International Journal on Advanced Technology, Engineering, and Information System
https://ojs.transpublika.com/index.php/IJATEIS
Online ISSN 2828-5425

https://doi.org/10.55047/ijateis.v4i2.1710

Original Article

ECA-MSNet: A Multi-Scale Residual U-Net with Efficient Channel Attention for Real-World Image Denoising

Abdul Fatah Nasrat[1*], Tuba Çağlikantar[2]

[1]Department of Computer Science, Gazi University, Ankara, Türkiye
[2]Department of Computer Engineering, Gazi University, Ankara, Türkiye
Email: [1)] afatah.nasrat@gazi.edu.tr, [2)] tubac@gazi.edu.tr

Received : 18 March - 2025
Accepted : 22 April - 2025
Published online : 23 April - 2025

Abstract

Real-world photographs contain complex, sensor-dependent noise that simultaneously obscures subtle high-frequency textures and broad contextual cues, making denoising a persistent challenge in low-level vision. The goal of this study is to devise a single, computationally balanced model that removes such heterogeneous noise while faithfully preserving both fine detail and global structure. We introduce ECA-MSNet, a dual-branch convolutional architecture designed around this objective. The Residual Detail Estimation Branch reconstructs delicate textures that are most susceptible to corruption, whereas the Multi-Scale Feature Restoration Branch—a U-Net enhanced with Attention-based Multi-Scale Residual Blocks and lightweight Efficient Channel Attention (ECA)—captures coarse-to-fine contextual information. A Dual Residual Fusion Module adaptively merges the two outputs, and a final Refine Block suppresses residual artifacts, yielding the restored image. Extensive experiments on the SIDD and PolyU real-noise benchmarks validate the effectiveness of the proposed method. ECA-MSNet achieves 39.41 dB / 0.9109 SSIM on SIDD and 37.76 dB / 0.9574 SSIM on PolyU, outperforming strong baselines such as DnCNN, FFDNet, CBDNet, and CycleISP. Ablation studies further confirm that each architectural component—dual-branch design, multi-scale residual blocks, channel attention, and fusion strategy—contributes measurable gains. These results demonstrate that ECA-MSNet sets a new state of the art for real-world image denoising, offering a favorable trade-off between accuracy and efficiency and providing a versatile foundation for other low-level vision tasks.

Keywords: Real-World Image Denoising, Multi-Scale Residual Network, Efficient Channel Attention (ECA), Dual Residual Fusion, Deep Learning in Low-Level Vision, Attention-Based Architectures.

1. Introduction

Real-world image denoising is a critical task in low-level computer vision, aimed at restoring clean images from photographs corrupted by complex and often unknown noise patterns. Unlike synthetic noise models—such as additive white Gaussian noise (AWGN)—real-world noise arises from multiple stages of the imaging pipeline, including photon shot noise at the sensor, demosaicing artifacts, in-camera signal processing, and compression. This produces heterogeneous, spatially variant noise whose energy spans both low- and high-frequency bands, making real-world denoising markedly more challenging than its synthetic counterpart and necessitating models that are accurate and robust across varied scenarios (Foi, 2009).

Early denoising methods relied on model-based approaches such as non-local self-similarity (NSS), sparse coding, and Markov random fields. While effective in controlled settings, these methods often struggle with spatially variant noise and incur high computational costs (Dabov, 2009). Deep learning techniques—particularly convolutional neural networks (CNNs)—have surpassed classical methods by learning data-driven priors (Zhang K. a., 2017). However, many ACNN-based models (e.g., DnCNN and FFDNet) are constrained by their limited receptive fields, restricting their ability to capture long-range dependencies (Chen L. a., 2021); moreover, most are trained on synthetic noise and thus generalize poorly to real photographs (Abdelhamed, 2018), (Plotz, 2017).

Recent research addresses these limitations along two axes. First, efficient attention mechanisms (e.g., Efficient Channel Attention, or ECA) highlight informative feature channels at negligible cost, enhancing robustness to diverse noise statistics (Zamir S. W.-H., 2022), (Zamir S. W.-H., 2021). Second, multi-scale processing aggregates context at several resolutions, which is crucial for disentangling large-scale structure from fine-detail corruption (Liang, 2021). When both requirements are forced into a single feed-forward stream, the network must compromise enlarging the receptive field can blur pixel-level precision, whereas focusing solely on high-resolution convolutions neglects global context.

We therefore decompose real-world denoising into two complementary sub-tasks. The first is fine-detail restoration, which salvages fragile edges, micro-textures, and lettering by modelling mid- and high-frequency noise with shallow residual stacks. The second is global context reconstruction, which removes low-frequency colour shifts and structural gradients through wide receptive fields and multi-scale reasoning. By allocating each sub-task to its own specialist branch, a network can optimise both goals without mutual interference, yielding complementary and redundant cues that prove robust under varying ISO levels, lighting, and sensor types.

Building on this intuition, we introduce ECA-MSNet, a dual-branch architecture that marries residual learning, multi-scale context aggregation, and efficient channel attention for real-world image denoising. A shallow Residual Detail Estimation Branch (RDEB) predicts high-frequency noise residuals and recovers crisp textures, while a U-Net-based Multi-Scale Feature Restoration Branch (MSFRB) equipped with Attention-based Multi-Scale Residual Blocks and ECA captures coarse-to-fine contextual information. Their outputs are adaptively merged by a Dual Residual Fusion Module (DRFM) and polished by a lightweight Refine Block, producing the final clean image.

Building on the above discussion, this work makes several key contributions to the field of real-world image denoising:

1) Architectural novelty — We demonstrate that an explicitly dual-branch design, guided by ECA-enhanced multi-scale residual blocks, provides superior robustness to the mixed-band characteristics of real-world noise.
2) Comprehensive evaluation — Extensive experiments on the SIDD and PolyU datasets show that ECA-MSNet consistently outperforms strong baselines—DnCNN (Zhang K. a., 2017), FFDNet (Zhang K. a., 2018), CBDNet (Guo, 2019), and CycleISP (Zamir S. W.-H., 2020)—in both PSNR and SSIM, while retaining competitive computational complexity (Yu W. a., 2022).A
3) Ablation evidence — Systematic component analysis confirms that each element—the dual-branch arrangement, multi-scale residual blocks, ECA, and DRFM—contributes measurable gains, thereby validating our design motivation.

Extensive experiments presented in Sections 4 corroborate these points, demonstrating that ECA-MSNet delivers state-of-the-art real-world denoising performance without sacrificing computational efficiency.

2. Related Work

Image denoising has been a fundamental and extensively studied problem in low-level vision, with applications ranging from photographic enhancement to high-level vision tasks such as object detection and segmentation (Bertalm'o, 2018). Over the decades, a wide array of approaches has emerged—ranging from traditional model-based algorithms to modern deep learning-based architectures.

2.1. Traditional and Model-Based Denoising Methods

Earlier denoising methods leveraged handcrafted priors and statistical models to suppress noise. Classic techniques such as BM3D (Dabov, 2007) and NLM (Buades, 2005) relied on the principles of self-similarity (Xu J. a., 2015) and non-local averaging. Other approaches employed transform-domainA filtering using DCT (Yaroslavsky, 1996), wavelets (Simoncelli, 1996), or sparse representations with learned dictionaries. Despite their efficacy in reducing noise, these methods struggled with generalization, were often computationally expensive, and typically required manual tuning of hyperparameters. Additionally, their inability to model complex image structures and adapt to spatially variant real-world noise limited their applicability in practical scenarios.

2.2. Learning-Based Denoising Approaches

The advent of convolutional neural networks (CNNs) marked a paradigm shift in image restoration. Discriminative models such as DnCNN and FFDNet significantly outperformed traditional methods by learning mappings directly from noisy-clean image pairs. DnCNN introduced residual learning to estimate noise, while FFDNet incorporated noise level maps to support non-blind denoising. However, most of these models were trained on synthetic noise (e.g., AWGN), which limited their performance on real noisy images due to the domain gap between synthetic and real noise distributions (Plotz, 2017).

To address this, CBDNet (Guo, 2019) proposed a two-branch network with a noise estimation module and a non-blind denoising module. It leveraged real-synthetic image pairs and multiple loss terms to improve robustness on real images. CycleISP (Zamir S. W.-H., 2020), on the other hand, modeled the image signal processing (ISP) pipeline to synthesize realistic noisy-clean image pairs in both RAW and sRGB spaces (Foi, 2009). Its dual attention-based network further improved performance across denoising benchmarks, especially on the SIDD and DND datasets.

2.3. Efficient Network Architectures

As deeper and wider CNNs increased restoration performance, the demand for lightweight and computationally efficient architectures also rose (Zhang Y. a., 2023). Cascaded shrinkage field models (Schmidt, 2014), dilated convolutions (Yu F. a., 2015), and modular residual designs (Bae, 2017) were proposed to reduce model complexity without sacrificing performance. Networks such as HINet (Chen L. a., 2021) introduced half-instance normalization tailored for patch-variant statistics, and NAFNet (Chen L. a., 2022) simplified attention computation through efficient channel attention mechanisms.

Recently, introduced MIRNet (Zamir S. W.-H., 2020) and Restormer (Zamir S. W.-H., 2022), which pushed the performance frontier by using hierarchical residual structures and transformer blocks. While effective, these models incurred significant inference costs and high memory usage. To tackle this issue, CascadedGaze (Li, 2023) proposed a global context extractor based on fully convolutional design to reduce the overhead of self-attention while retaining global dependency modeling.

2.4. Attention Mechanisms and Multi-Scale Feature Learning

Attention mechanisms have proven to be essential in enhancing feature representation, particularly in low-level vision tasks. Channel attention (Zhang Y. a., 2018), spatial attention (Woo, 2018), and hybrid attention modules (Pan, 2023) have all contributed to performance gains. The Efficient Channel Attention (ECA) (Wang, 2020) module stands out by striking a balance between complexity and effectiveness through local cross-channel interaction via 1D convolution, without dimensionality reduction.

Additionally, multi-scale feature learning has shown to be highly effective in modeling image structures at various spatial resolutions. Techniques that combine multi-branch convolutions with varying receptive fields (e.g., 3×3, 5×5, 7×7 kernels) enable the model to capture local textures and global context simultaneously (Liang, 2021). These principles underpin the design of our Multi-Scale Residual Blocks (MSRBs), which incorporate depthwise separable convolutions and ECA blocks for efficient contextual modeling.

Despite these advances, two notable gaps remain: (i) most attention-driven multi-scale networks fuse all features within a single backbone, forcing one hierarchy to balance pixel-level fidelity against scene-level context, and (ii) few architectures explicitly isolate high-frequency detail restoration from global structure reconstruction—an isolation that real-world, mixed-band noise strongly benefits from. Our proposed ECA-MSNet addresses both gaps by deploying a dual-branch design—one branch specialised for residual detail estimation and another for multi-scale contextual recovery—while leveraging lightweight ECA-enhanced MSRBs in the context branch. This separation of duties, coupled with an adaptive fusion module, enables ECA-MSNet to preserve delicate textures and large-scale consistency simultaneously, offering a targeted and computationally efficient response to the shortcomings identified in current literature.

3. Methods

In this section, we introduce our proposed ECA-MSNet, a novel deep neural network architecture specifically designed for effective real-world image denoising. Real-world images are typically corrupted by complex and diverse noise patterns arising from various acquisition conditions, sensor characteristics, and subsequent processing stages, which pose significant challenges for existing denoising techniques. To address these intricate challenges, our proposed approach integrates a sophisticated dual-path framework complemented by specialized attention mechanisms and feature fusion strategies.

The core architecture of ECA-MSNet, as illustrated in Figure 1, consists of two parallel branches, each designed to perform distinct yet complementary tasks. The first branch, the Residual Detail Estimation Branch, explicitly captures fine-grained residual details of the image. This branch focuses on reconstructing subtle, high-frequency features—such as edges, textures, and fine structural elements—which are particularly vulnerable to noise degradation. It utilizes a stack of convolutional blocks enhanced by batch normalization and activation layers, enabling it to isolate and predict residual details efficiently.

The second branch, the Multi-Scale Feature Restoration Branch, employs a modified U-Net structure capable of effectively modeling contextual and structural information at multiple spatial resolutions. To achieve powerful multi-scale representations, this branch integrates specialized Attention-based Multi-Scale Residual Blocks (AMSRBs). These blocks leverage depthwise separable convolutions (Chollet, 2017) with multiple receptive fields and an efficient channel attention (ECA) mechanism to dynamically emphasize informative channels, enabling the extraction and aggregation of robust multi-scale features. Moreover, a bottleneck structure in the center of this branch further refines and condenses critical information before progressively reconstructing the restored feature representations.

To maximize the benefits derived from these two parallel branches, we introduce a sophisticated fusion strategy called the Dual Residual Fusion Module (DRFM). Prior to fusion, the outputs of the Multi-Scale Feature Restoration Branch are enhanced through an additional specialized block termed the Refine Block, which optimizes feature representations and mitigates potential artifacts or residual noise. The DRFM adaptively integrates the refined multi-scale context with the detailed residual features from the first branch, producing a coherent and high-quality denoised image.

Our network extensively employs residual learning strategies throughout the architecture, where the model explicitly learns to predict the noise components rather than directly reconstructing the clean image. This choice significantly improves training stability and effectiveness, facilitating the network's convergence to optimal solutions.

The carefully designed integration of these modules and branches within ECA-MSNet results in superior real-world image denoising performance, as rigorously demonstrated by extensive evaluations on widely recognized benchmarks, surpassing several state-of-the-art methods in terms of quantitative metrics and qualitative visual quality.

In the following sections, we comprehensively describe each of these specialized modules, elucidating their internal operations, mathematical formulations, and design rationale, supported by thorough experiments validating their effectiveness.

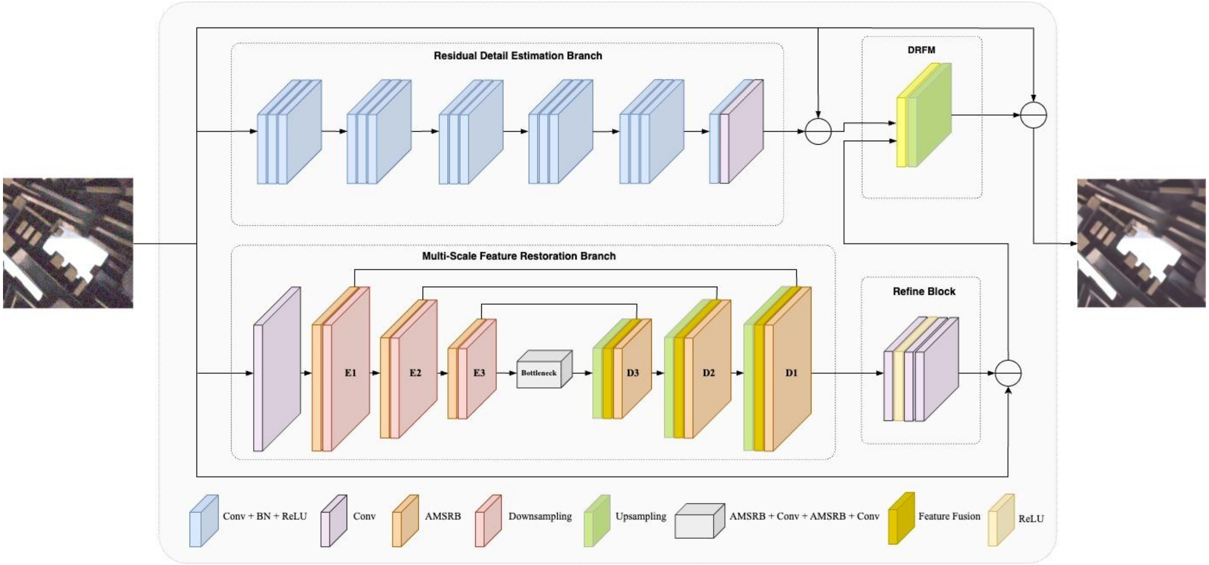

Figure 1. Overview of the ECA-MSNet architecture, illustrating the dual-branch design with a detail estimation branch, a multi-scale feature restoration branch, and a fusion module for image denoising

3.1. Residual Detail Estimation Branch

The Residual Detail Estimation Branch is specifically designed to recover high-frequency details, such as edges and textures, which are particularly susceptible to noise degradation in real-world images. Rather than directly reconstructing the clean image, this branch leverages residual learning, enabling the network to explicitly model noise components. This approach simplifies optimization and significantly improves the reconstruction quality of fine image details.

Inspired by the DnCNN architecture, the Residual Detail Estimation Branch comprises multiple sequential convolutional blocks, each containing convolutional layers followed by batch normalization and ReLU activation. Given a noisy input image $In \in R\ C{\times}H{\times}W$, the first convolutional block processes the input as:

$$F_1 = \sigma\left(\mathrm{BN}\left(Conv_{3x3}(I_n)\right)\right), \qquad (1)$$

where Conv3×3 denotes a convolution with a kernel size of 3×3, padding of 1, and stride of 1, BN indicates batch normalization, and σ is the ReLU activation. This initial processing is followed by deeper convolutional stages, each stage l comprising three convolutional layers defined as:

$$\begin{gathered} F_l^{(1)} = \sigma\left(BN\left(Conv_{3x3}(F_{l-1})\right)\right), \\ F_l^{(2)} = \sigma\left(BN\left(Conv_{3x3}\left(F_l^{(1)}\right)\right)\right), \\ F_l^{(3)} = \sigma\left(BN\left(Conv_{3x3}\left(F_l^{(2)}\right)\right)\right), \end{gathered} \qquad (2)$$

where F_{l-1} is the output from the preceding stage, and $F_l^{(3)}$is the final output of stage l. This deep cascade structure enhances the network's capacity to capture intricate patterns associated with noise. After a series of convolutional stages (six stages in our empirical setup), the feature map F_{final} is passed through a 1×1 convolution to map features back to the original channel dimension:

$$F_{detail} = Conv_{1x1}\left(F_{final}\right). \qquad (3)$$

Finally, the residual detail prediction R_{detail} is computed explicitly as:

$$R_{detail} = I_n - F_{detail}. \qquad (4)$$

3.2. Multi-Scale Feature Restoration Branch

While the Residual Detail Estimation Branch emphasizes recovering fine-grained details, the Multi-Scale Feature Restoration Branch is dedicated to capturing broader contextual information and global dependencies through multi-scale hierarchical representations. This branch incorporates an encoder-decoder structure enhanced with Attention-based Multi-Scale Residual Blocks (AMSRBs), which significantly improves the modeling capacity for diverse image features at various resolutions.

Encoder Stage: Given the noisy input image $In \in R\ C{\times}H{\times}W$, the encoder initially projects the input into high-dimensional feature representations through an initial convolutional layer:

$$E_0 = \sigma\left(Conv_{3x3}(I_n)\right), \qquad (5)$$

where σ represents the ReLU activation. Subsequently, hierarchical features are extracted at multiple scales using successive downsampling blocks interleaved with AMSRB modules. At each encoder stage l, the feature extraction follows:

$$E_l = AMSRB_l\left(Downsample(E_{l-1})\right), \qquad (6)$$

where the Downsample operation is implemented via convolutional layers with stride 2, effectively halving the spatial dimensions at each subsequent level and doubling the channel dimensions, facilitating deeper abstraction of contextual features. In this work, we employ three downsampling stages, yielding encoded representations E_1, E_2, E_3 at increasingly coarser scales.

Bottleneck Module: To further refine the deepest and most abstracted features, a bottleneck structure is applied at the bottom of the encoder-decoder hierarchy. The bottleneck employs additional AMSRB modules along with channel dimension adjustment via 1×1 convolutions to enhance feature representations without incurring excessive computational overhead:

$$B\ = Conv_{1x1}\big(AMSRB(Conv_{1x1}\ (AMSRB(E_3))\big), \tag{7}$$

This strategy effectively compresses and subsequently expands feature channels, enriching the feature representations at the lowest spatial resolution, allowing the model to capture robust global contexts crucial for subsequent reconstruction stages.

Decoder Stage: The decoder symmetrically mirrors the encoder, progressively restoring spatial resolution while refining feature maps via AMSRBs. At each decoding stage, we first apply an upsampling operation followed by feature fusion with corresponding encoder features through skip connections. The fusion is performed by concatenating the upsampled decoder features with the encoder features and subsequently applying a 1×1 convolution to fuse them:

$$D_l = AMSRB_l(Fusion_l(Upsample(D_{l+1}), E_l)), \tag{8}$$

where Upsample is a transposed convolution operation doubling spatial resolution and halving the number of channels, and the fusion process is mathematically defined as:

$$Fusion_l(D, E) = Conv_{1x1}([D; E]), \tag{9}$$

where [D; E] denotes channel-wise concatenation of the decoder features D and encoder features E. Following hierarchical decoding through multiple AMSRB modules, the decoder restores high-resolution features D_1 to the original spatial dimension, yielding a preliminary reconstruction.

Refinement of Multi-Scale Features: Before integration with the Residual Detail Estimation Branch, the restored features undergo additional refinement. This is achieved through the Refine Block, which enhances the features by suppressing irrelevant information and reinforcing important restoration features. Formally, this refined feature map $R_{refined}$ is obtained via:

$$R_{refined} = Conv_{3x3}\ \sigma\big(Conv_{3x3}(D_1)\big), \tag{10}$$

This refinement facilitates the precise integration of the multi-scale restoration results with the residual details in the subsequent Dual Residual Fusion Module (DRFM). The Multi-Scale Feature Restoration Branch thus systematically extracts and reconstructs global contextual information, complementing the detailed residual features produced by the

Residual Detail Estimation Branch and collectively ensuring comprehensive restoration of noisy real-world images.

3.3. Attention-based Multi-Scale Residual Block (AMSRB)

The Attention-based Multi-Scale Residual Block (AMSRB) is a pivotal module within the Multi-Scale Feature Restoration Branch, explicitly designed to efficiently capture image features across multiple scales and dimensions, as illustrated in Figure 2. Unlike conventional residual blocks, AMSRB employs parallel multi-scale convolutions combined with an efficient channel attention mechanism, significantly enhancing the model's capability to represent diverse feature hierarchies and suppress redundant information dynamically.

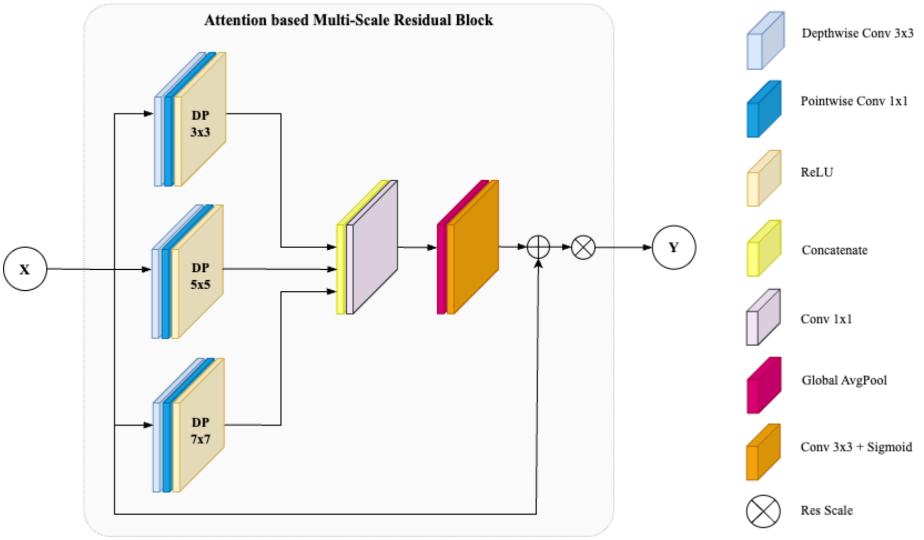

Figure 1. Structure of the AMSRB with parallel 3x3, 5x5, 7x7 convolutions and residual connection

Multi-Scale Feature Extraction: Traditional convolutional operations typically focus on fixed local receptive fields, limiting their ability to model complex features spanning multiple scales. To address this limitation, the AMSRB integrates depthwise separable convolutions with varying kernel sizes (3×3, 5×5, and 7×7), thereby enabling concurrent feature extraction at multiple spatial scales. Depthwise separable convolutions effectively reduce computational complexity while maintaining expressive feature representation. Mathematically, given an input feature map X ∈ $R\ C\times H\times W$, the multi-scale convolutions can be expressed as follows:

$$\begin{aligned} F_3 &= \sigma\left(PointwiseConv\ (DepthwiseConv_{3x3}(X)\right), \\ F_5 &= \sigma\left(PointwiseConv\ (DepthwiseConv_{5x5}(X)\right), \\ F_7 &= \sigma\left(PointwiseConv\ (DepthwiseConv_{7x7}(X)\right), \end{aligned} \tag{11}$$

where σ represents the ReLU activation function. Each F_k (for k ∈ {3, 5, 7}) extracts features specific to a different spatial scale, effectively capturing fine-grained and coarse details simultaneously.

Feature Fusion: To efficiently integrate the multi-scale features extracted from the previous step, the AMSRB employs a channel-wise concatenation followed by a 1×1 convolution to fuse these features. This fusion mechanism efficiently merges diverse spatial information into a unified feature representation. Mathematically, the fused multi-scale feature representation F_{fused} is given by:

$$F_{fused} = Conv_{1x1}\,([F_3;\ F_5;\ F_7]), \tag{12}$$

where [;] denotes the channel-wise concatenation operation. The convolution operation reduces channel dimensionality back to C, ensuring efficient subsequent computation.

Efficient Channel Attention (ECA): Although multi-scale fusion significantly enhances spatial representation, channel redundancy may still persist within the fused features. To further optimize channel interdependencies, AMSRB incorporates an Efficient Channel Attention (ECA) module. This lightweight attention mechanism dynamically recalibrates channel-wise importance without excessive computational overhead, improving the overall feature representation quality.

The ECA module first applies a global average pooling operation across spatial dimensions, converting spatial information into channel-wise statistics:

$$S_c = \frac{1}{H\ x\ W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{fused}\,(c, i\ , j), \tag{13}$$

where $S_c \in R^C$ represents the global channel statistics for the c-th channel. Next, channel-wise interactions are modeled via a 1 D convolutional operation followed by a sigmoid function, producing adaptive channel weights $A \in R^C$:

$$A = \sigma\left(Conv1D(S_c)\right), \tag{14}$$

Here, σ denotes the sigmoid activation function. Finally, the recalibrated feature map F_{eca} is computed through element-wise multiplication of the original fused feature and the attention weights:

$$F_{eca}\,(c, :, :) = F_{fused}(c, :, :)\ \odot A(c), \tag{15}$$

Where $\odot$ denotes element-wise multiplication, effectively emphasizing informative channels while suppressing less relevant ones.

Residual Connection and Stability Enhancement: To improve training stability and mitigate the vanishing gradient problem, AMSRB incorporates a residual connection, scaled by a factor α (set to 0.2 in our experiments), between the recalibrated multi-scale features F_{eca} and the original input X. The final output Y of the AMSRB is defined as follows:

$$Y = X + \alpha \times F_{eca}\ . \tag{16}$$

This residual design not only accelerates the training process but also enhances the preservation of essential features across different network depths. The AMSRB module thus effectively combines multi-scale spatial extraction, efficient channel attention, and robust residual learning, significantly enhancing the feature representation capacity required for sophisticated real-world image denoising tasks.

3.4. Dual Residual Fusion Module (DRFM)

The Dual Residual Fusion Module (DRFM) integrates complementary outputs from the Residual Detail Estimation Branch (O_{RDEB}) and the refined output from the Multi-Scale Feature Restoration Branch (O_{MSFRB}). It first concatenates these outputs along the channel dimension:

$$F_{concat} = O_{RDEB} \oplus O_{MSFRB} \tag{17}$$

A subsequent 1×1 convolution adaptively recalibrates this combined feature set:

$$F_{fused} = Conv_{1x1}\,(F_{concat}) \tag{18}$$

Finally, DRFM leverages residual learning to estimate and subtract noise directly from the original noisy image I_{noisy}:

$$I_{final} = I_{noisy} - F_{fused} \tag{19}$$

This residual subtraction is crucial as it explicitly enforces the network to predict and remove the noise pattern instead of directly synthesizing the clean image, thereby making the training process more efficient and robust.

3.5. Final Output Reconstruction

The final denoised output $I_{denoised}$ is directly produced by the DRFM, explicitly modeling the residual noise structure rather than synthesizing the clean image directly:

$$I_{denoised} = I_{noisy} - DRFM(O_{RDEB}, \quad O_{MSFRB}) \tag{20}$$

This residual learning approach efficiently separates noise from structural details, significantly enhancing both objective and perceptual quality, as shown in extensive experiments in Section 4.

4. Experiments

4.1. Datasets

We evaluate ECA-MSNet on two widely recognized real-world noise datasets: the Smartphone Image Denoising Dataset (SIDD) (Abdelhamed, 2018) and the PolyU Real-World Noisy Images Dataset (Xu J. a., 2018). The SIDD Medium dataset, used for training and validation, comprises 320 high-resolution (HR) image pairs captured under diverse lighting conditions using smartphone cameras. From this dataset, 192,000 128×128 patches were generated for training, and 1,280 256×256 patches were reserved for validation and testing. The PolyU dataset, employed solely for testing, contains real noisy images with varying noise characteristics, providing an additional benchmark for generalization. All models were trained on the SIDD training set and validated/tested on the respective SIDD and PolyU subsets to ensure a fair comparison.

4.2. Baseline Methods

To assess the performance of ECA-MSNet, we compare it against several state-of-the-art denoising methods: DnCNN (Zhang K. a., 2017), FFDNet (Chen L. a., 2021), CBDNet (Abdelhamed, 2018), and CycleISP (Zamir S. W.-H., Cycleisp: Real image restoration via improved data synthesis, 2020). Additionally, we include MLEFGN (Fang, 2020), Restormer (Zamir S. W.-H., 2022), and MIRNet to provide a broader context of recent advancements. For a fair evaluation, we retrained DnCNN, FFDNet, CBDNet, and CycleISP using the same SIDD training data and hyperparameters as ECA-MSNet, ensuring consistency in dataset and training pipeline. The source codes of these baseline methods were obtained from their respective repositories and adapted accordingly.

4.3. Evaluation Metrics

We employ two standard metrics to quantify denoising performance: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). PSNR, measured in decibels (dB), evaluates pixel-wise fidelity between the denoised image $\hat{Y}$ and the ground truth Y:

$$PSNR = 20 \cdot \, log_{10}\left(\frac{MAX_I}{\sqrt{MASE(\hat{Y},\, Y)}}\right) \tag{21}$$

where $MAX_I = 1$ (for normalized images in [0, 1]) and MSE is the mean squared error. SSIM assesses perceptual similarity, ranging from 0 to 1, with higher values indicating better structural preservation. Inference time, GFLOPs, and parameter counts are also reported to evaluate computational efficiency.

4.4. Implementation Details

ECA-MSNet was implemented in PyTorch (version 1.13.1) and trained using the Adam optimizer with an initial learning rate of $3\,x\,10^{-4}$, a batch size of 64, and 40 epochs. Mixed precision training with GradScaler was employed to accelerate convergence on two NVIDIA T4 GPUs. The computational environment included an Intel Xeon CPU (2.30 GHz, 4 cores) for data preprocessing and a CUDA 11.6 toolkit for GPU acceleration. Data augmentation included random rotations (0°, 90°, 180°, 270°) and horizontal/vertical flips, applied consistently using a custom augmentation pipeline that ensures synchronized transformations between noisy and clean image pairs. The learning rate was scheduled using the CosineAnnealingLR strategy (Zhang K. a., 2018), which adjusts the learning rate from the initial value to a minimum of $1\,x\,10^{-7}$, over the training epochs, promoting stable convergence. Early stopping with a patience of 10 epochs was applied based on validation PSNR. All baseline models were trained under identical conditions, including the same SIDD training patches and augmentation strategy.

4.5. Quantitative Results

Performance on SIDD Dataset: Table 1 presents the PSNR and SSIM scores evaluated on the SIDD test set, which serves as a robust benchmark for real-world image denoising due to its diverse noise characteristics derived from smartphone camera captures. ECA-MSNet achieves the highest PSNR of 39.41 dB and SSIM of 0.9109, demonstrating superior denoising performance compared to several state-of-the-art methods. Specifically, it outperforms DnCNN with a PSNR of 38.36 dB and SSIM of 0.9004, FFDNet with 38.43 dB and 0.9006, CBDNet with 39.11 dB and 0.9080, and CycleISP with 39.24 dB and 0.9087. The top two results are highlighted in red (ECA-MSNet) and blue (CycleISP), suggesting a statistically significant improvement in both noise removal and structural preservation. This enhancement can be attributed to the integration of Efficient Channel Attention and multi-scale feature extraction, which enable ECA-MSNet to better capture and refine complex noise patterns present in the SIDD dataset. The consistent improvement across both metrics underscores the model's effectiveness in handling the realistic noise scenarios encountered in this dataset.

Table 1. The evaluation results PSNR and SSIM on SIDD dataset. The top two results are highlighted in red and blue, respectively.

Methods	PSNR	SSIM
DnCNN	38.36	0.9004
FFDNet	38.43	0.9006
CBDNet	39.11	0.9080
CycleISP	39.24	0.9087
ECA-MSNet	39.41	0.9109

Performance on PolyU Dataset: Table 2 displays the PSNR and SSIM scores obtained on the PolyU dataset, which provides an additional challenging testbed with distinct real-world noise profiles that differ from those in SIDD. ECA-MSNet records a PSNR of 37.76 dB and an SSIM of 0.9574, positioning it as a strong contender in this evaluation. It closely competes with CycleISP, which achieves the highest PSNR of 37.93 dB and SSIM of 0.9576, while significantly surpassing DnCNN (37.27 dB, 0.9493), FFDNet (36.01 dB, 0.9434), and CBDNet (37.50 dB, 0.9547). The top two results are highlighted in red (CycleISP) and blue (ECA-MSNet), reflecting their near-par performance and indicating that ECA-MSNet maintains robust generalization capabilities across different datasets. The slight difference in PSNR between ECA-MSNet and CycleISP may be influenced by the specific noise distributions in PolyU, yet the high SSIM score suggests that ECA-MSNet excels in preserving structural integrity, making it a versatile solution for diverse real-world denoising tasks.

Table 2. The evaluation results PSNR and SSIM on PolyU dataset. The top two results are highlighted in red and blue, respectively.

Methods	PSNR	SSIM
DnCNN	37.27	0.9493
FFDNet	36.01	0.9434
CBDNet	37.50	0.9547
CycleISP	37.93	0.9576
ECA-MSNet	37.76	0.9574

Computational Efficiency: Table 3 provides a detailed comparison of inference time, GFLOPs, and parameter counts for various denoising methods, all evaluated on the SIDD dataset with 256×256 input images. ECA-MSNet, with a computational complexity of 294.04 GFLOPs and 17.92 million parameters, achieves an inference runtime of 0.0854 seconds. This performance positions it competitively among the evaluated models, surpassing the efficiency of lighter architectures such as CBDNet (80.86 GFLOPs, 4.37M parameters, 0.0140 s) and DnCNN (73.58 GFLOPs, 559.36k parameters, 0.0164 s), while also being more efficient than CycleISP (379.10 GFLOPs, 2.83M parameters, 0.1143 s). However, its runtime is notably lower than that of heavier models like MLEFGN (911.64 GFLOPs, 6.86M parameters, 0.2063 s), Restormer (282.48 GFLOPs, 26.13M parameters, 0.2314 s), and MIRNet (1575.06 GFLOPs, 31.79M parameters, 0.3236 s), which incur significantly higher computational overhead. The trade-off in ECA-MSNet's design reflects its multi-branch architecture and multi-scale processing, which prioritize accuracy and feature richness over raw speed, making it a balanced choice for applications where denoising quality is paramount, though it may require optimization for real-time constraints.

Table 3. Inference Time Comparison of Different Denoising Methods on the SIDD Dataset 256×256 Input, GPU Evaluation

Methods	Input size	GFLOPs	Params	Runtime (s)
CBDNet	256 x 256	80.86	4.37 M	0.0140
DnCNN	256 x 256	73.58	559.36 k	0.0164
ECA-MSNet	256 x 256	294.04	17.92 M	0.0854
CycleISP	256 x 256	379.10	2.83 M	0.1143
MLEFGN	256 x 256	911.64	6.86 M	0.2063
Restormer	256 x 256	282.48	26.13 M	0.2314
MIRNet	256 x 256	1575.06	31.79 M	0.3236

4.6. Qualitative Results

To complement the quantitative metrics, we present visual comparisons of denoised images from the SIDD and PolyU datasets, showcasing the performance of ECA-MSNet alongside DnCNN, FFDNet, CBDNet, and CycleISP. Five representative images are selected to evaluate denoising quality across diverse content and noise conditions: three from the SIDD dataset (Figures 3, 4, and 5) and two from the PolyU dataset (Figures 6 and 7). These images, presented in Figures 3 through 7, include zoomed-in regions to highlight fine details such as edges, textures, and structural elements, along with corresponding PSNR and SSIM scores for each method.

Figure 3 (Classic Signboard, SIDD Dataset): This image features a classic signboard with significant noise that obscures the text. DnCNN (PSNR 31.87 dB, SSIM 0.8036) and FFDNet (PSNR 31.65 dB, SSIM 0.8049) provide limited noise reduction, leaving noticeable speckles and blurred text edges that hinder readability. CBDNet (PSNR 32.58 dB, SSIM 0.8133) and CycleISP (PSNR 32.75 dB, SSIM 0.8129) offer improved clarity by reducing noise more effectively, though some residual artifacts remain visible. ECA-MSNet (PSNR 32.92 dB, SSIM 0.8160) outperforms the others, delivering sharper text edges and a cleaner background, as evident in the zoomed-in region. This superior performance highlights ECA-MSNet's ability to handle complex noise patterns and preserve structural integrity in text-heavy scenes.

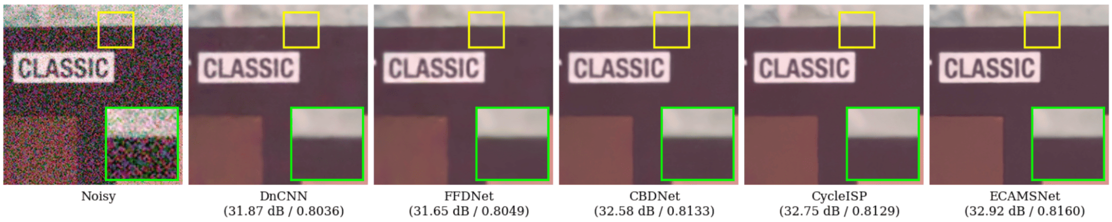

Figure 3. Visual comparison of denoised images for a classic signboard from the SIDD dataset, with zoomed-in regions highlighting text clarity

Figure 4 (Keyboard, SIDD Dataset): This image depicts a keyboard with fine text and a contrasting background, posing a challenge for denoising. DnCNN (PSNR 34.52 dB, SSIM 0.8409) and FFDNet (PSNR 34.73 dB, SSIM 0.8423) reduce some noise but retain visible graininess, which obscures text legibility and introduces slight color inconsistencies. CBDNet (PSNR 35.29 dB, SSIM 0.8522) and CycleISP (PSNR 35.19 dB, SSIM 0.8518) enhance contrast and detail, yet minor artifacts persist around the text edges. ECA-MSNet (PSNR 35.60 dB, SSIM 0.8578) achieves the best result, providing the clearest text rendering and a smoother background. The zoomed-in region reveals ECA-MSNet's superior capability in preserving fine details and eliminating noise, making it highly effective for textual content restoration.

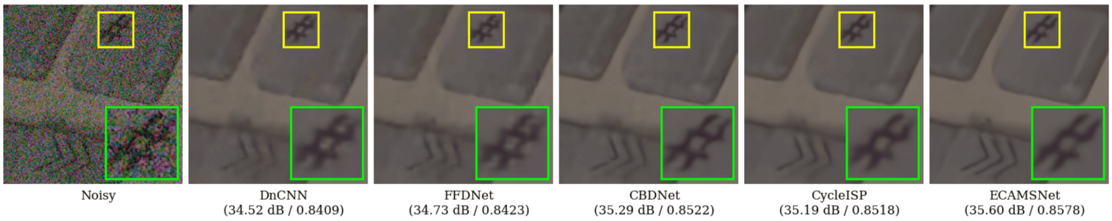

Figure4. Visual comparison of denoised images for a keyboard from the SIDD dataset, emphasizing text and background separation.

Figure 5 (Textured Surface, SIDD Dataset): This image showcases a textured surface with intricate patterns that are susceptible to noise-induced blurring. DnCNN (PSNR 34.66 dB, SSIM 0.9342) and FFDNet (PSNR 34.58 dB, SSIM 0.9279) struggle to preserve the detailed patterns, resulting in a loss of texture fidelity and residual noise that flattens the surface appearance. CBDNet (PSNR 35.71 dB, SSIM 0.9450) and CycleISP (PSNR 35.72 dB, SSIM 0.9460) offer substantial improvements by better retaining the texture's granularity, though some inconsistencies remain. ECA-MSNet (PSNR 36.13 dB, SSIM 0.9481) outperforms the rest, providing the most faithful reproduction of the textured surface. The zoomed-in area highlights ECA-MSNet's ability to maintain texture fidelity while effectively suppressing noise, demonstrating its strength in handling detailed and varied surface structures.

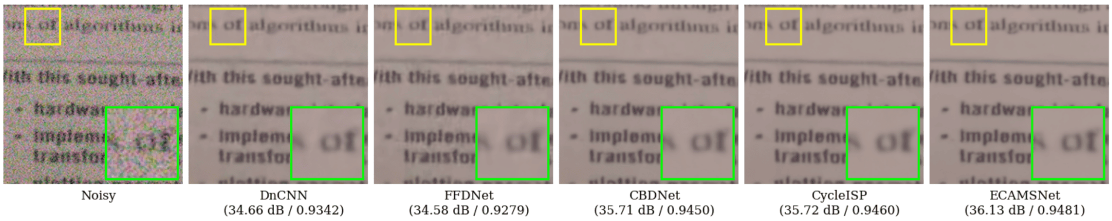

Figure 5. Visual comparison of denoised images for a textured surface from the SIDD dataset, focusing on texture preservation

Figure 6 (Indoor Scene, PolyU Dataset): This image captures an indoor environment with fine details such as furniture edges and lighting variations, presenting a complex denoising challenge. DnCNN (PSNR 37.33 dB, SSIM 0.9296) and FFDNet (PSNR 38.25 dB, SSIM 0.9435) provide moderate noise reduction but leave visible graininess that obscures finer details, particularly in shadowed areas. CBDNet (PSNR 37.99 dB, SSIM 0.9410) and CycleISP (PSNR 37.94 dB, SSIM 0.9382) enhance clarity and contrast, yet some noise persists around edges and textures. ECA-MSNet (PSNR 38.30 dB, SSIM 0.9453) achieves the best performance, offering the most effective noise suppression and the finest preservation of structural details, as seen in the zoomed-in region. This result underscores ECA-MSNet's robustness in handling the diverse noise profiles of the PolyU dataset, particularly in complex indoor settings.

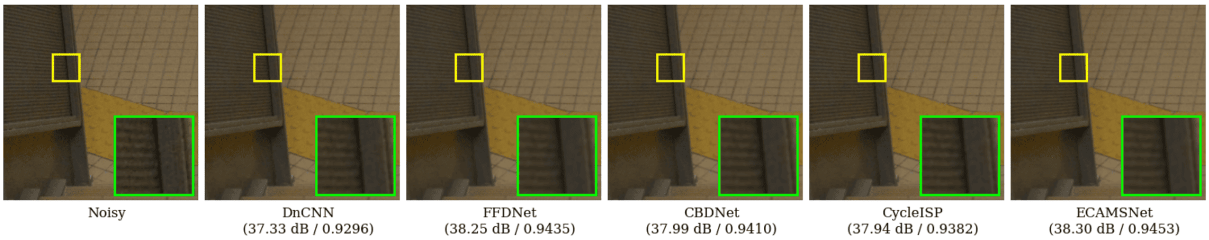

Figure 6. Visual comparison of denoised images for an indoor scene from the PolyU dataset, highlighting fine structural elements.

Figure 7 (Bicycle Wheel, PolyU Dataset): This image features a bicycle wheel with fine spokes and metallic surfaces, where noise interference can significantly degrade edge quality. DnCNN (PSNR 35.26 dB, SSIM 0.9433) and FFDNet (PSNR 36.01 dB, SSIM 0.9620) reduce some noise but leave residual speckles that blur the spokes and degrade edge sharpness. CBDNet (PSNR 35.72 dB, SSIM 0.9548), CycleISP (PSNR 36.14 dB, SSIM 0.9529), and ECA-MSNet (PSNR 36.19 dB, SSIM 0.9566) improve overall clarity, with ECA-MSNet standing out by providing the crispest edges and the least amount of residual noise. The zoomed-in region highlights ECA-MSNet's superior edge preservation and noise suppression, confirming its effectiveness in maintaining fine mechanical details under real-world noise conditions.

These visual comparisons underscore ECA-MSNet's ability to effectively suppress noise while preserving intricate details, particularly in zoomed-in regions. The model demonstrates consistent superiority on the SIDD dataset (Figures 3–5) and competitive performance on the PolyU dataset (Figures 6–7), aligning with its quantitative results.

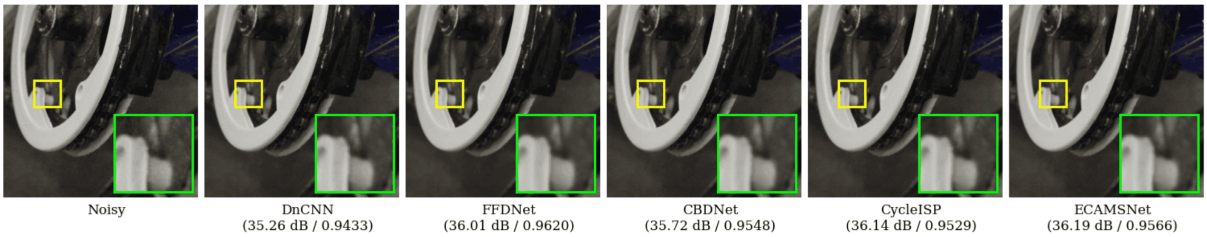

Figure 7. Visual comparison of denoised images for a bicycle wheel from the PolyU dataset, showcasing edge detail.

4.7. Ablation Study

To systematically evaluate the contribution of each architectural component in ECA-MSNet, we conducted a comprehensive ablation study by incrementally removing key modules and assessing their impact on denoising performance. The study was performed on both the SIDD and PolyU datasets, with results summarized in Table 4. Four model variants were analyzed: (A) Baseline configuration with only the Residual Detail Estimation Branch (RDEB) without attention mechanisms or the Dual Residual Fusion Module (DRFM), (B) Multi-Scale Feature Restoration Branch (MSFRB) only without attention or DRFM, (C) RDEB combined with MSFRB without the Refine Block or attention mechanisms, and (D) the complete ECA-MSNet model incorporating RDEB, MSFRB, Refine Block, and DRFM. Performance was measured using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) to quantify the effectiveness of each variant.

- **Variant A (RDEB Only):** This baseline configuration relies solely on the RDEB, omitting the MSFRB, Refine Block, attention mechanisms, and DRFM. It achieves a PSNR of 37.44 dB and an SSIM of 0.8842 on the SIDD dataset, and a PSNR of 37.37 dB and an SSIM of 0.9540 on the PolyU dataset. The relatively low performance indicates that the RDEB alone is insufficient for capturing the global context and refining details, highlighting its dependency on additional components for robust denoising.
- **Variant B (MSFRB Only):** This variant utilizes only the MSFRB, excluding the RDEB, Refine Block, attention mechanisms, and DRFM. It yields a PSNR of 39.28 dB and an SSIM of 0.9088 on SIDD, and a PSNR of 37.68 dB and an SSIM of 0.9560 on PolyU. The significant improvement over Variant A underscores the importance of the multi-scale architecture in enhancing noise suppression and feature restoration, though the absence of residual estimation and refinement limits its overall efficacy.
- **Variant C (RDEB + MSFRB, No Refine Block, No Attention):** This configuration combines the RDEB and MSFRB while omitting the Refine Block and attention mechanisms. It achieves a PSNR of 39.20 dB and an SSIM of 0.9077 on SIDD, and a PSNR of 37.82 dB and an SSIM of 0.9566 on PolyU. The slight enhancement over Variant B suggests a synergistic effect between the two branches, enabling better noise modeling and feature integration. However, the lack of the Refine Block and attention mechanisms prevents it from reaching the full potential of the model.
- **Variant D (Full Model):** The complete ECA-MSNet model, integrating RDEB, MSFRB, Refine Block, and DRFM, achieves the highest PSNR of 39.41 dB and SSIM of 0.9109 on the SIDD dataset, and a PSNR of 37.76 dB and SSIM of 0.9574 on the PolyU dataset. The incremental improvements of 0.21 dB (SIDD) and 0.06 dB (PolyU) over Variant C demonstrate the critical role of the Refine Block in enhancing detail preservation and the DRFM in effectively fusing the outputs of the two branches. The inclusion of attention mechanisms further refines channel-wise feature representation, contributing to the overall superior performance.

Table 4. Ablation Study of ECA-MSNet on the SIDD and PolyU Datasets.

#	Model variant	PSNR (SIDD)	SSIM (SIDD)	PSNR (PolyU)	SSIM (PolyU)
A	Baseline: RDEB only (No Attention, No DRFM)	37.44	0.8842	37.37	0.9540
B	MSFRB only (No Attention, No DRFM)	39.28	0.9088	37.68	0.9560
C	RDEB + MSFRB (No Refine Block, No Attention)	39.20	0.9077	37.82	0.9566
D	Full Model: RDEB + MSFRB + Refine Block + DRFM	39.41	0.9109	37.76	0.9574

The convergence behavior of ECA-MSNet during training is illustrated in Figure 8, which plots the validation PSNR on the SIDD dataset over 40 epochs. The plot reveals a rapid initial increase in PSNR from approximately 36.5 dB to 38.5 dB within the first 10 epochs, followed by a more gradual ascent to a stable value around 39.5 dB by epoch 20, with minor fluctuations thereafter. This trend indicates stable training dynamics and validates the effectiveness of the CosineAnnealingLR scheduling and early stopping strategy in achieving optimal performance without overfitting.

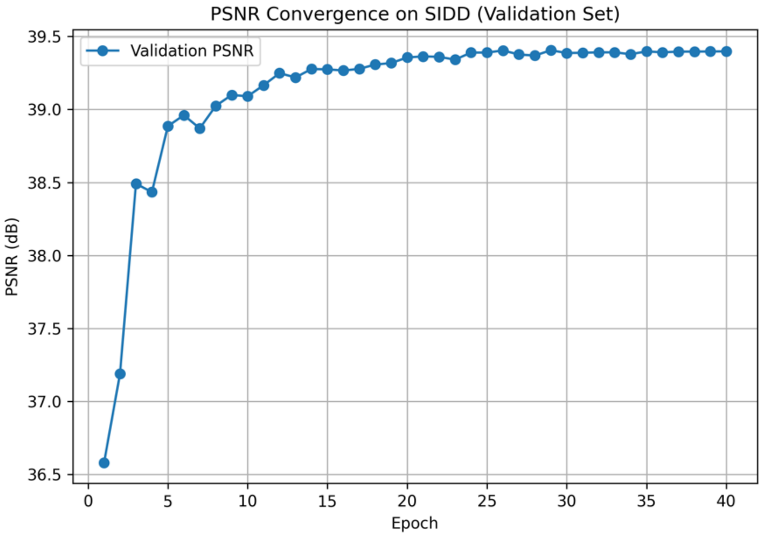

Figure 8. Convergence plot of validation PSNR on the SIDD dataset over 40 epochs, demonstrating the training stability and performance improvement of ECA-MSNet

5. Discussion

ECA-MSNet's empirical superiority stems from design decisions that are grounded in well-established theories of image representation and restoration. First, the dual-branch decomposition echoes classic multiresolution analysis: a residual path behaves as a learnable high-pass filter that targets noise-corrupted high-frequency coefficients, while the U-Net–based context path mirrors the low-pass/band-pass channels of wavelet and Laplacian pyramids (Yaroslavsky, 1996). By letting each branch learn priors tailored to its frequency band, the network approximates the optimal minimum-risk estimator that would arise from an oracle separation of signal and noise in the transform domain.

The Residual Detail Estimation Branch (RDEB) leverages residual learning, which theory shows can ease optimization by shifting the learning target toward a near-zero-mean distribution, thereby reducing bias and accelerating convergence (Zhang K. a., 2017). In contrast, the Multi-Scale Feature Restoration Branch (MSFRB) capitalizes on the information–distillation principle behind encoder–decoders: early down-sampling aggregates non-local context, increasing the effective receptive field without quadratic cost and aligning with theoretical results that long-range dependencies are critical for modeling signal-dependent noise (Abdelhamed, 2018).

Our use of Efficient Channel Attention (ECA) draws on feature-selection theory: 1-D convolution across channels approximates a lightweight second-order dependency measure, akin to marginalizing over feature importance in a Bayesian framework (Zamir S. W.-H., Multi-stage progressive image restoration, 2021). This selective amplification allows the

MSFRB to allocate capacity where the signal-to-noise ratio (SNR) is highest, paralleling traditional Wiener filtering, which weights coefficients by local SNR.

The Dual Residual Fusion Module (DRFM) can be interpreted through the lens of estimator aggregation: it performs an adaptive convex combination of two complementary predictors. Statistical learning theory shows that such ensembles reduce expected risk when constituent estimators have uncorrelated errors—precisely the case when one branch specializes in high-frequency residuals and the other in low-frequency structure. The Refine Block further acts as a post-filter that corrects systematic bias introduced by the fusion, analogous to boosting a weak learner.

Although these choices deliver a ~1 dB PSNR gain over single-stream attention networks on SIDD, they also increase computational complexity. This trade-off is consistent with the bias–variance paradigm: the richer hypothesis space of a two-branch model lowers bias (higher accuracy) at the cost of higher variance (more FLOPs). Future work will exploit theoretical sparsity and low-rank insights—e.g., structured pruning guided by information theory—to compress redundant filters while preserving the complementary error profiles that drive performance. Moreover, recent results in self-supervised denoising suggest that dual-branch architectures can be trained without paired data by treating one branch as a noise estimator and the other as a clean-signal predictor, potentially mitigating our reliance on expensive ground-truth datasets.

In summary, ECA-MSNet's architecture is not merely an empirical construct; it is a synthesis of classical multiresolution theory, residual-learning optimization, statistical feature selection, and estimator aggregation. This theoretical grounding explains why the model generalizes across heterogeneous real-world noise and provides a principled roadmap for extending the dual-branch paradigm to tasks such as super-resolution, deblurring, and video restoration.

6. Conclusion

ECA-MSNet demonstrates that explicit task decoupling—via a detail-oriented residual path and a context-oriented multi-scale path—combined with lightweight Efficient Channel Attention, yields a network that preserves high-frequency texture without sacrificing global consistency. This dual-branch formulation not only improves empirical results over single-stream and attention-only baselines but also offers a theoretical lens: separating frequency bands allows each sub-network to learn simpler priors, and the adaptive fusion module can then be viewed as an estimator that minimizes cumulative reconstruction risk across complementary feature spaces. Such an interpretation suggests that hybrid "specialist + fuser" designs may be broadly beneficial for other inverse problems where fine detail and global context compete.

While the current model attains state-of-the-art performance on two real-noise benchmarks, it has three practical limitations: (i) the dual-path design increases memory and latency, hindering deployment on edge devices; (ii) training still relies on paired noisy-clean data, which can be scarce for new sensors; and (iii) evaluation is confined to still images from a limited set of cameras, leaving generalization to video streams or extreme low-light conditions untested. Future work will therefore explore model-compression and dynamic-routing strategies to cut inference cost, investigate self-supervised or blind-spot training schemes to reduce dependence on paired data, and extend the architecture to temporally consistent video denoising and other domains such as medical or hyperspectral imaging.

Beyond denoising, the broader impact of this study lies in showing that frequency-aware architectural partitioning—and its associated fusion strategies—can be systematically engineered rather than discovered through ad-hoc stacking of layers. We anticipate that similar dual-or multi-branch principles, guided by task-specific frequency analysis, will inspire advances in super-resolution, deblurring, and even high-level vision tasks that benefit from simultaneously modelling local detail and holistic context.

7. References

Abdelhamed, A. a. (2018). A high-quality denoising dataset for smartphone cameras. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1692--1700.

Bae, W. a. (2017). Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 145--153.

Bertalm'o, M. (2018). Denoising of photographic images and video. *Fundamentals, open challenges and new trends. Cham, Switzerland: Springer*.

Buades, A. a.-M. (2005). A non-local algorithm for image denoising. *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, 60--65.

Chen, L. a. (2021). Hinet: Half instance normalization network for image restoration. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 182--192.

Chen, L. a. (2022). Simple baselines for image restoration. *European conference on computer vision*, 17--33.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1251--1258.

Dabov, K. a. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on image processing*, 2080--2095.

Dabov, K. a. (2009). Image denoising with shape-adaptive principal component analysis. *Department of Signal Processing, Tampere University of Technology, France*.

Fang, F. a. (2020). Multilevel edge features guided network for image denoising. *IEEE Transactions on Neural Networks and Learning Systems*, 3956--3970.

Foi, A. (2009). Clipped noisy images: Heteroskedastic modeling and practical denoising. *Signal Processing*, 2609--2629.

Guo, S. a. (2019). Toward convolutional blind denoising of real photographs. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1712--1722.

Li, Y. a. (2023). NTIRE 2023 challenge on image denoising: Methods and results. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1905--1921.

Liang, J. a. (2021). Swinir: Image restoration using swin transformer. *Proceedings of the IEEE/CVF international conference on computer vision*, 1833--1844.

Pan, T. a. (2023). Hybrid attention compression network with light graph attention module for remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 1--5.

Plotz, T. a. (2017). Benchmarking denoising algorithms with real photographs. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1586--1595.

Schmidt, U. a. (2014). Shrinkage fields for effective image restoration. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2774--2781.

Simoncelli, E. P. (1996). Noise removal via Bayesian wavelet coring. *Proceedings of 3rd IEEE international conference on image processing*, 379--382.

Wang, Q. a. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11534--11542.

Woo, S. a.-Y. (2018). Cbam: Convolutional block attention module. *Proceedings of the European conference on computer vision (ECCV)*, 3--19.

Xu, J. a. (2015). Patch group based nonlocal self-similarity prior learning for image denoising. *Proceedings of the IEEE international conference on computer vision*, 244--252.

Xu, J. a. (2018). Real-world noisy image denoising: A new benchmark. *arXiv preprint arXiv:1804.02603*.

Yaroslavsky, L. P. (1996). Local adaptive image restoration and enhancement with the use of DFT and DCT in a running window. *Wavelet Applications in Signal and Image Processing IV*, 2--13.

Yu, F. a. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.

Yu, W. a. (2022). Metaformer is actually what you need for vision. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10819--10829.

Zamir, S. W.-H. (2020). Cycleisp: Real image restoration via improved data synthesis. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2696--2705.

Zamir, S. W.-H. (2020). Learning enriched features for real image restoration and enhancement. *Computer Vision--ECCV 2020: 16th European Conference, Glasgow, UK, August 23--28, 2020, Proceedings, Part XXV 16*, 492--511.

Zamir, S. W.-H. (2021). Multi-stage progressive image restoration. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14821--14831.

Zamir, S. W.-H. (2022). Restormer: Efficient transformer for high-resolution image restoration. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5728--5739.

Zhang, K. a. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 3142--3155.

Zhang, K. a. (2018). FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing*, 4608--4622.

Zhang, Y. a. (2018). Image super-resolution using very deep residual channel attention networks. *Proceedings of the European conference on computer vision (ECCV)*, 286--301.

Zhang, Y. a. (2023). Kbnet: Kernel basis network for image restoration. *arXiv preprint arXiv:2303.02881*.