CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (1) (2026), Pages 505-523
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Complex-Valued Neural Networks with Adaptive Frequency Attention for Image Denoising

Marjono, Avin Maulana*, and Anggi Gustiningsih Hapsani

Department of Mathematics, Brawijaya University, Indonesia

Abstract

Image denoising encompasses various noise types; in this work, we focus specifically on periodic interference, which introduces coherent frequency-domain artifacts that are challenging to remove using conventional real-valued convolutional neural networks (CNNs). This paper introduces a Complex-Valued Neural Network with Adaptive Frequency Attention (CVNN-AFA) tailored to periodic noise removal, integrating complex-domain feature propagation with explicit radial frequency-band modulation. The proposed architecture employs complex convolutions, complex batch normalization, and ModReLU activations to jointly model amplitude and phase information. An Adaptive Frequency Attention (AFA) module operates in the Fourier domain and partitions the spectrum into low-, mid-, and high-frequency radial bands using distance-based masks, enabling adaptive band-wise reweighting aligned with interference characteristics. Experiments on the BSDS500 dataset augmented with synthetic periodic noise evaluate both low-noise and moderate-to-high noise regimes under matched training budgets and strong real-valued frequency-aware baselines. Results indicate that CVNN-AFA achieves competitive performance overall and provides consistent, moderate improvements in low-amplitude settings, while the real-valued frequency-aware baseline remains more robust under extreme corruption levels. Qualitative and spectral analyses suggest that the proposed approach offers incremental attenuation of periodic components while maintaining comparable detail preservation. These findings are specific to the controlled periodic noise scenarios evaluated in this study.

Keywords: Complex-Valued Neural Networks; Frequency Attention; Image Denoising; Periodic Noise; U-Net.

1. Introduction

Image denoising is a fundamental problem in computer vision, crucial for surveillance and medical imaging applications. The goal is to recover clean images from unwanted noise, improving visual quality and facilitating downstream analysis. Noise in images can generally be categorized into random and structured types. Random noise (Gaussian or photon noise) has been extensively studied, but structured noise such as periodic interference poses a unique challenge. Periodic noise manifests as recurrent patterns (lines or sinusoidal ripples) caused by electromagnetic interference, scanning errors or mechanical vibrations during image acquisition [1]. Because of its recurrent pattern, periodic noise is particularly damaging, leading to substantial degradation of image quality through loss of fine detail and distortion of global contrast [2]. Traditional methods

*Corresponding author. E-mail: avin_maulana@ub.ac.id

Submitted: October 10, 2025 Reviewed: October 19, 2025 Accepted: March 03, 2026
DOI: https://doi.org/10.18860/cauchy.v11i1.36851

address periodic noise by exploiting its frequency domain characteristics, for instance, applying notch filters at the noise frequencies, where the periodic artifacts appear as sharp spectral peaks in the Fourier domain. Frequency domain filtering can be effective when the noise frequency is known, but challenging if the noise pattern drifts or if its phase is indeterminate. In general, removing periodic interference without degrading image sharpness remains an important open problem in image restoration [3].

In the past ten years, deep learning methodologies have come to dominate image denoising, replacing many conventional model based strategies. Convolutional neural networks (CNNs) can learn direct mappings from noisy to clean images, encoding image priors from data [4]. Since the advent of early CNN based denoisers such as DnCNN [5] which surpassed traditional filter based techniques, the field has advanced rapidly toward more sophisticated architectures. Contemporary models employ residual learning, attention mechanisms, and transformer components to further elevate denoising performance [3]. Ren et al [6] proposed DeamNet, an Adaptive Consistency Prior network, that embeds model based optimization within a deep CNN and attained leading performance on both synthetic and real noise benchmarks. Transformer based models like Restormer [7] exploit global contextual modeling via self attention to advance high resolution image restoration. In parallel, simplified CNN frameworks such as NAFNet (Nonlinear Activation Free Network) deliver efficient and powerful baselines for image denoising. Modern denoising models have achieved notable improvements in PSNR and SSIM across Gaussian and camera noise datasets [3], [8]. However, most of these methods operate on real-valued images in the spatial domain and do not explicitly model the frequency phase structure of periodic noise.

Periodic noise is a phase-coherent phenomenon; its structure is encoded in both the magnitude and phase of the signal's Fourier transform. Conventional real-valued CNNs operate directly on spatial intensities and do not explicitly maintain complex-valued representations of amplitude and phase across layers. While such models can implicitly learn frequency-sensitive filters, they lack an architectural mechanism dedicated to structured spectral partitioning. As a result, periodic interference-characterized by localized symmetric peaks in the Fourier domain may not be explicitly isolated during feature processing. This limitation mirrors challenges in other specialized domains, where generic models fail to address nuanced problems like preserving spatial consistency in unaligned face images without tailored architectural designs or loss functions [9]. Consequently, even state-of-the-art real-valued networks may struggle when confronted with structured interference. These models can fail to completely eliminate the periodic pattern or over smoothing. Empirical comparisons report residual artifacts and texture loss for RIDNet, SADNet, and DeamNet [8], [10].

Standard CNNs process only real intensity values, making it hard to distinguish a shifted periodic pattern from legitimate image textures. Fourier analysis has shown that real-valued CNN denoisers tend to omit portions of high-frequency information, indicating a spectral bias toward lower frequencies [4]. As a result, residual periodic artifacts may persist or high-frequency details may be attenuated, particularly when the interference exhibits strong phase coherence. Recent studies have shown that discarding phase information can have significant consequences, as the phase (imaginary part) carries crucial structural information about the signal [11]. In other words, without a mechanism to handle phase, a CNN may suppress the amplitude of a periodic noise pattern but fail to achieve the requisite phase alignment for full cancellation, leaving residual striping. This limitation motivates exploring complex-valued neural networks (CVNN), which naturally incorporate both magnitude and phase. Despite recent interest, complex-valued CNNs had not been broadly adopted for image denoising, with the overwhelming majority of deep denoisers operating in the real-valued setting.

Motivated by these observations, we introduce a Complex-Valued Neural Network with Adaptive Frequency Attention (CVNN-AFA) for image denoising. The proposed method targets structured periodic noise by unifying spatial- and frequency-domain processing within a single model. First, we deploy complex valued convolutional layers across the network, representing each

feature as a complex number (real and imaginary components) rather than a single real value. This design enables filters to respond jointly to amplitude and phase components, providing a representation that is better aligned with the frequency-domain characteristics of periodic noise. We extend this concept to structured periodic image denoising via a complex-valued neural network that integrates complex convolutions, complex batch normalization, and suitable activation functions such as modReLU, thereby preserving phase across layers. Second, we propose an Adaptive Frequency Attention (AFA) module that operates in the Fourier domain. At the bottleneck, feature maps are transformed into the Fourier domain using a two-dimensional discrete Fourier transform. The proposed Adaptive Frequency Attention (AFA) module then partitions the spectrum into low-, mid-, and high-frequency radial bands using distance-based masks defined as fixed proportions of the feature-map resolution. Each band is reweighted through learned attention coefficients before inverse transformation. This explicit radial partitioning is motivated by the geometric structure of periodic noise, which concentrates energy at localized spectral peaks. The design is motivated by the observation that periodic interference localizes energy in distinct regions of the frequency spectrum.

While prior studies have explored complex-valued convolutional networks for image restoration [4], [12], existing approaches typically either operate globally in the frequency domain or employ complex convolutions without explicitly structuring spectral interactions. In contrast, the proposed CVNN-AFA integrates spatial-domain complex feature learning with a dedicated Fourier-domain attention module that explicitly partitions the spectrum into disjoint radial bands. This radial multi-band decomposition is motivated by the geometric structure of periodic interference, which manifests as localized spectral peaks at specific frequency radii. Unlike global frequency scaling or single-branch spectral attention mechanisms, the proposed Adaptive Frequency Attention independently reweights low-, mid-, and high-frequency regions using separate learned gates. By combining phase-aware feature propagation with structured band-wise spectral modulation inside a U-Net architecture, the proposed design enables targeted attenuation of periodic components while preserving spatial reconstruction capacity. To the best of our knowledge, such integration of complex-valued feature representation with explicit radial frequency partitioning has not been systematically evaluated for structured periodic noise removal under controlled multi-regime training.

In addition to proposing the CVNN-AFA architecture, we conduct a controlled empirical comparison against frequency-aware real-valued baselines under matched training budgets, allowing a systematic evaluation of when complex-valued representations provide advantages and where real-valued models remain competitive.

The remainder of this paper is organized as follows. Section 2 describes the proposed architecture and frequency attention mechanism. Section 3 presents experimental results and analysis. Section 4 concludes the paper.

2. Methods

The proposed framework aims to restore clean images from observations degraded by structured, periodic noise. Unlike stochastic noise such as Gaussian or Poisson distributions, periodic degradation manifests as frequency-domain interference patterns that are difficult to remove using conventional real-valued convolutional neural networks (CNNs). Motivated by these observations, we develop a Complex-Valued Neural Network with Adaptive Frequency Attention (CVNN-AFA) that explicitly models both the amplitude and phase of feature responses in the complex domain, enabling the network to learn frequency-aware denoising mappings.

Clean natural images from the BSDS500 dataset are synthetically degraded by periodic sinusoidal perturbations of varying frequencies and amplitudes to simulate real-world interference. Each noisy image is concatenated with a one-channel conditioning map encoding the noise intensity, forming a four-channel input tensor that supports non-blind denoising across different degradation levels. This input is processed by the proposed CVNN-AFA, which follows an

encoder-decoder (U-Net-like) topology built entirely with complex-valued convolutions, batch normalization, and ModReLU activations. The U-Net architecture's advantage in image denoising lies in its contracting and expanding paths, which effectively capture context and recover details, allowing for improved feature reuse and competitive performance in denoising tasks compared to traditional methods. [13] At the bottleneck, an *Adaptive Frequency Attention (AFA)* module implemented as a band-split attention mechanism-operates in the Fourier domain to dynamically reweight feature responses across frequency bands. This mechanism enhances the model's ability to suppress periodic components while preserving spatial details.

For fair comparison, we also construct a real-valued CNN (RealCNN) baseline using real-valued convolutions and identical training configurations. Both models are trained under the same supervision using the Charbonnier loss function, with optional spectral-domain regularization to encourage frequency consistency between restored and ground-truth images. Quantitative evaluation is performed using PSNR and SSIM metrics on multiple noise levels to demonstrate the effectiveness of the proposed complex-valued formulation.

Overall, the proposed methodology integrates three key ideas:

1. Complex-valued feature representation that captures phase-sensitive relationships crucial for frequency-dependent noise.
2. Adaptive frequency attention that modulates spectral responses according to learned amplitude distributions.
3. A unified denoising pipeline combining spatial and spectral learning within a compact encoder-decoder design.

This design provides a principled and interpretable approach to image denoising under structured noise, bridging the gap between spatial CNN filtering and frequency-domain processing.

2.1. Dataset and Preprocessing

Dataset. We employ the Berkeley Segmentation Dataset (BSDS500)[1] [14], a standard benchmark for low-level image processing and segmentation. The dataset contains 500 natural RGB images with diverse content and texture statistics. Following the official partition, we use 200 images for training, 100 for validation, and 200 for testing.

During training, images are randomly cropped into 128×128 patches. For each training image, four patches are sampled per epoch to increase data diversity. Random horizontal flipping and random 90° rotations are applied as data augmentation. During validation and testing, full-resolution images are processed directly without resizing or sliding-window tiling. No overlapping patch averaging or test-time augmentation is used.

Data Augmentation. To enhance generalization and prevent overfitting, several stochastic augmentation techniques are applied. Each patch undergoes random horizontal flip with probability 0.5 and random rotations by multiples of 90°. These geometric transformations preserve the underlying noise structure while promoting orientation invariance in the learned denoising function. No color jittering or intensity scaling is applied to ensure consistent frequency distribution across the dataset.

Noise Simulation. To evaluate the proposed model under structured degradation, we synthetically generate *periodic noise* on clean images *on the fly* during training. Given a clean patch I_c, a sinusoidal perturbation is added according to the formulation in Eq. (1):

$$I_n(x, y) = \text{clip}(I_c(x, y) + A \sin(2\pi(f_x x/W + f_y y/H) + \phi)), \tag{1}$$

where A denotes the amplitude of the sinusoidal interference, (f_x, f_y) are spatial frequencies, and $\phi \sim \mathcal{U}(0, 2\pi)$ is a randomly sampled phase. The spatial frequencies are randomly selected

[1] https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/

from $\{2, 3, 4, 6, 8, 10\}$, producing diverse stripe-like patterns that mimic electronic interference or scanning artifacts. The amplitude A is linearly scaled according to the designated noise level parameter σ. Two experimental regimes are considered:

- low-noise levels $\sigma \in \{5, 10, 15\}$,
- moderate-to-high noise levels $\sigma \in \{10, 30, 50\}$.

Larger values of σ correspond to proportionally larger sinusoidal amplitudes A, resulting in increasingly severe periodic corruption. This procedure generates paired clean–noisy samples (I_n, I_c) for supervised training.

Conditioning Map and Input Representation. To support non-blind denoising across multiple noise levels, each noisy image is augmented with an additional channel encoding the relative noise strength.

For a given noise level σ, a normalized conditioning scalar is defined as

$$a = \frac{\sigma}{\sigma_{\max}},$$

where $\sigma_{\max}$ is the maximum noise level within the corresponding training regime. Specifically, we set $\sigma_{\max} = 15$ for the low-noise regime ($\sigma \in \{5, 10, 15\}$) and $\sigma_{\max} = 50$ for the moderate-to-high regime ($\sigma \in \{10, 30, 50\}$). This yields $a \in [0, 1]$.

The scalar a is broadcast spatially to form a single-channel tensor $\mathbf{S} \in [0, 1]^{1\times H\times W}$, referred to as the *conditioning map*. The final network input becomes

$$\mathbf{X} = [I_n; \mathbf{S}] \in \mathbb{R}^{4\times H\times W}, \tag{2}$$

where $I_n \in [0, 1]^{3\times H\times W}$ is the RGB noisy image. The conditioning strategy is applied consistently to both CVNN-AFA and the real-valued baseline models to ensure a fair non-blind comparison.

Normalization. All pixel values are scaled to the range $[0, 1]$ before being converted into tensors. No mean-variance normalization is performed, as the model directly learns to operate on normalized luminance and chrominance values. The dataset pipeline is implemented using the PyTorch `Dataset` interface and supports multi-threaded data loading during training and validation.

2.2. Proposed Method: Complex-Valued Neural Network with Adaptive Frequency Attention (CVNN-AFA)

Real-valued CNNs are effective in spatial filtering but lack the ability to represent phase information, which is critical for periodic noise modeling. Periodic degradation appears as structured patterns in both the spatial and frequency domains, often encoded in the phase of the signal. In contrast, complex-valued neural networks (CVNNs) provide a natural framework for representing both amplitude and phase simultaneously. Given an input feature map $\mathbf{z} = \mathbf{x}_r + j\mathbf{x}_i$, the real part encodes spatial intensity correlations, while the imaginary part captures orthogonal phase components. This dual representation improves the network's sensitivity to directional and harmonic distortions that characterize periodic interference.

Moreover, operations in the complex domain preserve rotation equivariance in the frequency plane and maintain energy consistency under Fourier transforms. These properties allow the proposed CVNN-AFA to handle structured noise in a way that complements the spectral modulation performed by the attention module described below.

2.2.1. Complex Convolutional Blocks

Each layer in the proposed network performs convolution directly in the complex domain. Let $\mathbf{W}_r, \mathbf{W}_i$ denote the real and imaginary components of a complex filter. Given an input feature

pair $(\mathbf{x}_r, \mathbf{x}_i)$, the complex convolution is defined as shown in Eq. (3):

$$\begin{aligned} \mathbf{y}_r &= \mathbf{x}_r * \mathbf{W}_r - \mathbf{x}_i * \mathbf{W}_i, \\ \mathbf{y}_i &= \mathbf{x}_r * \mathbf{W}_i + \mathbf{x}_i * \mathbf{W}_r, \end{aligned} \tag{3}$$

where $*$ denotes 2D convolution.

Each convolution is followed by complex batch normalization (CBN) and a modulated ReLU activation (ModReLU). CBN is implemented as independent normalization of the real and imaginary components, as detailed in Section 2.5.5.

The ModReLU activation for each channel c is defined as:

$$\text{ModReLU}(z_c) = \frac{\max(0, |z_c| + b_c)}{|z_c| + \varepsilon} z_c, \tag{4}$$

where b_c is a learnable per-channel bias and ε ensures numerical stability. This formulation applies magnitude thresholding while preserving phase, since the scaling factor is real-valued. The bias parameter b_c is initialized to zero and broadcast spatially, allowing the activation to behave approximately as identity at the start of training. A small constant $\varepsilon = 1 \times 10^{-8}$ is used in the denominator for numerical stability.

To maintain stability and enable residual learning across scales, each complex block integrates a skip connection:

$$(\mathbf{y}_r, \mathbf{y}_i) = \mathcal{F}(\mathbf{x}_r, \mathbf{x}_i) + (\mathbf{x}_r, \mathbf{x}_i), \tag{5}$$

where $\mathcal{F}$ denotes two stacked complex convolutional layers with CBN and ModReLU. This structure corresponds to the `ComplexBlock` implementation.

2.2.2. Adaptive Frequency Attention (AFA)

At the bottleneck of the network, a specialized module named *Adaptive Frequency Attention (AFA)* is inserted to dynamically modulate feature responses in the Fourier domain. Given complex features $\mathbf{z} = \mathbf{x}_r + j\mathbf{x}_i$, a 2D orthonormal FFT is applied independently per channel, followed by a frequency shift operation:

$$\mathbf{Z}_f = \text{fftshift}(\mathcal{F}(\mathbf{z})).$$

This centers the zero-frequency component prior to radial partitioning. AFA partitions the spectrum into three radial bands: low, mid, and high frequencies, using precomputed masks $M_{\text{low}}, M_{\text{mid}}, M_{\text{high}}$. Let $\mathbf{Z}_f \in \mathbb{C}^{C \times H \times W}$ denote the centered Fourier spectrum after fftshift. For each frequency coordinate (u, v), we define the radial distance from the zero-frequency center as

$$d(u, v) = \sqrt{(u - H/2)^2 + (v - W/2)^2}. \tag{6}$$

Two radius thresholds are defined proportionally to the minimum spatial resolution:

$$r_{\text{low}} = \alpha \cdot \min(H, W), \quad r_{\text{high}} = \beta \cdot \min(H, W), \tag{7}$$

where α and β are radial scaling hyperparameters controlling the band boundaries. In all experiments, we set $\alpha = 0.25$ and $\beta = 0.5$. The radial masks are constructed as

$$M_{\text{low}}(u, v) = \begin{cases} 1, & d(u, v) < r_{\text{low}}, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

$$M_{\text{mid}}(u, v) = \begin{cases} 1, & r_{\text{low}} \leq d(u, v) \leq r_{\text{high}}, \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

$$M_{\text{high}}(u,v) = \begin{cases} 1, & d(u,v) > r_{\text{high}}, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

These masks are computed dynamically for each feature-map resolution during the forward pass and broadcast across channels. For each band b, the average log-magnitude response is computed to form a global descriptor:

$$g_{b,c} = \frac{1}{|M_b|} \sum_{u,v} M_b(u,v)\, \log(1 + |Z_{f,c}(u,v)|), \tag{11}$$

The descriptor is computed independently for each channel c, resulting in a C-dimensional global frequency representation per band. Each descriptor $\mathbf{g}_b$ from Eq. (11) is then passed through a two-layer perceptron (implemented as 1×1 convolutions) to obtain an attention gate, as formulated in Eq. (12):

$$\mathbf{a}_b = \tanh\big(W_{2,b}\, \text{ReLU}(W_{1,b}\, \mathbf{g}_b)\big). \tag{12}$$

The corresponding scaling factor is $\mathbf{s}_b = 1 + \gamma\, \mathbf{a}_b$, where γ is a gain parameter progressively increased during training (*AFCA warm-up* in code). The modulated spectrum is then reconstructed by applying the learned scaling factors to their respective bands, as shown in Eq. (13):

$$\tilde{\mathbf{Z}}_f = \mathbf{Z}_f \odot (M_{\text{low}}\mathbf{s}_{\text{low}} + M_{\text{mid}}\mathbf{s}_{\text{mid}} + M_{\text{high}}\mathbf{s}_{\text{high}}), \tag{13}$$

and finally transformed back to the spatial domain using an inverse FFT, $\tilde{\mathbf{z}} = \mathcal{F}^{-1}(\tilde{\mathbf{Z}}_f)$. This mechanism adaptively emphasizes or suppresses specific frequency bands depending on the learned amplitude statistics, effectively removing periodic noise components while preserving high-frequency textures. The process is summarized in Algorithm 1.

Algorithm 1 Adaptive Frequency Attention (AFA)

Require: Complex features $\mathbf{z} = \mathbf{x}_r + j\,\mathbf{x}_i \in \mathbb{C}^{C\times H\times W}$, gain γ, band masks $M_{\text{low}}, M_{\text{mid}}, M_{\text{high}}$
Ensure: Band-modulated complex features $\tilde{\mathbf{z}}$

$\mathbf{Z}_f \leftarrow \text{fftshift}(\mathcal{F}(\mathbf{z}))$ {Transform to centered frequency domain}
$\mathbf{M} \leftarrow \log(1 + |\mathbf{Z}_f|)$ {Compute log-magnitude response}
for $b \in \{\text{low},\ \text{mid},\ \text{high}\}$ **do**
  $\mathbf{g}_b \leftarrow \frac{1}{|M_b|} \sum_{u,v} M_b(u,v)\, \mathbf{M}(u,v)$ {cf. Eq. (11)}
  $\mathbf{a}_b \leftarrow \tanh\big(W_{2,b}\, \text{ReLU}(W_{1,b}\, \mathbf{g}_b)\big)$ {cf. Eq. (12)}
  $\mathbf{s}_b \leftarrow 1 + \gamma\, \mathbf{a}_b$
end for
$\mathbf{S} \leftarrow M_{\text{low}}\mathbf{s}_{\text{low}} + M_{\text{mid}}\mathbf{s}_{\text{mid}} + M_{\text{high}}\mathbf{s}_{\text{high}}$ {Assemble full scaling mask}
$\tilde{\mathbf{Z}}_f \leftarrow \mathbf{Z}_f \odot \mathbf{S}$ {cf. Eq. (13)}
$\tilde{\mathbf{z}} \leftarrow \mathcal{F}^{-1}(\text{ifftshift}(\tilde{\mathbf{Z}}_f))$ {Transform back to spatial domain}
return $\tilde{\mathbf{z}}$

Unlike conventional squeeze-and-excitation (SE) or channel-attention modules that operate purely in the spatial domain using global average pooling, the proposed AFA explicitly computes descriptors in the *Fourier domain*. More importantly, instead of producing a single global channel descriptor, AFA partitions the spectrum into three disjoint radial bands (low, mid, and high frequencies) defined by distance from the spectral center in Eq. (7) with $\alpha = 0.25$ and $\beta = 0.5$ in our implementation. This radial partitioning reflects the geometric structure of periodic interference, which appears as localized peaks in the frequency plane.

Each band is processed independently using a dedicated two-layer perceptron, producing three distinct channel-wise attention vectors. The final modulation mask is constructed by recombining the scaled bands:

$$\mathbf{S} = M_{\text{low}}\mathbf{s}_{\text{low}} + M_{\text{mid}}\mathbf{s}_{\text{mid}} + M_{\text{high}}\mathbf{s}_{\text{high}},$$

which differs fundamentally from single-branch frequency attention or spatial channel reweighting. This band-wise modulation enables selective amplification or suppression of specific frequency regions rather than uniform spectral scaling.

The two-dimensional Fourier transform is applied independently to each channel using orthonormal normalization, i.e., $\mathcal{F}(\cdot) = \mathrm{FFT2}(\cdot, \texttt{norm='ortho'})$. All Fourier-domain operations are implemented using PyTorch's complex FFT routines, which are fully differentiable and integrated with automatic differentiation (autograd). A frequency shift operation (fftshift) is applied to center the zero-frequency component prior to radial partitioning, and the inverse shift (ifftshift) is applied before the inverse FFT. No windowing, spectral smoothing, or additional frequency-domain preprocessing is introduced.

2.2.3. Overall Architecture of CVNN-AFA

The proposed CVNN-AFA follows a U-Net-like encoder-decoder architecture entirely built from complex-valued operations. The encoder comprises three stages of complex convolutional blocks, with the output of each stage $\mathbf{E}_k$ saved for a skip connection. Each stage is followed by stride-2 complex downsampling. At the bottleneck, a complex residual block, the AFA module, and another residual block together form the deepest feature representation. The decoder mirrors the encoder using complex transposed convolutions for upsampling. At each decoder stage k, the upsampled features $\mathbf{U}_k$ are concatenated with the corresponding skip features $\mathbf{E}_k$ from the encoder. The fused result is then processed by the decoder's complex block $\mathcal{G}_k$, as shown in Eq. (14):

$$\mathbf{B}_k = \mathcal{G}_k([\mathbf{U}_k, \mathbf{E}_k]), \tag{14}$$

where $[\cdot, \cdot]$ denotes channel-wise concatenation, and $\mathbf{B}_k$ is the refined output feature map that becomes the input for the next upsampling stage. Finally, a complex 3×3 convolution head reconstructs the real-valued residual, which is subtracted from the input noisy image to produce the clean estimate, as defined in Eq. (15):

$$\hat{I}_c = I_n - \mathrm{Re}(\mathbf{y}), \tag{15}$$

where I_n is the noisy RGB input, and $\mathrm{Re}(\mathbf{y})$ denotes the real component of the final complex output, representing the predicted residual noise. The complete forward pass is detailed in Algorithm 2.

2.3. Real-Valued CNN Baseline

To ensure a fair comparison and isolate the contribution of complex-valued representation and frequency attention, we implement a real-valued convolutional neural network (RealCNN). This baseline follows the same residual denoising principle as the proposed CVNN-AFA and replaces each complex operation with its standard real-valued equivalent.

Specifically, the network consists of an input stem convolution followed by multiple residual blocks, each implementing the function in Eq. (16):

$$\mathbf{y} = \mathbf{x} + \mathcal{F}_{\mathrm{real}}(\mathbf{x}), \tag{16}$$

where $\mathcal{F}_{\mathrm{real}}(\cdot)$ denotes a block containing two successive sets of Conv2D $\rightarrow$ BN $\rightarrow$ ReLU layers. This structure is analogous to the complex block in Eq. (5) but operates purely in the real domain. The output head of the network predicts a residual image $\mathbf{r}$. This residual is then subtracted from the noisy input to produce the final clean estimate, as shown in Eq. (17):

$$\hat{I}_c = I_n - \mathbf{r}. \tag{17}$$

Both RealCNN and CVNN-AFA share identical hyperparameters, optimizer settings, data augmentation, and conditioning-map inputs. In particular, the input tensor remains four-channel:

Algorithm 2 Forward propagation of the proposed CVNN-AFA model

Require: Noisy input $I_n \in [0,1]^{3\times H\times W}$, conditioning map $\mathbf{S} \in [0,1]^{1\times H\times W}$
Ensure: Restored clean estimate $\hat{I}_c$

$\mathbf{X}_r \leftarrow [I_n; \mathbf{S}]$; $\mathbf{X}_i \leftarrow \mathbf{0}$ {Initialize complex input}
Encoder:
for stage $k \in \{1,2,3\}$ **do**
  $(\mathbf{E}_{k,r}, \mathbf{E}_{k,i}) \leftarrow \text{ComplexBlocks}(\mathbf{X}_r, \mathbf{X}_i)$ {Save features for skip connection}
  $(\mathbf{X}_r, \mathbf{X}_i) \leftarrow \text{Downsample}(\mathbf{E}_{k,r}, \mathbf{E}_{k,i})$
end for
Bottleneck:
$(\mathbf{B}_r, \mathbf{B}_i) \leftarrow \text{ComplexBlock}(\mathbf{X}_r, \mathbf{X}_i)$
$(\mathbf{B}_r, \mathbf{B}_i) \leftarrow \text{AFA}(\mathbf{B}_r, \mathbf{B}_i)$ {Apply frequency attention}
$(\mathbf{B}_r, \mathbf{B}_i) \leftarrow \text{ComplexBlock}(\mathbf{B}_r, \mathbf{B}_i)$
Decoder:
for stage $k \in \{3,2,1\}$ **do**
  $(\mathbf{U}_{k,r}, \mathbf{U}_{k,i}) \leftarrow \text{Upsample}(\mathbf{B}_r, \mathbf{B}_i)$
  $(\mathbf{C}_r, \mathbf{C}_i) \leftarrow [\mathbf{U}_{k,r}; \mathbf{E}_{k,r}], [\mathbf{U}_{k,i}; \mathbf{E}_{k,i}]$ {Concatenate skip features}
  $(\mathbf{B}_r, \mathbf{B}_i) \leftarrow \text{ComplexBlock}(\mathbf{C}_r, \mathbf{C}_i)$ {Refine fused features}
end for
Output Reconstruction:
$\mathbf{R} \leftarrow \text{Re}(\text{Head}(\mathbf{B}_r, \mathbf{B}_i))$ {Compute final real residual}
$\hat{I}_c \leftarrow I_n - \mathbf{R}$ {Restore clean image via residual subtraction}
return $\hat{I}_c$

three for RGB and one for the noise-strength map, ensuring that both models are trained under the same supervision and data distribution. This controlled experimental design isolates the effect of complex-valued processing and adaptive frequency modulation, allowing for a direct and fair quantitative comparison in Section 3.

2.4. Frequency-Aware Real Baseline (RealCNN-Freq)

To provide a stronger and more competitive baseline, we further implement a frequency-aware real-valued CNN (RealCNN-Freq). This model extends the spatial RealCNN by incorporating a multi-band frequency attention module analogous to the AFA mechanism used in CVNN-AFA.

Specifically, feature maps at the bottleneck are transformed to the Fourier domain using a 2D orthonormal FFT. The spectrum is partitioned into low-, mid-, and high-frequency radial bands using the same radial distance formulation and band thresholds (Eqs. (6)–(7)). Each band is independently reweighted via learned attention coefficients before inverse transformation back to the spatial domain. Unlike CVNN-AFA, all operations are performed on real-valued feature representations.

By comparing RealCNN-Freq with CVNN-AFA under identical frequency modulation and training settings, we isolate the effect of complex-valued representation from that of frequency-aware gating. This ensures that any performance improvement cannot be attributed solely to spectral partitioning or attention mechanisms.

2.5. Training Scheme

2.5.1. Loss Function

The network is trained to minimize a hybrid loss function that operates in both the spatial and frequency domains. The primary spatial component is the Charbonnier loss. The Charbonnier loss function is a robust alternative to traditional loss functions like L2, providing better performance in image denoising by being less sensitive to outliers, thus enhancing the quality of reconstructed

images in the presence of noise [15]. A smooth approximation of the ℓ_1 norm defined in Eq. (18):

$$\mathcal{L}_{\text{char}}(\hat{I}_c, I_c) = \mathbb{E}_\Omega \left[\sqrt{(\hat{I}_c - I_c)^2 + \varepsilon^2} \right], \tag{18}$$

where $\hat{I}_c$ is the restored image, I_c is the clean reference, and we set $\varepsilon = 10^{-6}$. To provide direct supervision in the frequency domain, we add the spectral consistency loss shown in Eq. (19):

$$\mathcal{L}_{\text{spec}}(\hat{I}_c, I_c) = \big\| \, |\mathcal{F}(\hat{I}_c)| - |\mathcal{F}(I_c)| \, \big\|_1, \tag{19}$$

where $\mathcal{F}$ is the 2D FFT operator. The overall training objective, formulated in Eq. (20), is a weighted sum of the spatial loss from Eq. (18) and the spectral loss from Eq. (19):

$$\mathcal{L} = \mathcal{L}_{\text{char}} + \lambda_{\text{spec}} \mathcal{L}_{\text{spec}}, \tag{20}$$

with the spectral weight λ_{spec} serving as a key hyperparameter. This joint objective encourages the model to learn both pixel-level accuracy and frequency-level fidelity.

The spectral consistency loss is treated as an optional regularization term. When enabled, it is applied consistently to both the proposed CVNN-AFA and the Frequency-Aware Real Baseline (RealCNN-Freq), ensuring a fair comparison under matched spectral supervision. The shallow RealCNN baseline is trained using only the spatial Charbonnier loss, serving as a purely spatial reference model, and therefore excludes spectral supervision by design. In practice, the spectral term is implemented using an orthonormal FFT and may be linearly warmed up during early epochs to stabilize optimization. This design prevents improvements attributed to complex-valued processing from being conflated with additional spectral regularization.

2.5.2. Optimization and Regularization

The loss function from Eq. (20) is minimized using the AdamW optimizer. A cosine annealing schedule, given by Eq. (21), adjusts the learning rate over E total epochs:

$$\eta(t) = \eta_{\min} + \tfrac{1}{2}(\eta_0 - \eta_{\min})(1 + \cos \tfrac{\pi t}{E}), \tag{21}$$

where $\eta_{\min} = 10^{-6}$. Techniques utilizing cosine annealing have shown significant improvements in denoising performance, particularly in complex models like deep residual networks, which benefit from stable training dynamics [16]. To ensure stable convergence, several regularization techniques are employed. First, the AFA module's gain parameter γ is gradually warmed up according to Eq. (22):

$$\gamma(t) = \min\left(1, \frac{t}{T_{\text{afa}}}\right) \gamma_{\max}. \tag{22}$$

Second, the global ℓ_2 norm of the gradients is clipped to a maximum value of $\tau = 1.5$.

2.5.3. Evaluation Metrics

Model performance is evaluated using the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM). Given the mean squared error (MSE) between a clean target I_c and a prediction $\hat{I}_c$, and assuming a dynamic range $I_{\max} = 1.0$ (since all images are normalized to $[0, 1]$ prior to training and evaluation), PSNR is defined as:

$$\text{PSNR} = -10 \log_{10}(\text{MSE}). \tag{23}$$

SSIM provides a complementary measure of perceptual quality by comparing local statistics of luminance, contrast, and structure, formulated in Eq. (24):

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{24}$$

where local statistics are computed with an 11×11 Gaussian window ($\sigma = 1.5$). Scores are averaged across the test set for each noise level.

2.5.4. Experimental Configuration

All experiments are performed with a fixed random seed (42) for reproducibility. During training, images are augmented with random horizontal flips and rotations and dynamically cropped into 128×128 patches. Table 1 summarizes the key hyperparameter settings explored in our study.

Table 1: Key experimental parameters for model training and evaluation.

Parameter	Values / Description
Constant Setup	
Noise Type	Periodic (sinusoidal)
Image & Patch Size	128×128 pixels
Gradient Clip Norm (τ)	1.5
Random Seed	42
Hyperparameters Explored	
Model Architecture	RealCNN, CVNN-AFA
Epochs (E)	50, 80
Batch Size	2, 8
Learning Rate (η_0)	1×10^{-4}, 5×10^{-5}
Spectral Weight (λ_{spec})	0.0, 0.05, 0.1, 0.2, 0.3 (CVNN only)
AFA Warm-Up (T_{afa})	5, 10 epochs (CVNN only)
AFA Gain ($\gamma_{\max}$)	0.5, 0.6, 0.8 (CVNN only)

These configurations were designed to analyze the effect of spectral weighting and frequency attention scaling while maintaining a fair comparison with the real-valued baseline.

Noise-Level Protocol. Two distinct experimental regimes are considered:

- **Low-noise regime:** $\sigma \in \{5, 10, 15\}$
- **Moderate-noise regime:** $\sigma \in \{10, 30, 50\}$

For each regime, a separate model is trained from scratch using the corresponding set of noise levels. During training, noise levels are sampled uniformly from the specified set, resulting in joint multi-level supervision within that regime. Although $\sigma = 10$ appears in both regimes, the models are trained independently and do not share weights across regimes.

Training Budget Control. To ensure a fair comparison, all primary experiments are conducted under identical training budgets. Unless otherwise stated, models are trained for 50 epochs using the same optimizer (AdamW), learning rate schedule (cosine annealing), batch size, gradient clipping, conditioning strategy, and random seed. The total number of parameter updates (epochs $\times$ iterations per epoch) is therefore matched across models. All models, including the real-valued baselines and CVNN-AFA, receive identical non-blind conditioning maps encoding noise strength, ensuring that performance differences arise from architectural design rather than auxiliary information.

Additional 80-epoch experiments are conducted solely to analyze convergence behavior and extended training stability. These extended runs are reported separately and are not mixed with the primary 50-epoch comparison results. Validation PSNR curves across epochs are provided to illustrate convergence consistency.

2.5.5. Implementation Details

All models are optimized using AdamW with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and weight decay $= 0.0$. A cosine annealing learning rate schedule is applied with $T_{\max} = E$ and minimum learning rate 1×10^{-6}. Gradient clipping with norm threshold $\tau = 1.5$ is applied consistently unless otherwise stated. Complex-valued layers employ complex batch normalization (CBN) and

ModReLU activations. CBN is implemented by applying standard `BatchNorm2d` independently to the real and imaginary components of each feature map. For both components, we use $\varepsilon = 1 \times 10^{-5}$ and momentum $= 0.1$, with affine scaling enabled. Running statistics are tracked during training and used at inference time. No cross-covariance normalization between real and imaginary parts is performed. Weights are initialized using He initialization applied independently to real and imaginary components. All FFT computations are performed in single-precision (FP32). No mixed-precision training is used to maintain numerical stability for complex-valued operations. The official BSDS500 split (200 train / 100 validation / 200 test) is strictly followed without overlap. Model parameter counts are reported alongside quantitative results. All models are implemented in PyTorch and trained on a single NVIDIA GeForce RTX series GPU. All experiments use a fixed random seed (42) to ensure reproducibility.

3. Results and Discussion

This section presents the experimental evaluation of the proposed Complex-Valued Neural Network with Adaptive Frequency Attention (CVNN-AFA) against the RealCNN baseline. We report quantitative results across multiple periodic noise levels, analyze the impact of key hyperparameters, and discuss the qualitative and interpretative aspects of model behavior.

3.1. Quantitative Analysis

The experiments were conducted on two groups of noise intensities: low-to-moderate ($\sigma \in \{5, 10, 15\}$) and moderate-to-high ($\sigma \in \{10, 30, 50\}$). All primary comparisons are conducted under matched training budgets (50 epochs, identical optimizer and scheduler settings) to ensure fairness. Extended 80-epoch experiments are reported separately to analyze convergence behavior. Results are presented per noise regime to distinguish behavior under low-amplitude and high-amplitude periodic corruption.

3.1.1. Performance on Low-to-Moderate Noise

We first evaluate performance under low-amplitude periodic interference ($\sigma \in \{5, 10, 15\}$), where structured artifacts are subtle but perceptually noticeable. Table 2 reports the final PSNR and SSIM after 50 epochs under matched training budgets.

Both CVNN-AFA and the frequency-aware real baseline achieve strong restoration quality in this regime. At $\sigma = 5$, CVNN-AFA attains the highest PSNR of 29.98 dB, outperforming RealCNN-Freq by +0.73 dB. This indicates that complex-valued modeling is particularly effective when periodic corruption is weak and phase-aligned distortions remain subtle.

At $\sigma = 10$ and $\sigma = 15$, the performance gap narrows and becomes marginal, with both models achieving comparable results. The average PSNR difference across the three noise levels is approximately +0.07 dB in favor of CVNN-AFA, indicating that both architectures are capable of handling moderate low-amplitude interference.

Table 2: Denoising performance under low-amplitude periodic noise (50 epochs). Best PSNR values per noise level are shown in bold.

Model	Params (M)	$\sigma = 5$	$\sigma = 10$	$\sigma = 15$	Avg PSNR (dB)	Avg SSIM
RealCNN-Freq	0.45	29.25	**27.82**	**26.49**	27.85	0.9386
CVNN-AFA	20.88	**29.98**	27.68	26.14	**27.93**	**0.9393**

Although CVNN-AFA contains substantially more parameters (20.88M) than RealCNN-Freq (0.45M), the observed performance difference remains modest. This suggests that improvements in this regime are attributable to representational properties rather than model capacity alone. Overall, these results indicate that complex-valued representation provides a measurable but modest advantage in the low-noise regime. The benefit is most visible when interference is subtle and preserving fine structural details is critical.

3.1.2. Performance on Moderate-to-High Noise

We next evaluate the models under stronger periodic corruption ($\sigma \in \{10, 30, 50\}$). Results reported in Table 3 correspond to 80-epoch training to ensure convergence stability under higher noise amplitudes.

At $\sigma = 10$, CVNN-AFA achieves slightly higher PSNR (28.96 dB) than RealCNN-Freq (28.63 dB), indicating that complex-valued modeling remains beneficial under moderate structured interference. However, as the noise amplitude increases, the performance trend reverses.

At $\sigma = 30$, RealCNN-Freq outperforms CVNN-AFA by a noticeable margin (25.35 dB vs. 23.45 dB). The gap becomes substantial at $\sigma = 50$, where RealCNN-Freq achieves 24.30 dB while CVNN-AFA reaches 20.69 dB. This indicates that, under extremely strong periodic corruption, explicit multi-band frequency gating in the real domain provides more effective suppression than complex-valued representation alone. Notably, despite having a substantially larger parameter count, CVNN-AFA does not exhibit superior robustness under extreme noise levels. This observation reinforces that capacity alone does not determine performance, and that regime-dependent architectural properties play a critical role.

Table 3: Denoising performance under moderate-to-high periodic noise (80 epochs). Best PSNR per noise level is highlighted in bold.

Model	Params (M)	$\sigma = 10$	$\sigma = 30$	$\sigma = 50$	Avg PSNR (dB)	Avg SSIM
CVNN-AFA	20.88	**28.96**	23.45	20.69	24.37	0.8773
RealCNN-Freq	0.45	28.63	**25.35**	**24.30**	**26.09**	**0.9091**

We hypothesize that, at very high noise levels, the degradation dominates both magnitude and phase components of the signal, reducing the relative advantage of phase-aware modeling. In this regime, stable multi-band frequency attenuation appears more critical than fine-grained amplitude-phase interaction. Consequently, the frequency-aware real baseline exhibits superior robustness under extreme structured interference.

3.2. Convergence Behavior

To ensure that the reported performance differences are not artifacts of early stopping or mismatched training budgets, we analyze per-sigma convergence curves under strictly controlled optimization settings.

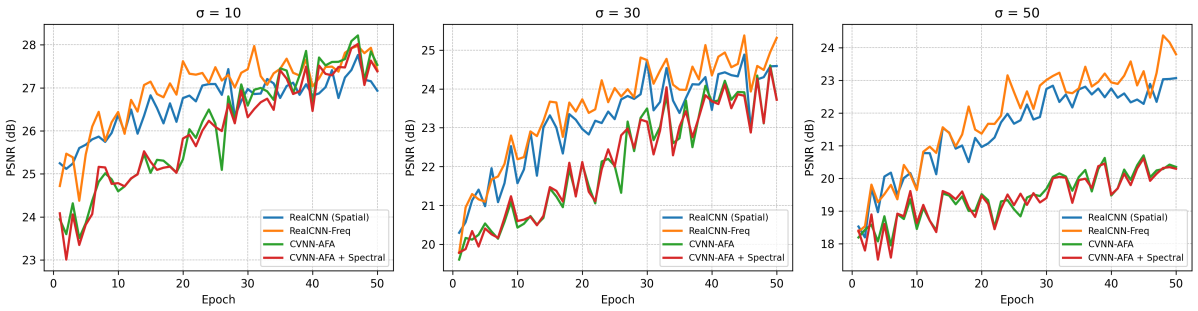

Fig. 1: Per-sigma PSNR convergence under moderate-to-high periodic noise ($\sigma = 10, 30, 50$) over 50 epochs. All models are trained under identical optimization settings and matched training budgets. At $\sigma = 10$, CVNN-AFA remains competitive with the frequency-aware real baseline. At higher noise levels ($\sigma = 30$ and $\sigma = 50$), RealCNN-Freq consistently achieves higher PSNR throughout training. The stable ranking across epochs indicates that the final performance differences are not due to early stopping effects.

Fig. 1 presents PSNR evolution over 50 epochs for $\sigma \in \{10, 30, 50\}$. All models are trained with identical optimizer configurations, learning-rate schedules, batch sizes, and numbers of parameter updates. Across all noise levels, the relative ranking between models stabilizes early

and remains consistent throughout training, indicating that the final performance differences are not caused by transient optimization dynamics.

Under moderate corruption ($\sigma = 10$), CVNN-AFA maintains a consistent advantage over the purely spatial RealCNN and remains competitive with the frequency-aware real baseline. However, as the noise amplitude increases ($\sigma = 30$ and $\sigma = 50$), RealCNN-Freq surpasses CVNN-AFA early in training and preserves this margin across epochs. Importantly, this behavior is observed from the initial convergence phase and does not reverse at later epochs. This confirms that the regression observed at $\sigma = 50$ is not due to insufficient training duration but reflects an intrinsic limitation of complex-valued frequency modulation under extreme periodic corruption.

To further validate this conclusion, we extend training to 80 epochs for selected models, as shown in Fig. 2. The extended curves demonstrate similar convergence trends and preserve the same performance ordering. The absence of late-stage crossover behavior indicates that increasing optimization budget does not alter the relative strengths of the models.

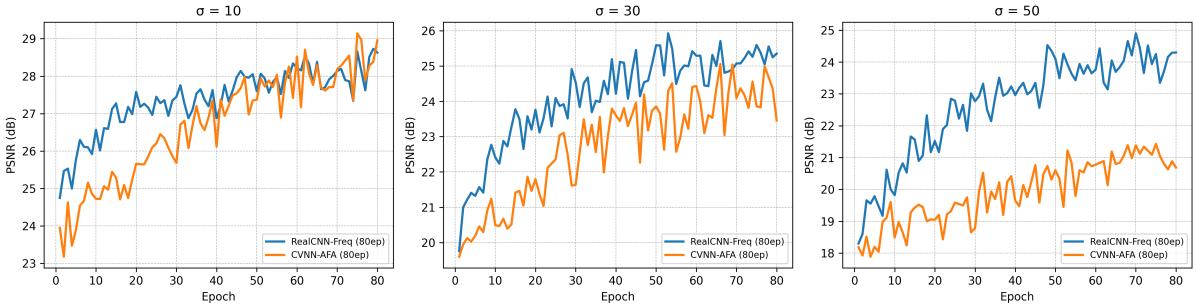

Fig. 2: Extended 80-epoch convergence curves under moderate-to-high periodic noise. The relative performance ordering remains unchanged compared to the 50-epoch setting. No late-stage crossover behavior is observed, confirming that increased training duration does not reverse the regression of CVNN-AFA at $\sigma = 50$.

Fig. 3 shows convergence curves for the low-noise regime ($\sigma \in \{5, 10, 15\}$). In this setting, CVNN-AFA consistently achieves equal or superior PSNR across epochs, particularly at $\sigma = 5$. The performance gap is modest but stable, and no instability or oscillatory behavior is observed. These results demonstrate reliable optimization of complex-valued parameters under realistic low-amplitude periodic distortions.

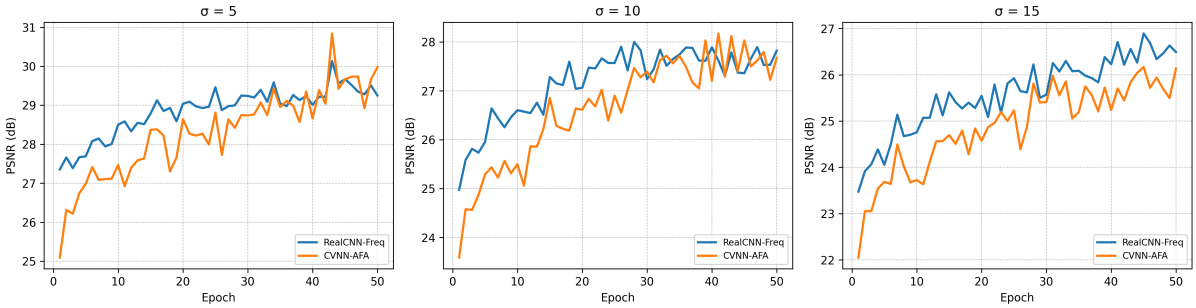

Fig. 3: Per-sigma PSNR convergence under low-amplitude periodic noise ($\sigma = 5, 10, 15$) over 50 epochs. CVNN-AFA achieves consistently equal or superior performance across epochs, particularly at $\sigma = 5$. The convergence curves exhibit smooth and stable optimization behavior without oscillatory dynamics, demonstrating reliable training of the complex-valued architecture in low-noise conditions.

3.3. Ablation Study

We conduct an ablation study to isolate the contributions of two key components: (i) the spectral regularization weight (λ_{spec}), and (ii) the maximum Adaptive Frequency Attention gain ($\gamma_{\max}$). As described in Subsection 2.5, λ_{spec} controls the strength of frequency-domain supervision, while

$\gamma_{\max}$ regulates the magnitude of band-wise spectral modulation.

Effect of Spectral Loss Weight. Introducing a small spectral regularization term ($\lambda_{\text{spec}} = 0.05$) yields marginal improvements in certain moderate-noise settings but does not consistently improve performance across all noise levels. Increasing the weight further ($\lambda_{\text{spec}} \geq 0.1$) produces no measurable gain and occasionally leads to slight degradation. These observations indicate that the primary supervision signal is already effectively captured by the spatial-domain Charbonnier loss in Eq. (18). Therefore, explicit spectral regularization plays a secondary role compared to the architectural frequency modulation performed by AFA.

Effect of AFA Gain. In contrast, increasing the maximum AFA gain to $\gamma_{\max} = 0.8$ consistently improves denoising performance in low-to-moderate noise regimes. Higher gain allows stronger reweighting of frequency bands, enabling more aggressive suppression of structured interference. However, under extreme corruption ($\sigma = 50$), the benefit of large gain diminishes, suggesting that excessive modulation cannot compensate for severe degradation of both magnitude and phase information. This behavior is consistent with the convergence analysis presented in Section 3.2.

Overall, the ablation results indicate that performance improvements primarily stem from the complex-valued representation and adaptive frequency attention mechanism itself, rather than from auxiliary spectral loss terms. The spectral loss provides limited complementary benefit, whereas adaptive frequency scaling constitutes the dominant factor in structured noise suppression.

3.4. Qualitative and Spectral Analysis

To complement the quantitative evaluation, we present a representative example at $\sigma = 10$ in Fig. 4. The noisy input exhibits visible vertical stripe artifacts that distort the background wall and slightly affect texture perception.

The frequency-aware real baseline (RealCNN-Freq) effectively attenuates the dominant stripe patterns. However, faint residual traces remain observable in relatively uniform regions, particularly along the upper background area. The proposed CVNN-AFA further reduces these residual patterns while maintaining comparable sharpness in object boundaries and textured regions. The visual improvement is moderate rather than dramatic, but it is consistent with the quantitative PSNR gain observed in the low-noise regime.

Fig. 4: Visual comparison under periodic noise with $\sigma = 10$. From left to right: (a) Ground Truth, (b) Noisy Input, (c) RealCNN-Freq baseline, and (d) CVNN-AFA (proposed). Both models reduce the stripe artifacts, with CVNN-AFA providing slightly cleaner reconstruction in uniform regions.

Further insight is provided by the corresponding log-magnitude Fourier spectra shown in Fig. 5. The noisy image exhibits symmetric high-energy peaks associated with the injected periodic

interference. Both restoration models suppress these peaks, confirming their ability to attenuate structured frequency components. The suppression achieved by CVNN-AFA is slightly stronger, although the overall spectral distributions of the two restored outputs remain similar. This observation reflects the strong frequency modeling capability of the real-valued baseline, while indicating that complex-valued processing provides incremental benefit in selectively attenuating structured periodic components.

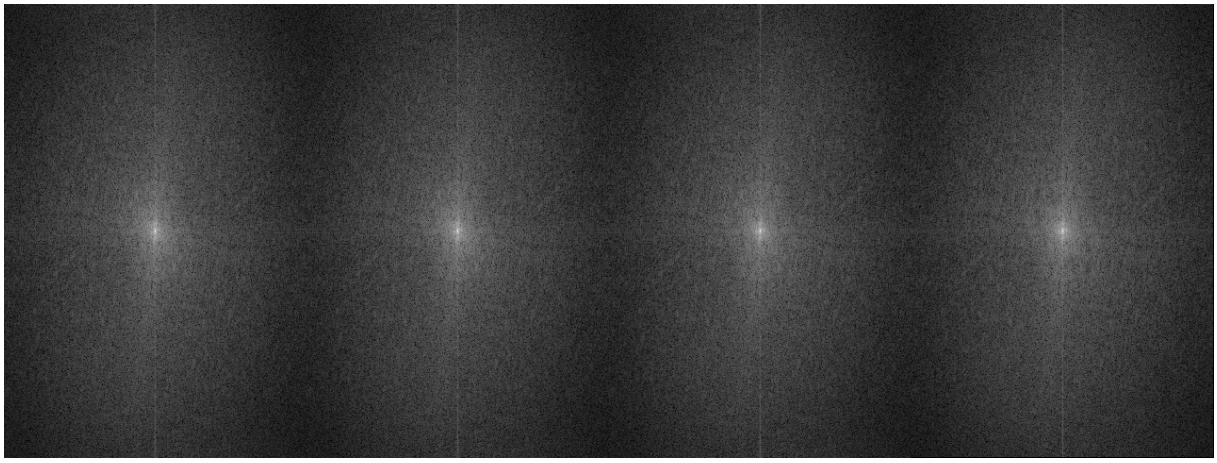

Fig. 5: Log-magnitude Fourier spectra corresponding to the spatial images in Fig. 4, displayed with shared normalization. From left to right: (a) Ground Truth, (b) Noisy Input, (c) RealCNN-Freq, and (d) CVNN-AFA. The noisy image exhibits symmetric high-energy peaks due to periodic interference. Both restoration models attenuate these peaks, with CVNN-AFA showing slightly stronger suppression while preserving the overall spectral structure.

Overall, the qualitative and spectral evidence aligns with the quantitative findings: CVNN-AFA offers consistent but moderate improvements under low-amplitude periodic corruption, rather than drastic spectral transformation.

3.5. Limitations and Future Work

Although the proposed CVNN-AFA demonstrates clear advantages in low-to-moderate noise regimes, several limitations remain.

First, as shown in Table 3 and supported by the convergence analysis in Subsection 3.2, the performance advantage of CVNN-AFA diminishes under extreme periodic corruption ($\sigma = 50$), where it is consistently outperformed by the frequency-aware real baseline. Importantly, extended training (80 epochs) does not alter this performance ordering, indicating that the regression is not caused by insufficient optimization but reflects a structural limitation of the current formulation.

Under very strong periodic interference, both magnitude and phase components of the input are severely distorted. In this regime, denoising shifts from selective frequency suppression to large-scale signal reconstruction. Because the proposed approach emphasizes adaptive spectral modulation and phase-aware filtering, its relative advantage becomes less pronounced when the underlying signal is heavily corrupted across all frequency bands. Future work may explore hybrid architectures that combine complex-valued processing with stronger spatial reconstruction modules or multi-scale residual refinement to improve robustness in extreme noise conditions.

Second, CVNN-AFA contains substantially more parameters than the frequency-aware real baseline. Despite this higher capacity, performance improvements are regime-dependent and do not consistently scale with model size. This indicates that architectural representation and inductive bias, rather than parameter count alone, govern effectiveness in structured noise removal. Nevertheless, the increased model complexity leads to higher computational cost and memory usage, which may limit deployment in resource-constrained environments. Designing more parameter-efficient complex-valued architectures remains an important direction for future

work.

Third, our experimental evaluation focuses exclusively on synthetically generated periodic noise. While this setting enables controlled analysis of structured frequency interference, it does not encompass the diversity of real-world degradation patterns. The generalization of CVNN-AFA to other noise distributions such as Gaussian noise, impulse noise, or mixed sensor artifacts remains an open question. Extending the adaptive frequency attention mechanism to dynamically respond to heterogeneous or unknown noise statistics represents an important direction for future research.

Overall, these limitations suggest that while complex-valued frequency-aware modeling is effective for structured periodic distortions, further investigation is required to enhance robustness under severe corruption and broaden applicability to more diverse real-world scenarios.

4. Conclusion

This work introduced a Complex-Valued Neural Network with Adaptive Frequency Attention (CVNN-AFA) for periodic noise removal. By integrating complex-valued feature representation with band-wise spectral modulation, the proposed architecture explicitly leverages the frequency-domain characteristics of periodic interference.

Experimental results indicate that CVNN-AFA provides consistent and moderate improvements in low-amplitude noise regimes, where interference remains spectrally localized. Under moderate corruption levels, the model remains competitive with a strong frequency-aware real-valued baseline. However, at extreme noise intensities, the relative advantage diminishes, and the real-valued frequency-aware model can achieve superior robustness. Convergence analysis confirms that these behaviors persist under extended training, suggesting regime-dependent modeling characteristics rather than optimization artifacts. Although CVNN-AFA employs substantially higher model capacity, the observed improvements are not uniformly capacity-driven but instead depend on representational alignment with the underlying noise structure. This highlights the importance of architectural inductive bias in structured noise suppression.

Overall, the study demonstrates that complex-valued representation combined with adaptive frequency attention offers a principled framework for periodic interference attenuation, particularly in low-to-moderate distortion settings. Future work may investigate hybrid reconstruction strategies, parameter-efficient complex architectures, stabilization techniques for high-noise regimes, and broader real-world noise distributions to further enhance robustness and generalization. It should be noted that the present study focuses on synthetically generated periodic noise under controlled experimental conditions. While this setting enables systematic analysis, real-world periodic interference may exhibit additional variability such as frequency drift and spatial non-stationarity. Extending the framework to real acquisition artifacts and mixed-noise scenarios remains an important direction for future investigation.

CRediT Authorship Contribution Statement

Marjono: Conceptualization, Formal Analysis, Methodology.
Avin Maulana: Software, Investigation, Validation, Visualization.
Anggi Gustiningsih Hapsani: Writing-Original Draft Preparation, Writing-Review & Editing.

Declaration of Generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Google's Gemini model for assistance with language editing, structuring the text, and refining the clarity of technical descriptions. The model was also used to correct LaTeX formatting and pseudocode algorithms based on the authors' original Python implementations. The core scientific contributions, including the proposed methodology, experimental design, data analysis, and interpretation of results, were solely conducted by the authors.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research was funded by the Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Brawijaya, through the PENELITIAN DANA INTERNAL FAKULTAS MIPA program for the 2025 fiscal year, under grant number 02115.39/UN10.F0901/B/PG/2025.

References

[1] L. Yan, M. Zhao, S. Liu, S. Shi, and J. Chen, "Cascaded Transformer U-Net for Image Restoration," *Signal Processing*, vol. 206, 2023. DOI: 10.1016/j.sigpro.2022.108902.

[2] A. S. Anandhi and M. Jaiganesh, "An enhanced image restoration using deep learning and transformer based contextual optimization algorithm," *Scientific Reports*, vol. 15, p. 10 324, 2025. DOI: 10.1038/s41598-025-94449-5.

[3] B. Li, Y. Zhang, W. Chen, Y. Ma, and L. Li, "Neural Network-Based Investigation of Periodic Noise Reduction Methods for High-Resolution Infrared Line Scanning Images," *Remote Sensing*, vol. 16, no. 5, p. 841, 2024. DOI: 10.3390/rs16050841.

[4] Z. H. Shah et al., "Image restoration in frequency space using complex-valued CNNs," *Frontiers in Artificial Intelligence*, vol. 7, p. 1 353 873, 2024. DOI: 10.3389/frai.2024.1353873.

[5] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017. DOI: 10.1109/TIP.2017.2662206.

[6] C. Ren, X. He, C. Wang, and Z. Zhao, "Adaptive Consistency Prior Based Deep Network for Image Denoising (DeamNet)," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2021, pp. 8596–8606. DOI: 10.1109/CVPR46437.2021.00849.

[7] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M. Yang, "Restormer: Efficient Transformer for High-Resolution Image Restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA: IEEE, Jun. 18–24, 2022, pp. 5728–5739. DOI: 10.1109/CVPR52688.2022.00564.

[8] Y. Gou, P. Hu, J. Lv, J. T. Zhou, and X. Peng, *Multi-scale adaptive network for single image denoising*, 2022. eprint: 2203.04313. https://arxiv.org/abs/2203.04313.

[9] A. Maulana, C. Fatichah, and N. Suciati, "Facial Inpainting Using Generative Adversarial Network with Feature Reconstruction and Landmark Loss to Preserve Spatial Consistency in Unaligned Face Images," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 6, pp. 219–228, 2020. DOI: 10.22266/ijies2020.1231.20.

[10] J. Varghese, S. Subash, K. Subramaniam, and K. P. Sridhar, "Adaptive Gaussian Notch Filter for Removing Periodic Noise from Digital Images," *IET Image Processing*, vol. 14, no. 8, pp. 1529–1538, 2020. DOI: 10.1049/iet-ipr.2018.5707.

[11] Q. Zeng, H. Qin, X. Yan, and T. Yang, "Fourier Domain Anomaly Detection and Spectral Fusion for Stripe Noise Removal of TIR Imagery," *Remote Sensing*, vol. 12, no. 22, p. 3714, 2020. DOI: 10.3390/rs12223714.

[12] S. Rawat, K. P. S. Rana, and V. Kumar, "A Novel Complex-Valued Convolutional Neural Network for Medical Image Denoising (CVMIDNet)," *Biomedical Signal Processing and Control*, vol. 69, p. 102 859, 2021. DOI: 10.1016/j.bspc.2021.102859.

[13] J. Gurrola-Ramos, O. Dalmau, and T. E. Alarcón, "A residual dense u-net neural network for image denoising," *IEEE Access*, vol. 9, pp. 31 742–31 754, 2021. DOI: 10.1109/ACCESS.2021.3061062.

[14] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011. DOI: 10.1109/TPAMI.2010.161.

[15] J. T. Barron, *A general and adaptive robust loss function*, 2019. arXiv: 1701.03077 [cs.CV]. https://arxiv.org/abs/1701.03077.

[16] G. Singh, A. Mittal, and N. Aggarwal, "Resdnn: Deep residual learning for natural image denoising," *IET Image Processing*, vol. 14, no. 11, pp. 2425–2434, 2020. DOI: 10.1049/iet-ipr.2019.0623. eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-ipr.2019.0623. https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-ipr.2019.0623.