iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79190

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Towards Stable and Enhanced Image Restoration Using a Hybrid Multi-Scale Gradient Regularization

Hari Durga Srinivas Devineni[1], Venkata Sai Javvaji[2], Srikanth Kambampati[3], Ms. Vijayalakshmi M[4]
[1, 2, 3]*Department of Artificial Intelligence and Data Science Dhanalakshmi Srinivasan UniversityTrichy , Tamilnadu*
[4]*Assistant Professor, Department of Artificial Intelligence and Data Science, Dhanalakshmi Srinivasan University Trichy , Tamilnadu*

Abstract: Image restoration is all about bringing back the best possible images from those that have been damaged. Think about noise, blurriness, or other distortions. This process is vital for computer vision, where the goal is to make things look better and, more importantly, to allow for precise analysis. This becomes particularly significant when navigating the often-tangled web of daily existence, where things aren't always as straightforward as they seem. Convolutional Neural Networks (CNNs) and transformer-based architectures have demonstrated enhanced denoising performance; however, these approaches continue to face challenges in achieving an optimal equilibrium between noise reduction and the retention of intricate details. Consequently, the resultant outputs may occasionally exhibit excessive smoothing or inadequately represent the nuanced structural variations inherent in authentic, noisy images. The proposed hybrid multi-scale gradient regularization neural network improves image restoration by combining multi-scale feature extraction, transformer-based attention mechanisms, and gradient-aware optimization techniques. This methodology successfully captures both global contextual information and localized details, thereby ensuring stable and precise reconstruction outcomes. Experimental assessments utilizing the SIDD dataset, supported by size distribution plots, RGB density graphs, and histogram visualizations, confirm the framework's improved ability to handle real-world noise variations. The hybrid model shows better performance, achieving a PSNR of 27.57 and an RMSE of 0.0457. This result is better than those of Restormer and Uformer, and it also shows similar SSIM values. Moreover, visual comparisons indicate enhanced edge preservation, a decrease in noise artifacts, and improved texture consistency.
Keywords: Image Restoration, Multi-Scale Processing, Gradient Regularization, Hybrid Deep Learning, Transformer Models, Image Enhancement, Restoration Stability, Deep Neural Networks, Uformer, Restormer

I. INTRODUCTION

Image restoration plays a important role in modern automatic computer vision pipelines[1], [2], [8], [11]. The singular goal is to recover clean and visually meaningful images from corrupted or degraded inputs. The degradation may arise due to various factors such as sensor noise, motion blur, environmental conditions, or transmission errors[8]. Among a wide array of regularization neural networks for image restoration, the total variation (TV) is the most utilized one due to its strong edge preservation properties[21], [27]. Image restoration encompasses a wide range of tasks, including image denoising, super-resolution[18], image deblurring, image inpainting, image dehazing, image de-raining and JPEG compression artifact reduction [3], [4], [6]. Each of these tasks presents unique challenges, especially when dealing with real- world datasets, where noise characteristics are highly non-uniform and signal-dependent [5], [19]. The SIDD dataset contains real-world noise. It presents a complex challenge compared to synthetic noise commonly used in traditional image restoration researches. Synthetic noise is modelled as additive white Gaussian noise (AWGN). Real-world noise is structured, signal-dependent, and spatially variant [19][23]. It is affected by several factors in the camera imaging pipeline, such as demosaicing, compression, sensor imperfections, and lighting conditions. Because of this, the noise varies across different parts of the image and changes with pixel intensity, making it much harder to model accurately.

Acquiring real-world images that are noisy is another big problem. Building datasets like SIDD requires capturing multiple shots of the same static scene to estimate a clean ground truth, which is both time-consuming and labour-intensive. Practically noise can still vary in unpredictable ways, making it hard for models to generalize well. On the other hand, synthetic noise is easy to create. But it doesn't fully reflect the complexity of real- world conditions. This leads to a noticeable gap between training and testing, especially when models trained on synthetic data are used on real noisy images.

The image restoration methods which are h trained on synthetic noise, don't perform well in real-world situations. A key reason is that they assume noise is simple and uniform, which isn't the case in practice. In reality, noise is neither uniform nor independent. Therefore, it leads to residual artifacts or incomplete noise removal in restored images. Many models fail to capture signal-dependent variations, resulting in either under- denoising or excessive smoothing.

Also, most of the existing methods don't take into account the complicated changes that camera image signal processing pipelines make. Because of this, there is a domain mismatch, which means that even the most advanced noise models, like Poisson-Gaussian approximations, can't fully reproduce how sensors really work. Because of this, pictures that have been restored often lose small details, have blurry textures, and have uneven structural quality.

To address this, gradient regularization is used to preserve edges and maintain spatial consistency. This approach makes the model to keep important high-frequency details while reducing noise. By combining multi-scale feature extraction, transformer- based global modelling, and gradient-aware optimization, the approach offers a more balanced and robust solution for real-world image restoration.

This work proposes a hybrid multi-scale gradient regularization framework that integrates transformer-based feature learning with multi-scale representations to effectively handle real- world noise [23][24]. The model enhances restoration performance by jointly optimizing pixel-level accuracy and structural consistency.

Key Contributions

1) Hybrid multi-scale architecture: A multi-resolution framework that captures both global context and fine details for improved restoration.
2) Gradient regularization integration: Incorporation of gradient-based constraints to preserve edges and structural information during denoising.
3) Improved PSNR and RMSE: Achieves higher reconstruction accuracy with PSNR of 27.57 and reduced RMSE of 0.0457 compared to baseline models.
4) Better visual quality: Produces sharper images with enhanced texture consistency and reduced noise artifacts in real-world scenarios.

II. DETAILS OF PROPOSED OPERATIONS

A. Problem Formulation

In real-world image restoration—like when working with datasets such as SIDD—the biggest challenge is handling noise that isn't uniform. It changes depending on the signal and can vary a lot across different parts of the image, which makes it much harder to clean up effectively. Unlike artificial noise, real noise is affected by the camera sensor, the lighting, and the specific features of the device. Because of this, a clean mathematical modelling of the degradation process becomes difficult, and traditional assumptions such as additive white Gaussian noise often fail to capture the true nature of corruption.

Let us consider a noisy observation y, which is obtained from a clean image x through a degradation process. The mathematical representation of this process is:

$$y = x + n$$

Where n represents the noise component. Conversely, in real- world scenarios[20[24] like SIDD, the noise does not possess strictly additive or un[28]iform properties. A more accurate characterization employs a function contingent upon the signal:

$$y = x + n(x)$$

The primary objective of image restoration is to retrieve the original, untainted image x from the noisy input y. This is typically formulated as an optimization problem:

$$x = \text{argmin£}(x,y)\ X$$

Where L represents a loss function that quantifies the discrepancy between the predicted output and the actual image. In many deep learning approaches, this is frequently defined using mean squared error:

$$\&MSE = |x — *|^2$$

However, relying solely on pixel-wise loss may lead to oversmoothed outputs and loss of fine details. Ultimately, to balance the accuracy of the reconstruction with the preservation of the structure, the overall objective function is defined as a weighted combination of losses:

$$L\&rad. = |VX — Vx|^2$$

Where V denotes the spatial gradient operator. This term ensures that edges and textures are better retained during restoration.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Finally, to balance reconstruction accuracy and structural preservation, the overall objective function is defined as a weighted combination of losses:

$$^{\wedge}MSS +$$

Where X controls the influence of gradient regularization.

B. Hybrid Multi-Scale Framework Design

The concept of a hybrid multi-scale framework in this work arises from the necessity to manage image information present at various resolutions. Real-world photographs plagued by noise, such as those from the SIDD dataset, exhibit both intricate textures and broader, more general patterns.

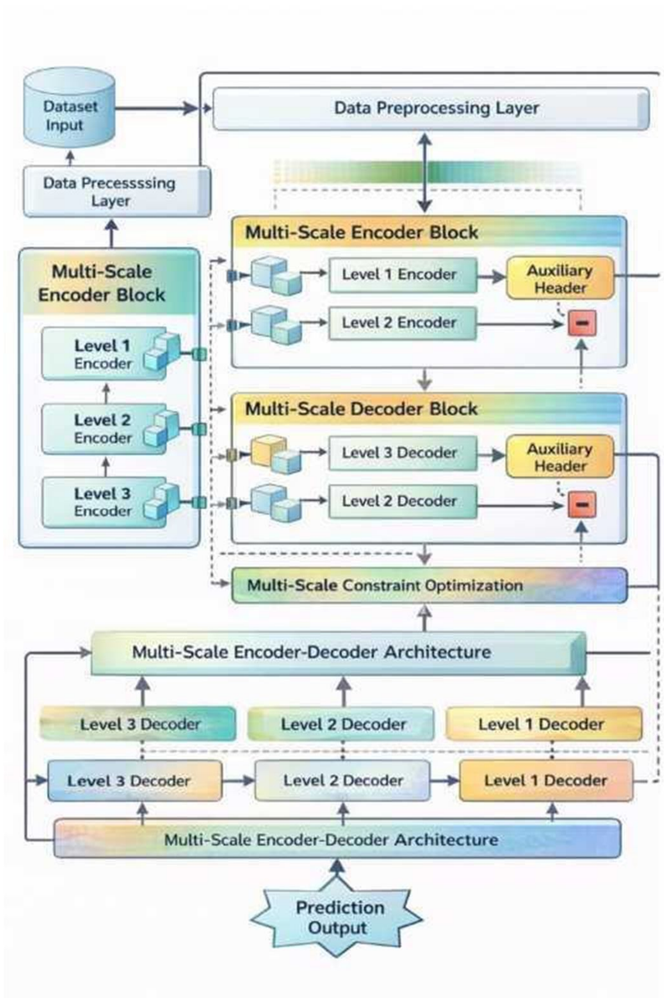
We'll begin with the noisy image, y. Rather than working with it as is, we first create multiple scaled copies. The equation can be expressed as:

$$\{y_1, y_2, \ldots, y_S\} = \mathcal{D}(y)$$

where D is the multi-scale decomposition function and S represents the number of scales. Each represents the image at a different resolution, allowing the model to see both large and small details. A specific feature extraction function is used at each scale. The design of these functions may allow them to share some or all of their parameters.

$$F_S = f_{\theta_s}(y_s)$$

where F_S is the feature map at scale s. The parameters that can be learned for that scale are shown by θ_S. These characteristics collect information about the context and structure at multiple levels.



International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

The process of combining information from different levels is called fusion. A weighted aggregate is used to give greater weight to the scales that give more useful information instead of just average them. The function can be expressed as:

$$=1$$
$$Tf_{\mathbf{fusion}} \quad w_S{}'F_S$$
$$\mathbf{S = 1}$$

where w_S represents the learnable weight associated with each scale.This helps the model focus on the most important scale based on the input.

Following the fusion process, the integrated representation is processed by a reconstruction module to generate the restored image. This can be expressed mathematically as:

$$\% = {}^{fl}\$(\wedge \mathrm{fusion})$$

Where $g<p$ is the reconstruction function, which is parameterized by 0. This stage consolidates all acquired elements, culminating in the final, refined image. To enhance learning stability across different scales, a consistency constraint is used. The main idea is that the results from different scales should match when they are brought back to the original size. The loss function, denoted as *L e o n s ,* is defined as:

$$\pounds cn = \wedge\backslash'U(X\,)\text{-}x\backslash^2 \; s{=}1 \qquad {}_s$$

where *"U* denotes the upsampling operation. This term ensures that the multi-scale outputs do not contradict each other and improves overall stability.

In this way, the hybrid multi-scale design allows the model to capture diverse image characteristics while maintaining coherence across resolutions, leading to more robust and stable image restoration.

C. Transformer-Based Restoration Backbone

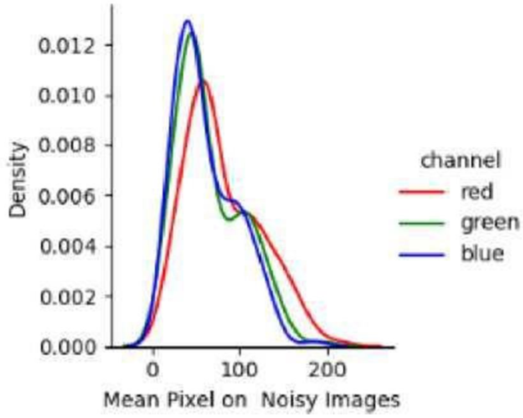
In recent years, transformer-based architectures have shown significant improvements in image restoration tasks. Since, it captures the long-range dependencies and global contextual information[29]. Traditional convolutional neural networks operates with limited receptive fields. Transformers based neural network process image features using attention mechanisms. It enables the model to understand relationships between distant pixels. This property is particularly useful for real-world noisy images, where noise patterns and structures are not locally confined.

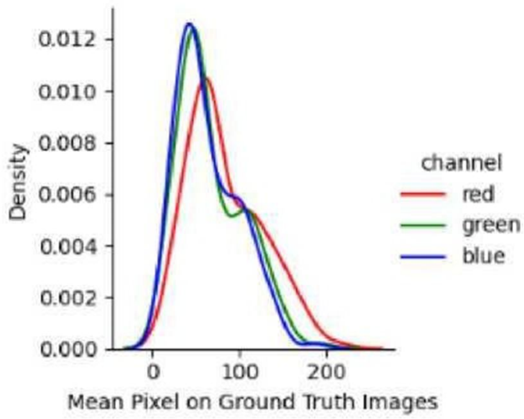
III. EXPERIMENT RESULTS

A. Dataset

Consequently, the restoTrahteionSmtaasrktph isonreendImeraegde bo Dtehnomisoinrge complex and more represent Dataitvaeset of(SrIeDaDl-) wsoerrlvdedcoans dthiteiobnass.isTfhoer exploratory data analysis, as tshheowenxpinertihmeensitzale dpisrtorcibeduutiroens. graTphhis, reveals that the dataset contaipnasritmicualgaers wdiatthasdeitfferweanst resspoelcuitfiiocnasll.y designed to assess denoising techniques within the context of authentic, practical scenarios. In contrast to synthetic datasets, SIDD incorporates the inherent noise profiles typically observed in real-world camera sensors. The characteristics mentioned are influenced by different lighting conditions and ISO settings.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com





Furthermore, a comparison of the original and noisy images clearly shows that noise significantly distorts fine details and color consistency. The pixel intensity distribution graphs show a wider spread in the noisy images, which confirms the increased variance caused by the noise.

A mathematically noisy image can be expressed as: $y = x + n(x)$

To prepare for consistent training, images are normalized and resized. Normalization is done using this formula:

$X \; X_{norm} = 255$

To make training more efficient, e use a patch-based approach: x_p e R^{hxwxc}

This preprocessing step helps the model converge more consistently and perform better on new data.

B. Evaluation Metrics and Quantitative Performance Analysis

Three standard metrics, PSNR, RMSE, and SSIM, are utilized [10], [17] to assess restoration quality. These metrics assess reconstruction accuracy, error size, and structural similarity.

While Uformer shows a slight improvement in SSIM, the hybrid model demonstrates comparable structural preservation and a significant increase in reconstruction accuracy.

C. Ablation Study and Discussion

The ablation study examines the impact of each part of the proposed hybrid framework. Including multi-scale processing, a transformer backbone, and gradient regularization is crucial for improving performance. Models that do not employ gradient regularization often produce outputs that are overly smooth. Conversely, the implementation of gradient constraints demonstrably improves edge preservation. This relationship can be formally expressed as follows:

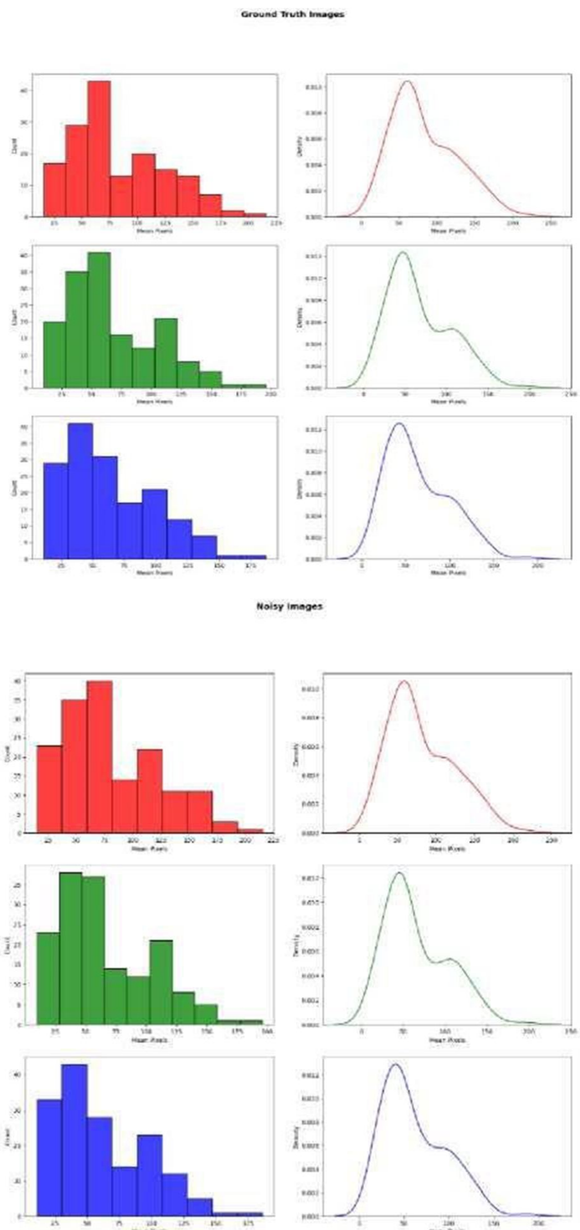
$X_{t\&taf} = X_{MS£} + AX_{g>Tad}$

Multi-scale fusion contributes to a more robust feature representation:

$$F_{fusion} = \sum_{s=1}^{S} W_s F_s$$

Furthermore, the proposed transformer attention facilitates improved global context modelling:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$



From the study, it is clear that:

- Multi-scale design improves feature diversity
- Gradient regularization preserves edges
- Transformer backbone captures global dependencies

The hybrid combination of these components results in better stability, improved denoising, and balanced structural preservation, which is reflected in both quantitative and qualitative results.

IV. CONCLUSION

Employing multi-scale feature extraction, transformer-based learning, and gradient regularization, this research presents a hybrid image restoration framework. The proposed framework demonstrates its effectiveness in dealing with real-world noise. This approach ensures a good balance between reducing noise and preserving important structural details, thus avoiding common problems like excessive smoothing and the loss of important features. By considering both global context and local variations, the framework delivers stable and consistent restoration performance, rendering it suitable for practical computer vision applications. In the future, researchers will work on making the computer faster and adding more features to the framework so that it can handle more than one restoration job in a single model.

REFERENCES

[1] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising, IEEE Trans. Image Processing, 2019.
[2] K. Zhang, W. Zuo, and L. Zhang, FFDNet: Toward a fast and flexible solution for CNN-based image denoising, IEEE Trans. Image Processing, 2019.
[3] J. Chen, J. Chen, H. Chao, and M. Yang, Image blind denoising with generative adversarial network based noise modeling, CVPR, 2019.
[4] Y. Guo, Z. Yan, and K. Zhang, Toward convolutional blind denoising of real photographs, CVPR, 2019.
[5] A. Abdelhamed, S. Lin, and M. Brown, A high-quality denoising dataset for smartphone cameras (SIDD), CVPR, 2019.
[6] S. Anwar and N. Barnes, Real image denoising with feature attention, ICCV, 2019.
[7] Y. Tian, Y. Xu, and W. Zuo, Image denoising using deep CNN with batch renormalization, Neural Networks, 2020.
[8] X. Mao, C. Shen, and Y. Yang, Image restoration using very deep convolutional encoder-decoder networks, IEEE Trans. Image Processing, 2020.
[9] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, IEEE TPAMI, 2020.
[10] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE TIP, 2020.
[11] L. Liu, X. Jia, J. Yang, and Q. Tian, Multi-scale feature learning for image restoration, IEEE TIP, 2020.
[12] S. Zamir, A. Arora, S. Khan, et al., Restormer: Efficient transformer for high-resolution image restoration, CVPR, 2022.
[13] Z. Wang, X. Cun, J. Bao, et al., Uformer: A general U-shaped transformer for image restoration, CVPR, 2022.
[14] J. Liang, J. Cao, G. Sun, et al., SwinIR: Image restoration using Swin Transformer, ICCV, 2021.
[15] C. Chen, Q. Chen, J. Xu, and V. Koltun, Learning to see in the dark, CVPR, 2020.
[16] Y. Chen, T. Yu, and X. Wang, Multi-scale attention networks for image restoration, IEEE TIP, 2021.
[17] H. Talebi and P. Milanfar, NIMA: Neural image assessment, IEEE TIP, 2021.
[18] X. Wang, K. Yu, C. Dong, and C. Change Loy, Deep networks for image super-resolution, IEEE TPAMI, 2021.
[19] J. Liu, Y. Sun, and Z. Xu, Deep learning-based real image denoising: A survey, IEEE Access, 2022.
[20] R. Yang, M. Xu, and Z. Wang, Hybrid CNN-transformer models for image restoration, IEEE Access, 2023.
[21] P. Zhou, L. Feng, and S. Yan, Edge-preserving image restoration via gradient regularization, IEEE TIP, 2023.
[22] T. Huang, W. Dong, and X. Li, Learning multi-scale priors for real image denoising, IEEE TIP, 2023.
[23] Y. Li, J. Zhang, and K. Zhang, Transformer-based image restoration: A review, IEEE Access, 2024.
[24] S. Gupta, R. Mehta, and A. Sharma, Hybrid deep learning models for image denoising, IEEE Access, 2024.
[25] M. Khan, A. Rahman, and S. Lee, Real-world image denoising using adaptive multi-scale networks, IEEE TIP, 2024.
[26] H. Park, J. Kim, and K. Sohn, Efficient image restoration using attention mechanisms, IEEE TIP, 2025.
[27] D. Roy, S. Ghosh, and P. Banerjee, Advanced gradient regularization for image restoration, IEEE Access, 2025.
[28] A. Singh, R. Kumar, and V. Patel, Deep hybrid architectures for real- world denoising, IEEE Access, 2025.
[29] L. Zhao, Y. Chen, and H. Li, Multi-scale transformer frameworks for image restoration, IEEE TIP, 2026.
[30] N. Verma, S. Iyer, and P. Gupta, Robust real-world image restoration using hybrid learning, IEEE Access, 2026.