Implicit Neural Representation with Multi-Scale Sine Activation

Jufeng Han[1,2], Shu Wei[1,3], Min Wu[1,2], Lina Yu[1,3*], WeiJun Li[1,2*], Linjun Sun[1,2*], Hong Qin[1,2], Yan Pang[1,3]

[1]AnnLab, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China
[2]School of Integrated Circuits, University of Chinese Academy of Sciences, Beijing 100049, China
[3]College of Materials Science and Opto-Electronic Technology, University of Chinese Academy of Sciences, Beijing 100049, China
hanjufeng@semi.ac.cn, yulina@semi.ac.cn

Abstract

Implicit Neural Representations (INRs) have become a powerful paradigm for modeling continuous signals in computer vision, graphics, and scientific computing. However, multilayer perceptrons (MLPs) generally suffer from severe spectral bias, which limits their ability to accurately model high-frequency details and multi-scale structures. To address this challenge, we propose a novel Multi-Scale Sine Activation (MSA), which explicitly introduces multi-scale frequency responses by incorporating multiple sets of sine activations with logarithmically spaced frequencies in parallel at each layer. MSA is further combined with an amplitude modulation mechanism to ensure numerical stability and robust optimization across different frequency channels. We conduct extensive experiments on a series of challenging tasks, including 1D multi-scale function fitting, image representation, video representation, 3D shape representation, and PDEs solving. Experimental results show that MSA outperforms existing state-of-the-art methods in terms of reconstruction accuracy, detail preservation, and training stability.

Code — https://github.com/JufengHan/MSA-INR

Introduction

In recent years, Implicit Neural Representations (INRs) have emerged as a powerful paradigm for signal modeling, demonstrating tremendous potential in computer vision and signal processing. Unlike traditional discrete representations such as pixel or voxel grids, INRs leverage multilayer perceptrons (MLPs) to directly map input coordinates in continuous space to signal values, enabling the continuous modeling of signals such as images, videos, and audio. This coordinate-based representation provides unique advantages, including analytic differentiability and resolution independence, offering new solutions for high-quality reconstruction, compression, and synthesis tasks. For example, INRs can directly represent an entire video sequence, enabling continuous video reconstruction and frame interpolation; by learning implicit fields of objects (such as signed distance functions or volumetric density fields), INRs achieve highly accurate representations of complex

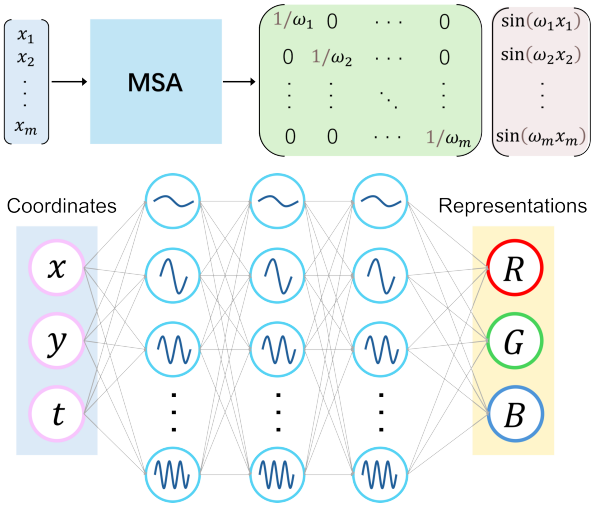


Figure 1: **MSA (top):** The input is activated by sine functions with different frequencies and amplitudes. **MSA for INRs (bottom):** In multi-channel scenarios such as images or videos, spatial and temporal coordinates are processed by MLPs with MSA, producing the corresponding signal representations

3D geometry and appearance (Park et al. 2019; Mescheder et al. 2019). Furthermore, when combined with differentiable rendering techniques, INRs are widely adopted for novel view synthesis tasks—for instance, Neural Radiance Fields (NeRF) generate continuous views via volumetric rendering, significantly advancing the development of photorealistic image synthesis (Mildenhall et al. 2021).

Despite the theoretical universal approximation capability of neural networks, they commonly exhibit a pronounced spectral bias during training: networks tend to fit low-frequency components first, while learning high-frequency components is markedly slower and may even fail to converge effectively (Rahaman et al. 2019; Xu et al. 2019). This phenomenon is particularly prominent when using traditional activation functions such as ReLU and Tanh, because these activations transmit high-frequency components inefficiently. In INRs tasks, many critical details—such as textures, edges, and the geometric structures of 3D shapes—are

*Corresponding author.

often rich in high-frequency information. Owing to the low-frequency preference of MLPs, the detailed representation of INRs is constrained, leading to over-smoothing, blurred textures, and ringing artifacts. Therefore, mitigating spectral bias and enhancing the high-frequency modeling capability of INRs has remained a key direction of continual optimization in this field.

To address these challenges, a variety of strategies have been proposed in recent years. One line of work is to transform the input coordinates, for example, by employing Fourier feature mapping, which enables the network to learn spectral information more effectively in the frequency domain (Tancik et al. 2020). Another approach enhances frequency representation by designing activation functions; for example, SIREN uses a single sinusoidal function as the activation (Sitzmann et al. 2020), which does not require additional coordinate transformations but does require specific weight-initialization strategies. There are also methods that use Gaussian functions (Ramasinghe and Lucey 2022), wavelet functions (Saragadam et al. 2023), and related choices as activations (Liu et al. 2024); although they reduce the dependence on initialization, they introduce additional computational overhead. Overall, current INRs still struggle to balance effectiveness and complexity, making it difficult to simultaneously achieve reconstruction accuracy and computational efficiency.

To this end, we propose a Multi-Scale Sine Activation (MSA) to further enhance the high-frequency modeling capability of INR. The core idea of MSA is to introduce multi-scale sine activations at each layer of the network, applying a set of sine transformations with different frequencies and amplitudes to the output of each linear transformation, thus enriching the spectral components of network features. It is worth emphasizing that MSA does not introduce any additional trainable parameters or alter the original network architecture; instead, it enhances spectral expressivity solely by modifying the activation function, ensuring both simplicity and efficiency. Across various challenging implicit representation tasks, MSA consistently achieves superior reconstruction accuracy and detail fidelity.

The main contributions of this paper are as follows:

- We propose a Multi-Scale Sine Activation (MSA), which introduces multi-scale sine transformations in each network layer, greatly expanding the spectral expressivity of implicit neural representations (INRs).
- We design MSA with adaptive amplitude modulation and a logarithmic frequency schedule that together reweight frequency components to curb gradient explosion and stabilize training while ensuring balanced low–to–high frequency coverage—ultimately improving spectral diversity, expressivity, and overall robustness.
- MSA attains superior performance on 1D function fitting, image representation, video representation, 3D shape representation, and PDEs solving, while reducing training time compared with strong baselines.

Related works

Spectral Bias

Spectral bias refers to the tendency of neural networks to prioritize learning the low-frequency components of a target function during training, while the learning of high-frequency components is significantly delayed. Rahaman et al. (2019) was the first to empirically demonstrate the ubiquity of this phenomenon through Fourier spectral analysis of deep ReLU networks. The Neural Tangent Kernel (NTK) theory provides a theoretical foundation for spectral bias: in the infinite-width limit, the training dynamics of the network are equivalent to kernel regression with a fixed NTK kernel, where gradient descent preferentially converges along the eigenfunctions associated with larger NTK eigenvalues (Jacot, Gabriel, and Hongler 2018; Cao et al. 2019). Since NTKs of typical networks assign larger eigenvalues to modes of lower complexity, the network can more easily fit these low-frequency components, thereby resulting in spectral bias.

Yang, Simon, and Bernstein (2023) proposed conditions for spectral norm scaling, revealing that appropriate initialization and training scales are critical for feature learning and high-frequency modeling, thus providing new theoretical support for overcoming spectral bias and improving the generalization ability of large models. Recent research shows that introducing PDE residuals in physics-informed neural networks does not substantially accelerate the decay of NTK eigenvalues (Gan et al. 2025).

Implicit Neural Representations

Implicit Neural Representations (INRs) are technique for parameterizing continuous signals or fields using neural networks (Fathony et al. 2020; Ramasinghe and Lucey 2022). The core idea is to employ MLPs to learn continuous mappings from coordinates to signal values. Unlike traditional discrete representations such as pixel grids or voxels, INRs enable signal representation at arbitrary resolutions and offer advantages.

INRs have been widely applied in image representation (Sitzmann et al. 2020), image and video compression (Dupont et al. 2021), 3D shape reconstruction (Park et al. 2019; Mescheder et al. 2019; Chen and Zhang 2019; Niemeyer et al. 2019), volume rendering and neural radiance fields (Mildenhall et al. 2021; Reiser et al. 2021; Gao et al. 2022), inverse problem modeling (Han, Jentzen, and E 2018), and generative modeling (Chan et al. 2021). One line of work focuses on the design of nonlinear activation functions, such as sinusoidal activations (Sitzmann et al. 2020; Liu et al. 2024; Morsali et al. 2025), wavelet-based activations (Saragadam et al. 2023), and hybrid combinations of multiple activations (Jayasundara et al. 2025), which have demonstrated improved high-frequency representation. Another line of work targets architectural enhancements, such as utilizing multiple sub-networks to process information at different scales (Liu, Cai, and Xu 2020), or applying Laplacian pyramid decomposition to model multi-scale signal structures (Saragadam et al. 2022).

Methods

The formulation of INR

Implicit Neural Representation (INR) is a method that leverages neural networks to model continuous signals or fields. The core idea is to map low-dimensional spatial or spatiotemporal coordinates to corresponding signal values, such as pixel intensities in images, RGB values in 3D scenes, or physical quantities in scientific computing. Specifically, INR employs a neural network to learn a continuous function $\Phi : \mathbb{R}^m \rightarrow \mathbb{R}^d$, enabling the approximation of $\Phi(\mathbf{x})$ over the entire target domain.

Unlike traditional explicit discretization methods, such as grid sampling or interpolation, INR focuses on learning a function $\Phi(\mathbf{x})$ that satisfies a set of complex implicit relationships $\{C_m\}_{m=1}^{M}$ with the target variables:

$$C_m\big(\mathbf{x}, \Phi(\mathbf{x}), \nabla\Phi(\mathbf{x}), \dots\big) = 0, \quad \forall \mathbf{x} \in \Omega_m, \quad m = 1, 2, ..., M \tag{1}$$

where $a(\mathbf{x})$ denotes the ground-truth signal value or observation at coordinate $\mathbf{x}$. C typically comprises algebraic relationships and differential operators, capturing the essential constraints of the signal as well as its underlying physical or mathematical laws.

The most common parameterization of INR is via a multilayer perceptron (MLP), with the network structure unified as follows:

$$\begin{aligned} \Phi_\Theta(\mathbf{x}) &= \mathbf{W}_n \left(\phi_{n-1} \circ \phi_{n-2} \circ \cdots \circ \phi_1\right)(\mathbf{x}), \\ \phi_i(\mathbf{x}) &= \sigma_i\left(\mathbf{W}_i\mathbf{x} + \mathbf{b}_i\right), \end{aligned} \tag{2}$$

where $\sigma_i(\cdot)$ is an element-wise nonlinear activation function, and $\mathbf{W}_i$ and $\mathbf{b}_i$ are the weight matrix and bias vector for the i-th layer, respectively. $\phi_i(\mathbf{x})$ denotes the transformation performed by each layer.

INR typically reformulates its objective as a loss minimization problem. In practice, the constraint residuals are computed at sampled points $\{\mathbf{x}_j\}_{j=1}^{N}$, and the optimization target is given by:

$$\mathcal{L} = \sum_{j=1}^{N}\sum_{m=1}^{M} \left\| C_m\big(\mathbf{x}_j, \Phi(\mathbf{x}_j), \nabla\Phi(\mathbf{x}_j), \dots\big) \right\| \tag{3}$$

where $\mathcal{L}$ quantifies the overall discrepancy between the network outputs and all constraint conditions. By minimizing this loss function via gradient-based optimization, the INR parameters $\Theta = \{\mathbf{W}^{(i)}, \mathbf{b}^{(i)}\}_{i=1}^{n}$ are optimized to learn a mapping $\Phi_\Theta(\mathbf{x})$ that jointly satisfies the set of complex implicit relationships defined by C_m.

Multi-Scale Sine Activation

The Multi-Scale Sine Activation (MSA) is designed to explicitly introduce structured frequency responses, enabling neural networks to efficiently model high-frequency components. MSA can be regarded as a structured sinusoidal activation comprising multiple frequency channels. Let $\mathbf{F} \in \mathbb{R}^{m\times m}$ be a diagonal matrix whose diagonal elements correspond to the activation frequencies of each neuron:

$$\mathrm{MSA}(\mathbf{x}) = \mathbf{F}^{-1}\sin(\mathbf{F}\mathbf{x}), \quad \mathbf{F} = \mathrm{diag}(\omega_1, \omega_2, ..., \omega_m) \tag{4}$$

We apply MSA after each hidden layer of the MLP. The overall architecture can be formulated as:

$$\begin{aligned} \Phi_\Theta(\mathbf{x}) &= \mathbf{W}_n \left(\phi_{n-1} \circ \phi_{n-2} \circ \cdots \circ \phi_1\right)(\mathbf{x}), \\ \phi_i(\mathbf{x}) &= \mathrm{MSA}\left(\mathbf{W}_i\mathbf{x} + \mathbf{b}_i\right), \end{aligned} \tag{5}$$

As shown in Figure 1, introducing multiple parallel channels with different frequencies in each layer allows MSA to simultaneously cover multiple frequency scales, thereby substantially enhancing its ability to model high-frequency and multi-scale tasks.

Amplitude Modulation Mechanism

To enhance the numerical stability and training robustness of the network across different frequency spectra, we propose an amplitude modulation mechanism. In MSA, this mechanism is implemented by an amplitude scaling matrix $\mathbf{F}^{-1}$, which is the inverse of the frequency matrix $\mathbf{F}$.

During the forward pass, the amplitude scaling matrix assigns different amplitude weights to the activations of different frequency channels. Specifically, the output of each neuron in the hidden layer can be expressed as $\mathrm{MSA}(x_i) = \frac{1}{\omega_i}\sin(\omega_i x_i)$. This structure naturally suppresses the amplitude of high-frequency channels, dynamically reducing the contribution of high-frequency activations to the overall network output. As a result, the adverse effects of high-frequency signals on downstream fitting are mitigated, which effectively reduces fluctuations and instability during training.

In the backward pass, the amplitude scaling matrix plays a crucial role in stabilizing the gradients. Since the high-frequency components of the sine activation can generate extremely large gradients during backpropagation, this easily leads to gradient explosion. Due to the presence of the amplitude scaling matrix, the derivative is given by $\mathrm{MSA}'(x_i) = \cos(\omega_i x_i)$, ensuring that the gradient of each neuron output with respect to its input is always constrained within the interval $[-1, 1]$. This significantly improves the stability of gradients during backpropagation.

Our ablation experiments further demonstrate the effectiveness of the amplitude modulation mechanism in alleviating high-frequency gradient explosion and enhancing both convergence efficiency and training stability of the network.

Frequency Setting Strategy

The selection of frequencies is one of the key factors for the success of MSA. We adopt a logarithmic setting strategy, selecting m frequencies evenly over the interval $[10^{s_{\min}}, 10^{s_{\max}}]$:

$$\lg\omega_i = s_{min} + \frac{i-1}{m-1}(s_{max} - s_{min}), \quad i = 1, 2, ..., m \tag{6}$$

here, $s_{\min}$ and $s_{\max}$ denote the lower and upper bounds of the frequency sampling range, respectively. Both are hyperparameters that need to be chosen according to the specific problem setting.

This approach generates a quasi-dyadic frequency sequence, achieving efficient and uniform spectral coverage.

Dense low-frequency activation helps to fit the global structure of the function, while the gradual sparsity of high-frequency activation avoids redundancy and reduces the risk of overfitting caused by interference between adjacent frequencies. Compared to linear spacing, logarithmic spacing exhibits greater robustness across different frequency scales, improves model generalization and numerical stability, and enables adaptation to a much broader range of frequency scales.

Discussion: Advantages of MSA

The MSA exhibits several notable advantages:

(1) No Additional Trainable Parameters and Low Computational Overhead. MSA achieves strong modeling capacity through a deterministic frequency design, without introducing any additional trainable parameters or significant computational burden. As a result, both training and inference with MSA are more efficient compared to methods that require complex architectures or learnable frequency parameters.

(2) Enhanced Spectral Expressivity via Fourier Series Structure. From the perspective of Fourier analysis, any continuous function can be approximated by a linear combination of sine and cosine waves of different frequencies (i.e., a Fourier series). In MSA, the output of each layer is inherently composed of multiple sine waves at different frequencies, meaning the input to the next layer is a linear combination of these sinusoidal components. This structure is equivalent to embedding a multi-scale Fourier expansion within the network, substantially enhancing the model's ability to approximate complex functions.

(3) Plug-and-Play Without Special Architecture or Initialization. Owing to rational frequency sampling and the amplitude scaling mechanism, MSA is highly modular and can be seamlessly integrated as an activation function into existing neural network architectures. It does not require any special parameter initialization, yet still ensures robust optimization and high-precision convergence during training, greatly facilitating practical adoption and deployment.

Additional analysis and proofs of the method are provided in the supplementary material.

Experiments

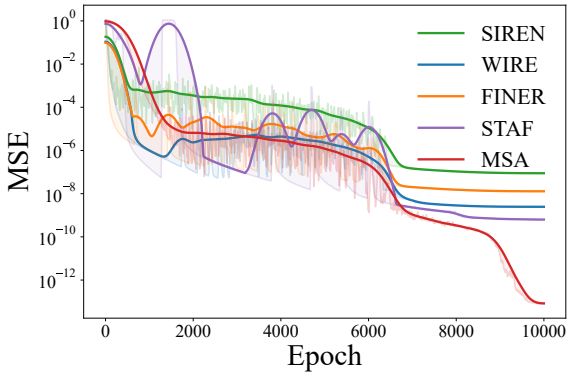
To comprehensively evaluate the effectiveness of the proposed MSA, we conduct systematic experiments across a series of challenging INRs tasks, including 1D function fitting, image representation, video representation, 3D shape reconstruction, and PDEs sovling. Additional analyses and experiments are provided in the supplementary material.

1D Fuction Fitting and NTK Analysis

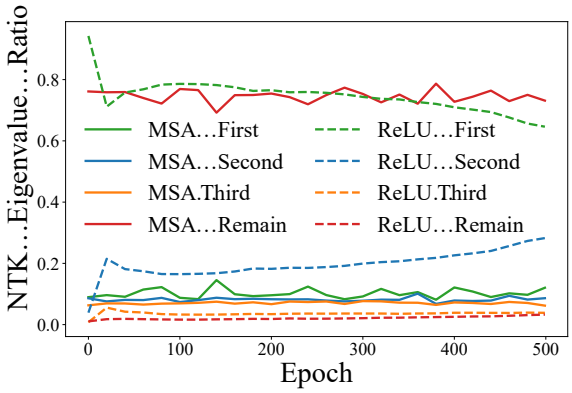
To evaluate the performance of MSA on 1D function fitting, we parameterize a continuous mapping $\Phi : \mathbb{R}^1 \to \mathbb{R}^1$ to represent a multi-scale target function:

$$y_4 = 0.5\sin(2x) + \sin(100x + 1.5)\cos(5x - 0.8) + \cos(200x)\cos(x) + \sin(300x + 0.3)\sin(5x)$$

This function is composed of the sum of four terms with different frequencies and phases, including low-frequency components as well as multiple high-frequency, modulated, and mixed-frequency components, thereby exhibiting clear multiscale characteristics. We compared several mainstream INR methods: (1) SIREN (Sitzmann et al. 2020); (2) WIRE (Saragadam et al. 2023); (3) FINER (Liu et al. 2024); (4) STAF (Morsali et al. 2025); and (5) our proposed MSA.



(a) Training loss curves



(b) NTK eigenvalue ratios

Figure 2: (a) Training loss comparison on 1D function fitting; (b) Temporal evolution of the empirical NTK spectrum for MSA and ReLU activations.

All models were implemented as MLPs with four hidden layers and 256 neurons per layer. Training was performed using the Adam optimizer for 10,000 epochs. The initial learning rate was set to 1×10^{-4} and gradually decayed to 1×10^{-7} via cosine annealing. For all models, 1,024 discrete points were randomly sampled within the interval $[-1, 1]$ for training. During training, the mean squared error (MSE) was used as the evaluation metric. As shown in Figure 2(a), all comparison methods yielded MSE above 1×10^{-8}, while MSA achieved an MSE as low as 7.96×10^{-14}, demonstrating its superiority in multiscale scenarios.

We conduct an empirical analysis of the Neural Tangent Kernel (NTK) on this task. The NTK reflects the learning dynamics of neural networks during training, and the distribution of its eigenvalues characterizes the model's ability

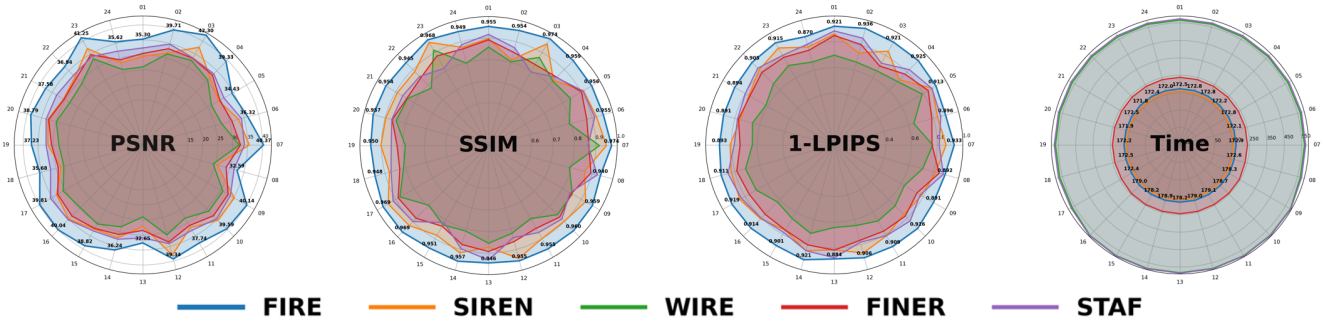


Figure 3: **Image representation on KODAK-24:** per-image PSNR (dB), SSIM, 1-LPIPS and Time (s) for MSA and baselines.

to fit different frequency components as well as its spectral bias.

During the experiment, we periodically compute the eigenvalues of $\mathbf{K}$, $\lambda_1, \lambda_2, \ldots, \lambda_N$, and track the normalized ratios of the largest three eigenvalues (First, Second, Third) and the sum of the remaining eigenvalues (Remain)

As shown in Figure 2(b), the NTK eigenvalues of the ReLU-MLP are always highly concentrated in the largest eigenvalue λ_1, with r_1 persistently exceeding 60%. This indicates that the network's learning is primarily focused on the low-frequency component, making it difficult to sufficiently activate other spectral directions and resulting in poor fitting for high-frequency or multi-scale signals.

In contrast, MSA exhibits a significantly different evolution of the spectrum. At the early stage of training, r_1 rapidly decreases, while r_{remain} increases quickly and eventually dominates (over 60%). This demonstrates that the network's learning capacity is allocated to a broader spectral range, effectively alleviating spectral bias.

Image Representation

We formulate image reconstruction as learning a continuous mapping from coordinates to colors: $\Phi : (x, y) \rightarrow (R, G, B)$, directly modeling a continuous field for the entire image without relying on convolutions or explicit interpolation. To ensure a fair comparison, MSA and the baselines SIREN, WIRE, FINER, and STAF are evaluated under an identical backbone and training protocol: an MLP with 4 hidden layers and 400 neurons per layer; the loss is MSE, the optimizer is Adam, and the learning rate decays from 1×10^{-3} to 5×10^{-4} via cosine annealing for a total of 1500 epochs. We reconstruct the 24 color images in the KODAK (Franzen 1999) dataset and compute and report PSNR, SSIM, and LPIPS (Zhang et al. 2018) (AlexNet backbone; lower is better) at the end of training.

As shown in Figure 3, MSA achieves the highest PSNR and SSIM values, along with the lowest LPIPS values across all 24 images, demonstrating a stable performance advantage. Although MSA is not the fastest in training time, it ranks second and is nearly as fast as SIREN, reflecting its competitiveness in efficiency. The amplitude scaling mechanism introduces an additional multiplication step, resulting in a slight increase in computational cost, which is still significantly lower than that of other baselines.

Methods	PSNR (dB)		SSIM		Time (min)
	Avg ↑	Std ↓	Avg ↑	Std ↓	
MSA	**42.74**	**4.63**	**0.980**	**0.006**	166.46
SIREN	40.19	5.59	0.967	0.012	**165.39**
WIRE	41.87	5.47	0.976	0.008	176.14
FINER	38.92	5.85	0.965	0.011	171.12
STAF	39.86	6.02	0.967	0.012	188.76

Table 1: Quantitative results of video representation for MSA and baselines. Avg and Std denote the mean and standard deviation over 100 frames, respectively; Time denotes the training time.

Video Representation

The video representation aims to evaluate the ability of INRs to fit and represent spatiotemporal high-frequency signals. In this task, the video is parameterized by a continuous function $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, which maps each three-dimensional coordinate (spatial position x, y and temporal frame index t) to its corresponding RGB value. This formulation enables the network to directly reconstruct the entire video sequence from spatiotemporal coordinates.

In this experiment, we use the first 100 frames of the publicly available color video "BigBuckBunny" as the regression target. Each frame has a spatial resolution of 180×320, resulting in tens of thousands of pixel samples for training. All models adopt MLPs architecture with three hidden layers, each containing 400 neurons, and are trained for 500 epochs. The learning rate is decayed from 1×10^{-3} to 5×10^{-4} using cosine annealing.

To quantitatively evaluate the fitting performance of each model, we compute the PSNR between the reconstructed and original videos on a frame-by-frame basis. As shown in Table 1, MSA achieves the highest PSNR and SSIM with the smallest variance, indicating not only high-fidelity reconstruction but also stable consistency across adjacent frames. Moreover, MSA requires only 1 more minute of training time than SIREN and less than the other baselines, demonstrating near-minimal training overhead while maintaining quality.

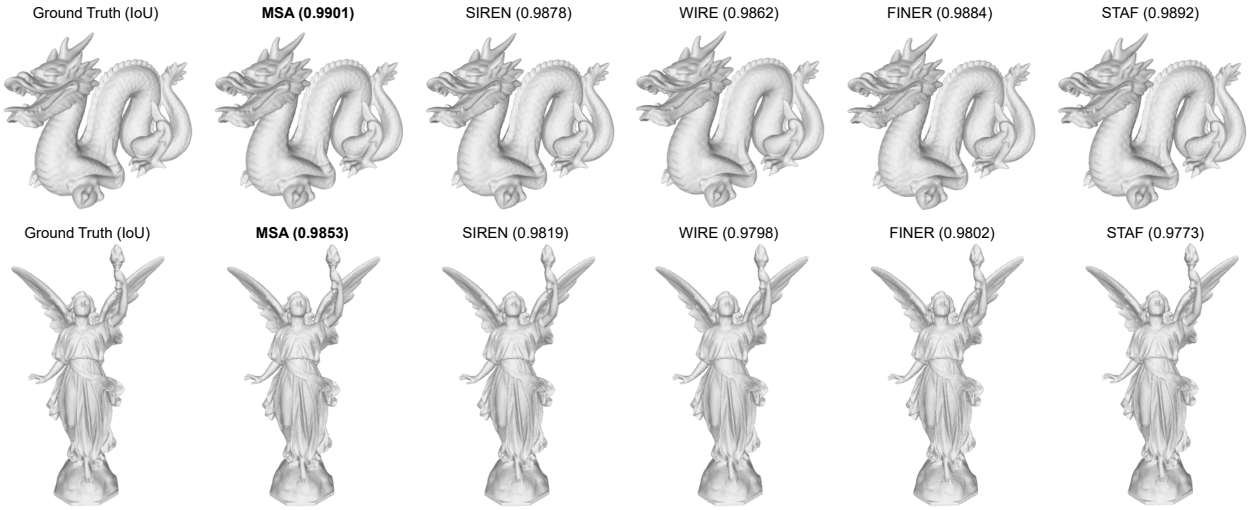


Figure 4: **Visual comparison of 3D shape (Dragon and Thai) reconstruction (IoU) by different INRs.**

Methods	Dragon	Lucy	Thai	Armadillo
MSA	**0.9901**	**0.9853**	**0.9825**	**0.9898**
SIREN	0.9878	0.9819	0.9770	0.9864
WIRE	0.9862	0.9798	0.9751	0.9838
FINER	0.9884	0.9802	0.9788	0.9872
STAF	0.9892	0.9773	0.9779	0.9855

Table 2: Quantitative results of 3D shape representation (IoU) for MSA and baselines.

3D Shape Representation

The 3D shape representation task aims to evaluate the ability of INR to represent and reconstruct complex geometric structures in three-dimensional space. In this experiment, we adopt the signed distance field (SDF) as the implicit representation of 3D objects (Jones, Baerentzen, and Sramek 2006). A neural network is used to parameterize a continuous function $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}$, which outputs the SDF value for each spatial point (x, y, z). The surface of the object is implicitly defined by the zero-level set where $\Phi(\mathbf{x}) = 0$.

We select four shapes from a public dataset (Stanford 2014) as reconstruction targets. All models employ MLPs with three hidden layers and 256 neurons per layer as the baseline model. The number of training epochs is set to 200, with an initial learning rate of 5×10^{-3}, which is gradually reduced using a piecewise decay schedule. The batchsize is set to 500,000.

Quantitative results are reported in Table 2, where MSA achieves the highest IoU across all four examples. Figure 4 visualizes the 3D shape (Dragon and Thai) reconstruction results of different methods, and the intersection over union (IoU) is used as the quantitative evaluation metric. Experimental results show that MSA achieves the highest IoU among all methods and is able to recover more geometric details, demonstrating its advantage in modeling high-dimensional complex structures.

PDEs Solving

Partial differential equations (PDEs) are fundamental mathematical tools for describing the evolution of continuous media systems. In recent years, physics-informed neural networks (PINNs) (Raissi, Perdikaris, and Karniadakis 2019) have emerged as a powerful deep learning framework for solving PDEs. However, traditional PINNs often struggle to efficiently learn solution functions in multi-scale or high-frequency problems due to spectral bias (Xu, Zhang, and Cai 2025). To address these challenges, we introduce the MSA into the PINN framework, resulting in MSA-PINN.

To comprehensively validate the effectiveness of the proposed method, we benchmarked the following representative approaches: (1) PINN (Raissi, Perdikaris, and Karniadakis 2019); (2) FF-PINN (Sallam and Fürth 2023); (3) sf-PINN (Wong et al. 2022); (4) GP-HM-StM (Fang et al. 2023); (5) TSA-PINN (Khademi and Dufour 2025); and (6) MSA-PINN (PINN using MSA as the activation function).

We benchmark our method on four canonical PDEs: 1D heat, 2D Poisson, 2D Allen–Cahn, and 1D advection, spanning linear and nonlinear regimes. The full equations and training details are given in the supplementary material.

We use the relative l_2 error as our evaluation metric. The results of the relative l_2 error for different methods on each PDE are reported in Table 3. For these high-frequency and multi-scale PDEs, conventional PINN struggles to overcome spectral bias and thus performs poorly. Both FF-PINN and sf-PINN are highly sensitive to the variance setting of frequency sampling. MSA-PINN effectively addresses this issue by employing fixed sampling frequencies. GP-HM-StM, a Gaussian process-based method, achieves performance comparable to MSA-PINN on certain equations but incurs substantial computational cost and may fail on specific PDEs. TSA-PINN introduces trainable sine and cosine activation neurons; it rarely produces satisfactory re-

sults, even with trainable frequencies. In the test case of the Advection equation, we set all loss weights to 1. The residual loss is significantly higher than the boundary and initial losses, posing a major optimization challenge for PINNs. As the results indicate, most baseline methods nearly fail in this scenario, while MSA-PINN successfully overcomes the optimization difficulty and yields highly accurate predictions.

Methods	Heat	Poisson	Allen-Cahn	Advection
MSA-PINN	**1.93e-04**	**1.03e-04**	**1.41e-03**	**3.35e-04**
PINN	1.00e+00	6.97e-02	9.59e-01	8.32e-01
FF-PINN	2.25e-03	2.01e-04	2.49e-03	2.16e-01
sf-PINN	2.38e-04	5.78e-04	3.27e-03	9.93e-01
GP-HM-StM	2.21e-04	1.26e-04	2.99e-03	1.00e+00
TSA-PINN	9.92e-01	2.70e-03	8.71e-01	1.41e-01

Table 3: The relative l_2 errors of different methods on various PDEs.

Ablation Study

Ablation on the Amplitude Modulation Mechanism In this section, we conduct an ablation study to specifically analyze the contribution of the amplitude modulation mechanism to the numerical stability and accuracy of MSA. Concretely, we remove the amplitude scaling matrix $\mathbf{F}^{-1}$ from MSA, i.e., we set $\text{MSA}(\mathbf{x}) = \sin(\mathbf{F}\mathbf{x})$ (denoted as MSA-NoAM), and compare its performance on PDE solving tasks.

The ablation experiment employs the Poisson equation, but with a different source term: $f(x, y) = \sin(\pi x)\sin(10\pi y)$ is used to construct the analytical solution and boundary conditions. Experimental results show that MSA achieves a highly accurate numerical solution for this task, with a relative l_2 error of only 3.49×10^{-5}; in contrast, MSA-NoAM completely fails to train, yielding an error as large as 1.31×10^{1}. The primary cause of this failure is the gradient explosion triggered by high-frequency channels. We verify this by logging and analysing the network gradients throughout training. For each layer's weights, we compute the logarithmic mean of their gradients across all epochs, defined as:

$$\text{Mean-Log}(l) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\log_{10}(g_{ij}^{(l)} + \epsilon) \qquad (7)$$

where m is the number of training epochs, and n is the number of parameter elements in the l-th layer. A larger Mean-Log value indicates a larger average gradient for that layer during training. As shown in Table 4, the gradients of all weights and biases in MSA are effectively controlled within a small range, while for MSA-NoAM, the gradients of all layers are abnormally large (exceeding 10^{20}).

Ablation on Frequency Setting Strategies In this section, we systematically compare the performance of three frequency setting strategies—logarithmic spacing, linear spacing, and uniform random sampling—across different values of $s_{\max}$. Experimental results on the 2D Helmholtz equation demonstrate that logarithmic sampling consistently

Activation	layer1	layer2	layer3	layer4
MSA-NoAM	29.01	27.08	25.12	22.41
MSA	-3.47	-5.31	-5.63	-5.26

Table 4: Comparison of the mean logarithmic gradient values for each layer's weights.

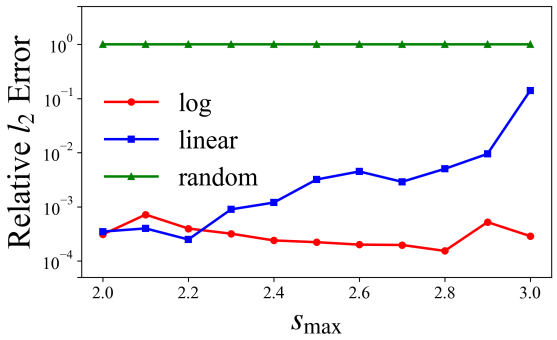


Figure 5: **Relative l_2 error of MSA-PINN versus $s_{\max}$ under different frequency sampling strategies.**

provides superior stability and accuracy throughout the entire $s_{\max}$ range (see Figure 5). Linear spacing achieves comparable accuracy to logarithmic spacing within a certain interval, but its performance is highly sensitive to the choice of $s_{\max}$ and deteriorates rapidly as $s_{\max}$ increases. In contrast, uniform random sampling fails to effectively solve the task when the frequency range is large. Overall, logarithmic spacing exhibits better robustness and generalization in multi-scale frequency modeling, making it the optimal frequency sampling strategy for MSA.

Conclusion

This paper proposes the Multi-Scale Sine Activation (MSA) to enhance the capability of INRs in modeling high-frequency details and multi-scale structures. MSA introduces multiple sine activation channels with different frequencies at each network layer, enabling the model to capture a rich spectrum of frequency features. The logarithmic frequency setting strategy ensures comprehensive coverage of both low and high frequency components while maintaining numerical stability. Combined with the amplitude modulation mechanism, MSA effectively balances the contributions of different frequency channels and the gradients during backpropagation, thereby improving training stability. Our experiments demonstrate that MSA consistently outperforms mainstream baselines in terms of fitting accuracy and detail preservation across a variety of INR tasks. Importantly, MSA does not introduce any additional trainable parameters and can be seamlessly integrated into existing architectures, providing an efficient and general solution for continuous modeling of high-frequency and complex structured signals. We believe that MSA holds great potential for broader applications in multimodal signal modeling and high-dimensional scientific computing.

Acknowledgments

This work was supported by the National Key R&D Program of China (Grant No. 2024YFF0618303) and the Scientific and Technological Innovation Project of the China Academy of Chinese Medical Sciences (Grant No. CI2023C001YG).

References

Cao, Y.; Fang, Z.; Wu, Y.; Zhou, D.-X.; and Gu, Q. 2019. Towards understanding the spectral bias of deep learning. *arXiv preprint arXiv:1912.01198*.

Chan, E. R.; Monteiro, M.; Kellnhofer, P.; Wu, J.; and Wetzstein, G. 2021. pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5799–5809.

Chen, Z.; and Zhang, H. 2019. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5939–5948.

Dupont, E.; Goliński, A.; Alizadeh, M.; Teh, Y. W.; and Doucet, A. 2021. Coin: Compression with implicit neural representations. *arXiv preprint arXiv:2103.03123*.

Fang, S.; Cooley, M.; Long, D.; Li, S.; Kirby, R.; and Zhe, S. 2023. Solving high frequency and multi-scale pdes with gaussian processes. *arXiv preprint arXiv:2311.04465*.

Fathony, R.; Sahu, A. K.; Willmott, D.; and Kolter, J. Z. 2020. Multiplicative filter networks. In *International conference on learning representations*.

Franzen, R. 1999. Kodak Lossless True Color Image Suite. http://r0k.us/graphics/kodak/. Accessed: 2025-06-15.

Gan, W.; Li, Y.; Lin, Q.; and Shi, Z. 2025. Neural Tangent Kernel of Neural Networks with Loss Informed by Differential Operators. *arXiv preprint arXiv:2503.11029*.

Gao, K.; Gao, Y.; He, H.; Lu, D.; Xu, L.; and Li, J. 2022. Nerf: Neural radiance field in 3d vision, a comprehensive review. *arXiv preprint arXiv:2210.00379*.

Han, J.; Jentzen, A.; and E, W. 2018. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34): 8505–8510.

Jacot, A.; Gabriel, F.; and Hongler, C. 2018. Neural tangent kernel: Convergence and generalization in neural networks. *Advances in neural information processing systems*, 31.

Jayasundara, D.; Zhao, H.; Labate, D.; and Patel, V. M. 2025. MIRE: Matched Implicit Neural Representations. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 8279–8288.

Jones, M. W.; Baerentzen, J. A.; and Sramek, M. 2006. 3D distance fields: A survey of techniques and applications. *IEEE Transactions on visualization and Computer Graphics*, 12(4): 581–599.

Khademi, A.; and Dufour, S. 2025. Physics-informed neural networks with trainable sinusoidal activation functions for approximating the solutions of the Navier-Stokes equations. *Computer Physics Communications*, 109672.

Liu, Z.; Cai, W.; and Xu, Z.-Q. J. 2020. Multi-scale deep neural network (MscaleDNN) for solving Poisson-Boltzmann equation in complex domains. *arXiv preprint arXiv:2007.11207*.

Liu, Z.; Zhu, H.; Zhang, Q.; Fu, J.; Deng, W.; Ma, Z.; Guo, Y.; and Cao, X. 2024. 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*, 2713–2722.

Mescheder, L.; Oechsle, M.; Niemeyer, M.; Nowozin, S.; and Geiger, A. 2019. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4460–4470.

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. *Communications of the ACM*, 65(1): 99–106.

Morsali, A.; Vaez, M.; Soltani, M.; Kazerouni, A.; Taati, B.; and Mohammad-Noori, M. 2025. STAF: Sinusoidal Trainable Activation Functions for Implicit Neural Representation. *arXiv preprint arXiv:2502.00869*.

Niemeyer, M.; Mescheder, L.; Oechsle, M.; and Geiger, A. 2019. Occupancy flow: 4d reconstruction by learning particle dynamics. In *Proceedings of the IEEE/CVF international conference on computer vision*, 5379–5389.

Park, J. J.; Florence, P.; Straub, J.; Newcombe, R.; and Lovegrove, S. 2019. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 165–174.

Rahaman, N.; Baratin, A.; Arpit, D.; Draxler, F.; Lin, M.; Hamprecht, F.; Bengio, Y.; and Courville, A. 2019. On the spectral bias of neural networks. In *International conference on machine learning*, 5301–5310. PMLR.

Raissi, M.; Perdikaris, P.; and Karniadakis, G. E. 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378: 686–707.

Ramasinghe, S.; and Lucey, S. 2022. Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps. In *European Conference on Computer Vision*, 142–158. Springer.

Reiser, C.; Peng, S.; Liao, Y.; and Geiger, A. 2021. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In *Proceedings of the IEEE/CVF international conference on computer vision*, 14335–14345.

Sallam, O.; and Fürth, M. 2023. On the use of Fourier Features-Physics Informed Neural Networks (FF-PINN) for forward and inverse fluid mechanics problems. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, 237(4): 846–866.

Saragadam, V.; LeJeune, D.; Tan, J.; Balakrishnan, G.; Veeraraghavan, A.; and Baraniuk, R. G. 2023. Wire: Wavelet

implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18507–18516.

Saragadam, V.; Tan, J.; Balakrishnan, G.; Baraniuk, R. G.; and Veeraraghavan, A. 2022. Miner: Multiscale implicit neural representation. In *European Conference on Computer Vision*, 318–333. Springer.

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.

Stanford, C. 2014. The Stanford 3D Scanning Repository. Http://graphics.stanford.edu/data/3Dscanrep/.

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.

Wong, J. C.; Ooi, C. C.; Gupta, A.; and Ong, Y.-S. 2022. Learning in sinusoidal spaces with physics-informed neural networks. *IEEE Transactions on Artificial Intelligence*, 5(3): 985–1000.

Xu, Z.-Q. J.; Zhang, L.; and Cai, W. 2025. On understanding and overcoming spectral biases of deep neural network learning methods for solving PDEs. *arXiv preprint arXiv:2501.09987*.

Xu, Z.-Q. J.; Zhang, Y.; Luo, T.; Xiao, Y.; and Ma, Z. 2019. Frequency principle: Fourier analysis sheds light on deep neural networks. *arXiv preprint arXiv:1901.06523*.

Yang, G.; Simon, J. B.; and Bernstein, J. 2023. A spectral condition for feature learning. *arXiv preprint arXiv:2310.17813*.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*.