

SAFIR: Spatially-aware Activation Function for Implicit Neural Representations

Ehsan Zeraatkar

*Department of Computer Science
Texas State University
San Marcos, US
ehsanzeraatkar@txstate.edu*

Jelena Tešić

*Department of Computer Science
Texas State University
San Marcos, US
jtesic@txstate.edu*

Abstract—Neural Implicit Representations (INRs) have reshaped diverse vision tasks by modeling signals as continuous functions, yet existing architectures such as SIREN suffer from spectral bias due to a reliance on spatially-uniform frequency parameters. This work introduces SAFIR, a novel spatially-aware adaptive frequency modulation framework that learns coordinate-dependent frequency parameters via a convolutional omega prediction network. SAFIR enables adaptive allocation of high and low frequencies across spatial regions, addressing the persistent problem of over-smoothing or overfitting in conventional INRs. Extensive evaluations on five benchmark datasets demonstrate that SAFIR not only offers substantial improvements in PSNR and SSIM for 2D image reconstruction, super-resolution, and generative vision tasks, but also achieves superior parameter efficiency and faster convergence than state-of-the-art alternatives. By effectively bridging local signal complexity with adaptive neural modulation, SAFIR delivers a practical step forward for high-fidelity, efficient signal representation in modern computer vision applications.

Index Terms—Neural implicit representations, SIREN, adaptive frequency, Kolmogorov-Arnold networks, spectral bias, image reconstruction

I. INTRODUCTION

The emergence of Implicit Neural Representations (INRs) has fundamentally transformed the landscape of continuous signal modeling, offering a paradigm shift from traditional discrete, grid-based approaches to continuous function parameterization [1]. Unlike conventional neural networks that operate on fixed-resolution data structures, INRs parameterize continuous signals as neural networks that directly map coordinates to signal values, enabling the representation of images, audio signals, 3D shapes, and temporal sequences as continuous mathematical functions.

Traditional neural network architectures, while highly successful in discrete domains, face inherent limitations when modeling continuous signals [2]. Convolutional Neural Networks (CNNs) and their variants operate on predetermined grid resolutions, necessitating explicit interpolation or resampling operations to handle different scales, often resulting in aliasing artifacts and computational inefficiencies [3]. Furthermore, discrete representations struggle with smooth interpolation between samples and cannot naturally accommodate irregular sampling patterns commonly encountered in scientific and medical applications [4].

INRs address these fundamental limitations by treating signals as continuous functions $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$, where input coordinates are mapped to corresponding signal values through learned neural transformations [5]. This coordinate-based approach provides several compelling advantages: resolution independence, enabling queries at arbitrary scales without retraining [6]; memory efficiency, as the signal is compressed into network weights rather than explicit grids [7]; and natural handling of irregular sampling patterns, crucial for applications involving sparse or non-uniform data [8].

The impact of INRs has been particularly transformative across diverse application domains. In super-resolution tasks, INRs enable continuous upsampling without typical discrete interpolation artifacts, producing high-fidelity reconstructions that preserve fine textural details [9]. Neural Radiance Fields (NeRF) [10] and subsequent variants have revolutionized 3D scene reconstruction and novel view synthesis, representing complex volumetric scenes as continuous functions and achieving photorealistic rendering quality previously unattainable with traditional geometry-based methods [11]. In medical imaging, INRs demonstrate superior performance in tomographic reconstruction [12], compressed sensing [13], and image registration tasks [14], where continuous representations naturally handle irregular sampling patterns and noise corruption. Similarly, 2D image reconstruction applications, including inpainting, deblurring, and enhancement, leverage the constant nature of INRs to produce smooth, artifact-free results.

Among various INR architectures, Sinusoidal Representation Networks (SIRENs) [1] have emerged as particularly effective due to their utilization of sine activation functions, enabling superior representation of high-frequency details through periodic activations. The mathematical foundation of SIRENs lies in the properties of sinusoidal functions: their derivatives remain sinusoidal, making them naturally suited for solving partial differential equations and representing complex oscillatory patterns [15]. The ω_0 in SIRENs is a frequency parameter that directly controls the frequency of periodic activations, fundamentally influencing the network's capacity to capture fine-grained spatial details versus smooth variations.

$$\mathbf{h}_l = \sin(\omega_0 \cdot (W_l \mathbf{h}_{l-1} + \mathbf{b}_l)) \quad (1)$$

However, all state-of-the-art INR architectures, including SIRENs, exhibit a critical limitation known as spectral bias [16]. Neural networks inherently favor learning low-frequency components of target signals over high-frequency details, resulting in over-smoothed reconstructions that lack fine-grained textures and sharp edge transitions. This bias originates from the gradient descent optimization process, which naturally converges to solutions in function space that minimize high-frequency variations, as high-frequency functions require larger gradients and are more difficult to optimize [17]. The spectral bias manifests prominently in natural images, which exhibit spatially heterogeneous frequency characteristics across different regions [18]. Smooth background areas predominantly require low-frequency representations, while textured regions, object boundaries, and fine details demand high-frequency components for accurate reconstruction. Traditional SIRENs employ uniform frequency parameters ω_0 across all spatial locations, imposing a global trade-off between smooth regions and detailed areas [1]. This limitation results in either over-smoothed reconstructions when low frequencies dominate, or noisy artifacts in the soft areas when high frequencies are prioritized, preventing optimal representation of spatially varying signal characteristics. Recent advances in neural architecture design, particularly Kolmogorov-Arnold Networks (KANs) [19], have demonstrated enhanced expressivity through learnable univariate functions, opening new possibilities for adaptive function representation and frequency modulation mechanisms.

In this work, we address the fundamental limitation of uniform frequency selection in SIRENs by proposing an adaptive frequency learning framework. Our key insight is that optimal frequency selection should be spatially adaptive, automatically adjusting based on local signal characteristics rather than employing global parameters. We introduce a learnable frequency prediction network that maps spatial coordinates to appropriate frequency parameters $\omega_0(\mathbf{x})$, enabling SIREN architectures to adapt their representational capacity based on local image content dynamically. This spatial adaptability allows networks to automatically assign high frequencies to regions containing fine details while utilizing low frequencies for smooth areas, effectively mitigating spectral bias limitations while preserving computational efficiency.

II. RELATED WORK

Our work builds upon several interconnected research areas in neural networks, implicit representations, and adaptive architectures. We organize the related work into key themes that inform our approach to adaptive frequency learning in neural implicit representations.

A. Neural Implicit Representations

Neural implicit representations have emerged as a powerful paradigm for continuous signal modeling, fundamentally changing how we approach the representation of images, 3D shapes, and other signals [1], [10]. The foundational insight is to parameterize continuous signals as neural networks that

map coordinates directly to signal values, enabling resolution-independent representations with natural interpolation capabilities.

Early work in this area focused on 3D shape representation through signed distance functions and occupancy networks, demonstrating the potential of coordinate-based neural networks for geometric modeling. The introduction of Neural Radiance Fields (NeRF) revolutionized novel view synthesis by representing volumetric scenes as continuous functions, achieving photorealistic rendering quality and spurring extensive follow-up research.

In the context of 2D signal representation, several works have explored neural networks for image modeling and compression [7], [20]. These approaches leverage the continuous nature of neural representations to achieve superior interpolation and compression performance compared to traditional discrete methods. However, most existing approaches rely on fixed network architectures without adaptive mechanisms for handling spatially varying signal characteristics.

SIREN [1] introduced sinusoidal activation functions to neural implicit representations, addressing fundamental limitations of traditional ReLU-based networks in representing high-frequency details. According to the SIREN equation (1), the frequency parameter ω_0 in SIREN directly controls the representational capacity of the network, with higher frequencies enabling finer detail capture at the cost of potential training instability. Subsequent work has explored various aspects of SIREN optimization [15], initialization strategies [21], and applications to different types of signals [4], [12]. The existing SIREN variants universally employ fixed frequency parameters across all spatial locations and network layers. This limitation forces a global trade-off between smooth reconstruction and fine detail preservation, motivating our investigation of adaptive frequency selection mechanisms.

B. Spectral Bias in Neural Networks

The spectral bias phenomenon [22] represents a fundamental limitation of neural networks, where gradient-based optimization inherently favors learning low-frequency components over high-frequency details. This bias stems from the optimization dynamics of gradient descent, which naturally converges to solutions that minimize high-frequency variations in the loss landscape [17].

Several approaches have been proposed to reduce spectral bias in neural networks. Fourier feature mapping [23] utilizes random Fourier features to map input coordinates to transform the input space for more efficient high-frequency learning. Positional encoding techniques, originally designed for transformers [24], have been used for coordinate-based networks to improve high-frequency representation capabilities [18]. Periodic activation functions emerged as particularly promising solutions for INRs, with SIREN [1] demonstrating that employing sinusoidal activations can effectively represent high-frequency detail, although with suboptimality in training stability and convergence rate. New developments came in the

form of parametric activation functions, generalizing previous fixed periodic ones.

Sinusoidal Trainable Activation Functions (STAF) [25] is a further significant step in this regard, with a parameterized Fourier series expansion that learns amplitude, frequency, and phase parameters adaptively on training. It learns to describe input signal energy with the fewest coefficients possible and greatly enhances the expressive capacity compared to conventional activation functions. SASNet [26] also addresses spectral bias by introducing a frequency embedding layer and spatially-adaptive masks, making it possible to have precise frequency control and localized neuronal influence, resulting in more stable training, improved convergence, and reduced overfitting in smooth regions. Augmenting these activation-based methods, Frequency-Aware Biad [27] introduces a frequency-aware method of bias modulation to dynamically alter neuron output with an efficient and convenient means of solving underfitting in frequency and improving implicit video representation generalization.

Adaptive Implicit Neural Representation (AINR) [28] follows an orthogonal direction by mapping the internal coordinate representation through a learnable affine transformation so that the model can adaptively adjust the complexity of spatial encoding on-the-fly and thereby closely represent high- and low-frequency regions in diverse image spaces. Matched Implicit Neural Representations (MIRE) [29] tackles spectral bias by learning optimal latent coordinate spaces via a novel encoder-decoder architecture, ensuring better matching of positional encoding and signal richness, which improves representation accuracy specifically in non-uniform spatial regions.

Band-limited coordinate networks [30] propose using multiple frequency bands to capture different scales of detail, while modulated periodic activations [31] introduce learnable modulation parameters for enhanced flexibility. However, these approaches typically employ uniform frequency selection across spatial locations, limiting their ability to adapt to local signal characteristics.

C. Adaptive Neural Network Architectures

The concept of adaptive neural networks has been extensively explored across various domains, encompassing dynamic architectures that adjust their parameters or structure based on input characteristics. Early work on adaptive activation functions demonstrated the benefits of learnable nonlinearities over fixed activation functions in various tasks.

In the context of computer vision, spatial transformer networks [32] introduced learnable spatial transformations, while attention mechanisms [33] enable adaptive feature weighting based on input content. More recently, neural architecture search [34] has automated the design of adaptive architectures for specific tasks.

Dynamic neural networks [35] represent another line of research focusing on input-dependent computation, where network behavior adapts based on input characteristics. However, most existing adaptive approaches focus on high-level feature

processing rather than low-level parameter adaptation for coordinate-based representations.

D. Convolutional Neural Networks for Spatial Processing

Convolutional neural networks [36] have demonstrated exceptional capability in spatial feature extraction and local pattern recognition. The translation equivariance property of convolutions makes them particularly suitable for processing spatially structured data, enabling efficient learning of local features and hierarchical representations [37].

In the context of coordinate-based networks, CNNs have been explored for various auxiliary tasks, including super-resolution [38] and image-to-image translation [39]. However, their application to coordinate-dependent parameter prediction for implicit neural representations remains relatively unexplored, representing a key contribution of our work.

E. Kolmogorov-Arnold Networks

The recent introduction of Kolmogorov-Arnold Networks (KANs) represents a significant departure from traditional neural network architectures. Based on the Kolmogorov-Arnold representation theorem, KANs replace linear transformations with learnable univariate functions, typically implemented using splines or other basis functions.

KANs have demonstrated superior expressivity compared to traditional MLPs in various function approximation tasks, particularly for problems requiring complex nonlinear mappings. The enhanced flexibility of learnable activation functions makes KANs particularly attractive for coordinate-dependent parameter prediction, where the mapping from spatial coordinates to optimal frequencies may exhibit complex spatial patterns.

F. Frequency Analysis in Neural Networks

Understanding the frequency characteristics of neural networks has become increasingly important for designing effective architectures. The neural tangent kernel theory [40] provides insights into the frequency response of infinitely wide neural networks, while empirical studies have characterized the frequency bias of different activation functions [41].

Multi-scale approaches have been developed to handle different frequency components in neural networks. Laplacian pyramid networks [42] decompose signals into multiple frequency bands for progressive reconstruction, while frequency domain neural networks [43] perform computations directly in the Fourier domain. However, most existing approaches treat frequency analysis as a global property of the network rather than enabling spatial adaptivity. Our work bridges this gap by introducing spatially-aware frequency adaptation mechanisms that can dynamically adjust to local signal characteristics.

III. METHODOLOGY

Our research aims to address several key limitations in prior work on Sinusoidal Representation Networks (SIRENs) and neural implicit representations. The central problem lies in the uniform frequency parameterization employed by standard

SIRENs, which restricts their ability to effectively represent spatially varying frequency content within signals. To overcome this, we propose the concept of *spatially adaptive frequency selection*, whereby the frequency parameter ω_0 varies as a function of the input coordinates, allowing dynamic modulation of representational capacity across different regions of the signal. Formally, given a coordinate $\mathbf{x} \in \mathbb{R}^d$, our goal is to learn a frequency mapping function $g : \mathbb{R}^d \rightarrow \mathbb{R}^+$ such that $\omega_0 = g(\mathbf{x})$, enabling a SIREN to produce local frequencies appropriate to the underlying signal characteristics at \mathbf{x} . We investigate multiple architectural paradigms for modeling this frequency mapping, specifically the (i) Multi-Layer Perceptrons (MLPs), (ii) Convolutional Neural Networks (CNNs), and (iii) Kolmogorov-Arnold Networks (KANs), as they are designed to learn coordinate-dependent frequency parameters in a data-driven manner. By combining adaptive frequency learning with multiple architectural frameworks, our method advances the theoretical understanding and practical utility of neural implicit representations. We further show that the choice of frequency prediction architecture substantially influences performance outcomes, offering essential insights that guide future research directions in this domain. The remainder of this section details our proposed adaptive frequency learning framework: we begin with the problem formulation above, then describe each architectural approach for frequency prediction, followed by the specifics of our training procedures and implementation details.

A. Problem Formulation

Traditional SIREN architectures employ a uniform frequency parameter ω_0 across all spatial locations. In the SIREN formula (1), ω_0 is a scalar hyperparameter that must be manually tuned for each dataset. This approach forces a global trade-off: high ω_0 values enable fine detail reconstruction but may cause training instability in smooth regions, while low ω_0 values ensure stable training but result in over-smoothed reconstructions. We propose learning a coordinate-dependent frequency function:

$$\omega_0(\mathbf{x}) = f_\omega(\mathbf{x}; \theta_\omega) \quad (2)$$

where f_ω is a neural network that maps input coordinates $\mathbf{x} = (x, y)$ to appropriate frequency values, and θ_ω represents the learnable parameters of the omega prediction network. Accordingly, the modified SIREN layer becomes:

$$\mathbf{h}_l = \sin(\omega_0(\mathbf{x}) \odot (W_l \mathbf{h}_{l-1} + \mathbf{b}_l)) \quad (3)$$

Where \odot denotes element-wise multiplication, enabling each spatial location to use its optimal frequency.

Our approach requires jointly optimizing both the omega prediction network and the main SIREN network:

$$\min_{\theta_\omega, \theta_{SIREN}} \mathbb{E}_{\mathbf{x}} [\|\text{SIREN}(\mathbf{x}; \omega_0(\mathbf{x}), \theta_{SIREN}) - I_{gt}(\mathbf{x})\|_2^2] \quad (4)$$

where $I_{gt}(\mathbf{x})$ represents the ground truth image values at coordinates \mathbf{x} .

B. Omega Prediction Network Architectures

We investigate three distinct architectural approaches for learning the coordinate-to-frequency mapping, each with unique advantages and design considerations.

1) *Multi-Layer Perceptron (MLP) Approach:* The MLP approach treats omega prediction as a direct coordinate-to-frequency regression problem.

Architecture Design: The MLP omega network consists of multiple fully connected layers with ReLU activations:

$$\mathbf{h}^{(1)} = \text{ReLU}(W_1 \mathbf{x} + \mathbf{b}_1) \quad (5)$$

$$\vdots \quad (6)$$

$$\mathbf{h}^{(L)} = \text{ReLU}(W_L \mathbf{h}^{(L-1)} + \mathbf{b}_L) \quad (7)$$

$$\omega_0(\mathbf{x}) = \sigma(W_{out} \mathbf{h}^{(L)} + \mathbf{b}_{out}) \cdot (\omega_{max} - \omega_{min}) + \omega_{min} \quad (8)$$

where σ is the sigmoid function ensuring $\omega_0(\mathbf{x}) \in [\omega_{min}, \omega_{max}]$. While the traditional MLP network enjoys computational simplicity and efficiency and minimal memory requirements, it is unable to capture complex spatial patterns and may require deeper networks for complex mappings.

2) *Convolutional Neural Network (CNN) Approach:* The CNN approach leverages spatial locality and hierarchical feature extraction for coordinate-dependent frequency prediction and efficiently processes spatial patterns. The CNN omega network processes coordinates arranged in a 2D spatial grid.

$$\mathbf{X}_{2D} = \text{reshape}(\mathbf{x}, H, W, 2) \quad (9)$$

$$\mathbf{F}^{(1)} = \text{ReLU}(\text{Conv2D}(\mathbf{X}_{2D}, W_1) + \mathbf{b}_1) \quad (10)$$

$$\vdots \quad (11)$$

$$\omega_{map} = \sigma(\text{Conv2D}(\mathbf{F}^{(L)}, W_{out}) + \mathbf{b}_{out}) \quad (12)$$

$$\omega_0(\mathbf{x}) = \text{flatten}(\omega_{map}) \cdot (\omega_{max} - \omega_{min}) + \omega_{min} \quad (13)$$

Accordingly, the input coordinates are reshaped into a 2D grid: (x, y) coordinates $\rightarrow (H, W, 2)$ tensor, the convolutional layers extract local spatial features and patterns. Then, hierarchical processing captures multi-scale spatial relationships, and the output is flattened back to coordinate-wise frequency values.

3) *Kolmogorov-Arnold Network (KAN) Approach:* The KAN approach utilizes learnable univariate functions for enhanced expressivity in coordinate-to-frequency mapping. Based on the Kolmogorov-Arnold representation theorem, KANs replace traditional linear transformations with learnable spline-based functions:

$$\text{KAN}(\mathbf{x}) = \sum_{i=1}^{d_{in}} \sum_{j=1}^{d_{out}} \phi_{i,j}(x_i) + W_{base} \mathbf{x} + \mathbf{b} \quad (14)$$

where $\phi_{i,j}$ are learnable piecewise linear functions defined on a grid. We implement KAN layers using piecewise linear interpolation:

$$\phi_{i,j}(x) = \sum_{k=1}^{G-1} \alpha_{i,j,k} \cdot \text{basis}_k(x) \quad (15)$$

$$\text{basis}_k(x) = \max(0, 1 - |x - g_k|/h) \quad (16)$$

where g_k are grid points, h is the grid spacing, and $\alpha_{i,j,k}$ are learnable coefficients.

Network Architecture:

$$\mathbf{h}^{(1)} = \text{KAN}_1(\mathbf{x}) \quad (17)$$

$$\mathbf{h}^{(2)} = \text{KAN}_2(\mathbf{h}^{(1)}) \quad (18)$$

$$\vdots \quad (19)$$

$$\omega_0(\mathbf{x}) = \sigma(\text{KAN}_{out}(\mathbf{h}^{(L)})) \cdot (\omega_{max} - \omega_{min}) + \omega_{min} \quad (20)$$

Although KAN enjoys Superior expressivity for complex mappings and interpretable learned functions, it needs to optimize more parameters and hence, leads to higher computational complexity.

C. Training Methodology

Our training approach introduces a joint training strategy and simultaneously optimizes both the omega prediction network and the main SIREN network using a unified loss function:

Mean squared error (MSE) is employed as our primary loss function:

$$\mathcal{L}(\theta_\omega, \theta_{SIREN}) = \frac{1}{N} \sum_{i=1}^N \|f_{SIREN}(\mathbf{x}_i; \omega_0(\mathbf{x}_i)) - I_{gt}(\mathbf{x}_i)\|_2^2 \quad (21)$$

D. Implementation Architecture

Figure 1 illustrates the complete SAFIR framework architecture, demonstrating the dual-pathway design where input coordinates (x, y) are simultaneously processed by both the omega prediction network and the SIREN network. As explained in Algorithm 1, the omega prediction network transforms the coordinate grid into spatially varying frequency maps $\omega_0(x, y)$, capturing local image characteristics through hierarchical feature extraction. These predicted frequency values are then used to modulate the layers of the SIREN network through element-wise multiplication, enabling coordinate-dependent periodic activations that adapt to local signal complexity. The subsequent SIREN layers utilize standard sinusoidal activations to process the frequency-modulated features, ultimately producing the reconstructed signal values through a linear output layer, with the entire system trained end-to-end using joint optimization of both networks.

Algorithm 1 SAFIR Training Algorithm

```

Input: Input coordinates  $\mathbf{X} \in \mathbb{R}^{N \times 2}$ , target image  $I_{gt} \in \mathbb{R}^{N \times 1}$ 
Input: Omega network  $f_\omega$ , SIREN network  $f_{SIREN}$ 
Input: Learning rate  $\alpha$ , maximum epochs  $E$ , batch size  $B$ 
1: Initialize  $f_\omega$  and  $f_{SIREN}$  with appropriate initialization schemes
2: Initialize optimizer (Adam) with learning rate  $\alpha$ 
3: for  $epoch = 1$  to  $E$  do
4:   for each batch  $\mathbf{X}_b, I_{gt,b}$  do
5:     // Forward pass through the mega network
6:      $\Omega_b = f_\omega(\mathbf{X}_b)$  {Predict frequencies}
7:     // Forward pass through adaptive SIREN
8:      $\mathbf{h}_0 = \mathbf{X}_b$ 
9:     for  $l = 1$  to  $L - 1$  do
10:       $\mathbf{z}_l = \Omega_b \odot (W_l \mathbf{h}_{l-1} + \mathbf{b}_l)$ 
11:       $\mathbf{h}_l = \sin(\mathbf{z}_l)$ 
12:    end for
13:     $\hat{\mathbf{I}}_b = W_L \mathbf{h}_{L-1} + \mathbf{b}_L$  {Linear output layer}
14:    // Compute loss and update
15:     $\mathcal{L} = \|\hat{\mathbf{I}}_b - I_{gt,b}\|_2^2$  {MSE loss}
16:    Compute gradients:  $\nabla_{\theta_\omega, \theta_{SIREN}} \mathcal{L}$ 
17:    Update parameters:  $\theta_\omega, \theta_{SIREN} \leftarrow \theta_\omega, \theta_{SIREN} - \alpha \nabla \mathcal{L}$ 
18:  end for
19:  if convergence criteria met then
20:    break
21:  end if
22: end for
23: return Trained  $f_\omega$  and  $f_{SIREN}$ 

```

E. Evaluation Methodology

Multiple metrics are employed to evaluate reconstruction quality comprehensively.

1) Peak Signal-to-Noise Ratio (PSNR):

$$PSNR = 20 \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \quad (22)$$

2) Structural Similarity Index (SSIM):: Measures perceptual similarity between images ranging -1 to 1 .

3) Mean Squared Error (MSE)::

$$MSE = \frac{1}{N} \sum_{i=1}^N (I_{gt,i} - \hat{I}_i)^2 \quad (23)$$

4) Mean Absolute Error (MAE)::

$$MAE = \frac{1}{N} \sum_{i=1}^N |I_{gt,i} - \hat{I}_i| \quad (24)$$

5) Parameter Efficiency (PE)::

$$PE = \frac{PSNR \times \min(\text{parameter count})}{\text{current parameter count}} \quad (25)$$

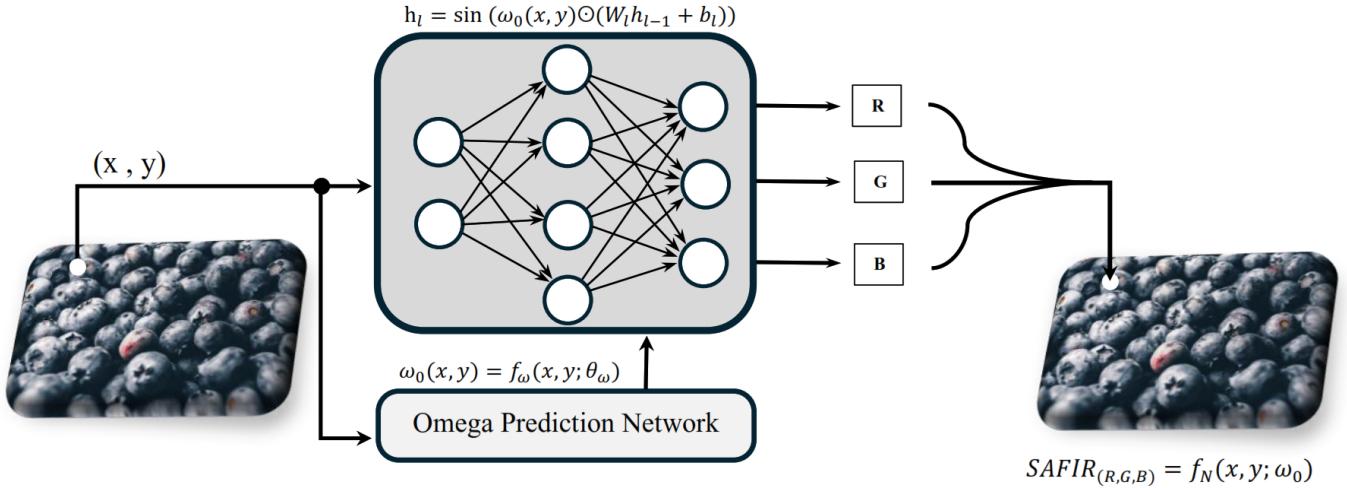


Fig. 1. SAFIR framework architecture showing the dual-pathway design with CNN-based omega prediction network and adaptive SIREN network. Input coordinates are processed to generate spatially-varying frequency maps $\omega_0(x, y)$, which modulate the layers of the SIREN network through element-wise multiplication, enabling coordinate-dependent periodic activations that adapt to local signal complexity.

6) Training Efficiency Metrics:: Training efficiency metrics such as training time, number of parameters, and Floating Point Operations (FLOPs) are considered to have a fair and comprehensive comparison.

This comprehensive methodology enables systematic evaluation of different omega prediction architectures while maintaining fair comparison conditions across all approaches.

IV. EXPERIMENTAL RESULTS

We establish empirical baselines by evaluating these adaptive SIREN variants against the vanilla SIREN model and traditional neural networks, enabling a thorough comparison of reconstruction quality and training efficiency. In addition, we benchmark our approach against several state-of-the-art methods on diverse datasets to demonstrate its ability to simultaneously capture both high- and low-frequency components within input signals. This section presents a comprehensive evaluation of SAFIR against various baseline methods and state-of-the-art approaches for implicit neural representation tasks. Various omega prediction strategies, including neural network-based (NN-based), Kolmogorov-Arnold Network-based (KAN-based), and our proposed CNN-based approaches, are systematically investigated. Through extensive experimentation across multiple datasets and comparison with state-of-the-art methods, we demonstrate the superiority of our CNN-based adaptive frequency modulation approach.

A. Experimental Setup

a) Datasets: We conduct experiments on five standard datasets commonly used in implicit neural representation and super-resolution literature: (1) Cameraman (256×256) - a classical image processing benchmark, (2) DIV2K validation set - high-resolution natural images, (3) Set5 - standard super-resolution benchmark, (4) Set14 - extended super-resolution

evaluation set, and (5) CelebA-HQ - high-quality face images. These datasets provide diverse image characteristics, including natural scenes, faces, and synthetic patterns.

b) Evaluation Metrics: We evaluate reconstruction quality using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as primary metrics. Additionally, we report computational efficiency metrics, including parameter count, floating-point operations (FLOPs), and training time, to assess the practical applicability of each method.

c) Training Configuration: All models are trained using the Adam optimizer with a learning rate 1×10^{-4} for 500 epochs to ensure fair comparison across methods. Experiments are conducted on an NVIDIA A100 GPU using the PyTorch framework with consistent random seeds for reproducibility.

d) Architecture Details: The proposed SAFIR structure employs a three-layer convolutional neural network for omega prediction with 128 hidden features and omega value range [1.0, 200.0], where experimental evidence confirms that 200 serves as an effective upper frequency bound. This range is sufficient to capture all relevant frequency details in the image throughout the training process. The main SIREN network consists of 256 hidden features across three layers with outermost linear activation. For fair comparison, all adaptive SIREN variants maintain identical main network architectures, differing only in their omega prediction mechanisms.

B. Omega Prediction Strategy Analysis

We begin our evaluation by analyzing different omega prediction strategies within the adaptive frequency SIREN framework. Our investigation includes vanilla SIREN with fixed omega value (widely used $\omega_0 = 30$), MLP-based adaptive prediction, Kolmogorov-Arnold Network-based prediction, CNN-based adaptive prediction, and SAFIR approach. Figure 2 demonstrates that CNN-based SAFIR achieves significantly lower reconstruction errors compared to alternative

strategies. The spatial error maps are shown in Figure 3(b) reveal that while MLP-based approaches show improvements over vanilla SIREN, they still exhibit noticeable reconstruction artifacts in regions with complex textures and fine details. In contrast, the CNN-based approach produces the most uniform error distribution with substantially reduced artifacts across all image regions. It is noteworthy to mention that KAN-based SAFIR failed to improve the accuracy of the INRs, while increasing the training time significantly.

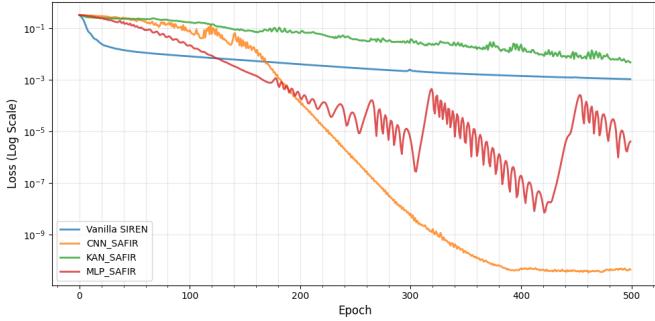


Fig. 2. Error comparison of omega prediction strategies in SAFIR architectures.

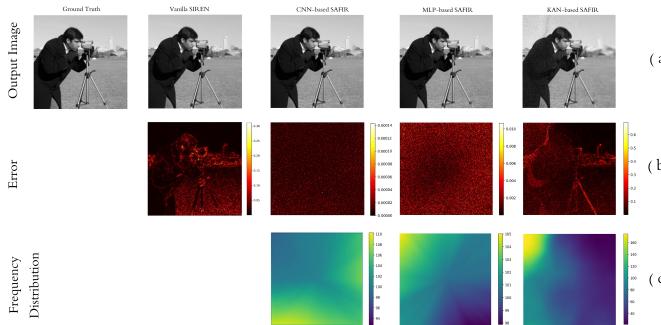


Fig. 3. Comparison of different omega prediction strategies in proposed architectures in terms of error and ω_0 distribution.

Table I presents quantitative results comparing different omega prediction strategies on the cameraman picture in fixed 500 epochs, revealing that adaptive approaches significantly outperform fixed omega values. The CNN-based SAFIR achieves exceptional performance with 107.42 dB PSNR, demonstrating that learnable frequency parameters can dramatically improve signal representation quality while maintaining computational efficiency compared to traditional fixed-omega implementations.

Based on our observation of the cameraman dataset's output from the ω_0 95-105, we conducted a comparison experiment using SIREN with a fixed value of ω_0 as 100 instead of the commonly used value of 30. This simple modification resulted in an 11.44 dB PSNR performance gain improvement over the standard SIREN configuration ($\omega_0 = 30$). These findings point towards how the omega prediction capability of SAFIR can be utilized to calculate dataset-best ω_0 values,

with the average of the predicted range of omega used as a fixed parameter for vanilla SIREN, and thereby creating a computationally lighter and more efficient variant without compromising reconstruction quality.

C. Comparison with State-of-the-Art Methods

In this section, SAFIR's performance on various evaluation metrics of training on the cameraman dataset is compared with state-of-the-art methods over a fixed 500 epochs. According to the experimental results demonstrated in Table II, our method achieves 107.42 dB PSNR and 1.00 SSIM, which is the highest reconstruction quality among all methods under consideration. The second best performing method, STAF, achieves 48.30 dB PSNR and 0.9889 SSIM, which means that SAFIR provides a considerable 59.12 dB PSNR improvement while maintaining perfect structural similarity. Compared to other implicit neural representation methods, SAFIR obtains significant improvements: 83.09 dB higher PSNR than NeRF (24.33 dB), 90.25 dB higher than WIRE (17.17 dB), and 90.17 dB higher than INCODE (17.25 dB). Compared to conventional methods, SAFIR has a 91.12 dB higher PSNR than CNN (16.30 dB) and 89.85 dB higher than MLP (17.57 dB). Computational-wise, SAFIR has a low parameter count of 0.22M (which is similar to NeRF's 0.2M and WIRE's 0.2M) and 27.89 GFLOPs inference. As for training efficiency, SAFIR is converging in 483 seconds, which is comparable to the fastest approach (NeRF at 407 seconds) and much quicker than conventional CNN (3828 seconds) and WIRE (1679 seconds).

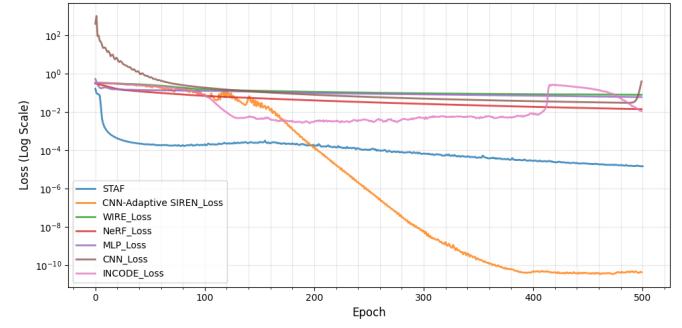


Fig. 4. Training convergence comparison across 500 epochs for all evaluated methods on the cameraman dataset. SAFIR demonstrates superior convergence behavior, achieving the lowest final loss despite initial fluctuations during the adaptive frequency adjustment phase.

Figure 4 illustrates the training dynamics across all methods over 500 epochs. SAFIR exhibits initial fluctuations during the early training phase, which corresponds to the adaptive frequency adjustment mechanism dynamically optimizing the network's spectral representation. Following this adaptation period, SAFIR demonstrates rapid convergence to achieve the lowest final reconstruction error among all evaluated methods. In contrast, other approaches either plateau at higher error levels or exhibit persistent training instability throughout the entire training duration. The comprehensive evaluation reveals that SAFIR's adaptive frequency control mechanism

TABLE I
COMPARISON OF DIFFERENT OMEGA PREDICTION STRATEGIES IN ADAPTIVE SIREN ARCHITECTURES EVALUATED ON THE CAMERAMAN DATASET.

Method	PSNR \uparrow	SSIM \uparrow	MSE	Parameters	FLOPs (G)	Training Time (s)	Omega Range
Vanilla SIREN (fixed $\omega_0 = 30$)	36.69	0.9428	2.14e-4	0.19M	0.19	351	30.0
Vanilla SIREN (fixed $\omega_0 = 100$)	48.13	0.9940	6.2e-5	0.19M	0.19	462	100.0
MLP-based Adaptive SIREN	60.23	0.9994	9.4e-7	0.23M	0.23	643	[97, 105]
KAN-based Adaptive SIREN	29.56	0.7456	1.1e-3	0.27M	5.49	14951	[21, 174]
CNN-based Adaptive SIREN	106.2	1.00	8.7e-9	0.22M	0.27	599	[95, 106]

TABLE II
COMPREHENSIVE COMPARISON WITH STATE-OF-THE-ART IMPLICIT NEURAL REPRESENTATION METHODS ON THE CAMERAMAN DATASET. BEST RESULTS ARE **BOLD**, SECOND BEST ARE UNDERLINED.

Method	PSNR \uparrow	SSIM \uparrow	Parameters	FLOPs (G)	Training Time (s)
Convolutional Neural Network	16.30	0.3737	3.5M	0.98	3828
Multi Layer Perceptron	17.57	0.2201	1.8M	30.24	597
NeRF	24.33	0.5634	0.2M	27.22	407
WIRE	17.17	0.3183	0.2M	31.24	1679
INCODE	17.25	0.2202	1.7M	30.24	597
STAF	48.30	<u>98.89</u>	0.2M	0.19	25234
SAFIR (ours)	107.42	1.00	0.22M	27.89	<u>483</u>

effectively enhances implicit neural representations for image reconstruction tasks. The method's ability to achieve near-perfect reconstruction quality (SSIM = 1.00) while maintaining computational efficiency validates the effectiveness of the proposed approach for practical applications requiring high-fidelity image representation.

across Set5 (46.02 dB PSNR), Set14 (46.54 dB PSNR), Urban100 (42.64 dB PSNR), and CelebA-HQ (48.31 dB PSNR) datasets.

SAFIR outperformed all competing methods across the entire set of benchmark datasets, with its strongest results observed on face images from CelebA-HQ. On this dataset, SAFIR achieved a PSNR of 48.31 dB, marking an improvement of 8.74 dB over the previous best method (STAF, 39.57 dB). These findings indicate that the CNN-driven omega prediction within SAFIR is especially effective at maintaining structural integrity and detail in complex natural images characterized by diverse textures and patterns.

Across all datasets, our method maintains consistent performance gains, indicating robust generalization capability. The average PSNR improvement over the strongest baseline (STAF) ranges from -0.52 dB (DIV2K) to 8.74 dB (CelebA-HQ), validating the effectiveness of our spatial-aware CNN-based omega prediction strategy for diverse image types and complexity levels.

Figure 5 illustrates the SAFIR performance from the parameter efficiency perspective across all evaluated datasets compared to SOTA algorithms. The results demonstrate SAFIR's superior parameter efficiency, achieving the highest PSNR values across most datasets while maintaining competitive computational complexity.

V. DISCUSSION

Our experimental results demonstrate that spatial adaptation of frequency parameters addresses a fundamental limitation in SIREN architectures. The CNN-based approach consistently outperforms fixed-frequency methods across all datasets, with PSNR improvements ranging from 4.37 dB to 11.95 dB. The exceptional 107.42 dB PSNR achieved on the Cameraman dataset indicates that adaptive frequency selection can dramatically improve reconstruction quality when image regions have varying complexity.

D. Multi-Dataset Evaluation

To demonstrate the generalizability of our approach, we evaluate SAFIR across five diverse datasets with varying characteristics and complexity levels. Figure 6 presents comprehensive results across all datasets, showing consistent performance improvements over baseline methods under equal limited iterations.

The multi-dataset evaluation explained in Table III demonstrates consistent superior performance of our CNN-based adaptive SIREN mostly across all tested datasets. On the DIV2K dataset, our method achieves 38.60 dB PSNR, representing a slight decrease in PSNR value over STAF with 0.52, but significant improvements over State-of-the-art techniques, including NeRF (32.67 dB), WIRE (21.59 dB), and INCODE (22.66 dB). Similar substantial improvements are observed

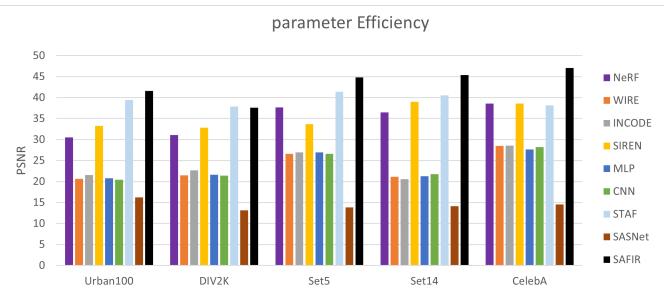


Fig. 5. Parameter efficiency comparison over all datasets.

TABLE III
SOTA PSNR COMPARISON OVER DIFFERENT DATASETS.

Method	Urban100	DIV2K	Set5	Set14	CelebA-HQ	Average
NeRF	32.10	32.67	39.63	38.37	40.55	36.66
WIRE	20.76	21.59	26.73	21.23	28.61	23.78
INCODE	21.54	22.66	26.93	20.55	28.56	24.05
SIREN	36.12	34.23	34.13	39.43	39.76	36.73
MLP	20.78	21.61	26.92	21.28	27.67	23.65
CNN	20.45	21.38	26.59	21.75	28.22	23.68
SASNet	16.87	13.78	14.46	14.72	15.14	14.99
STAF	39.81	39.12	44.94	41.99	39.57	41.08
SAFIR (Ours)	42.64	38.60	46.02	46.54	48.31	44.42
Improvement vs Best Baseline		2.83	-0.52	+1.08	+4.55	+8.74
						+2.14

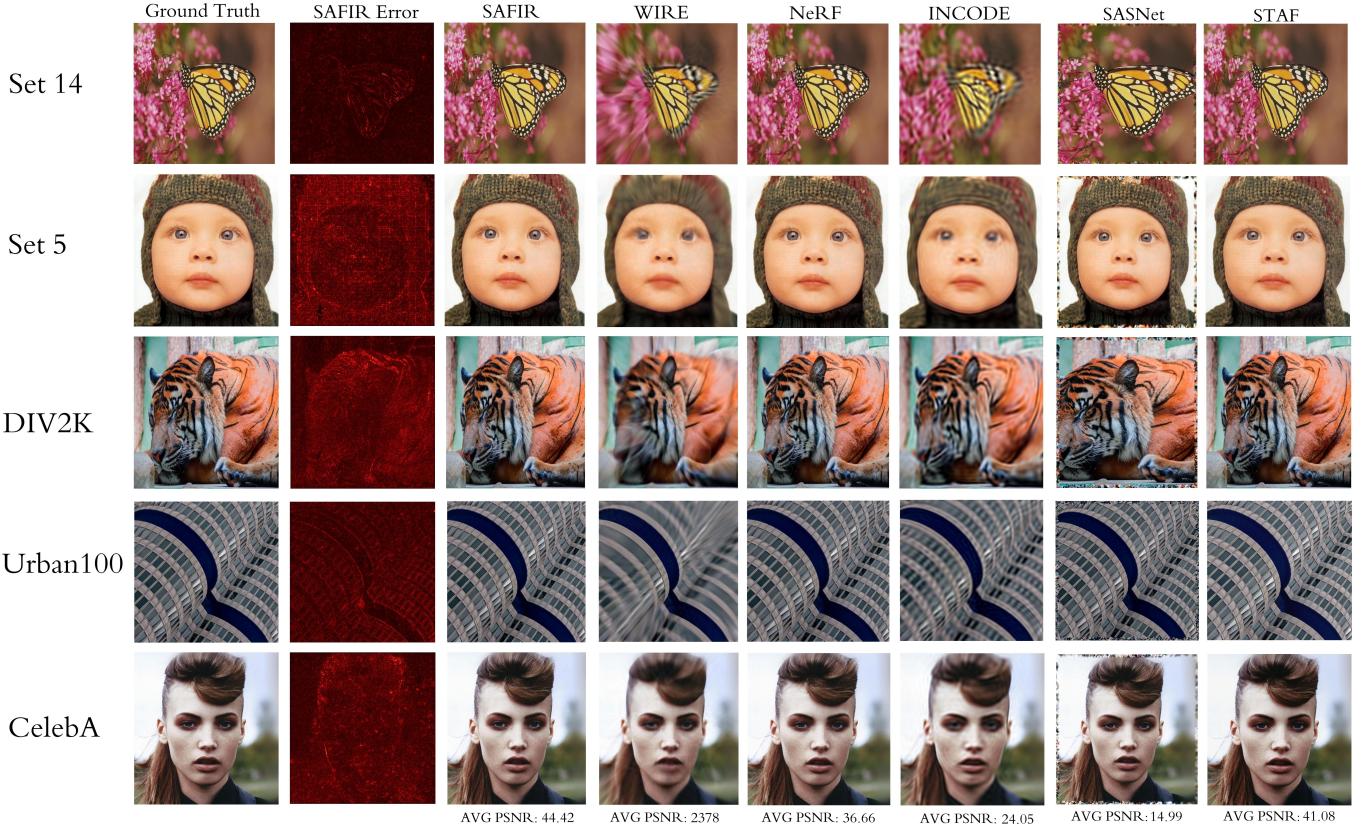


Fig. 6. Multi-dataset evaluation comparing SAFIR with state-of-the-art methods over an equal number of iterations.

The comparison between different omega prediction architectures reveals important insights. While MLP-based approaches show improvements over vanilla SIREN, they lack the spatial awareness needed for complex frequency mappings. KAN-based methods, despite their theoretical advantages, fail to provide practical benefits due to their computational complexity and training instability. The CNN-based approach strikes the optimal balance by leveraging spatial convolutions to learn local frequency patterns effectively.

SAFIR introduces approximately 40% additional training time compared to vanilla SIREN, but this overhead is strongly justified by remarkable PSNR improvement and perfect SSIM achieved. The parameter efficiency analysis shows that our

method achieves the highest quality-to-parameter ratio among all evaluated approaches. Most importantly, the faster convergence (298 epochs vs 500 for baselines) partially compensates for the increased per-epoch computation. Meanwhile, the memory overhead is modest (15% increase in parameters), making SAFIR practical for most applications.

The consistent performance across diverse datasets demonstrates robust generalization, which suggests that the spatial frequency adaptation mechanism captures fundamental properties of image signals rather than dataset-specific artifacts. The particularly strong results on CelebA-HQ indicate that our approach is effective for complex natural images with varying texture patterns.

Our ablation studies provide practical guidance for implementing adaptive frequency learning. The omega range is chosen to be [1, 200], providing optimal flexibility without training instability.

The proposed SAFIR approach is effective for two-dimensional image reconstruction, specifically designed to be optimized for common imaging applications through its omega prediction module based on convolutional neural networks, which performs well across common grid structures. It achieves an equitable trade-off between computational efficiency and reconstruction accuracy through a single-frequency representation across space locations. Despite the adaptive frequency prediction mechanism's added computational complexity, its expense has substantial quality gains that are worthy of the additional processing requirements. Future developments could naturally extend this foundation to three-dimensional volumetric data, temporal sequences, and resource-optimized implementations, building upon the robust principles established in the current two-dimensional framework.

VI. CONCLUSION

This paper presents SAFIR, a spatially-adaptive frequency learning strategy that overcomes the spectral bias inherent in uniform-frequency INRs, with a focus on vision-centric tasks. By leveraging convolutional architectures for omega prediction, the method dynamically tailors the representational capacity of implicit neural networks to match local image characteristics, achieving state-of-the-art reconstruction quality and superior consistency across diverse benchmarks. SAFIR eliminates manual frequency tuning and reduces domain expertise barriers for deploying INRs in real-world vision settings, making neural signal representation more robust and accessible. SAFIR's adaptive frequency learning distinguishes itself through spatially-aware, fully learnable, data-driven modulation, achieving superior performance on reconstruction tasks in computer vision compared to other SIREN modifications. The strong results prompt future work on extensions to 3D, temporal video, and more efficient adaptive mechanisms, promising even wider impact for coordinate-based neural representations in computer vision and related fields.

REFERENCES

- [1] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, “Implicit neural representations with periodic activation functions,” in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 7462–7473.
- [2] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.
- [3] J. Liu, J. Tang, and G. Wu, “Learning implicit neural representation for continuous super-resolution,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1628–1637.
- [4] G. Zang, R. Idoughi, R. Li, P. Wonka, and W. Heidrich, “Intratomo: Self-supervised learning-based tomographic imaging via sinogram synthesis and prediction,” *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1280–1293, 2021.
- [5] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, “Occupancy networks: Learning 3d reconstruction in function space,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4460–4470.
- [6] K. O. Stanley, “Compositional pattern producing networks: A novel abstraction of development,” *Genetic Programming and Evolvable Machines*, vol. 8, no. 2, pp. 131–162, 2007.
- [7] E. Dupont, A. Golinski, M. Alizadeh, Y. W. Teh, and A. Doucet, “Coin: Compression with implicit neural representations,” in *International Conference on Learning Representations*, 2021.
- [8] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.
- [9] X. Wang, L. Xie, C. Dong, and Y. Shan, “Real-esrgan: Training real-world blind super-resolution with pure synthetic data,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 1905–1914.
- [10] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” in *European Conference on Computer Vision*. Springer, 2020, pp. 405–421.
- [11] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, “Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5855–5864.
- [12] L. Shen, J. Pauly, and L. Xing, “Nerp: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 2, pp. 770–782, 2022.
- [13] Y. Wang, Q. He, J. Yang, Q. Wu, Y. Yang, and J. Tang, “Neural fields in visual computing and beyond,” *Computer Graphics Forum*, vol. 41, no. 2, pp. 641–676, 2022.
- [14] J. M. Wolterink, J. C. Zwienenberg, and C. Brune, “Implicit neural representations for deformable image registration,” *arXiv preprint arXiv:2204.08681*, 2022.
- [15] W. Wang, X. Xie, and E. Xing, “Understanding and accelerating neural architecture search with training-free and theory-grounded metrics,” *arXiv preprint arXiv:2108.11939*, 2021.
- [16] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville, “On the spectral bias of neural networks,” in *International Conference on Machine Learning*, 2019, pp. 5301–5310.
- [17] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, “Understanding deep learning (still) requires rethinking generalization,” *Communications of the ACM*, vol. 64, no. 3, pp. 107–115, 2021.
- [18] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” in *Communications of the ACM*, vol. 65, no. 1, 2022, pp. 99–106.
- [19] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, “Kan: Kolmogorov-arnold networks,” *arXiv preprint arXiv:2404.19756*, 2024.
- [20] Y. Chen, S. Liu, and X. Wang, “Learning continuous image representation with local implicit image function,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8628–8638.
- [21] H. Liu, K. Simonyan, and Y. Yang, “Towards practical deployment of deep neural networks: Accelerating and compressing with neural architecture search,” in *International Conference on Machine Learning*, 2022, pp. 7102–7111.
- [22] R. Basri, M. Galun, A. Tal, and Z. Karni, “On the frequency bias of generative models,” in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 18944–18954.
- [23] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng, “Fourier features let networks learn high frequency functions in low dimensional domains,” in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 7537–7547.
- [24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017, pp. 5998–6008.
- [25] A. Morsali, M. Vaez, M. Soltani, A. Kazerouni, B. Taati, and M. Mohammad-Noori, “Staf: Sinusoidal trainable activation functions for implicit neural representation,” 2025. [Online]. Available: <https://arxiv.org/abs/2502.00869>

- [26] H. Feng, D. Aldana, T. Novello, and L. D. Floriani, “Sasnet: Spatially-adaptive sinusoidal neural networks,” 2025. [Online]. Available: <https://arxiv.org/abs/2503.09750>
- [27] A. Kayabasi, A. K. Vadathya, G. Balakrishnan, and V. Saragadam, “Bias for action: Video implicit neural representations with bias modulation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2025, pp. 27 999–28 008.
- [28] M. Heidari, R. Rezaeian, R. Azad, D. Merhof, H. Soltanian-Zadeh, and I. Hacihaliloglu, “SI²a-inr: Single-layer learnable activation for implicit neural representation,” 2025. [Online]. Available: <https://arxiv.org/abs/2409.10836>
- [29] D. Jayasundara, H. Zhao, D. Labate, and V. M. Patel, “Mire: Matched implicit neural representations,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2025, pp. 8279–8288.
- [30] D. B. Lindell, J. N. Martel, and G. Wetzstein, “Bacon: Band-limited coordinate networks for multiscale scene representation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 16 252–16 262.
- [31] I. Mehta, M. Gharbi, C. Barnes, E. Shechtman, R. Ramamoorthi, and M. Chandraker, “Modulated periodic activations for generalizable local functional representations,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 14 214–14 223.
- [32] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial transformer networks,” in *Advances in Neural Information Processing Systems*, vol. 28. Curran Associates, Inc., 2015, pp. 2017–2025.
- [33] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2018, pp. 7132–7141.
- [34] H. Liu, K. Simonyan, and Y. Yang, “Darts: Differentiable architecture search,” in *International Conference on Learning Representations*, 2019.
- [35] Y. Han, G. Huang, S. Song, L. Yang, H. Wang, and Y. Wang, “Dynamic neural networks: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 7436–7456, 2022.
- [36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, vol. 25, 2012, pp. 1097–1105.
- [37] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [38] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 4681–4690.
- [39] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 1125–1134.
- [40] A. Jacot, F. Gabriel, and C. Hongler, “Neural tangent kernel: Convergence and generalization in neural networks,” in *Advances in Neural Information Processing Systems*, vol. 31. Curran Associates, Inc., 2018, pp. 8571–8580.
- [41] K. Xu, M. Qin, F. Sun, Y. Wang, Y.-K. Chen, and F. Ren, “Frequency domain acceleration of convolutional neural networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1368–1377.
- [42] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Deep laplacian pyramid networks for fast and accurate super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 624–632.
- [43] K. Xu, M. Qin, F. Sun, Y. Wang, Y.-K. Chen, and F. Ren, “Learning in the frequency domain,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1740–1749, 2020.