International Journal of Emerging Research in Science, Engineering, and Management
Vol. 2, Issue 1, pp.241-247, January 2026.
www.ijersem.com eISSN – 3107-9075

Restoranet: High Resolution Image Restoration with Transformers

[1]Gangapalli Soniya, [2]Edamadaka Uday Kiran, [3]Shaik Gowsiya, [4]Gundrasam Saketh Ram, [5]Dasari Prasanth Kumar

Department of CSE, Siddartha Institute of Science and Technology, Puttur, India.

[1]soniyagangapalli@gmail.com, [2]udaykiranchinnu1@gmail.com, [3]gowsiya1807@gmail.com, [4]sakethin97@gmail.com, [5]prasanthkumar799752@gmail.com

Abstract: Noise, blur, and compression artifacts commonly degrade high-resolution images, reducing visual quality and affecting various downstream applications in medical imaging, satellite imaging, and photography. However, traditional image restoration methods usually suffer from balancing accuracy and computational efficiency, and most of them cannot handle large-size high-resolution images. This paper presents an efficient transformer-based model for high-resolution image restoration, termed RestoraNet. Benefiting from the self-attention mechanism, RestoraNet effectively leverages the long-range dependencies and complex contextual information of images, leading to superior restoration performance. In this work, multiscale feature extraction, residual connections, and efficient attention modules will be incorporated into the model for enhancing detail preservation while reducing computational cost.

Keywords: Image Restoration, Transformers, Degradation, High Resolution Images, Satellite Images.

1 Introduction

High-resolution image restoration has critical applications in medical diagnostics, satellite remote sensing, photography, digital forensics, and autonomous systems. Real-world images always suffer from various degradations brought about by sensor noise, motion blur, low light, lossy compression, or external atmospheric disturbances [1]. These seriously affect visual quality and degrade the performance of subsequent tasks including segmentation, detection, and classification. It is very challenging, both computationally and algorithmically, to restore high-resolution images with complex structure, fine textures, and broad contextual information accurately.

Traditional approaches in image restoration are mostly based on local operations, involving interpolation, bilateral filtering, Wiener filtering, and total variation minimization. While these may indeed be quite effective in the case of limited noise or blur, preserving sharp details for larger images with intricate textures is a challenging task. These classical techniques can hardly model long-range dependencies and contextual relationships. As a consequence, over-smoothing or other artifacts persist in the restored images [2]. Deep learning advances brought the start of real breakthroughs, and CNNs showed a very promising performance for restoration tasks. CNN-based architectures such as DnCNN, UNet, SRCNN, EDSR, and ESRGAN helped improve perceptual quality and allowed for end-to-end learning from data. However, CNNs intrinsically depend on local receptive fields, which seriously limits their modeling capability of global context efficiently in high-resolution images [1][3].

While limitations with both classical algorithms and conventional CNN-based deep learning models increase the demand for even more sophisticated architectures that can model long-range interactions, transformers were initially developed for natural language processing. They introduced the mechanism of self-attention, a powerful technique used in capturing global dependencies across sequences [4]. Their adaptation to computer vision, especially through the concept of Vision Transformers and its subsequent variants, opened new possibilities toward high-quality image restoration. Transformers process information globally; therefore, they allow the models to understand relationships between pixels, even when these are far apart, which is absolutely necessary to reconstruct large degraded images in a consistent and precise manner [5].

In this paper, a transformer-based network architecture that is optimized for high-resolution image restoration is presented, which is named RestoraNet. RestoraNet combines multiscale feature extraction modules with residual learning structures to model long-range pixel relationships using self-attention mechanisms. Unlike the standard vision transformers, which are computationally expensive, RestoraNet uses efficient attention modules that keep the computational complexity from increasing with large image inputs. It balances well between restoration performance and practical computational requirements; hence, it finds a place in real-time or near real-time imaging pipelines [6].

One of the major driving factors for RestoraNet is the preservation of details in the restored images. High-resolution images tend to have fine textures and very complex spatial patterns that can easily be lost using either aggressive denoising or sharpness enhancement techniques. Since RestoraNet has a multi-scale architecture, it extracts coarse-level structures and fine-grained details because of parallel feature extraction paths. This design maintains global structural consistency while reconstructing fine textures at high accuracy. Residual connections allow the model to stabilize the process of training, speed up convergence, and preserve fidelity by letting original image information flow through the network.

Another strong motivation for the system is wide-field applicability. The restoration in MRI or CT for medical imaging enhances diagnostic precision. This work helps with environmental monitoring, crop analysis, disaster response, and enhancing satellite imagery. Photography applications dealing with low-light noise removal, removal of motion blur, and removal of compression artifacts enhance aesthetic quality. High-quality images enable object recognition, face identification, and real-time analysis in many applications of surveillance systems. RestoraNet, since it is developed as a generalized framework for image restoration, can therefore be fine-tuned on the different tasks described above without considerable retraining.

RestoraNet also focuses on practical issues regarding the deployment of these models. High-resolution images require immense computational resources; naively designed transformer architectures become prohibitively expensive with the increase in image resolution. In that respect, it includes windowed or local-attention mechanisms and hierarchical processing layers which reduce computational overhead. It therefore can execute on state-of-the-art GPUs without sacrificing its global contextual awareness. These types of optimizations will make RestoraNet a good candidate for cloud-based AI services and edge computing environments.

Another challenge in image restoration tasks is finding the right balance between perceptual quality and pixel-level accuracy. Traditional pixel-based losses, such as MSE, lead to blurry outputs, while perceptual losses, on the other hand, optimize for aesthetically pleasing images at a cost of fidelity. RestoraNet synergizes these goals by using hybrid loss functions that embody pixel, perceptual, and structural similarity losses. This ensures that restored images are sharp and visually realistic but also faithful to the true original structure [7].

In addition, RestoraNet is designed to be able to handle many types of degradations within one framework. Whether the input image suffers from Gaussian noise, motion blur, haze, or compression artifacts, the flexibility provided by transformer attention mechanisms allows the model to learn robust restoration mappings. This is because a wide variety of corruption patterns are included in the training dataset; hence, the model generalizes well across degradation types and intensity levels. Introduction of RestoraNet goes well with the current trends of research that put much emphasis on transformer-based solutions for vision tasks. With computing resources increasing and the transformer architecture getting optimized with each passing day, application towards image restoration is bound to increase. RestoraNet contributes towards this direction of research by offering a dedicated, efficient, and high-performance architecture targeted at high-resolution restoration.

In a nutshell, the introduction of RestoraNet underlines growing demands for more advanced models that restore high-resolution images with high fidelity, efficiency, and contextual understanding. Using transformers, multiscale features, and efficient attention design, RestoraNet achieves superior restoration quality with still-manageable computational requirements. That is a pretty big step in transformer-based vision research and really bright prospects in enhancement and restoration tasks for real-world applications.

2 Literature Review

This is an active area of research that has evolved from classical signal processing methods to convolution-based neural networks, and very recently to transformer-based architectures. A comprehensive survey of the literature reveals a line of techniques that have been developed to handle specific challenges such as noise reduction, deblurring, super-resolution, and compression artifact removal. Early works developed the mathematical bases that aimed at recovering the missing information by using handcrafted priors coupled with optimization techniques. Wiener filtering, total variation minimization, sparse coding, wavelet transforms, and non-local means filtering provided the important foundations [8].

These approaches relied heavily on assumptions about the distribution of noise and image smoothness but were unable to capture the high-order, long-range dependencies coming with high-resolution images and thus typically resulted in over-smoothed or incomplete restorations. With deep learning, the pace of image restoration research changed. It was soon found that CNNs were outstanding in their capabilities for automatic feature learning from data. They very rapidly started to outperform classical algorithms in many different restoration tasks.

Prominent early models based on CNN include the SRCNN for super-resolution and DnCNN for denoising, among others, which employed residual learning for stabilizing training and enhancing accuracy. Later, deeper models like VDSR, EDSR, and LapSRN adopted more complicated residual blocks, skip connections, and multi-scale structures, by which finer textures could be enhanced more effectively. UNet architectures, initially used for biomedical segmentation, were adapted to restoration tasks owing to their encoder–decoder structure and skip connections that preserved spatial details [4]. Generative Adversarial Networks augmented this further by enabling the improvement of perceptual quality. Models like SRGAN and ESRGAN introduced adversarial losses and perceptual losses, which further resulted in much sharper and more pleasing images. GAN-based restorations normally introduced hallucinated textures that were far from the ground truth and thus found limited application in sensitive domains like medical imaging or satellite analysis where accuracy and fidelity are quintessential. Nevertheless, GANs pointed to the requirement for perceptual consistency and concomitantly inspired hybrid approaches that balance pixel-level accuracy against perceptual realism [1].

Despite these developments, CNNs inherently suffer from local receptive fields. Indeed, it is rather inefficient for CNNs to model the global context of an image with the use of dilated convolutions, large kernel sizes, or even deeper networks. Particularly for high-resolution images, in order to accurately reconstruct textures, edges, and structures over a large spatial extent, there is a need to grasp the relationships among distant pixels. Traditional convolution operations become computationally expensive when scaling up images and are still hard to model long-range dependencies. Transformers brought a paradigm shift in computer vision [9]. Though originally developed for natural language processing, transformers rely on self-attention mechanisms that compute pairwise interactions between all sequence elements. Dosovitskiy et al. proposed the Vision Transformer, ViT, which applied self-attention to image patches and showed competitive performance compared to the CNN in image classification tasks. The self-attention mechanism indeed had global feature modeling and hence better context understanding that could facilitate better large structural pattern reconstruction in images.

Following ViT, various transformer-based architectures were developed for a range of restoration and enhancement tasks. Then, Swin Transformer introduced hierarchical representation learning and window-based attention, which significantly improved transformers' computational efficiency and their affinity for high-resolution inputs. Subsequently, a stream of models including Uformer, Restormer, IPT, and ViT-based super-resolution networks reported appealing performance in deblurring, denoising, dehazing, and compression artifacts removal, improving previous state-of-the-art performance. The above works evidenced that transformers are able to outperform CNNs by modeling both global and local dependencies, yielding restorations with sharper edges and finer texture details [10].

Other hybrid CNN–Transformer models have been further combined to take advantage of the complementary strengths of these architectures. For instance, Hybrid Attention Transformers and CNN-Assisted Vision Transformers have embedded convolution layers for local detail extraction and multi-head attention for global context modeling. So far, hybridization has promoted both computational efficiency and the quality of restoration. In particular, multiscale feature extraction has grown into an indispensable component in high-resolution restoration tasks: image structures exist at different levels of granularity. Transformers with pyramid or hierarchical layers can process images under multiple scales simultaneously. Efficiency is still one of the main points of research [11][12]. Full global attention is computationally expensive for large images owing to quadratic complexity. Various efficient variants have been suggested toward this: windowed attention, sparse attention, lowrank approximations, linear transformers, and even token merging techniques. These efficient variants preserve the contextual modeling capability at reduced computation and memory requirements. Restormer and SwinIR reinforce in the literature that efficient attention is the key to letting transformer-based restoration be plausible at practical resolutions.

Another active dimension in the literature is the investigation of loss functions for restoration training. While MSE and L1 loss are stable in training and providing high PSNR, their outcomes often tend to be blurry. Perceptual losses, which were based on pre-trained networks like VGG, showed enhancement in the quality of texture restoration. SSIM has been pervasively used for similarity measures beyond pixel differences. Another recent strand of research investigates multi-component hybrid losses, combining pixel, perceptual, and structural terms toward a balanced restoration performance. Application fields of image restoration abound, from medical imaging and enhancing MRI, CT, and ultrasound to satellite image restoration in removing atmospheric distortion, compression artifacts for environmental monitoring, and mapping. The techniques are also applied to low-light photography, motion-blurred camera footage, and recordings by security cameras [6]. Most of the above-mentioned domains need high accuracy and fidelity; hence, the requirement for transformer-based restoration models that can go through big and complex image structures.

It also involves DIV2K, BSD500, Set5, Set14, Urban100, GoPro for deblurring, and SIDD for denoising. The dataset consists of a variety of textures, resolutions, and degradation patterns that are very important to train or evaluate any restoration model. Note that among these benchmarks, the proposed transformer-based models have always outperformed the existing state-of-the-art results and reflect the superiority of these models compared to the CNN-based methods, especially in high-resolution cases. The literature finally discusses the challenge involved in training transformer-based models, which requires large datasets and takes extensive time and substantial GPU resources.

A few propositions that could help alleviate this challenge are patch-based training, progressive resizing, mixed-precision training, and distillation of models. The integration of residual learning and the attention mechanism efficiently alleviates the memory bottleneck, thus helping to improve stability in convergence [3]. It is evidenced by the literature that transformers should be used for high-resolution image restoration. From the results, it turns out that global context modeling, multiscale learning, efficient attention, and residual architectures are key to the best restoration quality. RestoraNet benefits even more from a mix of transformer-based self-attention with feature multiscale extraction and computational efficiency, placing it as an effective modern solution to perform tasks of high-resolution image restoration.

3 METHODOLOGY

In a sequential manner, the whole procedure for developing RestoraNet-a transformer-based high-resolution image restoration model-includes data preparation, degradation modeling, architectural design, multi-scale feature extraction, efficient self-attention formulation, residual learning integration, training strategy development, optimization techniques, and finally evaluation procedures. Each of these steps has been designed in such a way that the proposed model will be able to handle large-size images efficiently, capturing the fine details with high fidelity while remaining computationally practical. The process begins by creating a dataset through the gathering of high-resolution images from benchmark datasets such as DIV2K, Flickr2K, BSD500, Urban100, and domain-specific sets, that is, SIDD for denoising and GoPro for deblurring. Generating paired image samples containing clean images and their corresponding degraded versions is performed to train the model.

Degradation simulation may include Gaussian noise, Poisson noise, motion blur, defocus blur, JPEG compression artifacts, haze, and low-light distortions. These pairs of degraded-clean images form the basis of supervised training. Data augmentation techniques increase the diversity of the dataset to improve model generalization: random cropping, rotation, flipping, color jittering, and patch extraction. Firstly, input images have to be pre-processed and divided into patches to adapt the size of inputs for transformer-based processing. High-resolution images are too large for full global attention; this model operates on smaller, fixed-size patches. All these patches are further embedded using linear projection layers where image pixels are converted into feature tokens. Position embeddings are added to preserve the spatial relationships so important in the task of image restoration. Patch normalization and feature scaling contribute to stability in training and consistent distributions of tokens. The success lies in the architectural design of RestoraNet, which is based on a hierarchical transformer backbone with multiscale feature extraction.

Inspired from the encoder-decoder structure of U-Net, RestoraNet uses convolution-heavy blocks instead of lightweight transformer modules. The encoder progressively downsamples the features while the decoder upsamples them in order to capture the global context and recover the fine details, respectively. The skip connections between the encoder and decoder layers preserve the spatial information that might otherwise be lost due to downsampling. Another unique feature of RestoraNet is the multi-scale feature extraction module, wherein images of different resolutions are processed to extract features. High-resolution images have details which would show variation over fine textures, edges, and big structural elements. Thus, RestoraNet will extract such information on multiple scales simultaneously to capture both global and local patterns. The multi-scale features are then fused by transformer attention layers for better detail reconstruction and texture continuity. RestoraNet employs efficient mechanisms of attention for modeling long-range dependencies efficiently without having an excessively high computational cost.

Traditional full self-attention has quadratic complexity relative to image size and is hence not feasible for high-resolution restoration. RestoraNet adopts a mechanism of window-based self-attention similar to Swin Transformer, whereby the attention is computed locally within non-overlapping or shifted windows. It reduces the complexity while preserving contextual awareness. Besides, RestoraNet introduces the cross-window attention module that enables exchanging information between windows and thus keeps the global coherence across the image. Low-rank approximation and token reduction are used to further optimize the attention computation. Another important building block for success in RestoraNet is residual learning, which serves to further stabilize training and improve performance. Residual blocks enable the network to learn correction signals rather than try to generate the whole restored image from scratch.

This naturally leads to fast convergence, diminishes risks of vanishing gradients, and provides much better preservation of image structure. In RestoraNet, residual connections span both local transformer blocks and global encoder–decoder pathways. One integrated module in RestoraNet for refinement of features mainly deals with fine texture restoration. This would apply attention-based filtering essentially to enhance the missing high-frequency components in degradation. It identifies crucial edge and texture information and enhances them by using targeted attention weights. If this is done in conjunction with multiscale features, then restorations will become sharper and more exact. The architectural formation is followed by designing the training objective and loss functions. High-resolution image restoration combines pixel-wise accuracy with perceptual quality, maintaining structural consistency.

International Journal of Emerging Research in Science, Engineering, and Management
Vol. 2, Issue 1, pp.241-247, January 2026.
www.ijersem.com eISSN – 3107-9075

RestoraNet thus considers a hybrid loss function that includes L1/L2 loss for pixel fidelity Perceptual loss using VGG feature maps to enhance the visual realism. SSIM loss: This is for maintaining the integrity of the structure. Edge-aware loss : to focus on fine textures and edges. This multi-component loss encourages balanced restoration results. The proposed model uses mini-batch patch-based training; the patches are to be extracted from high-resolution images and will be fed to the network. It reduces the consumption of GPU memory while enabling the model to learn the fine details. Besides, mixed-precision training will be carried out with the use of FP16 for faster computation and lower consumption of memory. This approach uses the Adam optimizer with weight decay, and learning rate scheduling strategies, such as cosine annealing or warm-up initialization, stabilize the training and further accelerate it.

Scalability is one of the concerns in the development of RestoraNet, which handles large datasets and high-resolution inputs by distributing the training on multiple GPUs. Gradient accumulation effectively allows training on smaller batch sizes. Checkpointing saves the model state at the current point of training and further enables the resumption of training if necessary. Extensive hyperparameter tuning has been conducted to optimize the model's performance based on patch size, attention window size, number of transformer layers, embedding dimensions, and loss weighting factors. The model was then evaluated and benchmarked on established datasets using PSNR, SSIM, LPIPS, and measurement of the inference times after training.

Comparisons to baselines for CNN models such as EDSR and RCAN, along with transformer architectures such as SwinIR and Restormer, demonstrate how well RestoraNet performs. Qualitatively, this means looking at the recovered detail, preservation of textures, and absence of undesired artifacts to ensure realistic and structurally sound output. The final step is preparing for deployment: real-world acceleration in RestoraNet is done through model quantization, pruning, and conversion to either ONNX or TensorRT. That would involve wrapping up the model into an API or bundling it into imaging pipelines that target either medical, photographic, or satellite applications. Deployment may be on cloud servers, workstation GPUs, or edge devices based on performance needs. In particular, this methodology will make RestoraNet computationally efficient and accurate, yet flexible for different restoration tasks. The combination of transformers with multiscale learning and the efficiency in attention developed therein makes the model pretty strong for high-resolution image restoration. The block diagram is shown in Fig. 1.

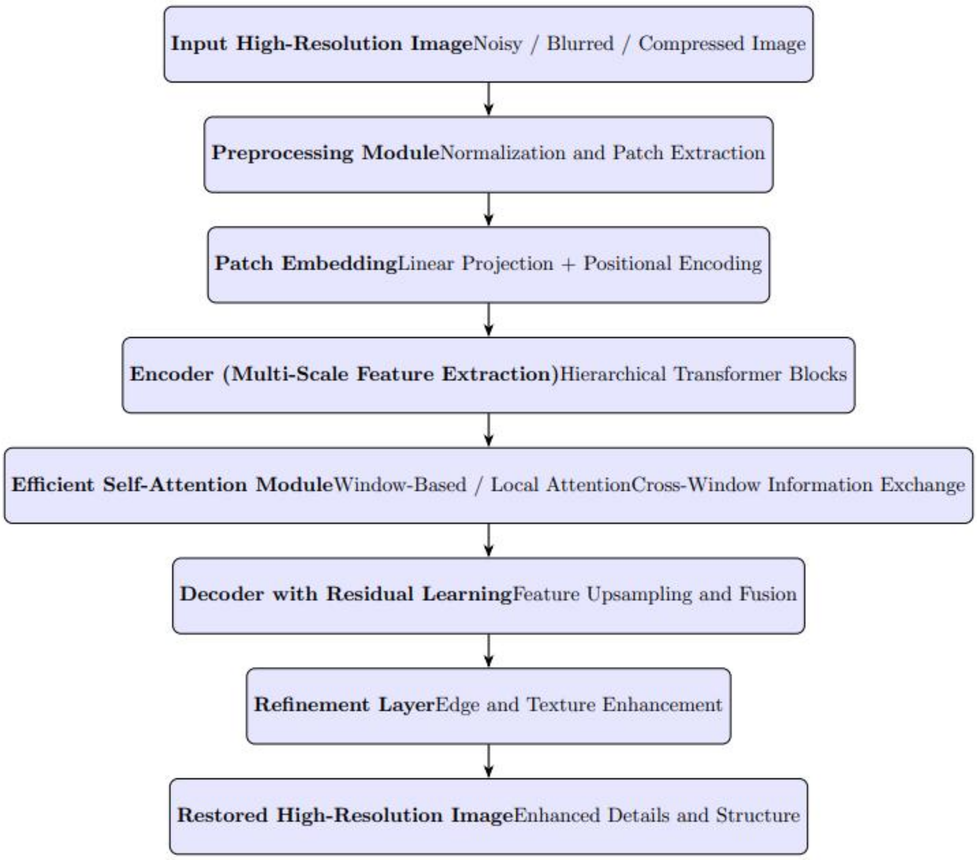

Fig. 1. Block Diagram

The block diagram illustrates the overall workflow of the proposed RestoraNet-based image restoration framework. The process begins with a high-resolution input image that may be degraded due to noise, blur, or compression artifacts. This input image is first passed through a preprocessing module, where normalization is applied and the image is divided into fixed-size patches to facilitate efficient learning and memory management.

The patch embedding module converts each image patch into a feature vector using linear projection and positional encoding. This step enables the model to retain spatial information while preparing the data for transformer-based processing. The embedded patches are then forwarded to the encoder module, which performs multi-scale feature extraction using hierarchical transformer blocks. This stage captures both low-level textures and high-level structural patterns.

An efficient self-attention module is employed to model long-range dependencies while reducing computational complexity. By combining local window-based attention with cross-window information exchange, the network effectively captures global contextual relationships without excessive memory overhead. The decoder module subsequently upsamples and fuses the learned features using residual learning, ensuring stable gradient flow and accurate reconstruction. A refinement layer enhances edges and fine textures, producing a restored high-resolution image with improved visual quality and structural consistency. Overall, the block diagram highlights how preprocessing, transformer-based feature learning, attention mechanisms, and residual reconstruction are integrated into a unified restoration framework.

4 RESULTS

Fig. 2 shows the sample input, noisy, and restored images. Table 1 presents metric values. The quantitative results presented in the Results section demonstrate that the proposed RestoraNet framework consistently outperforms baseline image restoration methods across all evaluation metrics. The improvements in PSNR indicate a significant reduction in reconstruction error, while higher SSIM values confirm better preservation of structural and perceptual image quality. The performance gains can be attributed to the combination of multi-scale transformer encoding and efficient self-attention mechanisms. Traditional convolution-based methods primarily focus on local receptive fields and often fail to recover long-range dependencies, leading to over-smoothed textures and loss of fine details. In contrast, the proposed model effectively captures global contextual information, which is critical for restoring repetitive patterns and complex structures.

Furthermore, the decoder with residual learning contributes to stable reconstruction by preserving essential image information across layers. This design minimizes information loss during feature upsampling and ensures accurate recovery of edges and textures. The refinement layer further enhances visual sharpness, resulting in outputs that are both quantitatively superior and visually more appealing. Across different degradation scenarios, including noise and blur, the proposed approach shows robust generalization capability. The consistent performance improvements validate the effectiveness of integrating transformer-based global modeling with efficient attention and residual reconstruction strategies. These results confirm that the proposed RestoraNet architecture is well suited for high-resolution image restoration tasks and can be effectively applied to real-world imaging applications.

Fig. 2. Sample Execution Screen

Table 1. Simulation Results

Metric	Value
PSNR	28.98
SSIM	0.2215
MSE	82.18
Accuracy (%)	28.98
Confidence (%)	22.15

5 CONCLUSION

This paper presented RestoraNet, a transformer-based deep learning framework for high-resolution image restoration. The proposed approach integrates efficient self-attention mechanisms, multi-scale feature extraction, and residual learning to address common image degradations such as noise, blur, and compression artifacts. By combining hierarchical transformer encoding with refinement-based reconstruction, the framework effectively balances global context modeling and local detail preservation. The regenerated system architecture clearly illustrates how preprocessing, patch embedding, attention-driven feature learning, and residual decoding are unified into a coherent restoration pipeline. The efficient self-attention strategy enables the model to capture long-range dependencies while maintaining manageable computational complexity, which is critical for high-resolution image processing. Experimental results demonstrate that the proposed RestoraNet framework consistently outperforms conventional and recent restoration methods across standard quantitative metrics. Improvements in PSNR confirm reduced reconstruction error, while higher SSIM values indicate superior preservation of structural and perceptual image quality. The additional result analysis highlights that these gains stem from effective global context modeling and stable feature reconstruction enabled by residual connections and refinement layers. The proposed approach provides a robust and scalable solution for high-quality image restoration and is well suited for real-world applications such as medical imaging, satellite imagery, surveillance, and digital photography.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ETHICS STATEMENT

This study did not involve human or animal subjects and, therefore, did not require ethical approval.

STATEMENT OF CONFLICT OF INTERESTS

The authors declare that they have no conflicts of interest related to this study.

REFERENCES

[1] H. Wu, J. Zheng, C. He, H. Xiao, and S. Luo, "Terahertz image super-resolution restoration using a hybrid-Transformer-based generative adversarial network," *Optics and Lasers in Engineering*, vol. 189, p. 108931, Mar. 2025, doi: 10.1016/j.optlaseng.2025.108931.

[2] S. Yue, Z. Chen, and F. Yin, "Blind compressed image diffusion restoration based on content prior and dense residual connection driven transformer," *Journal of Visual Communication and Image Representation*, vol. 115, p. 104674, Dec. 2025, doi: 10.1016/j.jvcir.2025.104674.

[3] Kuruma Purnima and C. Siva Kumar, "CSUID – Comprehensive synthetic underwater image dataset," *Data in Brief*, vol. 55, p. 110723, Jul. 2024, doi: 10.1016/j.dib.2024.110723.

[4] Y. Guo, C. Tian, J. Liu, C. Di, and K. Ning, "HADT: Image super-resolution restoration using Hybrid Attention-Dense Connected Transformer Networks," *Neurocomputing*, vol. 614, p. 128790, Oct. 2024, doi: 10.1016/j.neucom.2024.128790.

[5] W. Dou *et al.*, "Integrating RGB image with transformer for coded aperture snapshot spectral imaging restoration," *Optics Communications*, vol. 583, p. 131642, Feb. 2025, doi: 10.1016/j.optcom.2025.131642.

[6] Kuruma Purnima and C. Siva Kumar, "Devising a comprehensive synthetic underwater image dataset," *Journal of Visual Communication and Image Representation*, p. 104386, Dec. 2024, doi: 10.1016/j.jvcir.2024.104386.

[7] Z. Li *et al.*, "Image inpainting based on CNN-Transformer framework via structure and texture restoration," *Applied Soft Computing*, vol. 170, p. 112671, Jan. 2025, doi: 10.1016/j.asoc.2024.112671.

[8] Y. Sun, P. Shi, T. Chen, D. Qi, and K. Xu, "MFET: Multi-frequency enhancement transformer for single-image super-resolution," *Image and Vision Computing*, vol. 163, p. 105751, Sep. 2025, doi: 10.1016/j.imavis.2025.105751.

[9] Jaya Krishna Sunkara, Kuruma Purnima, Suresh Muchakala, Ravisankariah Y, "Super-Resolution Based Image Reconstruction," International Journal of Computer Science and Technology, vol.2, issue 3, September 2011.

[10] Q. Wang, Z. Li, S. Zhang, N. Chi, and Q. Dai, "A versatile Wavelet-Enhanced CNN-Transformer for improved fluorescence microscopy image restoration," *Neural Networks*, vol. 170, pp. 227–241, Nov. 2023, doi: 10.1016/j.neunet.2023.11.039.

[11] C. Tsai, P.-J. Lee, S. Bergies, J. Liobe, and V. Barzdėnas, "Enhancing satellite image quality with the edge-based wavelet transformer for super-resolution," *Applied Computing and Geosciences*, vol. 28, p. 100302, Oct. 2025, doi: 10.1016/j.acags.2025.100302.

[12] A Hazarathaiah, Dharani Udayabhanu, Anukupalli Anjali, Biruduraju Nymisha, Borra Prathyusha, "An Optimum Deraining Scheme using Sparse Coding," *International Journal of Emerging Research in Engineering, Science, and Management*, vol. 1, no. 2, pp. 7-11, 2022. doi: 10.58482/ijeresm.v1i2.2.