TSNE (van der Maaten and Hinton 2008) in Fig. 12. We also see semantic structures like lines, each of which is formed by embeddings learned in all iterations (the order is mapped from light to dark color) on an optimization path. All optimization paths start from a similar point, goes quite similar in the beginning and become diverse at the end which corresponds to different shapes.

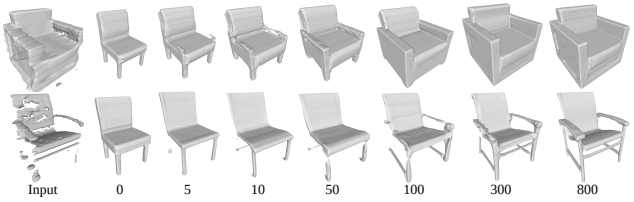

Figure 10: Visualization of the test-time optimization.

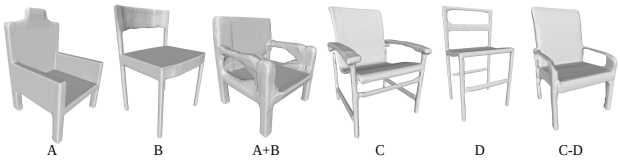

Figure 11: Embedding manipulations for shape generation.

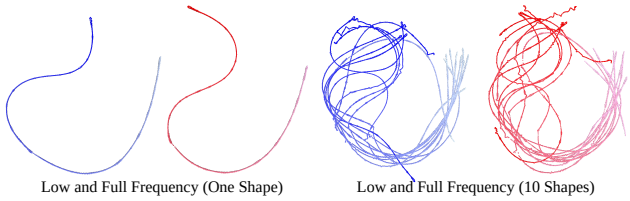

Figure 12: Embeddings for low and full frequency shapes.

Conclusion

We introduce frequency consolidation priors to sharpen neural implicit functions. We successfully learn the priors from an established set containing training pairs with low frequency components and full frequency coverage. The learned priors can seamlessly work with our novel ways of recovering full frequency coverage from a low frequency observation, which significantly increases the generalization ability of the learned priors. We show that the learned priors can recover high frequency components from the low frequency observation, which sharpens the surfaces but also completes some missing structures. Our numerical and visual comparisons with the latest methods on widely used shape or scene datasets show that our priors can recover geometries with higher frequencies by sharpening low frequency SDF observation than the latest methods.

References

Atzmon, M.; and Lipman, Y. 2020. SAL: Sign Agnostic Learning of Shapes From Raw Data. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Atzmon, M.; and Lipman, y. 2021. SALD: Sign Agnostic Learning with Derivatives. In *International Conference on Learning Representations*.

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 (CVPR)*.

Azinović, D.; Martin-Brualla, R.; Goldman, D. B.; Nießner, M.; and Thies, J. 2022. Neural RGB-D Surface Reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 6290–6301.

Benbarka, N.; Hofer, T.; ul-moqeet Riaz, H.; and Zell, A. 2022. Seeing Implicit Neural Representations as Fourier Series. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2283–2292.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

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

Chen, Z.; and Zhang, H. 2019. Learning Implicit Fields for Generative Shape Modeling. *IEEE Conference on Computer Vision and Pattern Recognition*.

Cherenkova, K.; Dupont, E.; Kacem, A.; Arzhannikov, I.; Gusev, G.; and Aouada, D. 2023. SepicNet: Sharp Edges Recovery by Parametric Inference of Curves in 3D Shapes. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2727–2735. Los Alamitos, CA, USA: IEEE Computer Society.

Cho, J.; Nam, S.; Rho, D.; Ko, J. H.; and Park, E. 2022. Streamable neural fields. In *European Conference on Computer Vision*, 595–612. Springer.

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*.

Feng, Y.-F.; Shen, L.-Y.; Yuan, C.-M.; and Li, X. 2023. Deep Shape Representation with Sharp Feature Preservation. *Comput. Aided Des.*, 157(C).

Fu, Q.; Xu, Q.; Ong, Y.; and Tao, W. 2022a. Geo-Neus: Geometry-Consistent Neural Implicit Surfaces Learning for Multi-view Reconstruction.

Fu, Q.; Xu, Q.; Ong, Y.-S.; and Tao, W. 2022b. 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*.

Gropp, A.; Yariv, L.; Haim, N.; Atzmon, M.; and Lipman, Y. 2020. Implicit Geometric Regularization for Learning Shapes. In *International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, 3789–3799.

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 (NeurIPS)*.

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*.

Jun, H.; and Nichol, A. 2023. Shap-e: Generating conditional 3d implicit functions. *arXiv preprint arXiv:2305.02463*.

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*.

Kong, X.; Liu, S.; Taher, M.; and Davison, A. J. 2023. vMAP: Vectorised Object Mapping for Neural Field SLAM. *arXiv preprint arXiv:2302.01838*.

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. 2023a. SECAD-Net: Self-Supervised CAD Reconstruction by Learning Sketch-Extrude Operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Li, Z.; Müller, T.; Evans, A.; Taylor, R. H.; Unberath, M.; Liu, M.-Y.; and Lin, C.-H. 2023b. Neuralangelo: High-Fidelity Neural Surface Reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Lin, C.-H.; Wang, C.; and Lucey, S. 2020. SDF-SRN: Learning Signed Distance 3D Object Reconstruction from Static Images. In *Advances in Neural Information Processing Systems*.

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, M.; Zhang, X.; and Su, H. 2020. Meshing Point Clouds with Predicted Intrinsic-Extrinsic Ratio Guidance. In *European Conference on Computer vision*.

Liu, S.; Saito, S.; Chen, W.; and Li, H. 2019. Learning to Infer Implicit Surfaces without 3D Supervision. In *Advances in Neural Information Processing Systems*.

Liu, S.; Zhang, Y.; Peng, S.; Shi, B.; Pollefeys, M.; and Cui, Z. 2020. DIST: Rendering Deep Implicit Signed Distance Function with Differentiable Sphere Tracing. In *IEEE Conference on Computer Vision and Pattern Recognition*.

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*.

Martel, J. N. P.; Lindell, D. B.; Lin, C. Z.; Chan, E. R.; Monteiro, M.; and Wetzstein, G. 2021. ACORN: Adaptive Coordinate Networks for Neural Scene Representation. *CoRR*, abs/2105.02788.

Matveev, A.; Rakhimov, R.; Artemov, A.; Bobrovskikh, G.; Egiazarian, V.; Bogomolov, E.; Panozzo, D.; Zorin, D.; and Burnaev, E. 2022. DEF: Deep Estimation of Sharp Geometric Features in 3D Shapes. *ACM Trans. Graph.*, 41(4).

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).

Mi, Z.; Luo, Y.; and Tao, W. 2020. SSRNet: Scalable 3D Surface Reconstruction Network. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Michalkiewicz, M.; Pontes, J. K.; Jack, D.; Baktashmotlagh, M.; and Eriksson, A. P. 2019. Deep Level Sets: Implicit Surface Representations for 3D Shape Inference. *CoRR*, abs/1901.06802.

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*.

Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *arXiv:2201.05989*.

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

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

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*.

Park, K.; Sinha, U.; Barron, J. T.; Bouaziz, S.; Goldman, D. B.; Seitz, S. M.; and Martin-Brualla, R. 2021. Nerfies: Deformable Neural Radiance Fields. *IEEE International Conference on Computer Vision*.

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

Peng, S.; Jiang, C. M.; Liao, Y.; Niemeyer, M.; Pollefeys, M.; and Geiger, A. 2021b. 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.

Rematas, K.; Martin-Brualla, R.; and Ferrari, V. 2021. Sharf: Shape-conditioned Radiance Fields from a Single View. In *International Conference on Machine Learning*.

Rosu, R. A.; and Behnke, S. 2023. PermutoSDF: Fast Multi-View Reconstruction with Implicit Surfaces using Permutohedral Lattices. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Rückert, D.; Franke, L.; and Stamminger, M. 2021. Adop: Approximate differentiable one-pixel point rendering. *arXiv:2110.06635*.

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*.

Saragadam, V.; Tan, J.; Balakrishnan, G.; Baraniuk, R. G.; and Veeraraghavan, A. 2022. MINER: Multiscale Implicit Neural Representations. In *European Conference on Computer Vision*.

Shayan, S.; David, L.; David, F.; and Marcus, B. 2022. Residual Multiplicative Filter Networks for Multiscale Reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Sitzmann, V.; Zollhöfer, M.; and Wetzstein, G. 2019. Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations. In *Advances in Neural Information Processing Systems*.

Songyou Peng, L. M. M. P. A. G., Michael Niemeyer. 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*.

Tang, J.; Lei, J.; Xu, D.; Ma, F.; Jia, K.; and Zhang, L. 2021. SA-ConvONet: Sign-Agnostic Optimization of Convolutional Occupancy Networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

Vallet, B.; and Levy, B. 2008. Spectral Geometry Processing with Manifold Harmonics. *Computer Graphics Forum*, 27: 251 – 260.

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. *Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition (CVPR)*.

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*, 27171–27183.

Wang, Y.; Skorokhodov, I.; and Wonka, P. 2022. HF-NeuS: Improved Surface Reconstruction Using High-Frequency Details.

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*.

Williams, F.; Schneider, T.; Silva, C.; Zorin, D.; Bruna, J.; and Panozzo, D. 2019. Deep Geometric Prior for Surface Reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Wu, Y.; and Sun, Z. 2020. DFR: Differentiable Function Rendering for Learning 3D Generation from Images. *Computer Graphics Forum*, 39(5): 241–252.

Yang, G.; Benaim, S.; Jampani, V.; Genova, K.; Barron, J.; Funkhouser, T.; Hariharan, B.; and Belongie, S. 2022. Polynomial Neural Fields for Subband Decomposition and Manipulation. In *Thirty-Sixth Conference on Neural Information Processing Systems*.

Yariv, L.; Gu, J.; Kasten, Y.; and Lipman, Y. 2021. Volume rendering of neural implicit surfaces. In *Advances in Neural Information Processing Systems*.

Yariv, L.; Kasten, Y.; Moran, D.; Galun, M.; Atzmon, M.; Ronen, B.; and Lipman, Y. 2020. Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance. *Advances in Neural Information Processing Systems*, 33.

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*.

Yu, Z.; Peng, S.; Niemeyer, M.; Sattler, T.; and Geiger, A. 2022. MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. *ArXiv*, abs/2022.00665.

Zakharov, S.; Kehl, W.; Bhargava, A.; and Gaidon, A. 2020. Autolabeling 3D Objects with Differentiable Rendering of SDF Shape Priors. In *IEEE Conference on Computer Vision and Pattern Recognition*.

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.

Zhao, W.; Lei, J.; Wen, Y.; Zhang, J.; and Jia, K. 2020. Sign-Agnostic Implicit Learning of Surface Self-Similarities for Shape Modeling and Reconstruction from Raw Point Clouds. *CoRR*, abs/2012.07498.

Zhou, J.; Ma, B.; Liu, Y.-S.; Fang, Y.; and Han, Z. 2022. Learning Consistency-Aware Unsigned Distance Functions Progressively from Raw Point Clouds. In *Advances in Neural Information Processing Systems*.

Zhu, Z.; Peng, S.; Larsson, V.; Xu, W.; Bao, H.; Cui, Z.; Oswald, M. R.; and Pollefeys, M. 2022a. NICE-SLAM: Neural Implicit Scalable Encoding for SLAM. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhu, Z.; Peng, S.; Larsson, V.; Xu, W.; Bao, H.; Cui, Z.; Oswald, M. R.; and Pollefeys, M. 2022b. NICE-SLAM: Neural Implicit Scalable Encoding for SLAM. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Reproducibility Checklist

1. **This paper:**
   - Includes a conceptual outline and/or pseudocode description of AI methods introduced (yes).
   - Clearly delineates statements that are opinions, hypothesis, and speculation from objective facts and results (yes).
   - Provides well marked pedagogical references for less-familiare readers to gain background necessary to replicate the paper (yes)
2. **Does this paper make theoretical contributions? (yes**

   If yes, please complete the list below.:
   - All assumptions and restrictions are stated clearly and formally.(Yes)
   - All novel claims are stated formally (e.g., in theorem statements). (yes)
   - Proofs of all novel claims are included. (yes)
   - Proof sketches or intuitions are given for complex and/or novel results. (yes)
   - Appropriate citations to theoretical tools used are given. (yes)
   - All theoretical claims are demonstrated empirically to hold. (yes)
   - All experimental code used to eliminate or disprove claims is included. (yes)
3. **Does this paper rely on one or more datasets? (yes)**

   If yes, please complete the list below.
   - A motivation is given for why the experiments are conducted on the selected datasets (yes)
   - All novel datasets introduced in this paper are included in a data appendix. (NA)
   - All novel datasets introduced in this paper will be made publicly available upon publication of the paper with a license that allows free usage for research purposes. (NA)
   - All datasets drawn from the existing literature (potentially including authors' own previously published work) are accompanied by appropriate citations. (yes)
   - All datasets drawn from the existing literature (potentially including authors' own previously published work) are publicly available. (yes)
   - All datasets that are not publicly available are described in detail, with explanation why publicly available alternatives are not scientifically satisficing. (NA)
4. **Does this paper include computational experiments? (yes)**

   If yes, please complete the list below.
   - Any code required for pre-processing data is included in the appendix. (no).
   - All source code required for conducting and analyzing the experiments is included in a code appendix. (partial)
   - All source code required for conducting and analyzing the experiments will be made publicly available upon publication of the paper with a license that allows free usage for research purposes. (yes)
   - All source code implementing new methods have comments detailing the implementation, with references to the paper where each step comes from (partial)
   - If an algorithm depends on randomness, then the method used for setting seeds is described in a way sufficient to allow replication of results. (NA)
   - This paper specifies the computing infrastructure used for running experiments (hardware and software), including GPU/CPU models; amount of memory; operating system; names and versions of relevant software libraries and frameworks. (yes)
   - This paper formally describes evaluation metrics used and explains the motivation for choosing these metrics. (yes)
   - This paper states the number of algorithm runs used to compute each reported result. (yes)
   - Analysis of experiments goes beyond single-dimensional summaries of performance (e.g., average; median) to include measures of variation, confidence, or other distributional information. (yes)
   - The significance of any improvement or decrease in performance is judged using appropriate statistical tests (e.g., Wilcoxon signed-rank). (no)
   - This paper lists all final (hyper-)parameters used for each model/algorithm in the paper's experiments. (yes)
   - This paper states the number and range of values tried per (hyper-) parameter during development of the paper, along with the criterion used for selecting the final parameter setting. (yes)