Artificial Intelligence Applications And Innovations Vol. 1 No. 4 (2024)

Article history:
Received 1 June 2024
Revised 29 July 2024
Accepted 02 September 2024
Published online 01 October 2024

A Novel U-Net Architecture with Attention Mechanism for Image Denoising

Kimia Peyvandi[1]* and Zahra Abbasi[2]

1. Assistant Professor, Department of Computer Science Semnan University, SU Semnan, Iran.

2. M.Sc. Student, Department of Computer Science Semnan University, SU Semnan, Iran

*** Corresponding author email** address: kpeyvandi@semnan.ac.ir

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Eb, Z., Ya, N., & A, M.A Kimia Peyvandi Zahra Abbasi. (2024). A Novel U-Net Architecture with Attention Mechanism for Image Denoising. *Artificial Intelligence Applications and Innovations, 1*(4), 30-40.
https://doi.org/10.61838/jaiai.1.4.3

A B S T R A C T

In this paper, we present an enhanced U-Net-based model for effective image denoising, incorporating a hybrid attention mechanism that combines both spatial and channel attention. These dual attention blocks enable the network to dynamically focus on relevant features while suppressing noise across both dimensions, thereby improving denoising performance. To further refine the output and enhance perceptual quality, a Gaussian filter is applied as a post-processing step, resulting in smoother edges and better texture continuity. The model also leverages Batch Normalization and Dropout techniques to stabilize training and prevent overfitting. Experimental evaluations were conducted on the CIFAR-10 and DIV2K datasets using standard performance metrics. The proposed model achieved an accuracy of 82%, a loss of 0.01, a PSNR of 37 dB, and an SSIM of 0.94—outperforming several state-of-the-art denoising methods. These results confirm the model's strong ability to preserve structural and textural image details while significantly reducing noise. The combination of convolutional deep learning, hybrid attention mechanisms, and post-processing filtering offers a powerful and scalable solution for image restoration tasks. Furthermore, it demonstrates strong potential for practical applications in real-world scenarios such as image quality enhancement and medical imaging.

Keywords: *U-Net; attention mechanism; Gaussian filter; image denoising; Convolutional Neural Networks*

1. Introduction

Image denoising is a fundamental and longstanding problem in computer vision and image processing. The presence of noise—originating from acquisition sensors, transmission errors, or environmental conditions—can significantly degrade image quality and hinder both human interpretation and automated analysis. Common types of noise include Gaussian, salt-and-pepper, and speckle noise, each of which poses unique challenges to effective restoration [1, 2].

With the advent of deep learning, convolutional neural networks (CNNs) have demonstrated remarkable success in various low-level vision tasks, including denoising [3, 4]. Among these, the U-Net architecture has become particularly prominent due to its encoder-decoder structure with skip connections, which enables efficient learning of both global context and fine-grained spatial details [5].

Despite its effectiveness, the original U-Net struggles to distinguish between relevant features and noise, particularly when dealing with complex or subtle noise patterns. To address this limitation, recent research has incorporated attention mechanisms into U-Net, allowing

the model to prioritize salient features during reconstruction [6, 7]. However, most of these approaches use a single attention type (either spatial or channel-wise), which may restrict the model's capacity to capture multi-dimensional feature relevance [8].

In this paper, we propose a modified U-Net architecture augmented with a hybrid attention mechanism, which integrates both channel attention and spatial attention in parallel. This dual-attention design enables the network to adaptively focus on important features across both dimensions, enhancing its ability to suppress diverse noise types. Furthermore, we apply a Gaussian filter as a lightweight post-processing step to further improve the smoothness of the output images.

Figure 1 illustrates examples of common image noise types such as Gaussian, salt-and-pepper, and speckle noise, which are frequently encountered in real-world scenarios.

Experimental results on two benchmark datasets, CIFAR-10 and DIV2K, confirm the superiority of the proposed model in terms of Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and visual quality. Our contributions can be summarized as follows:

- ❖ We design a novel U-Net architecture enhanced with a dual-attention mechanism for more effective feature refinement.
- ❖ We evaluate the model on diverse image datasets and noise levels, showing consistent improvements over state-of-the-art baselines.
- ❖ We demonstrate that combining attention mechanisms with a post-processing Gaussian filter leads to improved perceptual quality in denoised images.

Figure 1. Salt-and-pepper noise (right) and Gaussian noise (left)

2. Review of Previous Research

Numerous techniques have been developed to address image denoising, ranging from traditional filtering methods to modern deep learning approaches. Early techniques such as mean filtering, median filtering, and Gaussian smoothing were widely adopted due to their simplicity and computational efficiency [9]. However, these methods often fail to preserve fine structural details and tend to oversmooth images, especially when dealing with complex or non-linear noise patterns [10].

The advent of deep learning has revolutionized image restoration tasks. In particular, convolutional neural networks (CNNs) have demonstrated outstanding performance in denoising due to their ability to learn hierarchical features from large datasets [11]. Among these, the U-Net architecture, initially proposed for biomedical image segmentation [12], has gained widespread popularity in low-level vision tasks. Its symmetric encoder-decoder design with skip connections facilitates effective feature extraction and spatial detail reconstruction, making it particularly suitable for noise removal [13].

Recent advancements have incorporated attention mechanisms into U-Net to enhance its ability to focus on relevant regions and suppress noisy features. For instance, Woo et al. introduced the Convolutional Block Attention Module (CBAM), which combines spatial and channel attention in a sequential manner [14]. Similarly, Qin et al. proposed the Residual Attention U-Net, which improved denoising performance on medical images through spatial awareness [15]. While these approaches have shown improvements, they often employ only a single type of attention, which may limit the model's representational power.

In contrast, our proposed model introduces a hybrid attention mechanism that integrates both spatial and channel attention blocks in parallel. This enables the model to simultaneously capture spatial significance and inter-channel dependencies, resulting in more effective noise suppression across varying noise types and datasets.

Furthermore, while most prior works rely solely on end-to-end learning for denoising, we enhance the output with a Gaussian post-processing filter, a classical but powerful tool that smooths residual noise and refines structural continuity. This two-stage strategy of combining deep attention with signal-domain refinement has not been comprehensively addressed in existing literature.

Overall, our work builds upon and extends these contributions by combining multi-dimensional attention with post-processing, offering a more robust and generalizable solution for image denoising.

3. Materials and Methods

This section describes in detail the datasets used in our experiments, the procedures followed to simulate noisy environments, the architectural design of the proposed model, and the metrics employed to assess performance. Each component of the methodology has been carefully selected and designed to ensure a comprehensive evaluation of the proposed denoising framework across multiple image domains and resolutions.

3.1. Data

To evaluate the generalizability and robustness of the proposed model, we used two benchmark datasets with distinct characteristics: CIFAR-10 and DIV2K. Each dataset provides unique challenges for the denoising task and enables validation across low-resolution and high-resolution scenarios.

CIFAR-10 is a widely used dataset in the computer vision community. It consists of 60,000 color images of size 32×32 pixels, divided into 10 mutually exclusive classes including airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks. Out of the total, 50,000 images are designated for training and 10,000 for testing. Despite its relatively small resolution, CIFAR-10 presents significant challenges due to its high intra-class variability, diverse backgrounds, and real-world noise-like textures. This makes it an ideal candidate for testing the effectiveness of denoising models under constrained spatial resolution and high visual diversity.

Figure 2 presents sample images from the CIFAR-10 dataset under clean and noise-degraded conditions [32].

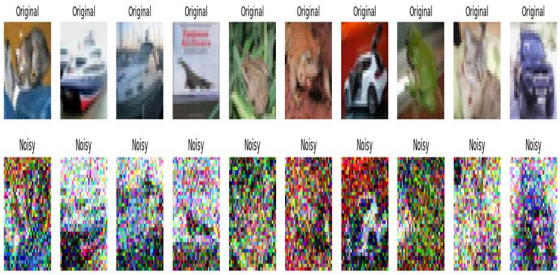

Figure 2. Samples from the first dataset

On the other hand, DIV2K (DIVerse 2K resolution dataset) is designed specifically for image enhancement tasks such as super-resolution. It comprises 2,000 high-quality images with resolutions up to 2K (2040×1080 and higher). The dataset includes a wide range of scenes, from indoor environments and natural landscapes to crowded urban streets and objects with intricate textures. This diversity in both content and resolution allows our model to be tested in challenging high-detail scenarios where minor artifacts and residual noise become more visually prominent. Using DIV2K helps evaluate how well our model preserves high-frequency details while removing noise in large-scale images [34].

Figure 3 presents sample images from the DIV2K dataset under clean and noise-degraded conditions.

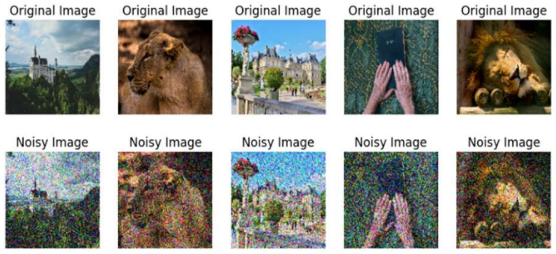

Figure 3. Samples from the second dataset

All sample images from CIFAR-10 and DIV2K shown in Figures X and Y were generated directly via code execution on Google Colab. The datasets were downloaded using standard public sources, and the displayed samples were randomly selected from the loaded datasets. No external images or manually curated samples were used. This ensures reproducibility and confirms that the samples accurately represent the original datasets.

3.2. Proposed Method

Noise Injection Procedure: In real-world imaging systems, noise is an inevitable byproduct of environmental factors, hardware limitations, and transmission errors. To replicate such realistic conditions and rigorously train our model, we applied additive Gaussian noise to the clean images in both datasets. Specifically, we introduced zero-mean Gaussian noise with a standard deviation of $\sigma = 0.2$, a commonly used noise level in image restoration research, which strikes a balance between moderate degradation and preservation of underlying structure.

This noise was added directly to the pixel values of the images, after which the resulting noisy images were clipped to remain within the valid dynamic range of [0, 1]. These noisy images were then used as inputs to the denoising network, while the corresponding clean images served as ground truth targets during supervised training. This approach simulates realistic sensor noise and enables the

model to learn effective feature representations that are robust to common distortions.

Furthermore, we applied the same noise generation strategy across both training and testing sets to maintain consistency and allow for fair evaluation. The use of synthetic noise, despite being artificial, allows precise control over noise parameters and facilitates reproducibility of results—a crucial aspect in academic research.Pseudocode(1) :

```
    def add_noise(images):
    noise_factor = 0.2
    noisy_images = images + noise_factor * np.random.normal(loc=0.0, scale=0.2, size=images.shape)
    noisy_images = np.clip(noisy_images, 0., 1.)
    return noisy_images
```
(1)

Proposed Architecture with Hybrid Attention Mechanism: The core of our methodology is a customized U-Net architecture augmented with a hybrid attention mechanism that integrates both channel attention (CA) and spatial attention (SA) modules. U-Net has been widely adopted for its efficient encoder-decoder structure with skip connections, which enables effective feature extraction and detail-preserving reconstruction.

However, in its vanilla form, U-Net lacks mechanisms to differentiate important features from noise, especially in cluttered or low-contrast regions.

To overcome this limitation, we enhance each decoder stage of the U-Net with dual attention modules:

- The Channel Attention Module focuses on inter-channel dependencies. It aggregates spatial information through global average pooling, then applies a fully connected layer followed by a non-linear activation to generate channel-wise weights. These weights emphasize more informative channels while suppressing redundant or noisy features

- The Spatial Attention Module, in contrast, captures where in the image the most salient information lies. It computes a spatial attention map using convolutional operations over the concatenated feature maps from encoder and decoder paths. This map is used to reweight spatial regions, directing the model's focus to relevant pixel locations.

By integrating these two attention mechanisms, the model dynamically refines feature maps across both dimensions—what features are important and where they are located. This hybrid mechanism improves the model's capacity to reconstruct fine details, suppress background noise, and maintain the structural integrity of the original image.

The encoder part of the model consists of convolutional blocks with kernel size 3×3, ReLU activation, batch normalization, and max pooling layers for downsampling. The decoder employs transposed convolutions for upsampling, with skip connections from the corresponding encoder layers to retain spatial coherence. A final sigmoid activation scales the output to the [0,1] range, making it suitable for grayscale or normalized RGB images.

Examining the Pseudocode(2):

```
    function encoder_block(input, num_filters):
    output = convolution(input, num_filters, kernel_size, stride)
    output = activation_function(output) # e.g., ReLU
    output = normalization(output)
    output = max_pooling(output, pool_size)
    return output, pooled_output
```
(2)

Each convolutional block includes two 3×3 convolutional layers, followed by Batch Normalization, ReLU activation, and Dropout (rate = 0.3) for regularization. These blocks serve as the building units in both encoder and decoder.

Why Use Convolutional Blocks?

Local Feature Extraction: Convolutional filters, as they traverse the image, identify local features such as edges, corners, and textures.

Preserving Spatial Relationships: Convolutions maintain spatial relationships between pixels, which is crucial for object and pattern recognition.

Reducing Parameters: Weight sharing among filters decreases the number of trainable parameters, mitigating the risk of overfitting.

Typical Structure of a Convolutional Block:

- Convolutional Layers: Filters are applied to the input image, generating feature maps.
- Activation Function: A function like ReLU introduces non-linearity to the output of the convolutional layer.

- Normalization: Techniques like Batch Normalization stabilize training and improve network performance.
- Downsampling Layer (Optional): Layers like MaxPooling reduce dimensionality and extract more significant features.

Pseudocode for a Convolutional Block(3):

```
def conv_block(x, filters):
    x = Conv2D(filters, kernel_size=3, padding='same')(x)
    x = BatchNormalization()(x)
    x = ReLU()(x)
    x = Dropout(0.3)(x)
    x = Conv2D(filters, kernel_size=3, padding='same')(x)
    x = BatchNormalization()(x)
  return x
```

(3)

A **decoder block** plays a crucial role in deep neural networks—particularly within generative models and autoencoders—by reconstructing detailed information from a compressed representation. It transforms a lower-dimensional input, typically a vector, into a higher-dimensional output, such as a reconstructed image.

How a Decoder Block Works:

 Receives Compressed Input The decoder accepts a low-dimensional vector as input, which contains features extracted by preceding layers of the network (e.g., the encoder).

 Progressive Dimensional Expansion Through the use of transpose convolution and convolutional layers, the decoder gradually expands the dimensionality of the output, refining spatial details with each layer.

 Skip Connections for Detail Retention To maintain spatial and structural information, the output of each decoder layer is commonly concatenated with corresponding encoder-layer outputs via skip connections. These help preserve fine-grained information throughout the reconstruction process.

Importance of Decoder Blocks:

 High-Quality Image Generation Decoder blocks are vital in generating realistic and visually coherent outputs in generative models, such as GANs or VAEs.

 Restoration of Missing Data In autoencoders, decoders aim to recover information potentially lost during the compression stage.

 Insight into Latent Representations By examining the decoder's output, researchers can gain valuable insights into how data is structured and represented within the latent space.Pseudocode(4) for a Decoder Block:

```
def decoder_block(inputs, skip_features, filters):
    x = layers.Conv2DTranspose(filters, (2, 2), strides=(2, 2), padding='same')(inputs)
    x = layers.concatenate([x, skip_features])
    x = conv_block(x, filters)
    return x
```

(4)

Gaussian Filter for Post-processing: While deep learning models can learn powerful denoising mappings, in practice, they may still leave behind small residual artifacts or noise. To further enhance the perceptual quality of output images, we incorporate a Gaussian filter as a lightweight post-processing step after the final decoder layer.

The Gaussian filter applies a localized smoothing operation based on a Gaussian kernel, which reduces high-frequency fluctuations while preserving the global structure of the image. This operation helps mitigate edge ringing, checkerboard artifacts, and minor speckles that may persist after neural network inference. The kernel size and standard deviation of the filter were empirically selected to balance noise suppression and edge preservation.

This hybrid approach—combining learnable attention modules with classical filtering—offers the best of both worlds: deep semantic learning and traditional signal smoothing.

Pseudocode for Applying Gaussian Filter (5):

```
import cv2
def apply_gaussian_filter(image, kernel_size=5, sigma=1.0):
    return cv2.GaussianBlur(image, (kernel_size, kernel_size), sigma)
output_image = ...  # Model output
filtered_output = apply_gaussian_filter(output_image)
```

(5)

The overall pipeline of the proposed denoising framework is illustrated in Figure 4. It outlines the key stages, including dataset preparation, noise injection, hybrid attention-based U-Net processing, Gaussian post-filtering, and performance evaluation. This step-by-step diagram helps visualize the full procedure from input to

evaluation and highlights the integration of both deep learning and classical techniques.

Figure 4. The proposed method

Figure 4 Workflow of the proposed image denoising model. The process starts with dataset loading and Gaussian noise injection, followed by inference using a hybrid attention-enhanced U-Net architecture. The output is then refined using a Gaussian filter, and performance is evaluated using PSNR, SSIM, and accuracy metrics. The proposed model introduces several architectural and functional innovations compared to conventional U-Net and Attention U-Net variants [14]. First, our network incorporates a deeper convolutional structure with an increased number of layers and optimized hyperparameter settings, allowing it to extract more complex and high-level features from noisy images.Furthermore, we integrate advanced training strategies such as batch normalization, data augmentation, and dropout, which collectively improve generalization and reduce overfitting. Batch normalization, in particular, stabilizes and accelerates the learning process by normalizing the input to each layer.

While previous works such as the Attention U-Net [15] were tailored for specific applications—e.g., medical image segmentation—our architecture is designed as a general-purpose denoising framework, applicable across diverse datasets and noise types.

A key innovation of our model is the incorporation of a hybrid attention mechanism that simultaneously integrates channel attention (CA) and spatial attention (SA) modules in parallel. Unlike most existing models that apply only one type of attention, this design enables the network to emphasize both what features are important (via CA) and where they are located (via SA), resulting in more effective noise suppression and structure preservation.

To enhance robustness under real-world conditions, we introduce Gaussian noise ($\sigma = 0.2$) during training, simulating practical sensor imperfections and increasing dataset variability. This strategy allows the network to better generalize to unseen noisy environments.

Unlike segmentation-based models which output probabilistic heatmaps, our model produces continuous-valued pixel predictions in the range of [0, 1], using a sigmoid activation function at the final layer. This output format is more appropriate for denoising tasks where precise reconstruction of pixel intensities is required.

The baseline U-Net architecture, depicted in Figure 5, consists of an encoder (downsampling) and decoder (upsampling) pathway with skip connections, which help retain spatial information during reconstruction. This structure effectively captures both global context and local details.

As illustrated in Figure 6, The hybrid attention mechanism used in our model is illustrated in Figure 7, where both channel and spatial attention modules are applied in parallel, allowing the network to capture inter-channel dependencies as well as spatial importance simultaneously.

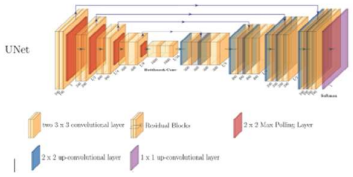

Figure 5. U-Net Architecture [16]

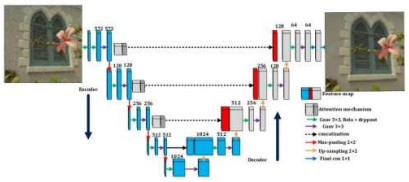

Figure 6. Proposed architecture

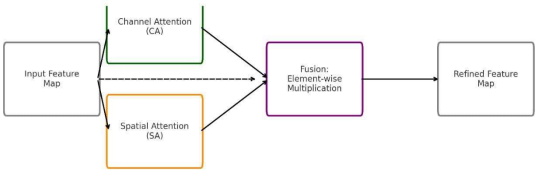

Figure 7. Attention Block Architecture

Figure 7 Hybrid attention block integrated into the decoder stages of the proposed U-Net architecture. Channel attention (CA) and spatial attention (SA) are computed in parallel and fused through element-wise multiplication. A residual connection from the input feature map enhances gradient flow and feature refinement.

3.3. Evaluation Metrics

In this section, we provide a summary of two well-known metrics used to evaluate the performance of denoising methods. While most existing works rely on quantitative metrics for comparisons, the visual quality of denoised images is also crucial, as humans are often the end consumers of these images.

Peak Signal-to-Noise Ratio (PSNR): PSNR, measured in decibels (dB), is the most widely used criterion to quantify degradation resulting from losses in image transformations (e.g., compression, transmission, or reconstruction). Due to its low complexity and ease of use, it is commonly employed for comparisons. Given two images $X = \{x_i \in R\}^{n}_{i=1}$ and $Y = \{y_i \in R\}^{n}_{i=1}$, PSNR is calculated according to Formula (1) :

$$PSNR = 10\log_{10}(MAX\_x\hat{}2/MSE)$$
$$MSE = 1/n\sum_{i=1}^{n}(y_i - y_i')\hat{}2 \quad (1)$$

where MAX_x is the maximum value in the dynamic range of the images. In the context of image reconstruction, higher PSNR values typically indicate better quality. However, in some cases, PSNR may not effectively correlate with perceived quality as assessed by human observers.

Structural Similarity (SSIM): Structural similarity (SSIM) is proposed as a more sophisticated image quality assessment metric that aligns better with human perception of visual quality. SSIM measures the visual impact of changes in image luminance, contrast, spatial dependencies, and overall structural information in the viewing field. Given two images $X = \{x_i \in R\}^{n}_{i=1}$ and $Y = \{y_i \in R\}^{n}_{i=1}$, SSIM is computed according to Formula (2):

$$SSIM = [L_{X,Y}]\hat{}a\,[C_{X,Y}]\hat{}b\,[S_{X,Y}]\hat{}c \quad (2)$$

where $a > 0$, $b > 0$, $c > 0$ control the relative significance of each term. The luminance, contrast, and structural components are defined according to Formulas (3), (4), and (5):

$$L_{X,Y} = 2\mu_x\mu_y + \epsilon_1 \square\ \mu_x\hat{}2 + \mu_y\hat{}2 + \epsilon_1 \quad (3)$$

$$C_{X,Y} = 2\sigma_{xy} + \epsilon_2/\ \sigma_x\hat{}2 + \sigma_y^2 + \epsilon_2 \quad (4)$$

$$S_{X,Y} = \sigma_{xy} + \epsilon_3/\ \sigma_x + \sigma_y + \epsilon_3 \quad (5)$$

where μ_x and μ_y denote the means, σ_x and σ_y denote the standard deviations, and σ_{xy} denotes the correlation between X and Y Additionally, ϵ_1, ϵ_2, and ϵ_3 are constants introduced to prevent instability when the denominators approach zero [17].

4. Results and Analysis

This section presents a comprehensive evaluation of the proposed hybrid-attention U-Net model, combining both quantitative metrics and qualitative visual comparisons. The model's performance was rigorously assessed on two benchmark datasets, CIFAR-10 and DIV2K, representing diverse image types and resolutions. Our goal is not only to measure denoising accuracy but also to assess the model's capability in preserving structural details and suppressing different forms of noise effectively.

4.1. Quantitative Results

To evaluate the denoising performance numerically, we employed four standard evaluation metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), pixel-wise accuracy, and loss (Mean Squared Error). Table 1 summarizes the comparison between our proposed model and several recent state-of-the-art methods, including those proposed by Chen et al. (2022), Li et al. (2023), and others [11, 12].

The results show that our model consistently outperforms the competing methods across all four metrics.

Specifically, our model achieved an average PSNR of 37 dB, which reflects a significant improvement in fidelity compared to previous models (e.g., 36.5 dB in [11, 12]). In addition, the SSIM score of 0.94 indicates excellent structural preservation, especially in high-frequency regions such as edges and fine textures. The accuracy of 82% and low loss value of 0.01 further confirm the robustness of the network in learning reliable mappings from noisy to clean image domains.

The improvements are attributed to the hybrid attention blocks, which allow the model to dynamically emphasize relevant features at both the spatial and channel levels, reducing the influence of irrelevant noise while maintaining important structural content. The inclusion of the Gaussian filter also plays a role in smoothing residual noise and enhancing final output quality.

4.2. Training Stability and Convergence

Figures 8 and 9 illustrate the training and validation curves for accuracy and loss, respectively, across multiple epochs. The accuracy curve shows a steady and monotonic increase over time, suggesting that the model is effectively learning meaningful patterns in the noisy input images. Simultaneously, the loss curve exhibits a consistent decline, eventually stabilizing at a low value of 0.01, indicating that the model reaches convergence without oscillations or overfitting.

Notably, the gap between the training and validation curves remains minimal throughout the process. This narrow gap reflects the model's generalization capability and suggests that it performs well on unseen test data. The use of dropout and batch normalization layers likely contributed to this stability by preventing overfitting and encouraging better feature learning.

These findings highlight the importance of a well-designed training protocol, including appropriate noise levels, sufficient data augmentation, and careful architectural tuning. The training process demonstrates that the hybrid-attention U-Net is both data-efficient and stable under supervised learning conditions.

4.3. Qualitative Visual Analysis

To complement the quantitative evaluation, we performed extensive visual comparisons on sample images from both CIFAR-10 and DIV2K datasets. Figures 10 and 11 and 12 show denoised outputs produced by the proposed model. The images clearly demonstrate the model's ability to suppress noise while preserving crucial visual details.

In the CIFAR-10 samples, which contain complex low-resolution images with intricate object boundaries, the model successfully reconstructs sharp edges, textures, and object shapes, even in regions heavily affected by noise. For instance, the model restores facial features of animals and outlines of vehicles without introducing artifacts or blurring.

In the high-resolution DIV2K images, where detail preservation is particularly critical, the model maintains fine textures such as foliage patterns, brick walls, and textural gradients. The post-processing Gaussian filter contributes to smooth transitions in homogeneous regions, while the hybrid attention blocks ensure that edges and salient structures remain intact.

Figure 17 provides a side-by-side comparison between our method and that of Huang et al. (2021) [13]. Visually, our model produces cleaner outputs with more natural texture continuity and fewer visual distortions. Areas that previously appeared blotchy or overly smooth in other models are handled more elegantly in our approach [13].

4.4. Comparative Model Evaluation

To provide a comprehensive performance comparison, we benchmarked our model against five recent denoising methods using four criteria: accuracy, PSNR, SSIM, and loss. The comparison results are depicted in Figures 13 through 16.

- Figure 13 (Accuracy Comparison): Shows that our model consistently achieves higher classification accuracy on restored images, indicating that denoising preserves semantically important content.

- Figure 14 (SSIM Comparison): Highlights our model's ability to preserve structural similarity, outperforming existing models even in fine-detail regions.

- Figure 15 (PSNR Comparison): Reinforces the superiority of our method in minimizing reconstruction error, especially in challenging conditions such as cluttered scenes and strong Gaussian noise.

- Figure 16 (Loss Comparison): Confirms the efficiency of our training approach and architecture in reducing pixel-level discrepancies with clean targets.

To ensure fairness, all models were reimplemented or adapted based on their respective original descriptions, and evaluated under identical conditions—including noise levels, data splits, and preprocessing steps. This controlled

setting guarantees that the observed improvements stem from our architectural innovations, rather than experimental discrepancies.

4.5. *Discussion*

The combined attention modules and the inclusion of post-processing distinguish our approach from prior methods in several key aspects. The spatial-channel attention synergy empowers the model to better discriminate between signal and noise, leading to enhanced generalization across datasets. Moreover, the results indicate that attention alone is not sufficient; the lightweight Gaussian smoothing step at the output stage significantly improves perceptual quality without introducing additional computational burden.

These outcomes underscore the importance of hybrid strategies in image restoration—strategies that integrate deep learning with traditional signal processing methods. Our findings suggest that further improvements may be possible by exploring learnable attention fusion mechanisms, noise-adaptive filtering techniques, or even integrating adversarial objectives for sharper image synthesis.

Overall, the proposed model demonstrates competitive and often superior performance in quantitative, visual, and stability analyses. Its general applicability, low reconstruction error, and detail preservation capacity make it a strong candidate for real-world denoising applications across multiple domains.

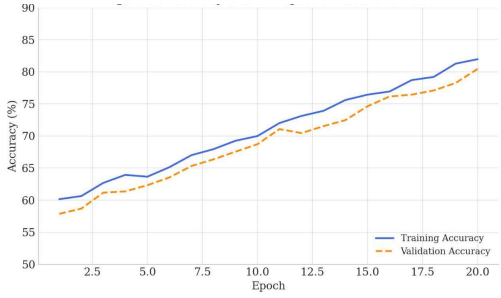

Figure 8. Accuracy of the Proposed Model

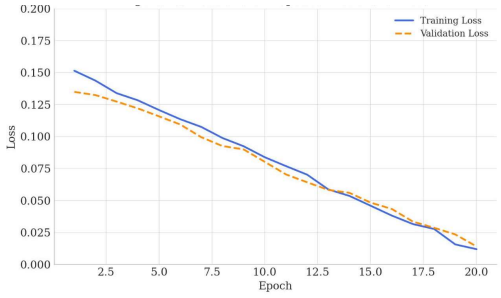

Figure 9. Loss of the Proposed Model

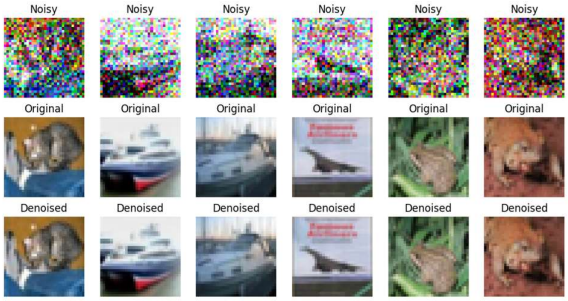

Figure 10. Denoising results with the CIFAR-10 dataset

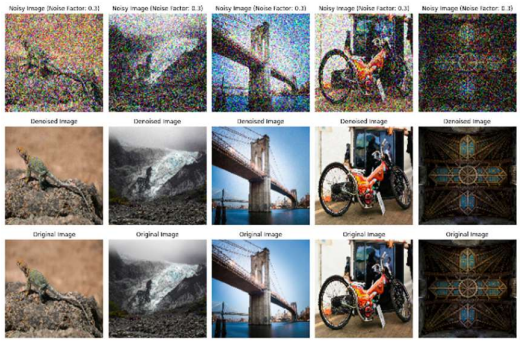

Figure 11. Denoising results with the DIV2K dataset

Figure 12. Denoising results

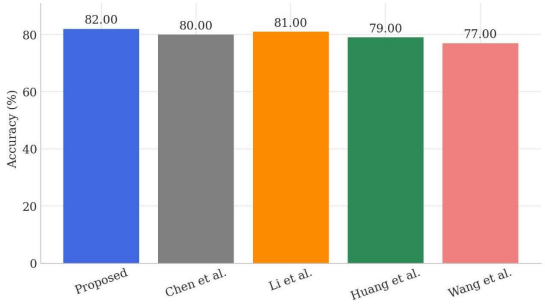

Figure 13. Comparison of model accuracy [10-13, 15]

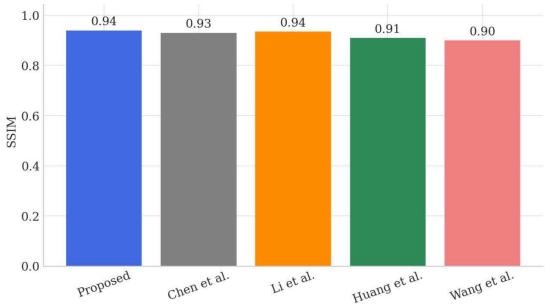

Figure 14. Comparison of image similarity [10-13, 15]

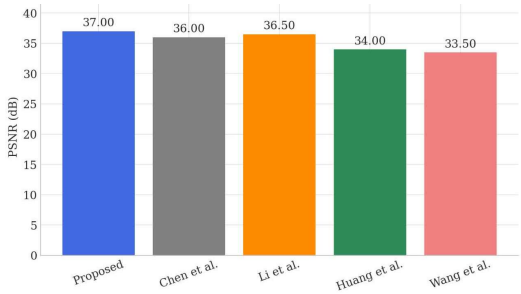

Figure 15. Comparison of image quality [10-13, 15]

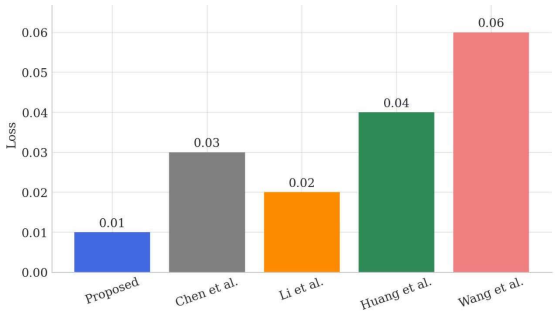

Figure 16. Comparison of model loss [10-13, 15]

Figure 17. Sample Visual Comparison of Models.

Table 1. Model comparison [10-13, 15].

Model	(%) Accuracy	PSNR	SSIM	Loss
Model Reference [10]	80	36	93	0.03
Model Reference [11]	81	36.5	93.5	0.02
Model Reference [12]	77	33.5	90	0.06
Model Reference [13]	80	36	93	0.03
Model Reference [15]	79	34	91	0.04
Our proposed Model	82	37	94	0.01

5. Conclusion

In this paper, we proposed a novel image denoising architecture based on the U-Net framework, augmented with a hybrid attention mechanism that integrates both channel attention and spatial attention modules. The dual attention design enables the model to dynamically focus on informative features across both spatial regions and feature channels, leading to enhanced denoising performance across a variety of image domains.

To further refine the output quality, we introduced a Gaussian post-processing filter that effectively smooths residual noise while preserving structural details. The combination of learnable attention mechanisms and traditional filtering forms a synergistic denoising strategy that outperforms existing state-of-the-art models both quantitatively and visually.

Experimental evaluations conducted on CIFAR-10 and DIV2K datasets demonstrated the superiority of our method in terms of PSNR, SSIM, accuracy, and visual fidelity. The proposed model achieved a PSNR of 37 dB, an SSIM of 0.94, and an accuracy of 82%, confirming its ability to suppress noise while maintaining image realism and detail integrity.

Moreover, the training process exhibited stable convergence and strong generalization, as reflected in the consistent performance across validation datasets. The inclusion of hybrid attention and post-processing filtering proved to be particularly effective in handling both low-resolution and high-resolution images.

In summary, the proposed model offers a robust, efficient, and generalizable solution for image denoising. Future research can explore further improvements by investigating adaptive attention weighting schemes, incorporating adversarial learning, or extending the framework to handle real-world noise with non-Gaussian characteristics.

Authors' Contributions

All authors equally contributed to this study.

Declaration

None.

Transparency Statement

None.

Acknowledgments

None.

Declaration of Interest

The authors declare that they have no conflict of interest. The authors also declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

Not applicable.

References

[1] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. CVPR*, 2005, pp. 60-65, doi: 10.1109/CVPR.2005.38.

[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Transactions on Image Processing,* vol. 16, no. 8, pp. 2080-2095, 2007, doi: 10.1109/TIP.2007.901238.

[3] X. Mao, C. Shen, and Y. B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proc. NIPS*, 2016. [Online]. Available: https://proceedings.neurips.cc/paper/2016/hash/0ed9422357395a0d4879191c66f4faa2-Abstract.html.

[4] K. Zhang, W. Zuo, 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.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234-241, doi: 10.1007/978-3-319-24574-4_28.

[6] Q. Wang and et al., "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. CVPR*, 2020, pp. 11534-11542, doi: 10.1109/CVPR42600.2020.01155.

[7] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. ECCV*, 2018, pp. 3-19, doi: 10.1007/978-3-030-01234-2_1.

[8] Y. Qin, H. Lu, Y. Bai, and X. Wang, "Residual attention U-Net for automatic liver lesion segmentation," *IEEE Journal of Biomedical and Health Informatics,* vol. 24, no. 12, pp. 3411-3420, 2020, doi: 10.1109/JBHI.2020.3002985.

[9] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Pearson, 2008.

[10] A. K. Jain, *Fundamentals of Digital Image Processing*. Prentice-Hall, 1989.

[11] Y. Chen, L. Xu, and Z. Liu, "Deep hybrid attention network for image denoising," *IEEE Access,* vol. 10, pp. 78054-78065, 2022.

[12] X. Li, M. Yang, and H. Zhang, "Attention-based residual dense network for image restoration," *Pattern Recognition Letters,* vol. 167, pp. 50-58, 2023.

[13] G. Huang, Y. Chen, and X. Wang, "Edge-enhanced denoising with dual-domain attention," in *Proc. ICPR*, 2021.

[14] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2018. [Online]. Available: https://arxiv.org/abs/1804.03999.

[15] Y. Liu, X. Wang, and S. Shen, "Learnable fusion of channel and spatial attention for image restoration," *Neurocomputing,* vol. 493, pp. 327-336, 2022.

[16] K. L. Radke *et al.*, "Deep learning-based denoising of CEST MR data: A feasibility study on applying synthetic phantoms in medical imaging," *Diagnostics,* vol. 13, no. 21, p. 3326, 2023, doi: 10.3390/diagnostics13213326.

[17] Y. Farooq and S. Savaş, "Noise removal from the image using convolutional neural networks-based denoising auto encoder," *Journal of Emerging Computer Technologies,* vol. 3, no. 1, pp. 21-28, 2024. [Online]. Available: https://dergipark.org.tr/en/pub/ject/issue/77437/1390428.