A Transformer-Driven Approach for Accurate and Efficient Single Image Deraining

Kommana Kusuma
ECE Department
GMR Institute of Technology
Rajam, Vizianagaram
kusumakommana2707@gmail.com

Mudadla Gowri Nithin
ECE Department
GMR Institute of Technology
Rajam, Vizianagaram
gowrinithin333@gmail.com

Jaggurothu Shyam Kumar
ECE Department
GMR Institute of Technology
Rajam, Vizianagaram
shyamshyam7564@gmail.com

Muthiki Naveen Kumar
ECE Department
GMR Institute of Technology
Rajam, Vizianagaram
naveenkumarmuthiki@gmail.com

*A. Sudhakar
ECE Department
GMR Institute of Technology
Rajam, Vizianagaram
sudhakar.a.@gmrit.edu.in

***Abstract*--Single image deraining is of primary importance in the enhancement of computer vision tasks in bad weather conditions. Traditional CNN-based image deraining methods are generally bad at modeling long-range persistence. In this regard, transformer-based methods provide global contextual modeling but lack specialized attention mechanism to handle localized patterns of rain in their removal. For this, Transformer-Driven Image Deraining Network (TDIDNet) is proposed, that effectively combines vision transformers (ViTs) with a convolutional attention mechanism to remove rain streams while preserving fine details. TDIDNet consists of two main steps, Feature Extraction and Feature Aggregation & Refinement. The first step analyses the input images through three parallel convolution paths, using Channel Attention Blocks and typical convolutions before passing their features through the ViT modules. The second step consists of using Image Original Resolution Block (IORB) to restore structural integrity. TDIDNet achieves superior PSNR, SSIM, and loss scores, outperforming state-of-the-art transformer-based methods.**

***KEYWORDS*-Image Deraining, Transformer-Driven Image Deraining Network (TDIDNet), vision transformer (ViT), Attention Module (AM), Channel Attention Block (CAB), Image Original Resolution Block (IORB).**

I.INTRODUCTION

Nowadays, image deraining—the process of removing rain streaks from images that are captured in rainy conditions, [1] has developed into an important task in computer vision. Rain severely degrades images, blurring scene details, degrading visibility, and introducing unwanted distortion. This degradation compromises the efficiency of many vision-based applications such as autonomous driving, surveillance, remote sensing, and military security. The complexity of single-image deraining arises from the varied characteristics of rain patterns, from light drizzles to downpours, which interact with the scene in diverse ways. The chief objective of image deraining, therefore, is the effective removal of the rain streaks whilst ensuring that the structure, color, and texture of the original image would be preserved in such a manner as to provide undistorted visual content.

Traditional image deraining approaches mostly rely on hand-crafted features and prior-based schemes, which mostly fail to generalize across different rain patterns. To deal with these challenges, the use of CNNs and deep learning approaches are best. The CNN model was fairly good in extracting features at several levels of the input image and could distinguish between the rain streaks and the background textures. Other CNN-based models have been developed like DATN, DIDMDN, RESCAN, and MSPFN which adopt a progressive approach to remove the rain while making sure that fine details are preserved.

Recently, Vision transformers (ViT) [2] have achieved remarkable success in the high-level vision tasks of object segmentation and recognition. Their self-attention mechanism enables to model dependencies, and therefore, they are suitable for modeling complex spatial correlations. Unlike the CNNs that focus on local pixel relationships, transformers analyze the whole image regions contextually. However, applying transformers straight onto less vision tasks like image deraining incorporates a few challenges:

To overcome above mentioned limitations, we introduce TDIDNet, a hybrid deep learning model that combines the advantages of both CNNs and transformers. CNN-based attention mechanisms fine-tune local spatial features, allowing us to ensure texture preservation perfectly. Vision transformer modules model long-range dependencies contributing to solid rain removal.

Image Original Resolution Block (IORB) [3] retains the structural integrity for high-fidelity image restoration. By bringing these two modalities together, TDIDNet effectively

DOI: 10.55041/IJSREM45570

provides rain streak removal while retaining image details present because of localized contextual augmentation via CNNs and global contextual understanding through transformers, which in turn allows better performance of vision models toward real-life applications.

TDIDNet improvements are as follows:

- Design a Novel Transformer Framework: It consists of parallel feature extraction pipelines that combine convolutional and transformer-based attention mechanisms, applied for better rain streak removal.
- Enhance Deraining Accuracy: Multi-path attention is applied to modulate how rain-relevant features are aggregated while preserving texture detail.
- High-quality image reconstruction: Feature maps are refined on the original resolution in the IORB module to prevent the loss of details and to maintain structural integrity.
- Best-performing approach: Experiments show that TDIDNet outperforms in both qualitative and quantitative evaluations with the existing models on real-world and synthetic datasets.

II.PROPOSED METHODOLOGY

The Transformer-Driven Image Deraining Network (TDIDNet) shown in Fig1. is proposed to effectively reduce rain drops from input image while preserving fine details and textures. The network integrates convolutional neural networks (CNNs) for localized feature extraction and vision transformers (ViTs) for global context modeling. TDIDNet consists of two main stages:

Feature Extraction: Multi-scale spatial and contextual features from the rainy image extracted using convolutional layers, channel attention blocks (CABs), and vision transformers.

Feature Aggregation & Refinement: the extracted features are aggregated, enhanced through attention mechanisms to reconstruct a high-quality rain-free image using an Image Original Resolution Block (IORB).

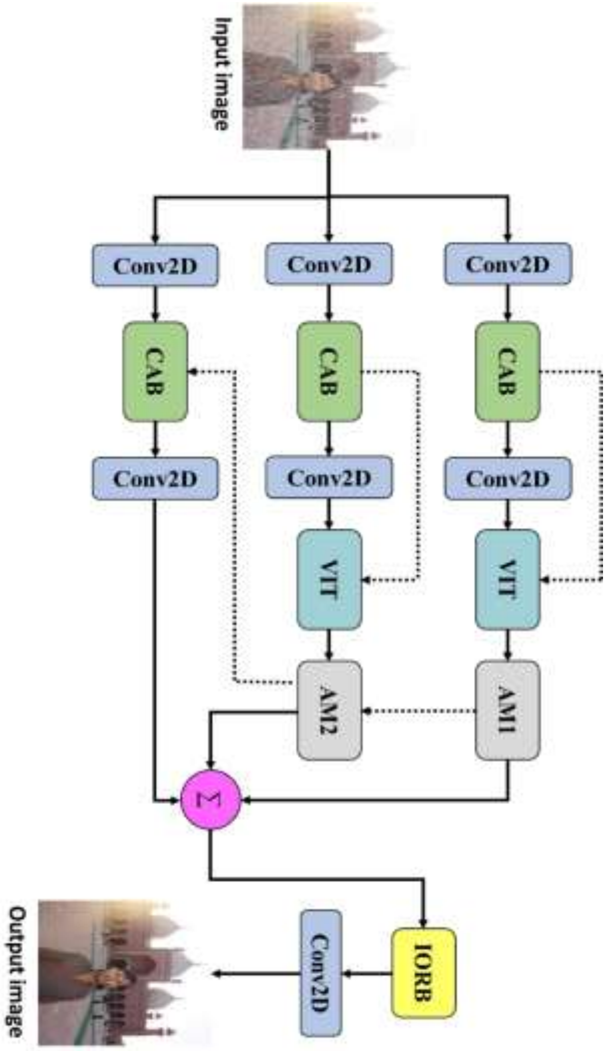


Fig1. Proposed Network TDIDNet Architecture

A. Channel Attention Block (CAB):

This method allows the network focus on important areas by suppressing irrelevant background noise. The CABs [4] will help the model to enhance the significant rain-related features while facilitating a contrast enhancement between the rain streaks and the background.

B. Vision Transformer (ViT):

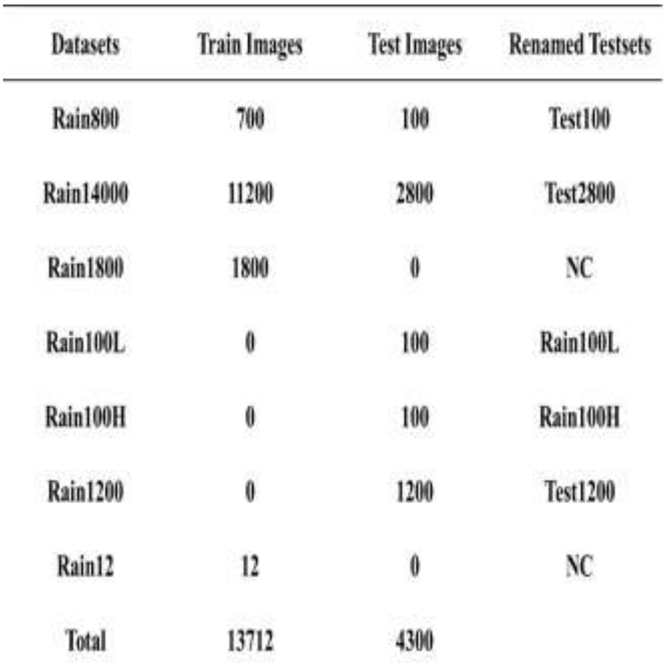
The transformer blocks capture global dependencies within the features by applying self-attention mechanisms that allow the network to recognize long-range rain streak structures while preserving high-level contextual details.

C. Attention Module (AM):

Attention mechanisms [5-7] are used to extract features by focusing on the region of rain-affected areas and enhancing the secession of rain streak from the background. This helps the network keep structural details while optimizing rain removal. TDIDNet implements several attention modules to improve the derailment performance by fulfilling specific purposes for feature refinement.

D. Image Original Resolution Block (IORB):

The Image Original Resolution Block ensures that the spatial resolution of the image will not be lost during the neural network propagation. IORB, unlike traditional downsampling algorithms that destroy fine features, allows feature refinement to be performed on relevant structures without introducing any noise or unwanted deformations.



Datasets	Train Images	Test Images	Renamed Testsets
Rain800	700	100	Test100
Rain14000	11200	2800	Test2800
Rain1800	1800	0	NC
Rain100L	0	100	Rain100L
Rain100H	0	100	Rain100H
Rain1200	0	1200	Test1200
Rain12	12	0	NC
Total	13712	4300	

Table1. Total Deraining Datasets used in TDIDNet

A. *Loss Function*

When we train the TDIDNet, we can minimize the loss function, so we can get a rain-free image without losing the original content. The loss function implements a combination of MSE, perceptual and adversarial friendly operable loss functions to encourage accurate removal of rain streaks, good retention of high-level features in the image as well as photorealism.

1) Mean Squared Error (MSE) Loss:

The MSE [8-9] loss seeks to minimize the sensor's disagreement between the derained output and the corresponding pixel of the given image. In this work, the class of MSE loss has also been outlined as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{I}_i - I_i\right)^2 \qquad (1)$$

Where $\hat{I}_i$ is represents the output image after executing the deraining algorithm on I, Ii represents the image under test in the algorithm, and N indicates the total number of pixels considered.

2) Perceptual Loss:

Perceptual loss [10] is introduced because of the need to maintain the high-level features existing in the image. In that, they measure the discrepancy between the derained and ground truth image compressing on the feature extraction of a pre-learned network such as VGG:

$$Perceptual\ loss = \frac{1}{N_1}\sum_{j=1}^{N_1}\left\|\phi_j(\hat{I}) - \phi_j(I)\right\|^2 \qquad (2)$$

where $\phi j(\hat{I})$ and $\phi j(I)$Are the feature maps of derained image $\hat{I}$ Nd ground truth image I extracted from the j-th layer of the pre-trained network and $N1$ is the number of layers used in calculations.

3) Adversarial Loss:

An adversarial loss [11] function is introduced to constrain the generated images to be visually acceptable:

$$Adversarial\ loss = E\left[log\left(D(\hat{I})\right)\right] \qquad (3)$$

D is a measure that tells whether an image comes from real images or generated ones in the GAN framework.

4) Total loss:

Combining all the above losses leads to the following total loss function which serves as a weighted ci-combination of the losses:

$$Total\ loss = \lambda_{MSE} \cdot MSE + \lambda_{Perceptual\ loss} \cdot Perceptual\ loss + \lambda_{Adversarial\ loss} \cdot Adversarial\ loss \quad (4)$$

B. End-to-End Training

TDIDNet had demonstrated an end-to-end training approach to the challenges of this task; thus, the model is essentially enabled to learn its feature representations and rain removal mechanisms directly from the data without requiring manually engineered feature extraction and thus is able to adapt to different rain patterns effectively. The network is trained over real-world and synthetic rainy image datasets to facilitate generalization in different deraining scenarios. Experimental results showed that TDIDNet clearly outperforms some of the obvious methods on both the real and synthetic benchmarking datasets gives normalized deraining performance. Both comparisons were based on standard visibility algorithms with minimal artifacts.

III. Experiments and Analysis

For the performance evaluation, TDIDNet was trained using both real and synthetic rain image datasets clearly showing precipitation. The performance was evaluated using SSIM and PSNR metrics, are used to calculate image quality. The model was trained on a combination of synthetic and real rainy image datasets for robust operation across different varying levels and patterns of rain intensity. The test datasets included Test2800, Test1200, Test100, Rain100L, and

Rain100H presented in Table1. [12] and they are commonly used datasets in prior deraining studies.

A. Implementation details:

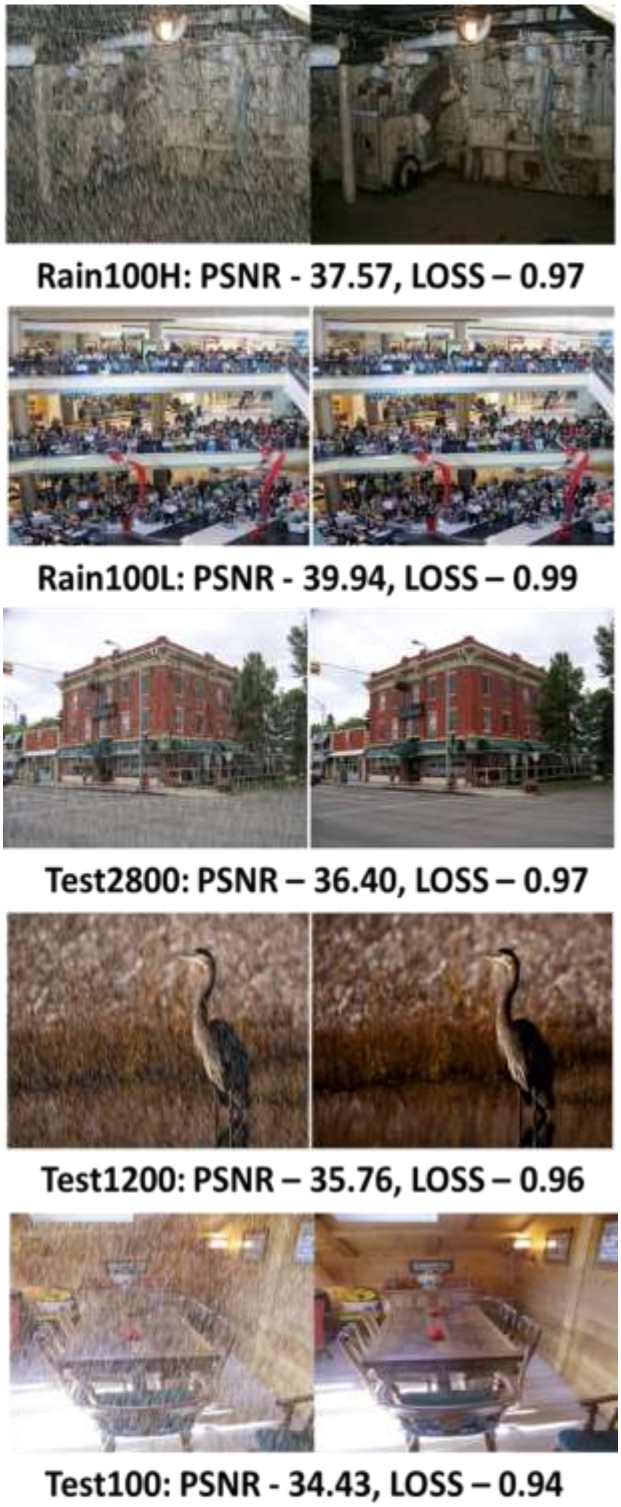
TDIDNet was trained without any pre-training, in an end-to-end manner. The architecture integrated two Channel Attention Blocks at each layer for better feature extraction. For down-sampling, the maximum of 2×2 size with a stride of 2 as an upsampling technique was used. The architecture included an Image Original Resolution Block that greatly improves rain streak removability while maintaining spatial fidelity during the deraining process. It was trained by using 8 batch size for the datasets that comprised Test2800, Test1200, Test100, Rain100L, and Rain100H with a resolution of 256×256 for 100 epochs [13-14].

IV. Results and Discussions:

This technique has thus been compared with a quantity of existing deraining networks such as RESCAN, DerainNet, DIDMDN, UMRL, MSPNet, SAPNet, and DATNet. The area under the curve (AUC) obtained indicated that with a measured PSNR and SSIM, Performance TDIDNet exhibited best performance in recovering the images and providing convincing evidence for the method based on TTLT as being very effective in rain streak separation with regard to different state-of-the-art methods.

Qualitative evaluation as shown in Fig2. further places TDIDNet in a very good light, as the derained images produced appear quite appealing, with minor artifacts. The network has succeeded in rain streak removal, and texturing is not blurred: Experimental comparisons with competitive models show this. The qualitative results are an additional validation of the quality image restoration that TDIDNet can produce owing to its feature extraction, attention, and reconstruction procedures.

Quantitative results for TDIDNet gives better results when compared to the other networks as shown in Table2. by using different datasets such as Test2800, Test1200, Test100, Rain100H, and Rain100L. This network finds some difficult in Rain100H but overall performance is very well compared with other networks.



Rain100H: PSNR - 37.57, LOSS – 0.97

Rain100L: PSNR - 39.94, LOSS – 0.99

Test2800: PSNR – 36.40, LOSS – 0.97

Test1200: PSNR – 35.76, LOSS – 0.96

Test100: PSNR - 34.43, LOSS – 0.94

Fig2. Qualitative results of our network for both real and synthetic datasets

NetworkMetrics/Year	Rain100L [13] PSNR, SSIM	Rain100H [13] PSNR, SSIM	Test100 [10] PSNR, SSIM	Test2800 [11] PSNR, SSIM	Test1200 [14] PSNR, SSIM
PRENet [9]/2019	32.44, 0.95	26.77, 0.85	24.81, 0.85	31.75, 0.91	31.36, 0.91
DerainNet [22]/2017	27.03, 0.84	14.92, 0.59	22.77, 0.81	24.31, 0.86	23.38, 0.83
DIDMDN [5]/2018	25.23, 0.74	17.35, 0.52	22.56, 0.82	28.13, 0.86	29.65, 0.90
UMRL [8]/2019	29.18, 0.92	26.01, 0.83	24.41, 0.83	29.97, 0.90	30.55, 0.91
MSPNet [10]/2020	32.44, 0.95	28.66, 0.86	27.50, 0.87	32.82, 0.93	32.39, 0.91
SAPNet [12]/2021	34.77, 0.97	29.46, 0.89	29.13, 0.88	32.18, 0.93	32.46, 0.91
DATNet [2]/2023	39.54, 0.95	31.09, 0.90	32.49, 0.91	34.62, 0.93	33.65, 0.93
Proposed Network	**39.94, 0.99**	**37.57, 0.97**	**34.43, 0.94**	**36.40, 0.97**	**35.76, 0.96**

Table2. Quantative results of our proposed network

V. Conclusion:

This paper proposed a novel Transformer Driven Image Deraining Network (TDIDNet), which effectively removes rain streaks from a single rainy image while keeping fine details and textures intact. TDIDNet achieves significantly better derained image quality by combining multi-scale convolutional feature extraction, attention mechanism and Transformer. The use of Channel Attention Blocks (CABs) guarantees stronger feature refinement, while the use of an Image Original Resolution Block (IORB) maintains high-resolution spatial information so that no details are lost while reconstructing. In addition, two Attention Modules (AM1 and AM2) further refine features to perform an accurate and effective removal of rain. Gradually refining the features selected thus causes the proposed network to offer superior visual quality and quantitative performance, gaining higher Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) than those produced by existing methods. Moreover, TDIDNet performs fairly well with a processing speed of one image every 0.12 seconds on an 18MB model size, lightweight and suitable for running on low-level edge devices without sacrificing performance.

REFERENCES

[1] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley, “Lightweight pyramid networks for image deraining,” IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 6, pp. 1794–1807, 2019.

[2] P. Zhao and T. Wang, “Degradation-aware transformer for single image deraining,” IEEE Access, vol. 11, pp. 145678–145690, 2023.

[3] Y. Cheng, J. Huang, H. Ren, W. Ran, and H. Lu, “Feature decoupling and reorganization network for single image deraining,” Multimedia Syst., vol. 30, no. 3, p. 154, 2024.

[4] L. Yu, B. Wang, J. He, G. S. Xia, and W. Yang, “Single image deraining with continuous rain density estimation,” IEEE Trans. Multimedia, vol. 25, pp. 443–456, 2021.

[5] S. Qin, S. Zhang, and Y. Zhang, “Using mask-based enhancement and feature aggregation for single image deraining,” IEEE Signal Process. Lett., vol. 30, pp. 1012–1016, 2023.

[6] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, “Deep joint rain detection and removal from a single image,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 1357–1366.

[7] Y. Wei, Z. Zhang, H. Zhang, R. Hong, and M. Wang, “A coarse-to-fine multi-stream hybrid deraining network for single image deraining,” in Proc. IEEE Int. Conf. Data Mining (ICDM), 2019, pp. 628–637.

[8] Y. Cui and A. Knoll, “Dual-domain strip attention for image restoration,” Neural Netw., vol. 171, pp. 429–439, 2024.

[9] Z. Jiang, S. Yang, J. Liu, X. Fan, and R. Liu, “Multi-scale synergism ensemble progressive and contrastive investigation for image restoration,” IEEE Trans. Instrum. Meas., vol. 73, pp. 1–13, 2023.

[10] R. Yasarla and V. M. Patel, “Uncertainty-guided multi-scale residual learning using a cycle spinning CNN for single image de-raining,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 8405–8414.

[11] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, “Removing rain from single images via a deep detail network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 3855–3863.

[12] M. Hema, R. Gurunadha, J. V. Suman, and M. Mallam, "Effective Image Reconstruction using Various Compressed Sensing Techniques," 2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE), 2024, pp. 1-6, doi: 10.1109/AMATHE61652.2024.10582191.

[13] Q. Qin, J. Yan, Q. Wang, X. Wang, M. Li, and Y. Wang, “ETDNet: An efficient transformer deraining model,” IEEE Access, vol. 9, pp. 119881–119893, 2021, doi: 10.1109/ACCESS.2021.3108516.

[14] E. Lee and Y. Hwang, “Decomformer: Decompose self-attention of transformer for efficient image restoration,” IEEE Access, vol. 12, pp. 38672–38684, 2024, doi: 10.1109/ACCESS.2024.3375360.