

Learning Compact Latent Space for Representing Neural Signed Distance Functions with High-fidelity Geometry Details

Qiang Bai¹, Bojian Wu², Xi Yang^{1*}, Zhizhong Han³

¹ Jilin University, Changchun, China

² Zhejiang University, Hangzhou, China

³ Wayne State University, Detroit, USA

baiqiang23@mails.jlu.edu.cn, ustcbjwu@gmail.com, yangxi21@jlu.edu.cn, h312h@wayne.edu

Abstract

Neural signed distance functions (SDFs) have been a vital representation to represent 3D shapes or scenes with neural networks. An SDF is an implicit function that can query signed distances at specific coordinates for recovering a 3D surface. Although implicit functions work well on a single shape or scene, they pose obstacles when analyzing multiple SDFs with high-fidelity geometry details, due to the limited information encoded in the latent space for SDFs and the loss of geometry details. To overcome these obstacles, we introduce a method to represent multiple SDFs in a common space, aiming to recover more high-fidelity geometry details with more compact latent representations. Our key idea is to take full advantage of the benefits of generalization-based and overfitting-based learning strategies, which manage to preserve high-fidelity geometry details with compact latent codes. Based on this framework, we also introduce a novel sampling strategy to sample training queries. The sampling can improve the training efficiency and eliminate artifacts caused by the influence of other SDFs. We report numerical and visual evaluations on widely used benchmarks to validate our designs and show advantages over the latest methods in terms of the representative ability and compactness.

Code — <https://github.com/eoozbq/Compact-SDF>

Introduction

Signed distance functions (SDFs) are a vital representation for 3D perception in computer vision and robotics. It predicts a signed distance at an arbitrary location, resulting in an implicit function which can be parameterized well by deep neural networks, dubbed neural SDFs (Park et al. 2019; Xie et al. 2022). To reconstruct a surface, an SDF predicts signed distances on all vertices of a volume grid, which is then used in the marching cubes algorithm (Lorensen and Cline 1987a). Although SDFs are good at representing geometry with arbitrary topology, training neural SDFs requires a large amount of forward and backward procedures, and is limited in recovering high-frequency geometry details, due to the bias of neural networks on low-frequency geometry.

To recover more geometry details, the latest methods employ 3D volume grids to represent signed distances or spatial

features that can be decoded into signed distances (Takikawa et al. 2021; Chen et al. 2021; Liu et al. 2021; Müller et al. 2022; Yariv et al. 2023). Unlike neural networks, signed distances or spatial features on vertices of a grid can get tuned without a constraint of consistency to the ones on the neighboring vertices, leading to a capability of recovering more high-fidelity geometry details, such as sharper edges and smoother surfaces with less artifacts (Jiang et al. 2020a). These methods usually employ an overfitting strategy to overfit one volume grid on a single shape. Due to the lack of a compact latent space to represent multiple volume grids, these methods can not generalize to other shapes like some priors trained in a data-driven manner. This also makes it challenging to analyze multiple SDFs which are expected to preserve high-fidelity geometry details.

To overcome this challenge, we introduce a method to learn a compact latent space for representing multiple SDFs, where more high-fidelity geometry details can be recovered. Our key idea is to take full advantage of the benefits from generalization-based neural networks and overfitting-based volume grids. Specifically, to learn SDF, we use a volume grid to model the signed distance field near the surface, while using a neural network to model the signed distance field far away from the surface, where the neural network and the volume grid share the same compact latent code representing the shape. This design aims to leverage the generalization ability of neural networks to remove artifacts caused by influence from different shapes, which removes the need of sampling dense queries in areas that are far away from the surface, while using the volume grid to improve the accuracy of the SDF near the surface, where we sample dense queries to significantly improve the geometry details. Based on this design, the learned latent codes are compact and more representative of shapes with high-fidelity geometry details. We report numerical and visual evaluations to show the superiority of our method over the latest methods on widely used benchmarks. Our main contributions are listed below,

- We introduce a novel method to represent multiple neural SDFs with high-fidelity geometry details, leading to a compact latent space for representing multiple SDFs.
- We introduce a novel sampling strategy to improve the efficiency and performance of learning implicit functions with volume grids.

*Corresponding author.

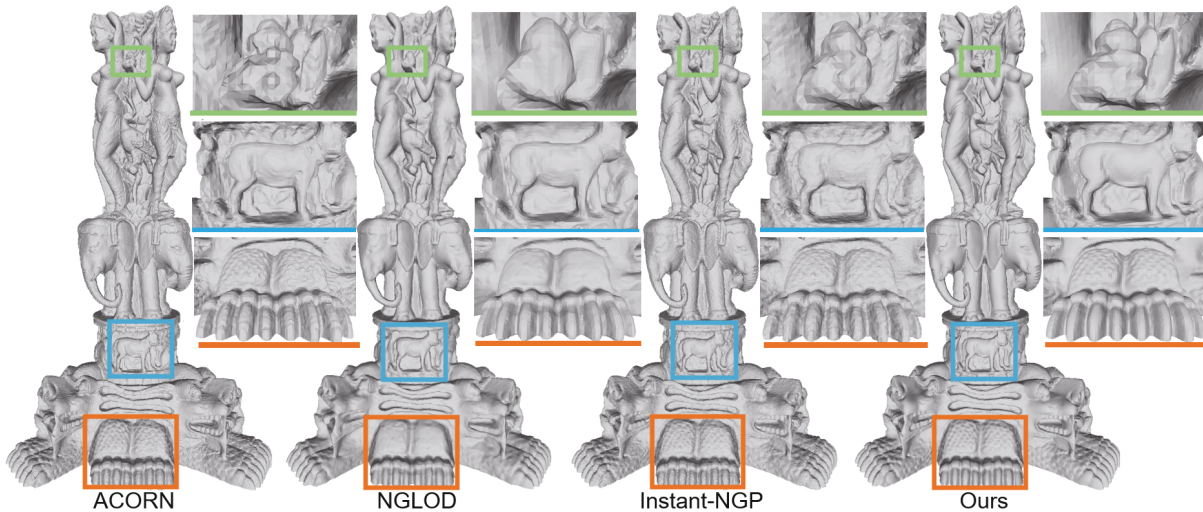


Figure 1: Reconstruction of a statuette using Marching Cubes at 512^3 resolution.

- We report the state-of-the-art performance and show advantages over the latest methods on widely used benchmarks in representative ability and compactness.

Related Work

Recently, there has been a growing interest in using neural fields (Xie et al. 2022), such as occupancy functions (Mescheder et al. 2019; Peng et al. 2020) or signed distance functions (SDFs) (Park et al. 2019), to represent shapes with neural networks, which allows mesh extraction through contouring methods, like marching cubes (Lorensen and Cline 1987a).

Early approaches used low-resolution occupancy grids (Choy et al. 2016; Girdhar et al. 2016; Wang et al. 2018), while later methods employed octrees (Tatarchenko, Dosovitskiy, and Brox 2017; Wang et al. 2017) or sparse grids (Museth 2013; Setaluri et al. 2014; Martel et al. 2021) to reduce memory costs, though discretization introduced artifacts. To enable continuous implicit representation, seminal works like DeepSDF (Park et al. 2019), and others (Chen and Zhang 2019; Mescheder et al. 2019; Michalkiewicz et al. 2019; Liu et al. 2019; Hertz et al. 2021; Lipman 2021; Remelli et al. 2020; Davies, Nowrouzezahrai, and Jacobson 2021; Erler et al. 2020; Genova et al. 2020a; Liu et al. 2024) encoded shapes into latent vectors using the MLPs, allowing distance field queries at any 3D point. Recent techniques improved surface smoothness via auto-decoding (Park et al. 2019), curriculum learning (Duan et al. 2020), adversarial training (Kleineberg, Fey, and Weichert 2020), or Lipschitz regularization (Liu et al. 2022). However, single global latent codes often lack fine details. Some methods address this by using periodic activations (Sitzmann et al. 2020) or Fourier features (Tancik et al. 2020) to enhance high-frequency detail recovery.

To better capture local details, recent methods partition 3D space into local grids (Jiang et al. 2020b; Chabra et al. 2020; Chibane, Alldieck, and Pons-Moll 2020; Genova et al. 2020b, 2019; Martel et al. 2021; Saragadam et al. 2022;

Yao et al. 2021), assigning each a dedicated latent code or using interpolated feature grids. DeepLS (Chabra et al. 2020) stores independent latent codes per grid cell, while NGLoD (Takikawa et al. 2021) uses a level-of-detail structure for progressive resolution. MDIF (Chen et al. 2021) employs hierarchical latent grids for multi-resolution decoding, and IMLS (Liu et al. 2021) defines implicit functions on point sets. Instant-NGP (Müller et al. 2022) introduces efficient multi-resolution hashing for compact representation, and MosaicSDF (Yariv et al. 2023) focuses grids near shape boundaries. By encoding only local geometry, these methods achieve finer details than global representations. Unlike existing SDF fitting techniques, our approach learns multiple SDFs in a shared latent space, enabling higher-fidelity details with more compact storage.

Additionally, in recent 3D generation tasks, encoding shapes into a shared latent space is key, as seen in Hyper-Diffusion (Erkoç et al. 2023), 3DILG (Zhang, Nießner, and Wonka 2022), and 3DShape2VecSet (Zhang et al. 2023b). While these methods focus on latent space mapping, they often overlook geometric details. In contrast, we identify the causes of detail loss and introduce dual branches to jointly encode high- and low-frequency information, enabling more precise geometric reconstruction.

Method

Overview. We aim to learn multiple SDFs with geometry details in a compact latent space. As illustrated in Fig. 2, our method is formed by two branches, one is a generalization-based network, the other is an overfitting-based network, which shares a learnable latent code that differentiates SDFs. We aim to learn parameters in the two branches, latent codes $\{z\}$ for representing multiple SDFs, and spatial features on vertices of a volume grid where queries can get spatial features c by trilinear interpolation. The latent code z is a sample in the learned latent code space. The first branch maps a coordinate with the latent code into a signed distance. We employ its generalization ability to generalize the prior to

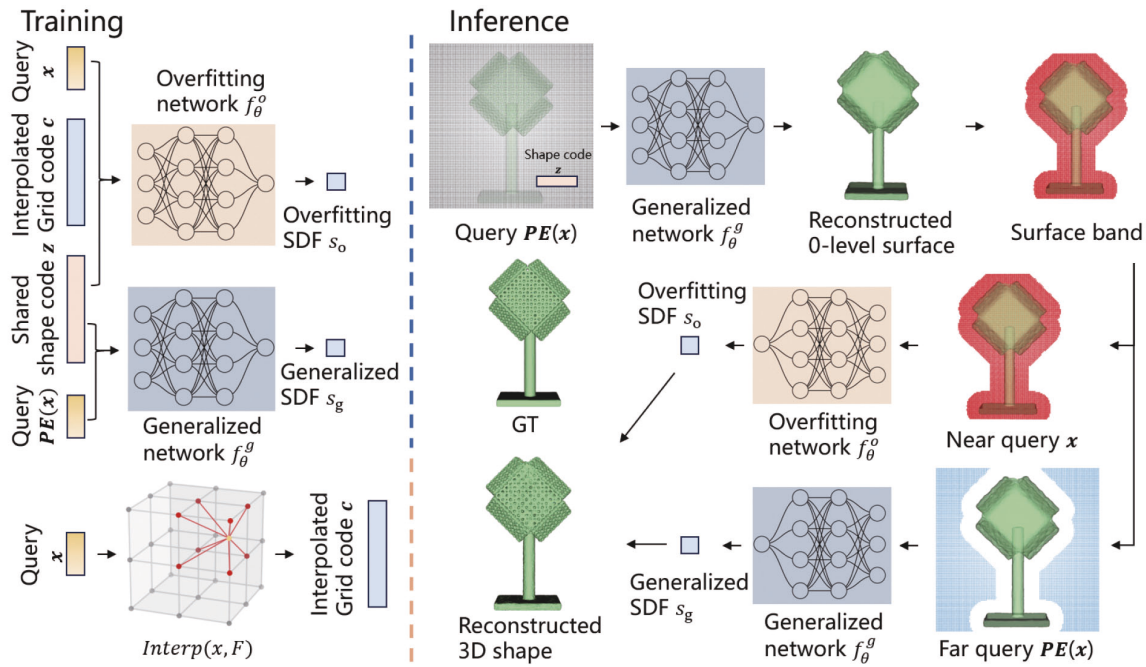


Figure 2: Overview. Our method fully leverages the advantages of both generalization-based and overfitting-based networks, combining their strengths to enhance training efficiency and performance.

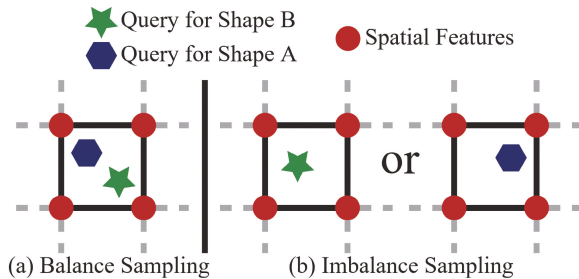


Figure 3: Imbalance query sampling.

predict signed distances in areas that are far away from a surface even we do not sample dense training samples. The second branch works with a spatial feature grid which is shared with different 3D shapes. We first interpolate the feature of a coordinate with the grid, and then maps the coordinate feature and the latent code into a signed distance. We leverage its overfitting of high-frequency 3D signals to preserve fine surface geometry using dense sampling samples.

Our dual branch architecture takes full advantages of the benefits of generalization-based and overfitting-based methods, prompting the strengths of both kinds of methods to improve the training efficiency and performance, aiming to represent multiple SDFs with geometry details using the latent codes. We fuse the signed distance prediction from both branches to reconstruct 3D surfaces using the Marching Cubes algorithm (Lorenzen and Cline 1987a).

Motivation. Although overfitting-based method works well on single objects with a learnable spatial feature grid,

multiple SDFs usually got influence from each other if they share a spatial feature grid with each other. The issue comes from the imbalanced feature queries sampled for different shapes. For example, one shape has much more sampled queries in one voxel in the spatial feature grid than other shapes. This fact makes features in this voxel lean to be more representative on the shape with more queries, resulting in poor detail recovery on some other shapes.

One solution is to ensure that all shapes have balanced queries sampled in each voxel on the grid, as shown in Fig. 3 (a). However, that would be a huge computational burden with a grid of cubic complexity, which is barely possible for training. To resolve this issue, we introduce a novel sampling strategy with a balance constraint. We find a trade off between the complexity and performance, i.e., we only ensure that all shapes have queries sampled in each voxel near their surfaces, which is a dense sampling in a bandwidth, while only sampling sparse points outside the bandwidth. Besides the overfitting branch trained by the dense sampling in the bandwidth, our generalization branch will generalize the prior of signed distances in areas outside the bandwidth, which does not produce artifacts even if we merely sample sparse queries outside the bandwidth. Our networks and sampling can resolve the imbalance issues.

Neural Signed Distance Function. An SDF is a continuous function that maps a given spatial point $\mathbf{x} \in \mathbb{R}^3$ to a scalar $s \in \mathbb{R}$, indicates the distance from the point to the surface and the sign of s determines whether the point lies inside or outside the surface, defined as $s = SDF(\mathbf{x})$. Then, its zero-level set implicitly defines the underlying 3D shape. Our due-branch network predicts signed distances s_g and s_o

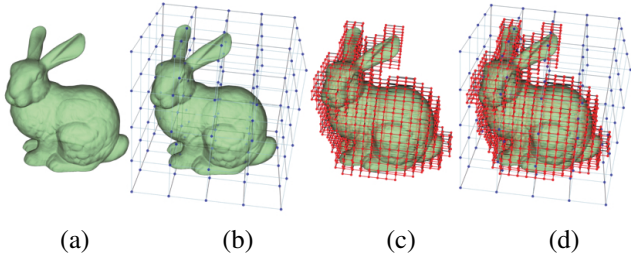


Figure 4: Sampling strategy of our method. Red points show dense samplings of the near surface, and blue points show sparse samplings of the far surface.

at \mathbf{x} , which are fused into one distance s during inference.

Generalized SDF. We employ a generalization-based network $s_g = f_\theta^g(PE(\mathbf{x}), \mathbf{z})$ as a branch to estimate a coarse surface, where PE denotes the positional encoding. The network takes a query point \mathbf{x} and a shape code \mathbf{z} as input, and predicts the signed distance s_g at \mathbf{x} to the surface of shape represented by the latent code \mathbf{z} . We apply positional encoding (Mildenhall et al. 2021) to improve boundary accuracy by introducing high-frequency details.

Overfitting SDF. To capture more high-fidelity geometric details, we employ an overfitting-based network $s_o = f_\theta^o(\mathbf{x}, \mathbf{c}, \mathbf{z})$ with a shared spatial feature grid. The vertices within the grid hold learnable features, which represent the shape characteristics of objects within that region. Given a 3D query point \mathbf{x} , we first interpolate its cube latent code \mathbf{c} from the grid using the trilinear interpolation. Then, the overfitting network takes the concatenation of query point \mathbf{x} , the shared shape code \mathbf{z} , and the interpolated feature \mathbf{c} as input, and predicts the signed distance s_o at \mathbf{x} .

Signed Distance Fusion. For inference, we fuse the SDF predictions s_g and s_o from both branches into a fused signed distance s . We first use the generalization network f_θ^g to reconstruct a coarse 3D shape using the marching cubes. Then, we extract the surface, find the voxels occupied by the surface, and expand voxels by n layers by both sides of the surface, forming a bandwidth \mathbb{B} of the surface. We rerun the marching cubes with a signed distance fusion below,

$$s = \begin{cases} s_o = f_\theta^o(\mathbf{x}, \mathbf{c}, \mathbf{z}), & \mathbf{x} \in \mathbb{B} \\ s_g = f_\theta^g(PE(\mathbf{x}), \mathbf{z}), & \text{others} \end{cases} \quad (1)$$

Sampling with a balance constraint. To work with our SDF representation, we also propose a sampling with a balance constraint to sample training queries for our overfitting-based network.

We use dense near-surface and uniform off-surface sampling in Fig. 4 to boost training efficiency. Specifically, we first divide the entire space at low resolution to obtain sparse uniform sampling, as shown in Fig. 4 (b). Next, we divide the space again at a higher resolution. We then find the voxels occupied by the surface of the object, expand them with n more layers of voxels to determine the bandwidth \mathbb{B} of the surface to obtain dense uniform sampling in bandwidth, as

shown in Fig. 4 (c). Finally, we merge the two sets and remove duplicates to obtain the training samples in Fig. 5(d).

Optimization. For training, we supervise the signed distance predictions s_g and s_o with the ground truth at the same time. We simultaneously optimize the two networks by minimizing a composite loss function that includes the generalized SDF reconstruction loss with shape code regularization, and the overfitting SDF reconstruction loss with grid code regularization. Thus, the overall loss is formulated as, where the SDF losses \mathcal{L}_{gen} and \mathcal{L}_{ovf} are defined as the mean squared error (MSE) between the predicted SDF values \hat{s} and the ground truth s' , measuring the discrepancy between the predicted and ground-truth signed distance values. N is the number of sampled query points.

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{\text{gen}} + \mathcal{L}_{\text{ovf}} + \lambda_1 \mathcal{L}_z + \lambda_2 \mathcal{L}_c \\ &= \frac{1}{N} \sum_{i=0}^N \|\hat{s}_i^g - s'_i\|_2^2 + \frac{1}{N} \sum_{i=0}^N \|\hat{s}_i^o - s'_i\|_2^2 \\ &\quad + \lambda_1 \frac{1}{\sigma_z^2} \|\mathbf{z}\|_2^2 + \lambda_2 \frac{1}{M} \sum_{j=0}^M \frac{1}{\sigma_{c_j}^2} \|\mathbf{c}_j\|_2^2, \end{aligned} \quad (2)$$

\mathcal{L}_z and \mathcal{L}_c are the regularization loss terms applied to the shape latent code \mathbf{z} and each grid latent code \mathbf{c}_j to prevent overfitting and encourage smoothness and compactness, where M is the number of grid code. λ_1 and λ_2 are coefficients to balance the contributions of each term.

Experiments

Implementation details

Dataset. We conduct experiments on three datasets: The Stanford Models (Curless and Levoy 1996), Shapenet-Core.v2 (Chang et al. 2015), and D-FAUST (Bogo et al. 2017). We select six detail-rich models from the Stanford 3D Scanning Repository for the reconstruction of single complex shapes. For ShapenetCore.v2, we select six object categories with the same train/test splits as DeepSDF (Park et al. 2019). On D-FAUST, we select models from three different human subjects, with each subject performing a variety of motion sequences, such as chicken_wings, jiggle_on_toes, and light_hopping_loose. We follow the train/test splits from (Zhang et al. 2023a). We preprocessed each mesh as follows: first, we scaled it to fit the unit cube $[-0.9, 0.9]^3$. Then, we used Manifold (Huang, Su, and Guibas 2018) to make the mesh watertight before sampling.

Metrics. We use Chamfer Distance (CD) \mathcal{L} , F-Score, precision, and recall to provide a comprehensive evaluation of our method. Chamfer distance captures overall accuracy, while F-Score, precision, and recall focus on the quality and completeness of surface predictions.

Setting. We employ a higher learning rate 1.0×10^{-1} to optimize the grid code (a 128^3 grid with 128-dim feature code for each grid), and a lower learning rate 1.0×10^{-3} for the shape code (a 256-dim code) to ensure stable convergence and prevent overfitting to early-stage representations. The bandwidth of $n=3$ defines the near and far surfaces, which are sampled at 512^3 and 128^3 , respectively.

Both networks use 256-dim codes and eight hidden layers of 512 units. The learning rate decays by a factor of 0.5 every 1000 epochs, with a total of 4000 training epochs.

Reconstruction

Single complex objects. We demonstrate the superiority of our method in preserving fine-grained geometric details in surface reconstruction on complex objects, compared to five SOTA approaches, including Instant-NGP (Müller et al. 2022), NGLOD (Takikawa et al. 2021), ACORN (Martel et al. 2021), MosaicSDF (Yariv et al. 2023), and HyperDiffusion (Erkoç et al. 2023). The results are reconstructed by marching cubes (Lorensen and Cline 1987b) with a uniform resolution of 512. As shown in Fig. 5, our method successfully reconstructs Lucy and Happy Buddha with high-fidelity geometry details and sharp edges, such as the hand, torch, skirt folds, petals, and prayer beads. In addition to visual comparison, we sample 500K points from the reconstructed surfaces for quantitative evaluation. As shown in Table 1, our method consistently outperforms existing approaches across all metrics, accurately recovering high-fidelity shape details that other methods fail to capture.

Method	CD(↓)	F-Score(↑)	Precision(↑)	Recall(↑)
ACORN	6.76e-05	0.982	0.967	0.998
NGLOD	6.77e-05	0.980	0.968	0.994
Instant-NGP	7.37E-05	0.976	0.962	0.990
M-SDF	1.30e-03	0.902	0.846	0.972
HyperDiffusion	1.20e-04	0.835	0.847	0.823
Ours	6.56e-05	0.983	0.969	0.998

Table 1: Quantitative comparison on the Stanford models.

Multiple 3D objects. As presented in Table 2, our approach consistently outperforms six state-of-the-art methods across all evaluation metrics on ShapeNet. Furthermore, Fig. 6 illustrates its ability to reconstruct complex lamp geometries without introducing artifacts in empty regions, while faithfully recovering fine details, particularly in objects containing small holes.

Compared to the existing methods, our method leverages spatially-aware grid features combined with global latent codes, enabling it to capture both coarse structure and fine-grained geometric details. Especially, since IF-Net relies on

Method	Bench	Chair	Plane	Table	Lamp	Sofa	Avg.
DeepSDF	4.890	8.630	2.660	6.330	14.63	5.040	7.030
IF-NET	1.340	1.000	0.225	0.857	0.817	1.100	0.890
3DILG	0.872	1.380	0.355	1.280	4.700	1.510	1.683
InsNGP	0.881	1.210	0.664	1.030	1.880	1.100	1.128
M-SDF	2.005	15.92	4.101	6.630	18.43	26.81	12.32
HyperDiff	1.640	1.490	2.310	1.470	6.990	2.830	2.788
Ours	0.463	0.898	0.223	0.790	0.517	0.751	0.607

Table 2: Comparison in multi-objects reconstruction on ShapeNet (Note: $CD \times 10^{-4}$, InsNGP denotes Instant-NGP, and HyperDiff denotes HyperDiffusion).

convolutional architectures that require dense spatial context, when we trained with global query points but tested on partial inputs (e.g., half of the query points), IF-Net struggles to extract meaningful features, and fails to recover complete human bodies.

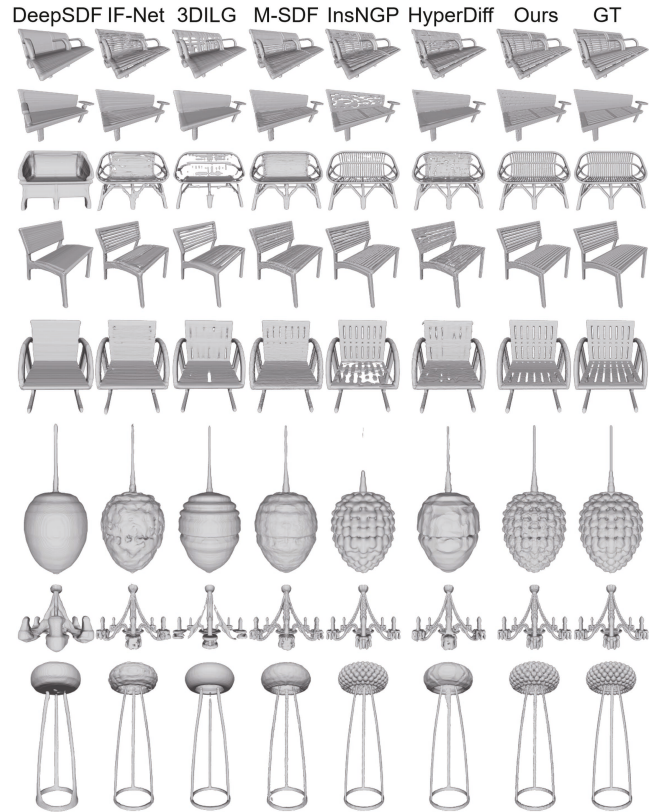


Figure 6: Comparison in multiple object reconstruction on ShapeNet.)

Like Instant-NGP (Müller et al. 2022), our method employs grid-based features to encode local shape information. However, while Instant-NGP demonstrates remarkable performance in single-object reconstruction, its effectiveness diminishes notably in multi-object reconstruction scenarios, even when trained with our sampled query points. This performance degeneration is caused by its limited capacity to distinguish between different shapes within a shared latent space. Specifically, hash collisions and the finite resolution of the hash table undermine the fidelity of feature representations, degenerating geometry details in reconstructions.

Shape completion

We evaluate our method on shape completion using a subset of the DFAUST dataset (Bogo et al. 2017). As shown in Table 3, our method achieves a superior quantitative performance compared to existing approaches, demonstrating its effectiveness in handling incomplete input data. Furthermore, Fig. 7 shows that our method successfully reconstructs partial human models while preserving fine-grained geometric details, such as the sharp structures of the fingers.

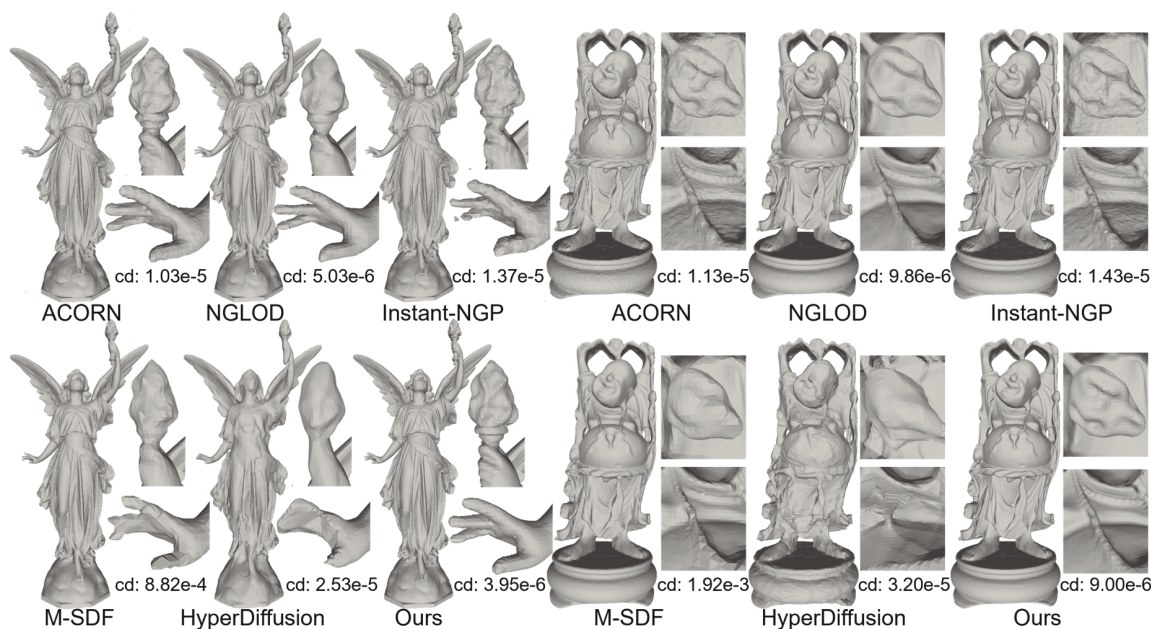


Figure 5: Comparison in single complex object reconstruction on the Stanford models.

Method	50002	50022	50025	Avg.
DeepSDF	3.87	3.39	3.48	3.58
IF-NET	391.79	486.36	544.60	474.25
Instant-NGP	0.342	0.523	0.345	0.403
Ours	0.134	0.124	0.115	0.124

Table 3: Shape completion results on DFAUST ($CD \times 10^{-4}$).

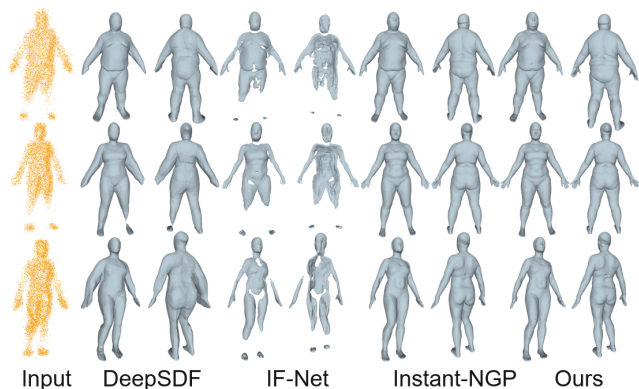


Figure 7: Comparison results in shape completion on DFAUST.

Shape interpolation

We report the results of shape interpolation between the latent codes of two shapes to demonstrate the better continuity and semantic significance of our learned shape embedding, compared to existing methods. The length of latent shape code used in our method, DeepSDF, and Instant-NGP is 256-dimensional, while HyperDiffusion adopts a 3-layer 128-dimensional MLP. As shown in Fig. 8,

compared to DeepSDF (Park et al. 2019) and Instant-NGP (Müller et al. 2022), our method produces more pronounced shape changes when interpolating between shapes. While DeepSDF has smoother transitions, especially in regions like airplane wings, our method allows for more noticeable shape transformations, capturing more complex geometry changes. For HyperDiffusion (Erkoç et al. 2023),

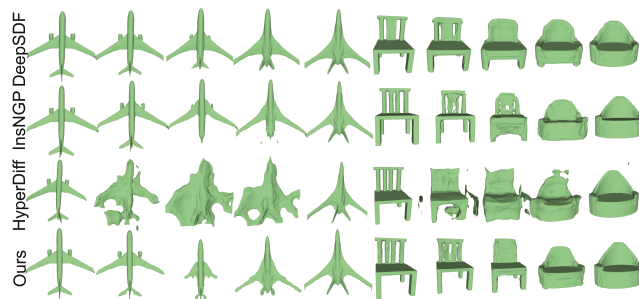


Figure 8: Comparison of shape interpolation on ShapeNet.

interpolating directly in their latent space results in unrealistic geometries or structural artifacts, since each object is represented by a separate MLP and the complexity of the MLP weights leads to nonlinear transitions. Moreover, our method learns more compact latent codes than HyperDiffusion. Each latent code is a one-column vector, and its dimension is flexible to set. While HyperDiffusion uses parameters of a neural network to represent a shape, which is not compact. While DeepSDF and Instant-NGP adopt similar codes, their bias toward low-frequency signals and resulting artifacts hinder accurate representation of multiple neural SDFs with detailed geometry.

Ablation study

Overfitting/Generalization. We report our ablation experiments on 100 randomly selected chair models. To validate the effectiveness of our dual-branch architecture, we separately evaluated the generalization-based and overfitting-based networks using the CD metric in Fig. 9. All experiments use the settings in Sec setting.

The generalization branch captures coarse geometric features to establish a reasonable shape prior, while the overfitting branch focuses on recovering high-frequency details. The reconstructions show that generalization branch can recover a surface with low-frequency geometry while the overfitting branch can recover a sharp surface but have some artifacts caused by imbalanced sampling outside the bandwidth. Eventually, the signed distance fusion can address the issues of reconstruction from both branches.

This demonstrates that combining signed distances from both branches not only removes artifacts of single-branch reconstruction but also captures fine surface geometry.

Bandwidth. We report the impact of bandwidth in Fig. 9. With the trained networks, we try other sizes of bandwidth in signed distance fusion, including 1, 3, and 6. The results show that the artifacts will inherit from the overfitting branch if the bandwidth is too large.

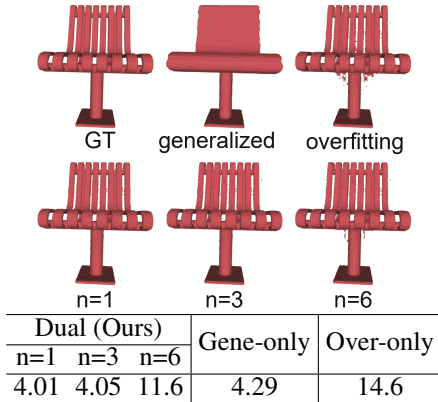


Figure 9: Comparison with single network under different bandwidths n in $CD \times 10^{-4}$.

Shared shape code dimensions. To explore the impact of the shape code dimensions on shape reconstruction fidelity, we compare the reconstruction results using different dimensions ($d = 8, 16, 32, 64, 128$) in Table 4. The performance keeps getting improved with higher dimensions until $d = 64$, and maintains afterward. Note that we only use 100 shapes in this experiment, and we may need higher dimensions for larger datasets.

Sampling strategy. We highlight our sampling by comparing with uniform sampling on grids. We try to use vertices of grids with different resolutions including 32^3 , 64^3 , 128^3 , and 256^3 . We report the numerical comparison with ours in Table 5. We see that the performance can get improved with higher resolutions but is still worse than ours. Our sampling can produce more samples near the surface,



Figure 10: Reconstruction results with different grid resolution sampling res .

which improves the recovery of geometry details, and relatively sparser sampling outside the bandwidth, which reduces the training burden by using much fewer samples and ensures the performance by the generalization of our method.

Dim.	8	16	32	64	128
Ours	0.443	0.433	0.410	0.397	0.409

Table 4: Impact of latent Code Dimensions on reconstruction in ($CD \times 10^{-4}$).

We compare reconstruction results in Fig. 10. At low resolution ($res=32$), the sampling density is insufficient to supervise the 128^3 spatial feature grid, leading to incomplete surfaces and poor reconstruction. As the resolution increases to 64 and 128, more query points fall within each grid cell, resulting in improved reconstructions with smoother surfaces and better preservation of fine details such as holes and sharp edges. In particular, our results show sharper surfaces.

Res.	32^3	64^3	128^3	256^3	Ours
CD	35.900	13.100	0.890	0.417	0.401
Samples	32^3	64^3	128^3	256^3	$\approx 220^3$

Table 5: Different grid resolution on reconstruction in ($CD \times 10^{-4}$).

Conclusion

We propose a method for learning a compact latent space to represent multiple neural SDFs with details. We leverage the strengths of both overfitting-based and generalization-based strategies for SDF representation. The generalization branch effectively captures the signed distance priors in regions far from the surface, even with sparse training samples. In contrast, the overfitting branch accurately reconstructs high-fidelity geometric details and mitigates interference from other shapes via a novel query sampling strategy, which achieves a favorable trade-off between learning efficiency and reconstruction quality. Extensive comparisons show our superiority over SOTA methods.

Acknowledgements

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 62206106).

References

- Bogo, F.; Romero, J.; Pons-Moll, G.; and Black, M. J. 2017. Dynamic FAUST: Registering Human Bodies in Motion. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.
- Chabra, R.; Lenssen, J. E.; Ilg, E.; Schmidt, T.; Straub, J.; Lovegrove, S.; and Newcombe, R. 2020. Deep Local Shapes: Learning Local SDF Priors for Detailed 3D Reconstruction. In *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX*, 608–625. Berlin, Heidelberg: Springer-Verlag. ISBN 978-3-030-58525-9.
- 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. arXiv:1512.03012.
- Chen, Z.; and Zhang, H. 2019. Learning Implicit Fields for Generative Shape Modeling. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5932–5941.
- Chen, Z.; Zhang, Y.; Genova, K.; Funkhouser, T.; Fanello, S.; Bouaziz, S.; Haene, C.; Du, R.; Keskin, C.; and Tang, D. 2021. Multiresolution Deep Implicit Functions for 3D Shape Representation. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV*. IEEE.
- Chibane, J.; Alldieck, T.; and Pons-Moll, G. 2020. Implicit Functions in Feature Space for 3D Shape Reconstruction and Completion. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
- 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 *Proceedings of the European Conference on Computer Vision (ECCV)*.
- Curless, B.; and Levoy, M. 1996. A volumetric method for building complex models from range images. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '96*, 303–312. New York, NY, USA: Association for Computing Machinery. ISBN 0897917464.
- Davies, T. R.; Nowrouzezahrai, D.; and Jacobson, A. 2021. On the Effectiveness of Weight-Encoded Neural Implicit 3D Shapes.
- Duan, Y.; Zhu, H.; Wang, H.; Yi, L.; Nevatia, R.; and Guibas, L. J. 2020. Curriculum deepsdf. In *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, 51–67. Springer.
- Erkoç, Z.; Ma, F.; Shan, Q.; Nießner, M.; and Dai, A. 2023. HyperDiffusion: Generating Implicit Neural Fields with Weight-Space Diffusion. arXiv:2303.17015.
- Erler, P.; Guerrero, P.; Ohrhallinger, S.; Mitra, N. J.; and Wimmer, M. 2020. Points2Surf: Learning Implicit Surfaces from Point Clouds. In Vedaldi, A.; Bischof, H.; Brox, T.; and Frahm, J.-M., eds., *Computer Vision – ECCV 2020*, 108–124. Cham: Springer International Publishing. ISBN 978-3-030-58558-7.
- Genova, K.; Cole, F.; Sud, A.; Sarna, A.; and Funkhouser, T. 2020a. Local Deep Implicit Functions for 3D Shape. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4856–4865.
- Genova, K.; Cole, F.; Sud, A.; Sarna, A.; and Funkhouser, T. 2020b. Local Deep Implicit Functions for 3D Shape. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4857–4866.
- Genova, K.; Cole, F.; Vlastic, D.; Sarna, A.; Freeman, W. T.; and Funkhouser, T. 2019. Learning shape templates with structured implicit functions. In *Proceedings of the IEEE/CVF international conference on computer vision*, 7154–7164.
- Girdhar, R.; Fouhey, D.; Rodriguez, M.; and Gupta, A. 2016. Learning a Predictable and Generative Vector Representation for Objects. In *ECCV*.
- Hertz, A.; Perel, O.; Giryès, R.; Sorkine-Hornung, O.; and Cohen-Or, D. 2021. Sape: Spatially-adaptive progressive encoding for neural optimization. *Advances in Neural Information Processing Systems*, 34: 8820–8832.
- Huang, J.; Su, H.; and Guibas, L. 2018. Robust Watertight Manifold Surface Generation Method for ShapeNet Models. arXiv preprint arXiv:1802.01698.
- Jiang, C.; Sud, A.; Makadia, A.; Huang, J.; Nießner, M.; and Funkhouser, T. 2020a. Local Implicit Grid Representations for 3D Scenes. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6000–6009.
- Jiang, C. M.; Sud, A.; Makadia, A.; Huang, J.; Nießner, M.; and Funkhouser, T. 2020b. Local Implicit Grid Representations for 3D Scenes. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.
- Kleineberg, M.; Fey, M.; and Weichert, F. 2020. Adversarial generation of continuous implicit shape representations. arXiv preprint arXiv:2002.00349.
- Lipman, Y. 2021. Phase Transitions, Distance Functions, and Implicit Neural Representations. arXiv:2106.07689.
- Liu, H.-T. D.; Williams, F.; Jacobson, A.; Fidler, S.; and Litany, O. 2022. Learning Smooth Neural Functions via Lipschitz Regularization. In *ACM SIGGRAPH 2022 Conference Proceedings, SIGGRAPH '22*. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393379.
- Liu, L.; Wang, B.; Xie, H.; Liu, D.; Liu, L.; Tian, Z.; Yang, K.; and Wang, B. 2024. SurroundSDF: Implicit 3D Scene Understanding Based on Signed Distance Field. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 21614–21623.
- Liu, S.; Saito, S.; Chen, W.; and Li, H. 2019. *Learning to infer implicit surfaces without 3D supervision*. Red Hook, NY, USA: Curran Associates Inc.

- Liu, S.-L.; Guo, H.-X.; Pan, H.; Wang, P.-S.; Tong, X.; and Liu, Y. 2021. Deep Implicit Moving Least-Squares Functions for 3D Reconstruction. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1788–1797.
- Lorensen, W. E.; and Cline, H. E. 1987a. Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21(4): 163–169.
- Lorensen, W. E.; and Cline, H. E. 1987b. Marching cubes: A high resolution 3D surface construction algorithm. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '87, 163–169. New York, NY, USA: Association for Computing Machinery. ISBN 0897912276.
- Martel, J. N.; Lindell, D. B.; Lin, C. Z.; Chan, E. R.; Monteiro, M.; and Wetzstein, G. 2021. Acorn: adaptive coordinate networks for neural scene representation. *ACM Transactions on Graphics (TOG)*, 40(4): 1–13.
- Mescheder, L.; Oechsle, M.; Niemeyer, M.; Nowozin, S.; and Geiger, A. 2019. Occupancy Networks: Learning 3D Reconstruction in Function Space. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4455–4465.
- Michalkiewicz, M.; Pontes, J. K.; Jack, D.; Baktashmotlagh, M.; and Eriksson, A. 2019. Deep level sets: Implicit surface representations for 3d shape inference. *arXiv preprint arXiv:1901.06802*.
- Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. NeRF: representing scenes as neural radiance fields for view synthesis. *Commun. ACM*, 65(1): 99–106.
- Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.*, 41(4): 102:1–102:15.
- Museth, K. 2013. VDB: High-resolution sparse volumes with dynamic topology. *ACM transactions on graphics (TOG)*, 32(3): 1–22.
- Park, J. J.; Florence, P.; Straub, J.; Newcombe, R.; and Lovegrove, S. 2019. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Peng, S.; Niemeyer, M.; Mescheder, L.; Pollefeys, M.; and Geiger, A. 2020. Convolutional Occupancy Networks. In *European Conference on Computer Vision (ECCV)*.
- Remelli, E.; Lukoianov, A.; Richter, S.; Guillard, B.; Bagautdinov, T.; Baque, P.; and Fua, P. 2020. MeshSDF: Differentiable Iso-Surface Extraction. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M. F.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 22468–22478. Curran Associates, Inc.
- Saragadam, V.; Tan, J.; Balakrishnan, G.; Baraniuk, R. G.; and Veeraraghavan, A. 2022. MINER: Multiscale Implicit Neural Representations. In *European Conf. Computer Vision (ECCV)*.
- Setaluri, R.; Aanjaneya, M.; Bauer, S.; and Sifakis, E. 2014. SPGrid: A sparse paged grid structure applied to adaptive smoke simulation. *ACM Transactions on Graphics (TOG)*, 33(6): 1–12.
- Sitzmann, V.; Martel, J.; Bergman, A.; Lindell, D.; and Wetzstein, G. 2020. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33: 7462–7473.
- 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 *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11353–11362.
- Tancik, M.; Srinivasan, P.; Mildenhall, B.; Fridovich-Keil, S.; Raghavan, N.; Singhal, U.; Ramamoorthi, R.; Barron, J.; and Ng, R. 2020. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33: 7537–7547.
- Tatarchenko, M.; Dosovitskiy, A.; and Brox, T. 2017. Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3D Outputs. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2107–2115.
- Wang, H.; Schor, N.; Hu, R.; Huang, H.; Cohen-Or, D.; and Huang, H. 2018. Global-to-local generative model for 3d shapes. *ACM Transactions on Graphics (TOG)*, 37(6): 1–10.
- Wang, P.-S.; Liu, Y.; Guo, Y.-X.; Sun, C.-Y.; and Tong, X. 2017. O-cnn: Octree-based convolutional neural networks for 3d shape analysis. *ACM Transactions On Graphics (TOG)*, 36(4): 1–11.
- Xie, Y.; Takikawa, T.; Saito, S.; Litany, O.; Yan, S.; Khan, N.; Tombari, F.; Tompkin, J.; Sitzmann, V.; and Sridhar, S. 2022. Neural fields in visual computing and beyond. In *Computer Graphics Forum*, volume 41, 641–676. Wiley Online Library.
- Yao, S.; Yang, F.; Cheng, Y.; and Mozerov, M. G. 2021. 3D Shapes Local Geometry Codes Learning with SDF. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2110–2117.
- Yariv, L.; Puny, O.; Neverova, N.; Gafni, O.; and Lipman, Y. 2023. Mosaic-SDF for 3D Generative Models. *arXiv*.
- Zhang, B.; Li, J.; Deng, X.; Zhang, Y.; Ma, C.; and Wang, H. 2023a. Self-supervised Learning of Implicit Shape Representation with Dense Correspondence for Deformable Objects. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 14222–14232.
- Zhang, B.; Nießner, M.; and Wonka, P. 2022. 3DILG: Irregular Latent Grids for 3D Generative Modeling. In Oh, A. H.; Agarwal, A.; Belgrave, D.; and Cho, K., eds., *Advances in Neural Information Processing Systems*.
- Zhang, B.; Tang, J.; Niessner, M.; and Wonka, P. 2023b. 3dshape2vecset: A 3d shape representation for neural fields and generative diffusion models. *ACM Transactions On Graphics (TOG)*, 42(4): 1–16.