

MetricGrids: Arbitrary Nonlinear Approximation with Elementary Metric Grids based Implicit Neural Representation

Shu Wang^{1*} Yanbo Gao^{1*} Shuai Li^{1†} Chong Lv¹ Xun Cai¹
 Chuankun Li² Hui Yuan¹ Jinglin Zhang¹
¹Shandong University ²North University of China

Abstract

This paper presents MetricGrids, a novel grid-based neural representation that combines elementary metric grids in various metric spaces to approximate complex nonlinear signals. While grid-based representations are widely adopted for their efficiency and scalability, the existing feature grids with linear indexing for continuous-space points can only provide degenerate linear latent space representations, and such representations cannot be adequately compensated to represent complex nonlinear signals by the following compact decoder. To address this problem while keeping the simplicity of a regular grid structure, our approach builds upon the standard grid-based paradigm by constructing multiple elementary metric grids as high-order terms to approximate complex nonlinearities, following the Taylor expansion principle. Furthermore, we enhance model compactness with hash encoding based on different sparsities of the grids to prevent detrimental hash collisions, and a high-order extrapolation decoder to reduce explicit grid storage requirements. experimental results on both 2D and 3D reconstructions demonstrate the superior fitting and rendering accuracy of the proposed method across diverse signal types, validating its robustness and generalizability. Code is available at <https://github.com/wangshu31/MetricGrids>.

1. Introduction

Implicit neural representations (INRs) [49] have emerged as a novel paradigm for signal representation, demonstrating superior performance in reconstructing various signals, including 2D images, 3D videos, 3D shapes, and radiance fields [9–11, 34, 48]. INRs leverage neural networks to establish a mapping between continuous coordinates and target outputs by employing regularly sampled signal values as training data and optimizing the network parameters.

*Equal contributions.

†Corresponding author.

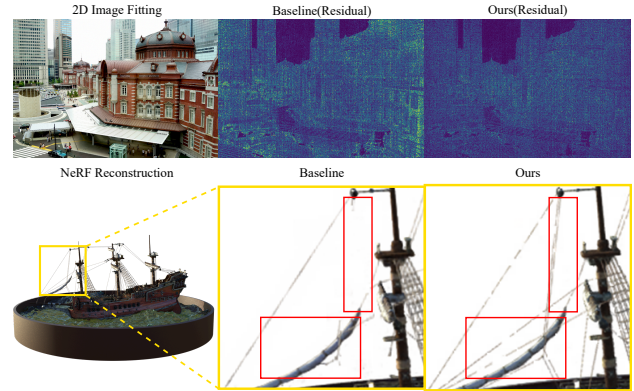


Figure 1. Illustration of the nonlinear fitting capabilities for various type of signals using the existing INR method and our MetricGrids.

The original INRs utilize a fully connected network with parameters shared across the entire input domain, resulting in a compact model [5, 36]. However, its expressive ability is constrained by spectral bias [3, 38, 53] and the capacity of the network [20], making it challenging to fit high-frequency and complex signals. To address this limitation, hybrid representations have been proposed [47, 51, 52], which store latent space features at discrete vertex coordinates and use a decoder for reconstruction. Features are typically stored in grid-based structures, such as feature maps for 2D signals and feature volumes for 3D signals [21, 28]. The input coordinates can be used to directly retrieve the corresponding features through indexing rules.

In order to represent complex details, hybrid representations need to establish a dense feature grid in the entire signal input domain to ensure that the feature at continuous-space points can be obtained with high accuracy based on the stored points. Especially considering that different signals show different complexity and sparsity among the space, a regular grid-based representation often leads to large storage and redundancy. Various methods have been proposed for efficient grid structure. Early methods focused on modifying grid structures through heuristic modifications such as pruning and merging to create signal-

specific data structures [32, 52, 61]. More recent advancements have simplified grid structures to reduce storage demands [7, 8, 13]. There are also methods combining the grid with different representations, such as NeuRBF [12] using both adaptive radial bases and grids. However, these methods still require dense features in the complex signal regions, and enhancing the representation accuracy of grid-based methods remains challenging. Especially how to effectively approximate a continuous-space nonlinear signal with a discrete grid representation has not been thoroughly investigated yet.

This paper first investigates the mechanism of using grid structure for signal approximation, and shows that the regular grid relying on linear interpolation implicitly constrains the latent feature space to a linear form, which is a degenerate case of the nonlinear signal. Consequently, for complex nonlinear signals, the features at continuous-space points that are not stored in the grid cannot be accurately represented with the grid, leading to representation errors as shown in Fig. 1. To solve this problem, we propose MetricGrids, which leverages the concept of Taylor expansion by constructing multiple elementary metric grids in various metric spaces. Each elementary metric grid serves as a high-order approximation of the latent feature space. Subsequently, a compact representation of the elementary metric grids is developed to explore the sparsity of high-order derivatives, thereby maintaining the model compactness. Furthermore, to reduce the number of required grids while achieving complex nonlinear transformations in the latent space, a high-order extrapolation decoder is used to progressively generate higher-order terms based on the learned metric grids.

The contributions of this paper can be summarized as follows.

- We propose a novel method, MetricGrids, to enhance the grid-based neural representations. It approximates complex nonlinearities of latent feature spaces using elementary metric grids in the form of Taylor expansion.
- A hash encoding based compact representation of the elementary metric grids is provided to improve parameter efficiency, which explores the sparsity of the high-order derivatives.
- We propose a high-order extrapolation decoder to predict the high-order terms based on the low-order approximation obtained from the learned elementary metric grids.
- Extensive experiments on multiple fundamental INR tasks have been conducted and the proposed MetricGrids achieves state-of-the-art performance while keeping the model compactness.

2. Related Work

Implicit Neural Representations. Implicit neural representations (also called neural fields) use neural networks to

represent signals such as the 3D shape [11, 33, 36]. INRs typically take 3D coordinates as input and produce signed distance function (SDF) or occupancy values. In [49], Sitzmann *et al.* introduced SIREN, extending implicit neural representations to 2D image representation. Later Mildenhall *et al.* proposed Neural Radiance Fields (NeRF) [34] to render a complete 3D scene with INR and volumetric rendering. As the complexity of represented signals increases, many methods have been proposed to improve the representation accuracy, including using different activation functions [29, 39, 43], altering the positional encoding basis function [53, 55], or progressively increasing the positional encoding frequency [19]. Different neural network architectures have also been developed. In [16], Multiplicative Filter Networks (MFN) was used to represent the reconstructed signal by multiplying together sinusoidal or Gabor wavelet functions applied to the input. Based on this, BA-CON [27], PNF [59], and RMFN [44] were proposed by carefully controlling the bandwidth of each filter layer to achieve multi-scale signal reconstruction. Fourier reparameterization [45] and batch normalization [4] were further proposed to improve the weight composition and learning process, in order to mitigate the spectral bias of networks. A comprehensive explanation of the fully implicit representation approach from the perspective of structured signal dictionaries was presented in [62], establishing a connection between network parameters and frequency support.

Hybrid Representations. Another approach to address spectral bias is to localize certain network parameters by utilizing a feature grid to capture and store local features of specific signal coordinates. Since these features are directly related to the spatial location of the signal, this approach is also referred to as hybrid or grid-based representation [21, 37, 51]. Such methods acquire features of any continuous point through indexing the grid with linear interpolation and use them as the latent vector for MLP decoder. However, its expressiveness is closely tied to grid size, often resulting in high memory consumption. Various techniques have been developed to reduce the grid size, including multi-resolution orthogonal planes [6, 17, 46, 65], sparse octree structures [60, 61], hash encoding [35], tensor decomposition [7] and wavelet transform [41], etc. Despite these advancements, most methods only focus on reducing grid size and lack the study on improving the fitting accuracy.

There are also methods investigating using different types of local representations, instead of a regular grid. Some methods employ MLPs to learn local structures. Kilo-NeRF [40] proposed to directly replace the feature vectors stored in the grid with local MLPs. In [42] and [18], a pyramid structure and levels of experts were proposed to represent different scales of the local features, respectively. Different from using MLPs, 3DGS [22] employed explicitly

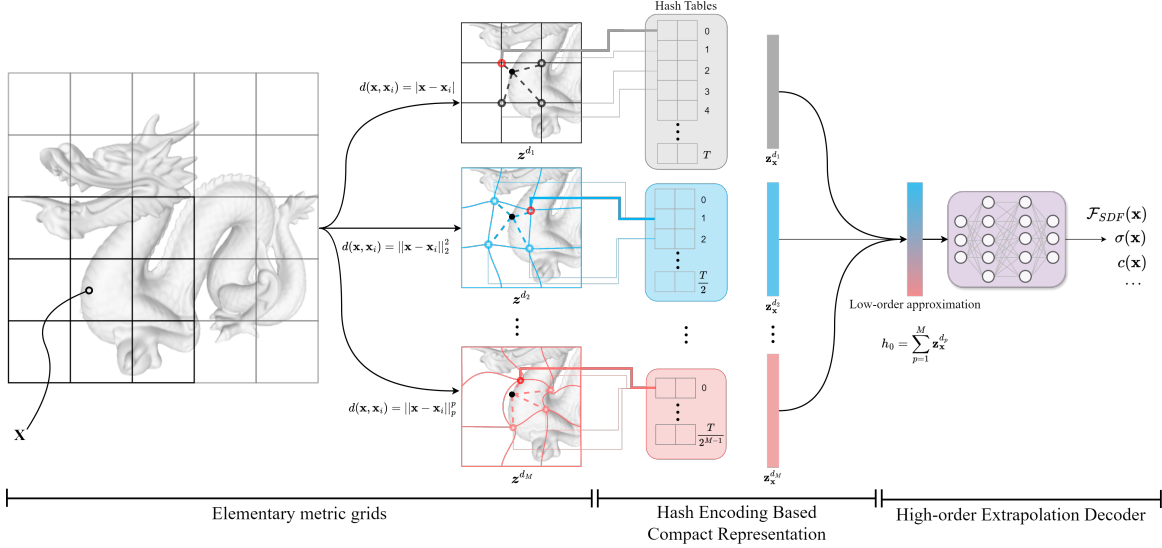


Figure 2. Illustration of the proposed **MetricGrids**. Given an input coordinate $\mathbf{x}$, features are indexed in multiple metric grids, representing the different-order approximations as Taylor expansion. Hash encoding is used based on the sparsity of each metric grid to form a compact representation. A high-order extrapolation decoder is used to further estimate the remaining high-order terms not stored in the metric grids, thereby enhancing the nonlinear fitting capability.

learned Gaussian primitives at different locations for representation, which has emerged as a pioneering paradigm in the field of radiance fields and achieves remarkable outcomes. NeuRBF [12] also proposed to use adaptive radial basis function based representation to enhance the grid. On the other hand, there are also methods of relaxing the grid structures by utilizing point clouds [26, 30, 58, 64] to store scene features. In contrast to the above methods, this paper aims to improve the standard grid-based paradigm in terms of fitting accuracy, and develops a novel MetricGrids framework inspired by the Taylor expansion to improve the nonlinear approximation accuracy.

3. Proposed Method

Our goal is to represent a continuous signal with a discrete grid-based neural representation. Naturally, it is an approximation problem and in this paper, we propose MetricGrids, by combining a series of elementary metric grids to solve this problem.

3.1. Formulation of MetricGrids

Given a signal $\mathbf{y}$ with a set of observations, an implicit neural representation is defined as a mapping from input to output, parameterized using a neural network: $\hat{\mathbf{y}} = f_{\Theta}(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}^{out}$, where $\mathbf{x} \in \mathbb{R}^d$ and $\hat{\mathbf{y}} \in \mathbb{R}^{out}$ denote the input (typically coordinates) and reconstructed output, respectively, Θ denotes the learnable parameters. The objective is to reconstruct the signal, *i.e.*, minimize the distance $\|\mathbf{y} - \hat{\mathbf{y}}\|$ [31, 57]. For fully implicit representation, a neural network like MLP is used to map the coordinates directly to reconstruct the signal, usually exhibiting a strong

spectral bias [3, 38, 53] towards the low frequency.

The grid-based representation, also known as the hybrid representation, mitigates this problem by localizing most of the parameters as a feature grid. It can be expressed as a two-stage model:

$$\begin{aligned} \mathbf{z}_{\mathbf{x}} &= g(\mathbf{x}, \mathcal{Z}), \\ \hat{\mathbf{y}} &= f_d(\mathbf{z}_{\mathbf{x}}, \theta_d), \end{aligned} \quad (1)$$

where learnable parameters contain a feature grid and a decoder $\Theta = \{\mathcal{Z}, \theta_d\}$. The grid $\mathcal{Z} \in \mathbb{R}^{N \times C_h}$ is defined on the discrete coordinates of the input domain consisting of learned features on the vertex. $g(\cdot)$ is a distance-dependent (typically $L1$ distance $\|\mathbf{x} - \mathbf{x}_i\|$) indexing function to interpolate the feature at any point $\mathbf{x}$ in the continuous signal. The decoder f_d typically utilizes a compact MLP, similar to fully implicit methods. The grid employs an explicit geometric data structure to store local information of the signal, thereby avoiding the spectral bias problem. It also accelerates the inference, since the indexing function usually does not require network inference [15, 51, 56].

From the perspective of signal representation [54], the grid can be considered as a discrete feature representation of a continuous signal. The indexing process is to generate the feature at any continuous location based on the discrete grid, and the linear interpolation based generation indicates a linear approximation of the signal. However, natural scenes are usually complex and highly nonlinear, and thus correspond to nonlinear feature spaces. The conventional grid based paradigm only employs a feature grid constructed in a degenerate space that is a linear approximation of the scene. In regions with complex signals, this degener-

ate representation presents challenges, and cannot achieve good performance as shown in Fig. 1.

To solve the above problem, the proposed MetricGrids uses a series of elementary metric grids to generate complex nonlinear representations. In view that any point of a continuous function can be approximated at its neighboring point with a polynomial containing different order terms using Taylor expansion, any point in a grid can also be approximated by a series of elementary metric grids. Each elementary metric grid contains the derivative-related information as the different order terms in Taylor expansion and the distance between points are calculated with metrics of different orders. By learning such an elementary metric grid, it can provide a form of $\frac{z^{(n)}(x_i)}{n!}(x - x_i)^n$, where x and x_i represent a point and its neighboring vertex point stored in a grid, respectively, $z^{(n)}(x_i)$ represent the n -th order derivative at the vertex point x_i and $(x - x_i)^n$ can be conveniently generated by a n -th order metric defined on the grid. For high-dimensional space, it can be similarly deduced by approximating the sum of polynomials in each direction with the defined metric $\sum_j \frac{z^{(n)}(x_i)}{n!}(x_j - x_{ij})^n \approx C_n \|\mathbf{x} - \mathbf{x}_i\|^n$ since the n -th order metric accounts for the distances of varying orders in all dimensions j .

By summarizing the elementary metric grids, the proposed MetricGrids can provide accurate features for continuous-space that are not directly stored in the grid. In order to reduce the storage of multiple grids, a hash encoding based compact representation is used to fully exploit the sparsity of high-order terms. Moreover, to further reduce the number of required elementary metric grids, the high-order grids are no longer explicitly stored and instead a high-order extrapolation decoder is designed to approximate them based on the stored lower-order grids.

In the following, Sec. 3.2 introduces the elementary metric-defined derivative grids, Sec. 3.3 describes the hash encoding based compact representation, and Sec. 3.4 explains the high-order extrapolation based decoding. Fig. 2 illustrates the proposed framework.

3.2. Elementary Metric Grids

To obtain a nonlinear approximation of a point in the scene, multiple elementary grids are used in different metric spaces, resulting in a series of elementary metric grids. Specifically, the distances in various nonlinear metric spaces are used as proxies of the polynomial basis $(\mathbf{x} - \mathbf{x}_i)^n$, while the grids are learned to store different order derivatives related features.

Accordingly, a series of elementary metric grids are constructed with different metrics as

$$\mathcal{Z}^{metrics} = [\mathcal{Z}^{d_1}, \mathcal{Z}^{d_2}, \dots, \mathcal{Z}^{d_M}] \in \mathbb{R}^{M \times N \times C_h}, \quad (2)$$

where $\mathcal{Z}^{d_p}, d_p(\mathbf{x}, \mathbf{x}_i) = \|\mathbf{x} - \mathbf{x}_i\|_p^p, \forall \mathbf{x}, \mathbf{x}_i \in \mathbb{R}^d$,

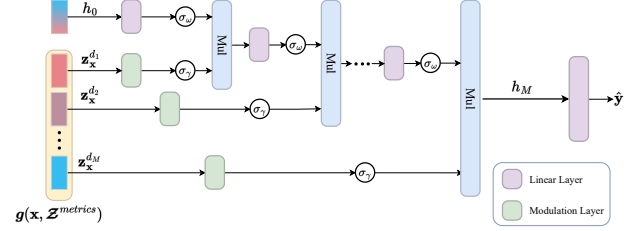


Figure 3. Illustration of the high-order extrapolation decoder. The ‘Mul’ represents the Hadamard product.

where $\mathcal{Z}^{d_p}$ represents a grid structure with discrete vertex coordinate, $d_p(\cdot)$ represents the metric defined on the grid. To align with the Taylor expansion, metrics that are the exponential form of the p -norm (to the power of p) can be used.

For a point $\mathbf{x}$, our method indexes the metric grids and produces a feature based on the vertex features stored in the grids and distance under diverse metrics as:

$$g(\mathbf{x}, \mathcal{Z}^{metrics}) = [\mathbf{z}_x^{d_1}, \mathbf{z}_x^{d_2}, \dots, \mathbf{z}_x^{d_M}], \quad (3)$$

where $\mathbf{z}_x^{d_p} = g(\mathbf{x}, \mathcal{Z}^{d_p}; d_p(\mathbf{x}, \mathbf{x}_i))$.

With the learning-based grid generation, a high order derivative term $\frac{z^{(n)}(x_i)}{n!}(x - x_i)^n$ in the Taylor expansion can be directly generated with the grids learning the different derivatives induced features. As a special case, the feature $\mathbf{z}_x^{d_1}$ extracted from the first elementary metric grid aligns with the existing grid-based model paradigm by providing a linear approximation of the latent space. In contrast, $\mathbf{z}_x^{d_m}$ ($m \geq 2$) represents high-order derivative induced features to provide a nonlinear approximation.

Together, these elementary grids enable latent space features to approximate arbitrary nonlinearities across different locations, thereby facilitating the representation of complex signals. In addition to mimicking the Taylor expansion, different types of metric grids can also be defined as long as the metrics provide a nonlinear distance measure. In such a case, it can also provide a nonlinear approximation of any point within a regularly stored feature grid.

3.3. Hash Encoding Based Compact Representation

In this section, a compact representation of the elementary metric grids is proposed to improve the efficiency of storing and learning multiple grids. Motivated by Instant-NGP [35], for each vertex coordinate $\mathbf{x}_i$, multiple feature vectors from the elementary metric grids, representing the linear approximation and high-order terms of the latent space need to be stored. Simply taking the elementary metric grids as an extra dimension in addition to the spatial dimensions and creating hash encoding together can cause hash collisions not only in space but also in different elementary metric grids. Especially considering different elementary metric grids containing features of different



Figure 4. Qualitative Comparison on Kodak dataset. Zoomed-in views of windows and roofs illustrate the ability of the baseline model and our proposed MetricGrids in fitting complex details.

natures, such collision among elementary metric grids may significantly reduce the representation accuracy. Therefore, the elementary metric grids are encoded separately to avoid collision among grids.

On the other hand, with the elementary metric grids representing the information of linear approximation, and different-order derivatives, the sparsity of the grids is different. In smooth regions of the signal, the different-order derivatives are all zero, making them much sparser than the linear first-order approximation. Therefore, the elementary metric grids are encoded differently by setting a shorter hash table length T for higher-order metric grids to encourage the fusion of similar features in high-order items. Furthermore, with the nonlinear representation using the elementary metric grids, the high-resolution linear grids are no longer necessary for fitting complex signals. Consequently, the maximum resolution of the grids is also reduced to minimize the number of parameters used in the final representation. For a vertex coordinate $\mathbf{x}_i$, the hash mapping at each level can be expressed as:

$$H(\mathbf{x}_i) = [h^{d_1}(\mathbf{x}_i, T), \dots, h^{d_M}(\mathbf{x}_i, \frac{T}{2^{M-1}})],$$

$$\text{where } h^{d_m}(\mathbf{x}, T) = \left(\bigoplus_{j=1}^d x_j \pi_j \right) \bmod T, \quad (4)$$

where $\oplus$ denotes the bitwise XOR operation and π_i are unique, large prime numbers, as in Instant NGP [35]. With the above hash encoding based compression of the elementary metric grids, a compact representation, of the same size as the conventional grid, can be obtained while multiple-order features are extracted to represent each point for better nonlinearity representation.

3.4. High-order Extrapolation Decoder

To reduce the number of learned grids and improve the representation efficiency of the elementary metric grids, only a few grids are explicitly learned in the experiments, three elementary metric grids to be specific. In order to further reduce the approximation error due to the limited number of grids, a high-order extrapolation decoder is developed to generate high-order terms based on the linear and relatively lower-order terms stored in the grids, as shown in Fig. 3.

First, the linear and lower-order feature of point $\mathbf{x}$ $g(\mathbf{x}, \mathcal{Z}^{metrics}) = [\mathbf{z}_x^{d_1}, \mathbf{z}_x^{d_2}, \dots, \mathbf{z}_x^{d_{M-1}}]$ which indexed from the grid are extracted and summarized as the initial approximation of the nonlinear feature. Then a hierarchical extrapolation decoder is developed to estimate the higher-order terms by leveraging the initial approximation and different-order features. Specifically, the features are first modulated with a linear layer and then multiplied with the Hadamard product [14, 15, 24, 27] to generate a higher-order term, and then progressively updated the approximation. For the output features of the decoder layer ℓ , this process can be expressed as follows:

$$h_{\ell+1} = \omega_{\ell}(h_{\ell}) \circ \gamma_{\ell}(\mathbf{x}), \quad (5)$$

$$\omega_{\ell}(h_{\ell}) = \sigma_{\omega}(\mathbf{W}_{\ell} h_{\ell-1} + \mathbf{b}_{\ell}), \quad (6)$$

$$\gamma_{\ell}(\mathbf{x}) = \sigma_{\gamma}(\mathbf{W}_{\ell}^{\gamma} \mathbf{z}_x^{d_{\ell-1}} + \mathbf{b}_{\ell}^{\gamma}), \quad (7)$$

where $h_{\ell} \in \mathbb{R}^N$ represents the gradually updated approximation, $\omega_{\ell}(\cdot)$ and $\gamma_{\ell}(\cdot)$ are backbone layers and modulation layers with weight $\mathbf{W}$ and bias $\mathbf{b}$ that adjust the current approximation and higher-order terms. σ_{ω} and σ_{γ} represent nonlinear activation functions, $\sin(\cdot)$ and $\sin^2(\cdot)$ are used in the experiments, as the sine activation function has been demonstrated to outperform common activation functions (e.g. ReLU) in INR tasks [49].

Without considering the activation function, the theoretical highest-order term achievable after M decoder layers is $m + 1 + 2 + \dots + m = m + \frac{m(m+1)}{2} = \frac{m(m+3)}{2}$, due to the repeated multiplication operations within each layer. Finally, the reconstructed features are converted to the signal domain through a linear layer:

$$\hat{\mathbf{y}} = \mathbf{W}_o h_M + \mathbf{b}_o. \quad (8)$$

This method hierarchically integrates features from each derivative space and extrapolates higher-order terms based on the approximation and lower-order terms, allowing for more complex nonlinear transformations in the latent space.

4. Experiments

4.1. Implementation Details

The proposed MetricGrids framework are evaluated across various Implicit Neural Representation tasks, including 2D

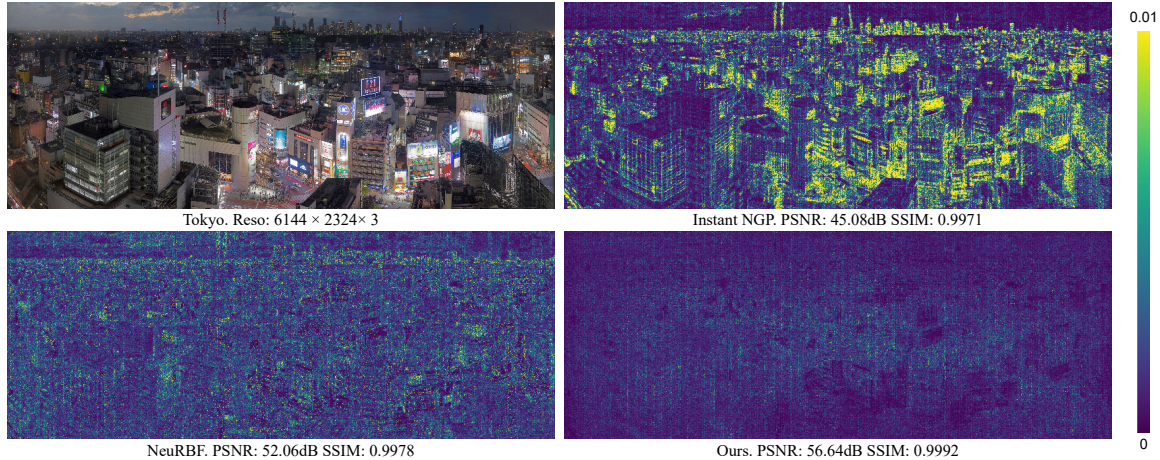


Figure 5. Qualitative Comparison on Gigapixel Image. The upper-right shows the ground truth image, containing numerous intricate details. The upper right, lower-left and lower-right images show L2 error maps for the baselines and our method, respectively. All experiments use consistent hyperparameter settings.

image fitting, 3D signed distance function fitting, and 5D neural radiance field reconstruction for novel view synthesis. In the elementary metric grids implementation, three grids are used with the first grid using linear metric as a regular metric while the other two grids with nonlinear metrics. In addition to the nonlinear metrics $[x^2, x^3]$, the other nonlinear metrics including $[\sin(\cdot), \arcsin(\cdot)]$ and $[\sin(\cdot), \cos(\cdot)]$ [34] are also evaluated for 2D image and 3D SDF fitting, and for neural radiance field reconstruction, respectively. Such nonlinear metrics can actually provide more nonlinear properties in experiments, since they can be further approximated with the polynomial terms. The high-order extrapolation decoder architecture comprised five layers with 64 hidden units, with the number of layers directly corresponding to the established Elementary Metric Grids. More details can be found in supplementary. The training was conducted on a single NVIDIA RTX 3090 GPU using the Adam [23] optimizer ($\beta_1 = 0.9, \beta_2 = 0.99, \epsilon = 10^{-15}$) with cosine annealing for learning rate decay. L2 loss was employed for 2D image fitting and NeRF reconstruction, while MAPE loss as in [35] was utilized for 3D SDF reconstruction. The same point sampling strategy as in the Instant NGP [35] and NeuRBF [12] is adopted for SDF fitting experiments.

4.2. 2D Image Fitting

We first evaluate the effectiveness of fitting 2D images using the Kodak dataset [1], which consists of 24 images with a resolution of 768×512 . PSNR(dB) and SSIM are used to assess the reconstructed image quality. The performance is first compared with the existing methods with a similar model size (around 207K) and the same training iterations (20,000 steps), the results are shown in Tab. 1. It can be seen that the proposed MetricGrids achieves the best performance compared with other methods. Moreover, it can

Method	Rep.	PSNR $\uparrow$	SSIM $\uparrow$	Params(K)
WIRE [43]	Implicit	37.59	0.9596	373
SIREN [49]	Implicit	38.70	0.9512	207
SCONE [24]	Implicit	40.29	0.9690	207
I-NGP [35]	Hybrid	37.06	0.9386	206
NFFB [56]	Hybrid	38.77	0.9537	208
NeuRBF [12]	Hybrid	38.70	0.9488	207
Ours(p-norm)	Hybrid	40.63	0.9720	207
Ours	Hybrid	41.57	0.9735	207

Table 1. Result comparisons of the proposed method against the existing methods under a similar model size, on 2D Image Fitting using Kodak Dataset. **Best** and **second best** results are highlighted.

be seen that while MetricGrids with the p -norm based nonlinear metrics $[x^2, x^3]$ improves the results, the highly nonlinear metrics $[\sin(\cdot), \arcsin(\cdot)]$ perform even better. It is mostly because such nonlinear metrics themselves can be considered as a combination of different-order terms under Taylor expansion, thus help the grid to learn nonlinear approximation. Fig. 4 further shows some visual examples.

Comparison with recent work GaussianImage [63] using Gaussian primitives to represent images and other methods with different numbers of parameters are also performed. The results are shown in Tab. 2, metrics of the comparison methods are taken from GaussianImage [63]. It can be seen that the proposed method with a medium size (m) and large size (l) both significantly improve the performance over GaussianImage [63], especially with the similar number of parameters.

A notable advantage of the grid-based model over the fully implicit representation is the capacity to accommodate gigapixel images [32]. To evaluate this scalability, our method is tested on the publicly available gigapixel images, and compared with InstantNGP [35] and NeuRBF [12] as representative state-of-the-art baselines, using the same

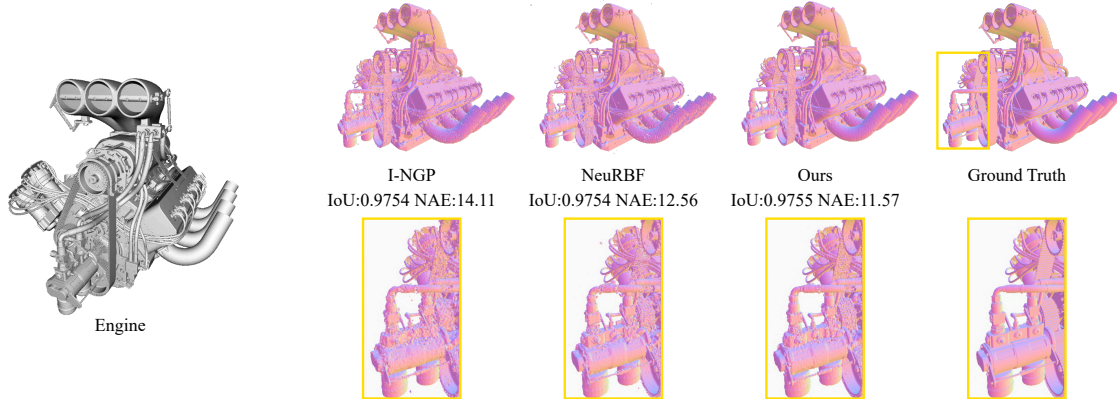


Figure 6. Qualitative Comparison on 3D Signed Distance Field Reconstruction. Details are best viewed with zoom-in.

Method	Rep.	PSNR $\uparrow$	MS-SSIM $\uparrow$	Params(K)
I-NGP [35]	Hybrid	43.88	0.9976	300
NeuRBF [12]	Hybrid	43.78	0.9964	337
3D GS [22]	Explicit	43.69	0.9991	3540
GaussianImage [63]	Explicit	44.08	0.9985	560
Ours-m	Hybrid	46.97	0.9986	335
Ours-l	Hybrid	52.22	0.9987	533

Table 2. Result comparison with Gaussian primitive based methods, on 2D Image Fitting using Kodak Dataset.

Method	IOU $\uparrow$	NAE $\downarrow$	Chamfer $\downarrow$	Params
NGLOD5 [52]	0.99991	4.21	0.0022	10.15M
I-NGP [35]	0.99991	3.42	0.0023	950K
NeuRBF [12]	0.99988	3.20	0.0021	856K
Ours	0.99991	3.09	0.0021	905K

Table 3. Result comparison on 3D Signed Distance Function Reconstruction using Stanford 3D Dataset.

number of parameters. The results are shown in Fig. 5, with an illustration of the error map. It can be seen that our method demonstrates superior fidelity in gigapixel image reconstruction compared to InstantNGP [35] and NeuRBF [12]. Moreover, the error distribution of the proposed method is relatively uniform across the whole image, indicating that the proposed method can efficiently fit complex regions such as edges with the higher-order grids. Additional results for other gigapixel images are provided in the supplementary material.

4.3. 3D Shape Reconstruction

To assess the quality of 3D shape representation, 7 scenes from the Stanford 3D Scanning Repository [50] are evaluated, following the hyperparameter configurations established in NeuRBF [12]. The evaluation employed multiple metrics: Intersection over Union (IoU), Normal Angular Error (NAE), and Chamfer Distance (CD). The results are demonstrated in Tab. 3, and comparative analysis against baseline and state-of-the-art methods reveals that the proposed approach achieves superior performance across all

Methods	Steps	Params	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS _{VGG} $\downarrow$
Mip-NeRF 360 [2]	250k	3.23M	33.25	0.962	0.039
TensoRF [7]	30k	17.95M	33.14	0.963	0.047
K-Planes [17]	30k	33M	32.36	0.962	0.048
I-NGP [35]	35k	12.21M	33.18	0.963	0.051
NerfAcc [25]	20k	12.21M	32.55	-	0.056
NeuRBF [12]	30k	17.74M	34.62	0.975	0.034
Ours	30k	13.11M	35.02	0.976	0.033

Table 4. Result comparison on Neural Radiance Field Reconstruction using Blender Dataset [34], the symbol “-” indicates that specific information is unavailable in the respective paper.

metrics while maintaining comparable model compactness. The qualitative comparisons are presented in Fig. 6. It can be seen that our method yields a more distinct representation in intricate areas than the other methods.

4.4. Neural Radiance Field Reconstruction

In the Neural Radiance Field setting, a volumetric shape is represented by a spatial (3D) density function and a spatial-directional (5D) emission function. The Blender dataset [34] which consists of eight object-centric scenes is widely used as a standard benchmark. We explore using NerfAcc [25] as the baseline implementation, while utilizing the same decoder as NebRBF. The experimental settings and hyperparameter selections are outlined in NerfAcc [25]. The results are shown in Tab. 4. It can be seen that under the same number of iterations, our method effectively enhances the rendering quality of new viewpoints while maintaining model compactness. Some visual comparisons are shown in Fig. 7. It can be seen that the proposed method demonstrates consistent improvements in both complex structures and smooth regions.

4.5. Ablation Study

To validate the effectiveness of each proposed module, ablation studies are conducted on both Kodak dataset [1] for 2D image fitting and Stanford 3D Scanning Repository [50] for SDF reconstruction.

Evaluation on elementary metric grids. The use of

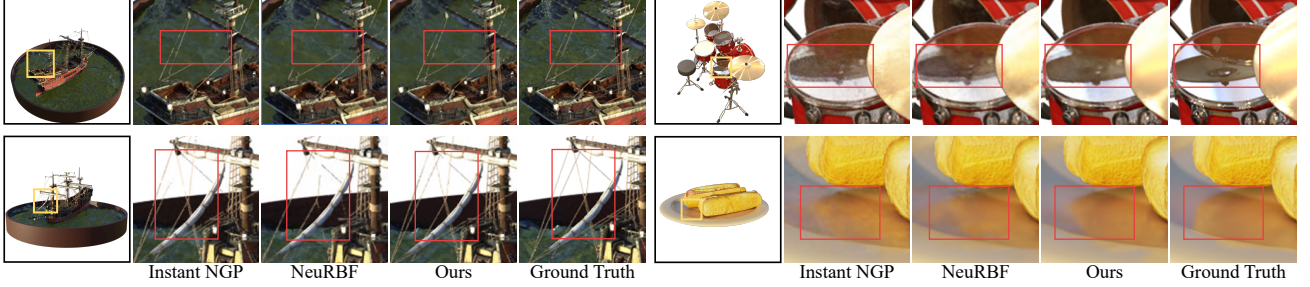


Figure 7. Qualitative Comparison on Neural Radiance Field Reconstruction. Our method consistently outperforms baselines in capturing both complex and fine details in scenes, as well as in smooth regions with reflections.

Settings	Decoder	2D image				3D SDF	
		PSNR	SSIM	Time	Params	CD	NAE
Baseline	MLP	37.06	0.9386	195s	206k	0.0023	3.43
M=2 [†]	MLP	39.45	0.9580	336s	234k	0.0022	3.34
M=3 [†]	MLP	40.85	0.9684	385s	258k	0.0022	3.20
M=4 [†]	MLP	41.91	0.9737	576s	279k	0.0023	3.15
M=2	MLP	38.71	0.9519	326s	206k	0.0022	3.40
M=3	MLP	38.95	0.9538	428s	206k	0.0022	3.35
M=4	MLP	38.98	0.9540	534s	206k	0.0025	3.35
Ours	w/o hierarchy	40.23	0.9668	1136s	212k	0.0022	3.29
Ours Full	High-order	41.57	0.9735	1148s	207k	0.0021	3.09

Table 5. Ablation Study. M indicates the number of elementary metric grids, aligned with the definition in Eq. (2). The symbol [†] indicates that the overall model size is not restricted with the increase of grids.

elementary metric grids was first evaluated to show the effectiveness of nonlinear approximation. It is built upon the baseline method, using InstantNGP [35]. We set the highest resolution of the metric grids to half of that used in the I-NGP and reduced the length of the hash table to maintain parameter counts comparable, resulting in metric grids with lower spatial resolution and increased sparsity as described in Sec. 3.3. The results are shown in the first section of Tab. 5. We have evaluated the effect of different grid numbers (M) by selecting 2, 3, and 4 metric grids based on the baseline. It can be seen that the elementary metric grids defined in nonlinear metric spaces significantly enhance the performance across all tasks, validating its effectiveness as high-order term approximations. Among the different configurations, M=2 yielded the largest improvement, further underscoring the importance of introducing nonlinear metrics for complex signal fitting. Additionally, experiments with restricting the total parameters when adding grids are also conducted, as detailed in the central section of Tab. 5. Under the same number of parameters, adding a higher-order term grid effectively improves performance, demonstrating that nonlinear interpolation via metric grids provides a better inductive bias compared to linear interpolation in terms of representation quality. However, performance gains diminish when the number of higher-order terms exceeds three, indicating explicit storage of higher-order terms proves computationally inefficient. Thus, Met-

ricGrids with M=3 is applied in our implementation.

Evaluation on high-order extrapolation decoder.

Based on the previous analysis, we set the order of the elementary metric grid to 3 and incorporate the proposed high-order extrapolation decoder for the implicit estimation of high-order terms at different positions, thereby achieving a balance between fitting accuracy and model compactness, which constitutes our full model. On one hand, it adds more nonlinearity layers to the decoder, improving the decoding accuracy. On the other hand, it generates higher-order terms based on the initial low-order approximation and different grids. To validate this, we masked the hierarchical structure of different grid feature inputs in the decoder (denoted as w/o hierarchy) and demonstrated significantly degraded decoder performance without hierarchical feature integration. In contrast, the high-order extrapolation decoder effectively leverages multiple metric grids to enhance the final performance.

5. Conclusion

In this paper, we have proposed MetricGrids, which provides accurate neural representations for various signal types. Our analysis reveals that the linear indexing function used to obtain continuous-space point features in the current grid-based neural representation leads to latent space degradation, thus resulting in parameter inefficiency and approximation error on nonlinear signals. Drawing inspiration from Taylor expansion, elementary metric grids defined in nonlinear metric spaces are proposed as high-order terms to capture complex latent space nonlinearity, thereby enhancing representation capabilities. A compact representation is further designed using hash encoding and a high-order extrapolation decoder to improve fitting accuracy without increasing parameter and storage requirements. Experimental results demonstrated that our method significantly outperforms baseline models and achieves state-of-the-art performance in 2D image fitting, 3D shape fitting, and neural radiance field reconstruction tasks.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China (No. 62271290 and 62001092).

References

- [1] Kodak lossless true color image suite. <https://r0k.us/graphics/kodak/>, 1999. 6, 7
- [2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 7
- [3] Alberto Bietti and Julien Mairal. On the inductive bias of neural tangent kernels. *Advances in Neural Information Processing Systems*, 32, 2019. 1, 3
- [4] Zhicheng Cai, Hao Zhu, Qiu Shen, Xinran Wang, and Xun Cao. Batch normalization alleviates the spectral bias in coordinate networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25160–25171, 2024. 2
- [5] Rohan Chabra, Jan E Lenssen, Eddy Ilg, Tanner Schmidt, Julian Straub, Steven Lovegrove, and Richard Newcombe. 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 16*, pages 608–625. Springer, 2020. 1
- [6] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. Efficient geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16123–16133, 2022. 2
- [7] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European conference on computer vision*, pages 333–350. Springer, 2022. 2, 7
- [8] Anpei Chen, Zexiang Xu, Xinyue Wei, Siyu Tang, Hao Su, and Andreas Geiger. Factor fields: A unified framework for neural fields and beyond. *arXiv preprint arXiv:2302.01226*, 2023. 2
- [9] Hao Chen, Bo He, Hanyu Wang, Yixuan Ren, Ser Nam Lim, and Abhinav Shrivastava. Nerv: Neural representations for videos. *Advances in Neural Information Processing Systems*, 34:21557–21568, 2021. 1
- [10] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8628–8638, 2021.
- [11] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5939–5948, 2019. 1, 2
- [12] Zhang Chen, Zhong Li, Liangchen Song, Lele Chen, Jingyi Yu, Junsong Yuan, and Yi Xu. Neurf: A neural fields representation with adaptive radial basis functions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4182–4194, 2023. 2, 3, 6, 7
- [13] Julian Chibane, Thimo Alldieck, and Gerard Pons-Moll. Implicit functions in feature space for 3d shape reconstruction and completion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6970–6981, 2020. 2
- [14] Grigorios G Chrysos, Stylianos Moschoglou, Giorgos Bouritsas, Jiankang Deng, Yannis Panagakis, and Stefanos Zafeiriou. Deep polynomial neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 44(8): 4021–4034, 2021. 5
- [15] Yishun Dou, Zhong Zheng, Qiaoqiao Jin, and Bingbing Ni. Multiplicative fourier level of detail. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1808–1817, 2023. 3, 5
- [16] Rizal Fathony, Anit Kumar Sahu, Devin Willmott, and J Zico Kolter. Multiplicative filter networks. In *International Conference on Learning Representations*, 2020. 2
- [17] Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12479–12488, 2023. 2, 7
- [18] Zekun Hao, Arun Mallya, Serge Belongie, and Ming-Yu Liu. Implicit neural representations with levels-of-experts. *Advances in Neural Information Processing Systems*, 35:2564–2576, 2022. 2
- [19] Amir Hertz, Or Perel, Raja Giryes, Olga Sorkine-Hornung, and Daniel Cohen-Or. Sape: Spatially-adaptive progressive encoding for neural optimization. *Advances in Neural Information Processing Systems*, 34:8820–8832, 2021. 2
- [20] Kurt Hornik. Approximation capabilities of multilayer feed-forward networks. *Neural networks*, 4(2):251–257, 1991. 1
- [21] Chiyu Jiang, Avneesh Sud, Ameesh Makadia, Jingwei Huang, Matthias Nießner, Thomas Funkhouser, et al. Local implicit grid representations for 3d scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6001–6010, 2020. 1, 2
- [22] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 2, 7
- [23] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6
- [24] Jason Chun Lok Li, Chang Liu, Binxiao Huang, and Ngai Wong. Learning spatially collaged fourier bases for implicit neural representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 13492–13499, 2024. 5, 6
- [25] Ruilong Li, Hang Gao, Matthew Tancik, and Angjoo Kanazawa. Nerfacc: Efficient sampling accelerates nerfs. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18537–18546, 2023. 7
- [26] Tianyang Li, Xin Wen, Yu-Shen Liu, Hua Su, and Zhizhong Han. Learning deep implicit functions for 3d shapes with dynamic code clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12840–12850, 2022. 3

- [27] David B Lindell, Dave Van Veen, Jeong Joon Park, and Gordon Wetzstein. Bacon: Band-limited coordinate networks for multiscale scene representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16252–16262, 2022. 2, 5
- [28] Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. *Advances in Neural Information Processing Systems*, 33:15651–15663, 2020. 1
- [29] Zhen Liu, Hao Zhu, Qi Zhang, Jingde Fu, Weibing Deng, Zhan Ma, Yanwen Guo, and Xun Cao. Finer: Flexible spectral-bias tuning in implicit neural representation by variable-periodic activation functions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2713–2722, 2024. 2
- [30] Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20654–20664, 2024. 3
- [31] Shishira R Maiya, Anubhav Gupta, Matthew Gwilliam, Max Ehrlich, and Abhinav Shrivastava. Latent-inr: A flexible framework for implicit representations of videos with discriminative semantics. *arXiv preprint arXiv:2408.02672*, 2024. 3
- [32] Julien NP Martel, David B Lindell, Connor Z Lin, Eric R Chan, Marco Monteiro, and Gordon Wetzstein. Acorn: Adaptive coordinate networks for neural scene representation. *arXiv preprint arXiv:2105.02788*, 2021. 2, 6
- [33] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2
- [34] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1, 2, 6, 7
- [35] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 2, 4, 5, 6, 7, 8
- [36] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019. 1, 2
- [37] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 523–540. Springer, 2020. 2
- [38] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International conference on machine learning*, pages 5301–5310. PMLR, 2019. 1, 3
- [39] Sameera Ramasinghe and Simon Lucey. Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps. In *European Conference on Computer Vision*, pages 142–158. Springer, 2022. 2
- [40] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 14335–14345, 2021. 2
- [41] Daniel Rho, Byeonghyeon Lee, Seungtae Nam, Joo Chan Lee, Jong Hwan Ko, and Eunbyung Park. Masked wavelet representation for compact neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20680–20690, 2023. 2
- [42] Vishwanath Saragadam, Jasper Tan, Guha Balakrishnan, Richard G Baraniuk, and Ashok Veeraraghavan. Miner: Multiscale implicit neural representation. In *European Conference on Computer Vision*, pages 318–333. Springer, 2022. 2
- [43] Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G Baraniuk. Wire: Wavelet implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18507–18516, 2023. 2, 6
- [44] Shayan Shekarforoush, David Lindell, David J Fleet, and Marcus A Brubaker. Residual multiplicative filter networks for multiscale reconstruction. *Advances in Neural Information Processing Systems*, 35:8550–8563, 2022. 2
- [45] Kexuan Shi, Xingyu Zhou, and Shuhang Gu. Improved implicit neural representation with fourier reparameterized training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25985–25994, 2024. 2
- [46] J Ryan Shue, Eric Ryan Chan, Ryan Po, Zachary Ankner, Jiajun Wu, and Gordon Wetzstein. 3d neural field generation using triplane diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20875–20886, 2023. 2
- [47] Vincent Sitzmann, Justus Thies, Felix Heide, Matthias Nießner, Gordon Wetzstein, and Michael Zollhofer. Deepvoxels: Learning persistent 3d feature embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2437–2446, 2019. 1
- [48] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Advances in Neural Information Processing Systems*, 32, 2019. 1
- [49] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020. 1, 2, 5, 6
- [50] Stanford University. The Stanford 3d scanning repository. <https://graphics.stanford.edu/data/3Dscanrep/>. 7
- [51] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields

- reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5459–5469, 2022. [1](#), [2](#), [3](#)
- [52] Towaki Takikawa, Joey Litalien, Kangxue Yin, Karsten Kreis, Charles Loop, Derek Nowrouzezahrai, Alec Jacobson, Morgan McGuire, and Sanja Fidler. Neural geometric level of detail: Real-time rendering with implicit 3d shapes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11358–11367, 2021. [1](#), [2](#), [7](#)
- [53] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020. [1](#), [2](#), [3](#)
- [54] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017. [3](#)
- [55] Peng-Shuai Wang, Yang Liu, Yu-Qi Yang, and Xin Tong. Spline positional encoding for learning 3d implicit signed distance fields. *arXiv preprint arXiv:2106.01553*, 2021. [2](#)
- [56] Zhijie Wu, Yuhe Jin, and Kwang Moo Yi. Neural fourier filter bank. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14153–14163, 2023. [3](#), [6](#)
- [57] Shaowen Xie, Hao Zhu, Zhen Liu, Qi Zhang, You Zhou, Xun Cao, and Zhan Ma. Diner: Disorder-invariant implicit neural representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6143–6152, 2023. [3](#)
- [58] Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. Pointnerf: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5438–5448, 2022. [3](#)
- [59] Guandao Yang, Sagie Benaim, Varun Jampani, Kyle Genova, Jonathan Barron, Thomas Funkhouser, Bharath Hariharan, and Serge Belongie. Polynomial neural fields for subband decomposition and manipulation. *Advances in Neural Information Processing Systems*, 35:4401–4415, 2022. [2](#)
- [60] Runzhao Yang. Tinc: Tree-structured implicit neural compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18517–18526, 2023. [2](#)
- [61] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. Plenotrees for real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5752–5761, 2021. [2](#)
- [62] Gizem Yüce, Guillermo Ortiz-Jiménez, Beril Besbinar, and Pascal Frossard. A structured dictionary perspective on implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19228–19238, 2022. [2](#)
- [63] Xinjie Zhang, Xingtong Ge, Tongda Xu, Dailan He, Yan Wang, Hongwei Qin, Guo Lu, Jing Geng, and Jun Zhang. Gaussianimage: 1000 fps image representation and compression by 2d gaussian splatting. In *European Conference on Computer Vision*, pages 327–345. Springer, 2025. [6](#), [7](#)
- [64] Yanshu Zhang, Shichong Peng, Alireza Moazeni, and Ke Li. Papr: Proximity attention point rendering. *Advances in Neural Information Processing Systems*, 36, 2024. [3](#)
- [65] Zi-Xin Zou, Zhipeng Yu, Yuan-Chen Guo, Yangguang Li, Ding Liang, Yan-Pei Cao, and Song-Hai Zhang. Triplane meets gaussian splatting: Fast and generalizable single-view 3d reconstruction with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10324–10335, 2024. [2](#)