FLAIR: Frequency- and Locality-Aware Implicit Neural Representations

Sukhun Ko[1], Seokhyun Youn[1], Dahyeon Kye[1], Kyle Min[2], Chanho Eom[1], Jihyong Oh[1†]
[1] Chung-Ang University [2] Oracle
{looloo330, hisn16, rpekgus, cheom, jihyongoh}@cau.ac.kr kylemin@umich.edu

https://cmlab-korea.github.io/FLAIR/

Abstract

Implicit Neural Representations (INRs) leverage neural networks to map coordinates to corresponding signals, enabling continuous and compact representations. This paradigm has driven significant advances in various vision tasks. However, existing INRs lack frequency selectivity and spatial localization, leading to an over-reliance on redundant signal components. Consequently, they exhibit spectral bias, tending to learn low-frequency components early while struggling to capture fine high-frequency details. To address these issues, we propose FLAIR (Frequency- and Locality-Aware Implicit Neural Representations), which incorporates two key innovations. The first is Band-Localized Activation (BLA), a novel activation designed for joint frequency selection and spatial localization under the constraints of the time-frequency uncertainty principle (TFUP). Through structured frequency control and spatially localized responses, BLA effectively mitigates spectral bias and enhances training stability. The second is Wavelet-Energy-Guided Encoding (WEGE), which leverages the discrete wavelet transform to compute energy scores and explicitly guide frequency information to the network, enabling precise frequency selection and adaptive band control. Our method consistently outperforms existing INRs in 2D image representation, as well as 3D shape reconstruction and novel view synthesis.

1. Introduction

Traditional explicit representations, including discrete grid-based methods [10, 17, 29] such as volumes and unstructured formats like point clouds, have contributed to solving vision problems [53, 54]. However, these approaches have increasingly shown limitations in addressing the diverse and ill-posed inverse problems [12, 60]. To overcome these limitations, implicit neural representations (INRs) [48] model signals as continuous mappings from input coordinates to their corresponding values via neural networks, thereby supporting continuous attribute queries and facilitating the seamless integration of differentiable physical processes. This coordinate-based paradigm has enabled notable advances in previously challenging tasks such as super-resolution (SR) [8] and denoising [5]. More recently, it has been extended to higher-dimensional problems including 3D shape reconstruction [32, 39] and neural radiance fields [34].

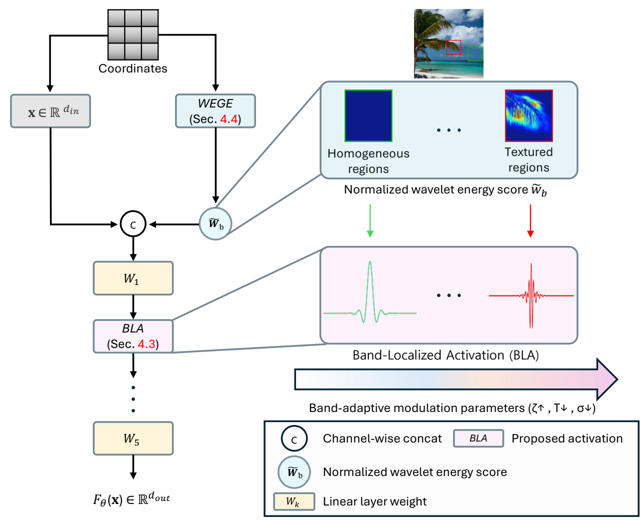


Figure 1. **Overall architecture of FLAIR.** FLAIR consists of two complementary components: Band-Localized Activation (BLA) and a Wavelet-Energy-Guided Encoding (WEGE) module. **Right:** WEGE computes normalized wavelet-energy scores $\tilde{w}_b$ over the input coordinates, highlighting spatial frequency characteristics by assigning lower wavelet-energy scores to homogeneous regions (green box) and higher scores to textured regions (red box). **Left:** Wavelet-energy scores $\tilde{w}_b$ are channel-wise concatenated with input coordinates and passed through Band-Localized Activation (BLA). BLA modulates the signal representation via learnable band-adaptive parameters (ζ, T, σ), enabling frequency shifting and band-limiting across low- and high-frequency components.

However, existing INR methods commonly suffer from the spectral bias problem [40], where networks tend to learn low-frequency components first. This leads to an inability to recover fine details, a challenge that remains unre-

[1] Department of Imaging Science and Arts & Metaverse Convergence
† Corresponding author

solved. To mitigate the spectral bias, various positional encoding (PE) schemes have been proposed, such as Sinusoidal PE [52] and Wavelet PE [64], which embed input coordinates into a higher-dimensional space using a fixed set of basis functions. Although these approaches embed coordinate information into a fixed set of M bases, their representational capacity is fundamentally limited by the number of bases M, as well as by their learnable range. These limitations constrain their expressive capacity and often lead to imperfect representations [41].

Beyond predefined frequency bases, recent studies [45, 48] have attempted to mitigate spectral bias by enhancing the expressiveness of activation functions. However, these methods have primarily focused on broadening the domain of activation functions [30, 42], rather than enabling precise frequency selection. As a result, networks may construct overlapping bases, lacking the ability to selectively target the necessary frequency bands. This limitation impairs their effectiveness in real-world inverse problems, such as denoising [61], where the suppression of irrelevant frequency bands (*e.g.*, noise) is essential for accurate signal recovery.

To address these limitations, we propose **FLAIR** (Frequency- and Locality-Aware Implicit Neural Representations), as illustrated in Fig. 1. FLAIR integrates two complementary components: (i) Band-Localized Activation (BLA) for joint frequency-domain selection and time-domain localization, and (ii) Wavelet-Energy-Guided Encoding (WEGE) for region-adaptive frequency guidance.

Specifically, we adopt a band-limited formulation that enables explicit frequency selection for signal modeling. However, increasing the degree of band limitation inevitably induces oscillations in the time domain, leading to training instability and degraded time-domain localization [16]. To mitigate these issues, instead of directly employing the ideal band-limited function such as `sinc` [44], we newly propose the Band-Localized Activation (BLA). Our design is based on the band-limiting term for frequency control and on transition-smoothing and localization terms, which are responsible for improving training stability and adaptively governing spatial localization, as illustrated in Fig. 4 (c). Although the time–frequency uncertainty principle (TFUP) states that perfect localization in both domains is unattainable, our formulation allows the model to adaptively learn the optimal trade-off between frequency selectivity and spatial localization via learnable parameters.

As the second key component of FLAIR, we propose Wavelet-Energy-Guided Encoding (WEGE), which adaptively provides explicit information to quantify continuous-frequency components, indicating whether a region is high- or low-frequency. This allows the model to perform region-adaptive frequency selection, thereby enhancing its representational flexibility. Notably, WEGE requires only 0.1K additional parameters and is compatible with plug-and-play integration into other activation-based architectures. In particular, it complements BLA by providing explicit frequency information that enhances its frequency selection capability. In summary, our contributions are as follows:

- We propose FLAIR, a novel INR architecture that integrates two complementary components: (i) Band-Localized Activation (BLA) and (ii) Wavelet-Energy-Guided Encoding (WEGE).
- We introduce BLA, a novel activation function that mitigates spectral bias through frequency selection, while suppressing oscillations and enhancing time localization under TFUP.
- We introduce WEGE, which provides quantified continuous-frequency information, enabling BLA to perform region-adaptive frequency selection. Importantly, WEGE requires only 0.1K additional parameters and is compatible with plug-and-play integration.
- We achieve state-of-the-art performance across diverse INR tasks, including 2D image fitting and restoration, 3D signed distance field, and 5D neural radiance field.

2. Related Work

Implicit Neural Representations. Standard MLPs with ReLU activations [1] exhibit a strong bias towards learning low-frequency content due to the non-periodic nature of ReLU, limiting their ability to capture fine details, a phenomenon known as spectral bias [40]. To address this spectral bias [7, 46, 47], a series of periodic activation functions has been proposed, most notably SIREN [48] and FINER [30]. SIREN employs sinusoidal activations to model high-frequency components, but its performance is highly dependent on the choice of hyperparameters and is sensitive to initialization, necessitating careful design choices to avoid suboptimal results. FINER introduces a variable-periodic sine activation and tunes the supported frequency set by adjusting the initialization range of the bias vector. However, in practice, both SIREN and FINER remain limited in addressing spectral bias due to their reliance on fixed global harmonic bases, which restrict flexible spectral adaptation during training.

Alternatively, wavelet-based activations such as WIRE [43] have aimed to address spectral bias through spatial compactness [9], offering improved interpretability and spatial control. Nonetheless, these approaches remain limited in their ability to fully overcome spectral bias, as they do not offer explicit frequency selection.

More recently, filter-based activation functions have been actively investigated in the context of INRs. A recent theoretical analysis [44] has suggested that `sinc` functions, as a type of filter-based activation, can serve as an optimal choice for INR tasks. Despite their theoretical advantages, these filter-based functions exhibit inherent limi-

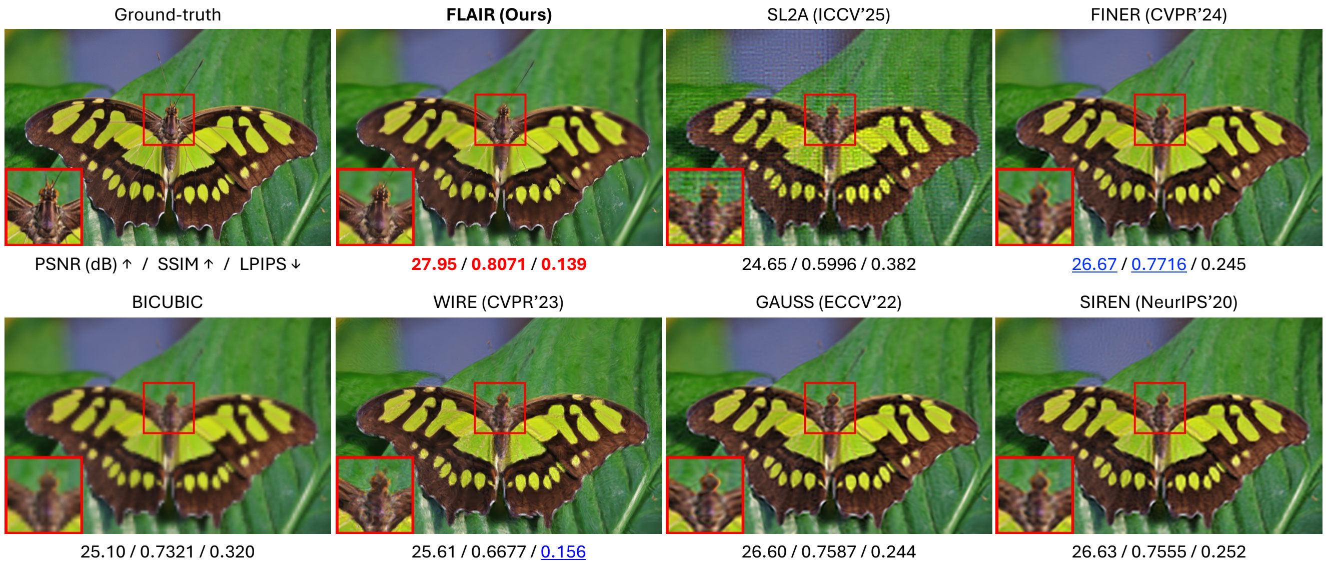


Figure 2. **Qualitative comparisons of ×4 super-resolution.** Red and blue denote the best and second-best performances, respectively.

tations. In the spatial domain, their infinite oscillations can induce training instability and visual artifacts [16, 21], and in the absence of frequency-shift mechanisms, they only act as low-pass [51], resulting in limited ability to generalize across diverse vision tasks [8, 34, 39].

To address these challenges, we propose BLA (Sec. 4.3), a key component of FLAIR that enables joint spatial localization and frequency selection. By operating under the TFUP, BLA allows the model to adaptively balance spatial localization and frequency selectivity during training, thereby mitigating spectral bias and overcoming the aforementioned limitations.

Frequency-Guided Representations and Conditioning. Frequency-guided representations and conditioning methods have been actively explored to enhance the modeling of task-relevant frequency components, which are essential for preserving fine details in low-level vision tasks [26, 56]. For example, Local Texture Estimator for Implicit Representation Function (LTE) [25] employs dedicated modules composed of convolutional neural networks (CNN) [24] and fully connected (FC) layers [4] to separately estimate key frequency components, such as amplitude, frequency, and phase, allowing the network to infer dominant frequencies and the corresponding Fourier coefficients.

In parallel, the Local Implicit Wavelet Transformer (LIWT) [13] leverages DWT to decompose input features into low- and high-frequency components, which are further refined to enhance frequency-aware representation. More recently, Zhao *et al*. [64] introduce a method that adaptively places local wavelet bases at high-frequency regions, enabling efficient encoding of fine details.

Building on this line of research, we propose a novel encoding, WEGE (Sec. 4.4), specifically designed to work with activation functions in INRs. Unlike methods [13, 25] that introduce additional auxiliary networks and increase optimization complexity, which hinders fast convergence required for INR settings, our approach leverages a non-parametric DWT-based prior to guide region-adaptive frequency modulation, with the entire WEGE module introducing only 0.1K parameters.

3. Preliminary: Time-Frequency Uncertainty Principle (TFUP)

The mathematical foundation of the time-frequency uncertainty principle (TFUP) in signal processing originates from the Heisenberg uncertainty principle [6] in quantum mechanics, which states that there is a fundamental limit to simultaneously and precisely measuring a particle's position and momentum given by $\Delta x \, \Delta p \geq \hbar/2$, where Δx and Δp are the standard deviations of position and momentum, and $\hbar = h/(2\pi)$ is the reduced Planck constant.

Extending this concept to the domain of signal processing [38], the TFUP states that for any signal $x(t)$, the product of its temporal standard deviation σ_t and frequency standard deviation σ_f is bounded below by $\sigma_t \sigma_f \geq \frac{1}{4\pi}$, where σ_t and σ_f quantify the temporal and frequency uncertainty of $x(t)$, respectively. This relationship implies that a signal highly localized in time (small σ_t) must necessarily have a broad frequency spectrum (large σ_f), and vice versa. Thus, the uncertainty principle imposes a fundamental limit on the simultaneous localization of a signal in both time and frequency domains.

4. Proposed Method: FLAIR

In this section, we first provide a brief formulation of an Implicit Neural Representations (INRs). Grounded in

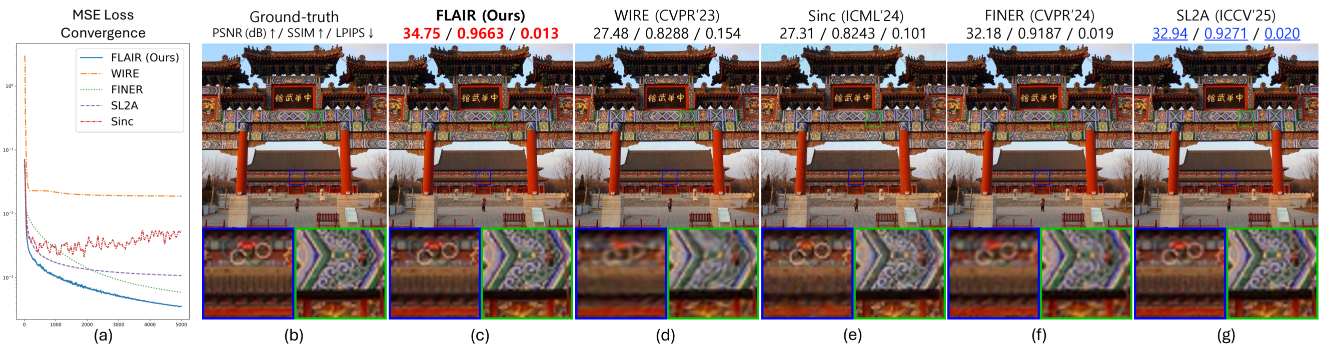


Figure 3. **(a) MSE loss convergence. (b) Ground-truth. (c)–(g) Comparison between FLAIR and other methods on image fitting.**

the Riesz basis framework, `sinc` functions are optimal band-limited bases, yet their global support introduces instability, motivating localized designs for INRs. Building on these foundations, we present a novel framework, FLAIR, composed of two key components: (i) Band-Localized Activation (BLA) (Sec. 4.3), a band-limited activation for frequency-domain selection and time-domain localization, and (ii) Wavelet-Energy-Guided Encoding (WEGE) (Sec. 4.4) for region-adaptive frequency guidance.

4.1. Formulation of an INR

An Implicit Neural Representation (INR) aims to represent a continuous signal by directly mapping input coordinates to their corresponding signal values through a neural network F_θ. Formally, the function is defined as $F_\theta : \mathbb{R}^{d_{\text{in}}} \to \mathbb{R}^{d_{\text{out}}}$, $\mathbf{x} \mapsto F_\theta(\mathbf{x})$, where $\theta \in \mathbb{R}^M$ denotes all learnable parameters of the network. The input $\mathbf{x} \in \mathbb{R}^{d_{\text{in}}}$ corresponds to a spatial coordinate, such as (x, y) in 2D or (x, y, z) in 3D. The output $F_\theta(\mathbf{x}) \in \mathbb{R}^{d_{\text{out}}}$ represents the signal value at that location (e.g., RGB color or volumetric density σ). The training data is given as a set of pairs $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, where $\mathbf{y}_i$ denotes the ground-truth signal at coordinate $\mathbf{x}_i$. The network is optimized by minimizing a standard mean squared error (MSE) loss:

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \|F_\theta(\mathbf{x}_i) - \mathbf{y}_i\|_2^2, \tag{1}$$

where N denotes the number of training samples. This simple yet general formulation enables INRs to model complex signals with high fidelity. However, accurate reconstruction of visual signals requires stronger control over both frequency selection and spatial localization. To address this, we optimize F_θ by introducing the Band-Localized Activation (BLA), which achieves these properties.

4.2. Instability of Sinc Activations

Building on Riesz basis framework [11, 55], Saratchandran *et al.* [44] show that any signal $s \in L^2(\mathbb{R})$ can be approximated within an error bound ϵ, and further establish that the `sinc` function is theoretically optimal.

Despite this theoretical optimality, such guarantees do not translate into stable optimization. The instability appears in the gradient of the loss $\mathcal{L}$ with respect to the weights W_{ij}, driven by the residual and the oscillatory term induced by the derivative of the sinc activation:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial W_{ij}} = -\frac{2}{\pi (W_i x)^2} \int \underbrace{[g(x) - f(x; W)]}_{\textit{residual}} \cdot x_j \\ \cdot \underbrace{\left[\frac{\pi W_i x}{T} \cos\left(\frac{\pi W_i x}{T}\right) - \sin\left(\frac{\pi W_i x}{T}\right) \right]}_{\textit{oscillatory}} dx. \end{aligned} \tag{2}$$

Here, $g(x)$ is the target function and $f(x; W)$ is the network output. Intuitively, the residual term inevitably arises during training, and the derivative of the sinc function (oscillatory term), having global support, propagates the residual error across the entire gradient field. This interaction amplifies instability throughout optimization, as shown in Fig. 3 (a).

4.3. Proposed Band-Localized Activation (BLA)

To mitigate the aforementioned oscillatory instability, we propose the Band-Localized Activation (BLA), which suppresses the global oscillation behavior in Eq. (2) by imposing the exponential bound $\exp\left(-\frac{(W_i x)^2}{T^2 \sigma^2}\right)$. Conceptually, BLA attenuates residual error propagation through localized gradient responses, resulting in improved stability. Full derivations are provided in the *Supplementary*.

Beyond the gradient-level analysis, the time–frequency domain offers an intuitive view of the underlying behavior. As shown in Fig. 4 (c) red box, BLA exploits the sharp cut-off behavior for frequency selectivity while suppressing infinite oscillations (blue box), thereby enabling improved time localization. Formally, BLA is expressed as a modulated basis function:

$$\psi^{\text{BLA}}(x) = \phi^{\text{basis}}(x) \cdot \underbrace{\exp(2\pi \zeta x j)}_{\textit{frequency-shifting}}, \tag{3}$$

where $\zeta \in \mathbb{R}$ denotes the modulation frequency, and j is the imaginary unit. Note that all the above functions are

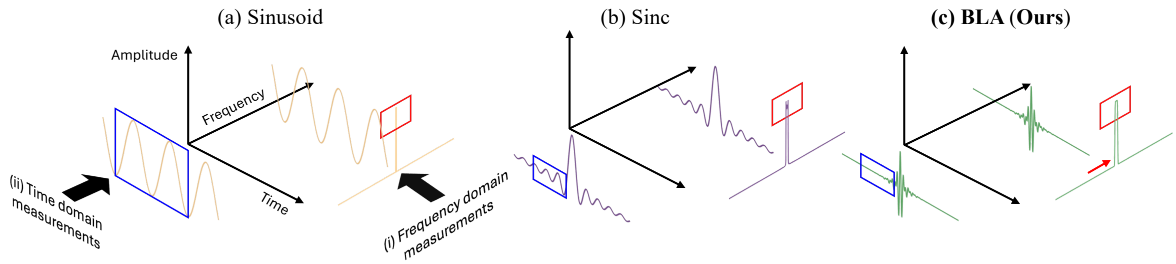


Figure 4. **Both precise frequency selection and time localization under the TFUP.** In domain (i), Sinusoid (a) produces only two fixed frequency components, so representing composite signals with diverse spectra requires many hidden dimensions. `Sinc` (b) provides sharp band selectivity. Our BLA (c) retains similar selectivity to (b) and further introduces a learnable frequency shift parameter ζ to access higher frequency bands (red box). In domain (ii), while (a) and (b) exhibit global support and oscillations (blue box), BLA (c) yields localized responses and reduces oscillations, mitigating noise propagation. Finally, our BLA (c) jointly modulates the frequency–time trade-off under the TFUP through its learnable parameters (ζ, T, σ), achieving both precise frequency selectivity and time localization.

defined in the time domain. Importantly, multiplying by $\exp(2\pi j\zeta x)$ in the time domain is equivalent to shifting the frequency support of the basis function to ζ in the Fourier domain. This allows the basis to be centered at any desired frequency, enabling the network to adaptively capture both low- and high-frequency components as needed. In practice, ζ is treated as a learnable parameter. The basis function ϕ^{basis} is defined as:

$$\phi^{\text{basis}}(x) = \underbrace{\frac{\texttt{sinc}\left(\frac{x}{T}\right)}{T}}_{\textit{band-limiting}} \cdot \underbrace{\frac{\cos\left(\frac{\pi\beta x}{T}\right)}{1-\left(\frac{2\beta x}{T}\right)^2}}_{\textit{transition-smoothing}} \cdot \underbrace{\exp\left(-\frac{x^2}{2\sigma^2}\right)}_{\textit{localization}}. \tag{4}$$

Here, $T > 0$ is the bandlimit parameter that controls the overall bandwidth, and β is the roll-off factor in the transition-smoothing term, fixed at 0.05 to preserve the inherent sharpness of the band-limited cutoff. Finally, $\sigma > 0$ is the scale parameter in the localization term, adaptively governing spatial localization. We treat T, ζ, and σ as learnable parameters while keeping β fixed, and the effect of each parameter on frequency and spatial localization is illustrated in Fig. 9. For our BLA, we initialize T=1.0, ζ=1.0, and σ=2.0, with further initialization details provided in the *Supplementary*.

We employ this function as the activation in each layer of our implicit neural representation (INR) network:

$$\begin{aligned} \vec{z}^{\,0} &= \vec{x}, \\ \vec{z}^{\,l} &= \psi\big(W^l \vec{z}^{\,l-1} + \vec{b}^{\,l}\big), \quad l = 1, 2, \ldots, L-1, \\ f(\vec{x};\theta) &= W^L \vec{z}^{\,L-1} + \vec{b}^{\,L}. \end{aligned} \tag{5}$$

$\vec{x}$ denotes the input coordinate to the network and $\psi(\cdot)$ is our proposed activation function, BLA. $\vec{z}^{\,l}$ indicates the output of layer l. L denotes the total number of layers in the network, and $\theta = \{W^l, b^l \mid l = 1, \ldots, L\}$ denotes the set of learnable parameters. This structure enables each layer to adaptively exploit both frequency selectivity and spatial localization, resulting in a more expressive INR.

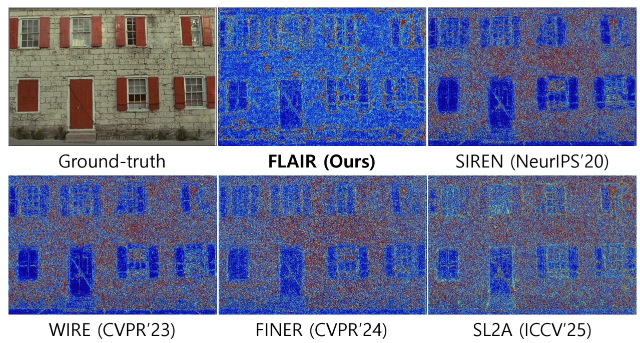


Figure 5. **Residual error heatmaps on Kodak 01.** Per-pixel absolute errors (RGB-averaged) are normalized to [0, 1]. Blue indicates lower error, and red indicates higher error. FLAIR consistently achieves lower reconstruction error than other methods.

4.4. Wavelet-Energy-Guided Encoding (WEGE)

To reduce artifacts in smooth regions from frequency leakage and to alleviate spectral bias by explicitly learning only the target frequency bands in detailed areas, we introduce Wavelet-Energy-Guided Encoding (WEGE). In INR settings where fast convergence is required, WEGE avoids additional auxiliary networks that increase optimization complexity, instead utilizing non-parametric discrete wavelet transform (DWT) and guided filtering [18]. Specifically, WEGE adaptively provides explicit information that quantifies continuous-frequency components, effectively indicating the relative dominance of high- or low-frequency content in each region. Our method first computes a pixel-wise energy score R_b for each spatial coordinate (x, y) of the input image by applying DWT. Unlike the Fourier transform, DWT jointly captures spatial and spectral information, enabling position-aware scoring. Formally, we decompose the input image I_k into high- and low-frequency components via DWT and compute R_b as:

$$\begin{aligned} D_k^H,\ D_k^L &= \text{DWT}(I_k), \\ R_b &= \text{IDWT}(D_k^H) - \text{IDWT}(D_k^L). \end{aligned} \tag{6}$$

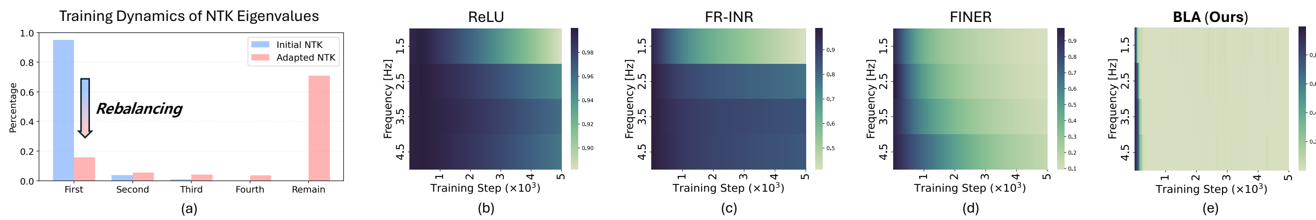


Figure 6. **Eigenvalue distribution of NTKs (a) and frequency-specific error analysis (b)–(e).** In (a), the training dynamics of BLA are illustrated via its empirical NTK. The x-axis enumerates the ranked eigenvalues, with "First" denoting the percentage contribution of the largest eigenvalue, and the y-axis indicates the percentage assigned to each component. In (b)-(e), we visualize the frequency-specific approximation error, where the x-axis corresponds to the training step and the y-axis to the frequency (Hz). (b)-(d) reduce low-frequency error early while learning higher-frequency components slowly, whereas our BLA (e) captures multiple frequency bands early.

The components of D_k^H correspond to the high-frequency sub-bands $\{HL, LH, HH\}$, while D_k^L corresponds to the low-frequency sub-band $\{LL\}$. The residual map R_b emphasizes spatial structures such as edges by subtracting the low-frequency components. To obtain a pixel-wise energy score map E_b, we compute it by averaging the squared R_b across channels. We then normalize E_b using its $E_{\min}$ and $E_{\max}$ values over the image, yielding a normalized energy score w_b that preserves edge-aware structures.

Although the w_b-based normalized energy score map exhibits edge-aware properties, pixel-wise scoring often results in discontinuities, consequently producing noisy artifacts in the RGB output domain (details in *Supplementary*). To address this, we apply a filtering operation to the normalized scores in order to suppress such score discontinuities.

Accordingly, the guided filtering [18] operation $\mathcal{G}$ is applied to the normalized energy score map w_b to mitigate discontinuities while preserving edges, formulated as:

$$\tilde{w}_b(x) = \mathcal{G}(I(x), w_b(x); r, \epsilon), \tag{7}$$

where $I(x)$ is the gray-scale guidance image, $w_b(x)$ is the normalized wavelet energy score at pixel x, r is the window radius, and ϵ is the regularization parameter. The effects of these hyperparameters (r and ϵ) are evaluated in Table 5. The resulting filtered energy score map $\tilde{w}_b$, obtained by applying the guided filtering in Eq. (7), is channel concatenated with the original spatial coordinates (x, y) and fed as input to the INR network $S(x, y) = f_\theta(x, y, \tilde{w}_b)$, where f_θ denotes the INR network parameterized by θ and internally employs the proposed activation BLA to adaptively select frequencies based on the local characteristics inferred from WEGE. This frequency-aware modulation enables capturing both high-frequency details and low-frequency structures of the target signal as shown in Fig. 5.

4.5. Neural Tangent Kernel Perspective

Neural Tangent Kernel (NTK) theory [19] provides a principled framework to interpret neural network training as kernel regression. The convergence dynamics are fundamentally governed by the distribution of NTK eigenvalues, where larger eigenvalues correspond to components that are learned more rapidly and thus fitted more accurately [3].

However, the assumptions of standard NTK theory do not directly extend to practical training regimes [19]. We therefore adopt the empirical NTK formulation commonly employed in recent studies [37, 46, 62]. For samples x_i and x_j, the empirical NTK is defined as $k'_{\text{NTK}}(x_i, x_j) = J_{f_\Theta}(x_i)\, J_{f_\Theta}(x_j)^\top$. Here, $J_{f_\Theta}(x_i)$ denotes the Jacobian matrix of the network output f_Θ with respect to its parameters at the i-th sample x_i, and $k'_{\text{NTK}}(x_i, x_j)$ corresponds to the (i, j)-entry of the resulting empirical NTK matrix.

We use this empirical NTK to analyze how BLA influences the distribution of its NTK eigenvalues during training. Following [42, 46], we use the 1D target function $f(x) = 2R\Big(\frac{\sin(3\pi x)+\sin(5\pi x)+\sin(7\pi x)+\sin(9\pi x)}{2}\Big)$, where $R(\cdot)$ is a rounding function that increases the approximation complexity. Fig. 6 (a) shows the resulting NTK eigenvalue distributions of our proposed BLA.

At initialization, the blue bars reveal a strong concentration on the "First" eigenmode due to the band-limiting term in Eq. (4), capturing over 90% of the spectral mass. During training, however, the learnable frequency-shifting term in Eq. (3) gradually moves the frequency band toward the target components. As a result, BLA redistributes the spectral mass across the "Remain" eigenmodes (red).

The broadened eigenvalue distribution observed in Fig. 6 (a) suggests that BLA may mitigate the spectral bias by distributing its capacity more evenly across frequencies. To verify this, we further examine the frequency-domain behavior of BLA. In Fig. 6 (b)–(e), we follow the observation in prior work [59] that MLPs exhibit a spectral bias, where low-frequency errors decrease much faster than high-frequency ones. We quantify this effect using the relative discrepancy $\Delta_k = \frac{|\mathcal{F}_D[g](k) - \mathcal{F}_D[f_\Theta](k)|}{|\mathcal{F}_D[g](k)|}$. As shown in Fig. 6 (e), our BLA achieves consistently low approximation errors across the entire frequency spectrum in contrast to other activations. This behavior aligns with the NTK analysis in (a), where BLA's eigenvalue redistribution mitigates the spectral bias, as reflected in (e).

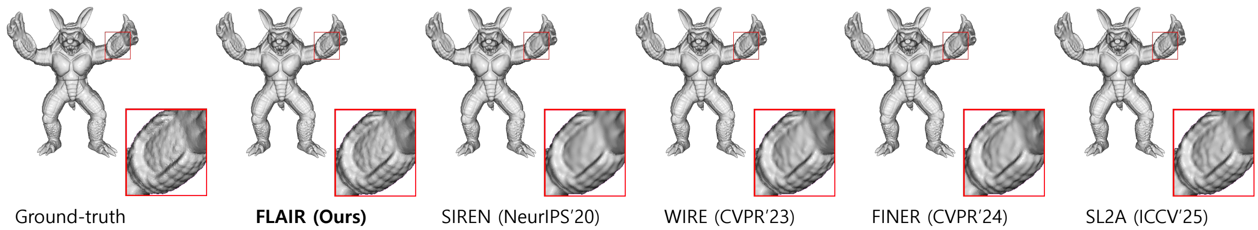


Figure 7. **Qualitative comparisons on representing the signed distance field of Armadillo.**

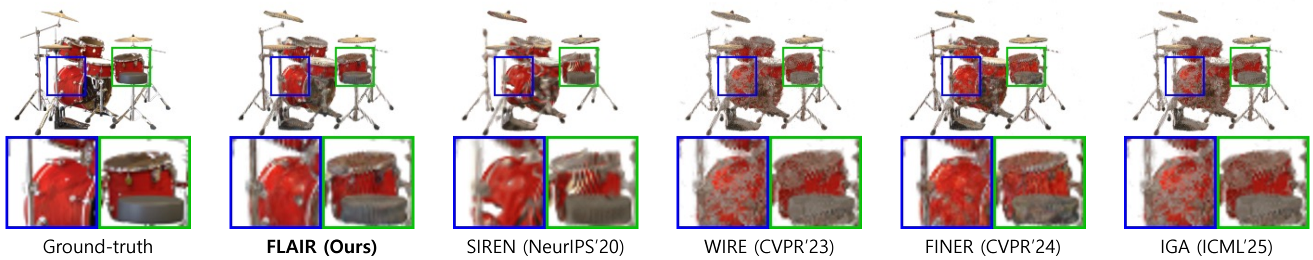


Figure 8. **Qualitative comparisons between FLAIR and the baselines on NeRF.** We follow WIRE [43], downsampling images to 200×200 and training the radiance field using only 25 input views instead of the default 100 images.

Table 1. **Quantitative image-fitting comparisons on Kodak (24 images) and DIV2K (16 images).** Red and blue indicate best and second-best per column.

Methods	#Params (K)↓	Kodak (24 images)			DIV2K (16 images)		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SL2A (ICCV'25)	330.2	36.14	0.9304	0.060	36.26	0.9481	0.034
IGA (ICML'25)	205.1	34.49	0.9596	0.056	33.51	0.9742	0.022
FR (CVPR'24)	6299.9	27.62	0.8781	0.151	26.22	0.9001	0.124
FINER (CVPR'24)	198.9	35.46	0.9246	0.065	35.01	0.9457	0.039
WIRE (CVPR'23)	91.6	28.80	0.7928	0.333	32.17	0.8874	0.107
SIREN (NeurIPS'20)	198.9	28.64	0.7792	0.369	34.36	0.9388	0.050
FLAIR (Ours)	199.0	**37.12**	**0.9644**	**0.054**	**38.19**	**0.9754**	**0.016**

Table 2. **Quantitative comparisons on representing signed distance fields.** We report Chamfer Distance ↓ and IoU ↑.

	Methods	Armadillo	Dragon	Lucy	Thai Statue	BeardedMan	Avg.
Chamfer ↓	SL2A (ICCV'25)	3.235e-6	2.987e-5	1.972e-6	2.988e-6	3.615e-6	8.336e-6
	FINER (CVPR'24)	3.368e-6	2.846e-5	2.526e-6	3.931e-6	3.838e-6	8.425e-6
	WIRE (CVPR'23)	3.278e-6	2.884e-5	2.039e-6	3.143e-6	3.778e-6	8.216e-6
	SIREN (NeurIPS'20)	3.838e-6	2.698e-5	2.540e-6	3.358e-6	4.338e-6	8.212e-6
	FLAIR (Ours)	**3.207e-6**	**2.656e-5**	**1.906e-6**	**2.972e-6**	**3.547e-6**	**7.637e-6**
IoU ↑	SL2A (ICCV'25)	9.899e-1	9.849e-1	**9.810e-1**	9.657e-1	9.925e-1	9.828e-1
	FINER (CVPR'24)	9.897e-1	9.749e-1	9.701e-1	9.545e-1	9.902e-1	9.759e-1
	WIRE (CVPR'23)	9.893e-1	9.821e-1	9.778e-1	9.687e-1	9.908e-1	9.817e-1
	SIREN (NeurIPS'20)	9.828e-1	9.709e-1	9.657e-1	9.669e-1	9.888e-1	9.750e-1
	FLAIR (Ours)	**9.903e-1**	**9.851e-1**	9.802e-1	**9.688e-1**	**9.945e-1**	**9.838e-1**

5. Experiments

Implementation Details. We thoroughly explore all possible combinations of learning rates (1×10^{-2}, 5×10^{-3}, 1×10^{-3}, 5×10^{-4}, 1×10^{-4}) for each of the state-of-the-art (SOTA) models to choose each best configuration for fair comparisons. For our model, we use a learning rate of 5×10^{-4} in all experiments. For activation functions that are sensitive to initialization, such as SIREN (where the initial ω_0 is set to a large value, *e.g.*, 30), we follow each method's recommended setting. All models are trained with 4 hidden layers and 256 hidden features per layer, and the number of training iterations is kept consistent across all models. All experiments are conducted on an RTX A6000 GPU. To ensure fair comparison, results are averaged over 5 random seeds. The variance is consistently small across activations, so standard deviations are omitted for brevity.

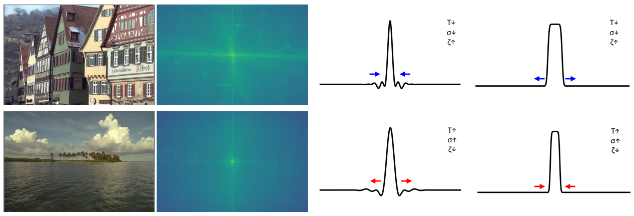
5.1. Signal Representation

2D Image Representation. For the 2D image representation task, we use the widely adopted Kodak [14] dataset and DIV2K [2], which are widely recognized benchmarks for image representation. Results after 5,000 training iterations are summarized in Table 1, where our model achieves the best average scores across all metrics, including PSNR, SSIM [57], and LPIPS [63], while being significantly more parameter-efficient than the second-best methods [42, 47]. As shown in Fig. 3, our method achieves faster convergence and better reconstruction of fine structural details. We attribute this performance advantage to the model's strong joint frequency selection and spatial localization capabilities. Comprehensive per-scene qualitative and quantitative results are provided in the *Supplementary Material*.

3D Shape Representation. Signed distance function (SDF) represents the distance from any 3D point to the closest surface using a continuous function, where the sign indicates whether the point lies inside (negative) or outside (positive) the object boundary [22]. We follow the baseline setup [28] and evaluate on five shapes from the Stanford 3D Scanning Repository [49]. Per-scene quantitative results in Table 2 show that FLAIR outperforms competing approaches. As shown in Fig. 7, FLAIR also reconstructs fine-grained ge-

Table 3. **Average results on the NeRF synthetic dataset** [34]. Per-scene results are provided in the Supplementary.

Metrics	IGA (ICML'25)	FINER (CVPR'24)	WIRE (CVPR'23)	SIREN (NeurIPS'20)	**FLAIR (Ours)**
PSNR↑	28.52	28.97	27.20	25.60	**29.31**
SSIM↑	0.9217	0.9261	0.9009	0.8857	**0.9325**
LPIPS↓	0.063	0.044	0.065	0.105	**0.041**



(a) RGB image (b) FFT magnitude (c) Time domain (d) Frequency domain

Figure 9. **Ablation study on scene adaptivity.** Comparison between a complex scene (top row) and a homogeneous scene (bottom row).

ometry with sharper and more faithful surface details than other methods.

5.2. Inverse Problems of 2D Images

Arbitrary-Scale Super-Resolution. INRs inherently act as continuous interpolants, enabling super-resolution beyond fixed scales. Accordingly, we evaluate our method under arbitrary-scale settings (*e.g.*, $\times 4$, $\times 6$, $\times 8$), with further results in the *Supplementary*. As shown in Fig. 2, FLAIR achieves the most fine-grained reconstruction at $\times 4$, particularly around the butterfly's head (red box).

Image Denoising. Denoising is a particularly challenging problem that requires not only effective noise suppression but also the preservation of fine details. To address this challenge, precise frequency selection in regions where noise and structure interact is essential. FLAIR meets this requirement by effectively removing noise while preserving fine textures, as shown in the *Supplementary*.

5.3. Neural Radiance Fields

Neural Radiance Fields. NeRFs [34] synthesize novel views of complex 3D scenes by learning a continuous volumetric radiance field that maps 3D coordinates and viewing directions to color and density. Following prior work [30, 43], we adopt a challenging setting using only 25 training images, instead of the standard 100. As shown in Fig. 8, FLAIR remains robust under this limited-view regime and preserves structural details. Table 3 reports results on the Blender benchmark [34], where FLAIR achieves the best average performance across scenes. Additional per-scene results are provided in the *Supplementary Material*.

5.4. Ablation Study

Scene-Adaptive Behavior of BLA. To validate whether the learnable parameters of BLA adapt to scene complexity, we conduct an image fitting experiment on two representative scenes, a complex scene (top row) and a homogeneous scene (bottom row), as shown in Fig. 9. We fit BLA to each image and report the learned parameters (T, σ, ζ) across layers in Table 4. The complex scene exhibits smaller average values of T and σ, indicating stronger time-domain localization and broader frequency-domain support, as reflected in the top row (c) and (d) of Fig. 9. In addition, its larger ζ induces a stronger spectral shift observed in Fig. 9 (d), revealing increased emphasis on high-frequency components compared to the bottom row.

Table 4. **Layer-wise parameters in BLA.** Each parameter (T, σ, ζ) is a learned scalar, and the columns list the values learned in each layer l.

Scene	Parameter	Layer index l 1	2	3	4	**Average**
Complex (Kodak 08)	T	0.713	1.024	1.046	0.674	**0.864**
	σ	1.955	2.011	2.041	1.468	**1.868**
	ζ	1.178	1.005	0.993	1.166	**1.086**
Homogeneous (Kodak 16)	T	1.033	1.026	0.962	1.008	**1.005**
	σ	1.964	2.016	2.136	2.010	**2.015**
	ζ	1.114	0.999	0.989	1.021	**1.031**

Table 5. **Ablation of WEGE.** We evaluate the effect of the DWT level J and the guided filter parameters (window radius r and regularization ϵ) within WEGE. Fitting experiments are conducted on five randomly selected DIV2K scenes (0, 5, 8, 9, 13). We report the average performance for brevity, while the relative ordering across individual scenes and various tasks remains consistent, indicating that the chosen configuration maintains generality.

WEGE	r	ϵ	J = 1 PSNR↑ / SSIM↑ / LPIPS↓	J = 2 PSNR↑ / SSIM↑ / LPIPS↓	J = 3 PSNR↑ / SSIM↑ / LPIPS↓
×	×	×	33.26 / 0.9479 / 0.050 *(baseline without DWT)*		
✓	×	×	33.48 / 0.9494 / 0.031	30.88 / 0.9171 / 0.037	30.61 / 0.9152 / 0.046
✓	2	1e-8	35.11 / 0.9659 / 0.020	33.58 / 0.9540 / 0.033	33.53 / 0.9527 / 0.036
✓	6	1e-5	**36.39 / 0.9779 / 0.014**	34.92 / 0.9635 / 0.033	34.36 / 0.9591 / 0.035
✓	10	1e-2	36.01 / 0.9717 / 0.023	34.14 / 0.9595 / 0.045	32.95 / 0.9431 / 0.040

Hyperparameter Analysis of WEGE. As shown in Table 5, we adopt the Daubechies-3 wavelet basis for its smoothness and compact support, and we vary the DWT level J. Among all settings, $J = 1$ performs best, effectively preserving fine structures. We also examine the guided filter, which smooths edge discontinuities in the raw score and yields more stable representations. We set $r = 6$ and $\epsilon = 1e{-5}$ to balance excessive sharpness and blur, reducing score discontinuities and preserving score accuracy. Additional score maps are provided in the *Supplementary*.

6. Conclusion

We have proposed FLAIR, a Frequency- and Locality-Aware Implicit Neural Representation framework that unifies two complementary components. BLA is a novel activation function enabling precise frequency selection while suppressing excessive spatial oscillations, thereby facilitating effective signal modeling. WEGE is a lightweight module tailored to INR architectures, adding only 0.1K parameters. Our method consistently alleviates spectral bias and demonstrates superior fine-detail reconstruction across various tasks and datasets.

7. Acknowledgement

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2025S1A5C3A04022639). This research was supported by Culture, Sports and Tourism RD Program through the Korea Creative Content Agency grant funded by Ministry of Culture, Sports and Tourism in 2024 (Project Name : Developing Professionals for RD in Contents Production Based on Generative Ai and Cloud, Project Number : RS-2024-00352578, Contribution Rate: 50%). This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Graduate School of Virtual Convergence support program(IITP-2024-RS-2024-00418847) supervised by the IITP(Institute for Information Communications Technology Planning Evaluation).

References

[1] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv*, 2018. 2

[2] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPRW*, pages 126–135, 2017. 7, 1

[3] Jinshuai Bai, Gui-Rong Liu, Ashish Gupta, Laith Alzubaidi, Xi-Qiao Feng, and YuanTong Gu. Physics-informed radial basis network (pirbn): A local approximating neural network for solving nonlinear partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 415: 116290, 2023. 6

[4] SH Shabbeer Basha, Shiv Ram Dubey, Viswanath Pulabaigari, and Snehasis Mukherjee. Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing*, 378:112–119, 2020. 3

[5] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005. 1

[6] Paul Busch, Teiko Heinonen, and Pekka Lahti. Heisenberg's uncertainty principle. *Physics reports*, 452(6):155–176, 2007. 3

[7] Zhicheng Cai, Hao Zhu, Qiu Shen, Xinran Wang, and Xun Cao. Batch normalization alleviates the spectral bias in coordinate networks. In *CVPR*, pages 25160–25171, 2024. 2

[8] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *CVPR*, pages 8628–8638, 2021. 1, 3, 8

[9] Shin-Fang Chng, Sameera Ramasinghe, Jamie Sherrah, and Simon Lucey. Gaussian activated neural radiance fields for high fidelity reconstruction and pose estimation. In *ECCV*, pages 264–280, 2022. 2

[10] Christopher B Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In *ECCV*, pages 628–644, 2016. 1

[11] Ole Christensen et al. *An introduction to frames and Riesz bases*. Springer, 2003. 4

[12] Guy Demoment. Image reconstruction and restoration: Overview of common estimation structures and problems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12):2024–2036, 2002. 1

[13] Minghong Duan, Linhao Qu, Shaolei Liu, and Manning Wang. Local implicit wavelet transformer for arbitrary-scale super-resolution. *arXiv*, 2024. 3

[14] E.Kodak. Kodak lossless true color image suite. `https://r0k.us/graphics/kodak/`, 1999. Accessed: June 1, 2025. 7

[15] Pierre Foret, Ariel Kleiner, Hossein Mobahi, and Behnam Neyshabur. Sharpness-aware minimization for efficiently improving generalization. *arXiv preprint arXiv:2010.01412*, 2020. 1

[16] David Gottlieb and Chi-Wang Shu. On the gibbs phenomenon and its resolution. *SIAM Review*, 39(4):644–668, 1997. 2, 3

[17] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. *IEEE TPAMI*, 43(12):4338–4364, 2020. 1

[18] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. *IEEE TPAMI*, 35(6):1397–1409, 2012. 5, 6

[19] Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. *NeurIPS*, 31, 2018. 6

[20] Jinho Jeong, Sangmin Han, Jinwoo Kim, and Seon Joo Kim. Latent space super-resolution for higher-resolution image generation with diffusion models. In *CVPR*, pages 2355–2365, 2025. 7

[21] Abdul J. Jerri. The shannon sampling theorem—its various extensions and applications: A tutorial review. *Proceedings of the IEEE*, 65(11):1565–1596, 1977. 3

[22] Mark W Jones, J Andreas Baerentzen, and Milos Sramek. 3d distance fields: A survey of techniques and applications. *IEEE TVCG*, 12(4):581–599, 2006. 7

[23] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*, 2016. 1

[24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *NeurIPS*, 25, 2012. 3

[25] Jaewon Lee and Kyong Hwan Jin. Local texture estimator for implicit representation function. In *CVPR*, pages 1929–1938, 2022. 3

[26] Fei Li, Linfeng Zhang, Zikun Liu, Juan Lei, and Zhenbo Li. Multi-frequency representation enhancement with privilege information for video super-resolution. In *ICCV*, pages 12814–12825, 2023. 3

[27] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. Visualizing the loss landscape of neural nets. *NeurIPS*, 31, 2018. 1

[28] David B Lindell, Dave Van Veen, Jeong Joon Park, and Gordon Wetzstein. Bacon: Band-limited coordinate networks for multiscale scene representation. In *CVPR*, pages 16252–16262, 2022. 7

[29] Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. *NeurIPS*, 33: 15651–15663, 2020. 1

[30] Zhen Liu, Hao Zhu, Qi Zhang, Jingde Fu, Weibing Deng, Zhan Ma, Yanwen Guo, and Xun Cao. Finer: Flexible spectral-bias tuning in implicit neural representation by variable-periodic activation functions. In *CVPR*, pages 2713–2722, 2024. 2, 8

[31] Julien NP Martel, David B Lindell, Connor Z Lin, Eric R Chan, Marco Monteiro, and Gordon Wetzstein. Acorn: Adaptive coordinate networks for neural scene representation. *arXiv preprint arXiv:2105.02788*, 2021. 7

[32] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *CVPR*, pages 4460–4470, 2019. 1

[33] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (ToG)*, 38(4):1–14, 2019. 7

[34] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1, 3, 8

[35] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 7

[36] Jong Kwon Oh, Hwijae Son, Hyung Ju Hwang, and Jihyong Oh. Sofono: Arbitrary-scale image super-resolution via sobolev fourier neural operator. *Neurocomputing*, page 131944, 2025. 8

[37] Guillermo Ortiz-Jiménez, Seyed-Mohsen Moosavi-Dezfooli, and Pascal Frossard. What can linearized neural networks actually say about generalization? *NeurIPS*, 34: 8998–9010, 2021. 6

[38] Reza Parhizkar, Yann Barbotin, and Martin Vetterli. Sequences with minimal time–frequency uncertainty. *Applied and Computational Harmonic Analysis*, 38(3):452–468, 2015. 3

[39] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *CVPR*, pages 165–174, 2019. 1, 3

[40] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning*, pages 5301–5310, 2019. 1, 2

[41] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. *NeurIPS*, 20, 2007. 2

[42] Reza Rezaeian, Moein Heidari, Reza Azad, Dorit Merhof, Hamid Soltanian-Zadeh, and Ilker Hacihaliloglu. Sl2a-inr: Single-layer learnable activation for implicit neural representation. In *ICCV*, pages 26065–26074, 2025. 2, 6, 7

[43] Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G Baraniuk. Wire: Wavelet implicit neural representations. In *CVPR*, pages 18507–18516, 2023. 2, 7, 8

[44] Hemanth Saratchandran, Sameera Ramasinghe, Violetta Shevchenko, Alexander Long, and Simon Lucey. A sampling theory perspective on activations for implicit neural representations. In *International Conference on Machine Learning*, pages 43422–43444, 2024. 2, 4

[45] Danzel Serrano, Jakub Szymkowiak, and Przemyslaw Musialski. Hosc: A periodic activation function for preserving sharp features in implicit neural representations. *arXiv*, 2024. 2

[46] Kexuan Shi, Xingyu Zhou, and Shuhang Gu. Improved implicit neural representation with fourier reparameterized training. In *CVPR*, pages 25985–25994, 2024. 2, 6

[47] Kexuan Shi, Hai Chen, Leheng Zhang, and Shuhang Gu. Inductive gradient adjustment for spectral bias in implicit neural representations. In *International Conference on Machine Learning*, 2025. 2, 7, 6

[48] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *NeurIPS*, 33:7462–7473, 2020. 1, 2

[49] Stanford University Computer Graphics Laboratory. The stanford 3d scanning repository. https://graphics.stanford.edu/data/3Dscanrep/, 1994. Accessed: 2025-06-01. 7

[50] Elias M Stein and Rami Shakarchi. *Fourier analysis: an introduction*. Princeton University Press, 2011. 5

[51] Frank Stenger. Numerical methods based on whittaker cardinal, or sinc functions. *Siam Review*, 23(2):165–224, 1981. 3

[52] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS*, 33:7537–7547, 2020. 2

[53] Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, Wang Yifan, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. Advances in neural rendering. *Comput. Graph. Forum*, 41(2):703–735, 2022. 1

[54] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *CVPR*, pages 9446–9454, 2018. 1

[55] Michael Unser. Sampling-50 years after shannon. *Proceedings of the IEEE*, 88(4):569–587, 2002. 4

[56] Fei Wang, Dan Guo, Kun Li, Zhun Zhong, and Meng Wang. Frequency decoupling for motion magnification via multi-level isomorphic architecture. In *CVPR*, pages 18984–18994, 2024. 3

[57] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004. 7, 2

[58] Zhijie Wu, Yuhe Jin, and Kwang Moo Yi. Neural fourier filter bank. In *Proceedings of the IEEE/CVF Conference*

on Computer Vision and Pattern Recognition, pages 14153–14163, 2023. 7

[59] Zhiqin John Xu. Understanding training and generalization in deep learning by fourier analysis. *arXiv preprint arXiv:1808.04295*, 2018. 6

[60] Yanyang Yan, Wenqi Ren, Yuanfang Guo, Rui Wang, and Xiaochun Cao. Image deblurring via extreme channels prior. In *CVPR*, pages 4003–4011, 2017. 1

[61] Zipei Yan, Zhengji Liu, and Jizhou Li. Boosting of implicit neural representation-based image denoiser. In *ICASSP*, pages 4295–4299, 2024. 2

[62] Gizem Yüce, Guillermo Ortiz-Jiménez, Beril Besbinar, and Pascal Frossard. A structured dictionary perspective on implicit neural representations. In *CVPR*, pages 19228–19238, 2022. 6

[63] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. 7, 2

[64] Hongxu Zhao, Zelin Gao, Yue Wang, Rong Xiong, and Yu Zhang. Adaptive wavelet-positional encoding for high-frequency information learning in implicit neural representation. In *AAAI*, pages 10430–10438, 2025. 2, 3