Raindrop removal from a single image using Generative Adversarial Network

Varshaa Puvvada Srinath
Scope
Vellore Institute of Technology
Chennai, Tamil Nadu
varshaa.puvvada2022@vitstudent.ac.in

Bhuvana Prabha B
Scope
Vellore Institute of Technology
Chennai, Tamil Nadu
bhuvanaprabha.b2022@vitstudent.ac.in

Vatsal Ojha
Scope
Vellore Institute of Technology
Chennai, Tamil Nadu
vatsalumeshraj.ojha2022@vitstudent.ac.in

Abstract -Raindrop-induced visual degradation on camera lenses has the potential to severely degrade image quality, obstructs perception of the scene. We first considered a baseline model that employed, RaindropRemovalNet, which utilizes FFT-based residual blocks to improve frequency-domain features and Global Context (GC) blocks to learn long-range spatial relations. Although efficient, this architecture comes with high computation expenses because it performs frequency-space processing. We mitigate this through a more superior GAN-based model that maximizes performance while enhancing efficiency. Our architecture employs a Generator-Discriminator combination that is trained with a hybrid loss function merging L1 loss, perceptual loss through the use of a pre-trained VGG network, and adversarial loss. This methodology allows the Generator to generate visually consistent, raindrop-free images that are close to the ground truth. Paired rainy-clean image datasets experiments prove that our GAN-based approach can achieve better qualitative performance with better PSNR and SSIM scores by eliminating both coarse and fine raindrops with detailed background restoration. This research reveals the possibilities of the frequency-domain processing, global attention mechanisms, and adversarial learning in reliable image restoration. The full code can be found on GitHub at [GithubLink], and there is an extensive walkthrough in the accompanying video at [link].

Keywords – Raindrop removal, deep learning, GAN, perceptual loss, VGG, PyTorch

I. Introduction

Images taken outside tend to be impacted by poor weather conditions, one of which is raindrops on camera lenses, a notable problem. Raindrops create structured occlusions that deform image content, reduce visibility, and considerably degrade the performance of computer vision algorithms such as object detection, segmentation, and self-driving navigation. Artifact removal from such images is therefore critical for maintaining image sharpness and maintaining essential scene information.

Conventional image processing methods have difficulty with the irregular patterns and changing transparency of raindrops. Recent developments in deep learning have indicated potential in restoring images degraded by different types of noise, such as rain.

To mitigate the above limitations and deliver better performance, we propose a second model founded on Generative Adversarial Networks (GANs). The GAN-based model has two main elements they are Generator and Discriminator. The Generator is an encoder-decoder network with residual connections which takes rainy images as the input and produces raindrop free and clean images at the same time. The Discriminator is a convolutional classifier that asseses whether the generated output is real or fake by differentiating it from ground-truth clean images.

Training this architecture includes a multi-objective loss function in addition to Generator and Discriminator which is comprised of three main components L1 Loss,Perceptual Loss, Adversarial Loss.L1 Loss promotes pixel-wise similarity between the output image and the ground truth, allowing for low-frequency accuracy. Perceptual Loss is derived from features extracted using a pre-trained VGG-19 network, this loss promotes the Generator to retain high-level structural and texture information, ensuring visual realism and semantic consistency.Adversarial Loss is given by the Discriminator, this part encourages the Generator to generate images not only correct but also not distinguishable from actual clean images.

For the sake of stable training, we utilize a warm-up stage where adversarial loss is progressively added. The model is trained on a dataset of rainy-clean pairs of images with PyTorch with data from Qian et al raindrop datatset. While testing, both quantitative measurements (PSNR, SSIM) and visual observations attest the better performance of the GAN-based model. It exhibits improved clarity, improved texture recovery, and more uniform raindrop removal under different conditions in images.

This work provides both a real-world baseline and a high-quality GAN model for image deraining, blending spatial and frequency domain processing with perceptual and adversarial learning techniques.

II. Literature Survey

Photographs captured through glass surfaces are frequently degraded by artifacts such as raindrops or dirt, particularly in scenarios involving vehicle interiors or enclosed outdoor surveillance systems. While optical defocus at capture time can mitigate these issues, it requires specific camera placement and depth-of-field conditions, limiting its practicality. To address this, a post-capture image restoration approach is proposed in [2], utilizing a dataset of clean and corrupted image pairs to train a convolutional neural network (CNN) that learns to reconstruct clean patches from contaminated inputs. This method effectively identifies and removes localized rain and dirt artifacts by learning their visual signatures. Notably, prior work by Eigen et al. represents one of the few efforts dedicated to single-image raindrop removal. Their approach involves a shallow CNN architecture comprising three layers with 512 neurons each, trained on paired data. While effective for small and sparse droplets, the method underperforms in cases of large or dense raindrops and tends to produce blurred outputs, likely due to limited network capacity and insufficient loss constraints. Comparative results underscore the improvements achieved by more specialized or deeper network architectures in handling complex degradation patterns.

Conditional adversarial networks have emerged as a versatile framework for tackling image-to-image translation tasks by jointly learning both the mapping from input to output images and the loss functions that guide this process. Unlike traditional approaches that rely on task-specific loss formulations, the method implemented in [3] enables a unified strategy applicable across diverse applications such as photo synthesis from label maps, edge-to-object reconstruction, and image colorization. The success and widespread adoption of the pix2pix software underscore the practical utility and accessibility of this approach, with minimal need for manual parameter tuning. Conditional GANs extend traditional GANs by conditioning the generative process on an observed input image alongside random noise, enabling context-aware generation. The adversarial training setup—where the generator strives to produce outputs indistinguishable from real images and the discriminator aims to detect synthetic ones—drives the model toward realism. The network architectures, adapted from prior work, employ standard convolutional blocks with Batch Normalization and ReLU activation, providing a robust foundation for high-quality image generation across various domains.

Raindrops adhered to glass surfaces introduce significant visual artifacts in images captured during rainy conditions, posing a considerable challenge for post-processing restoration. A key difficulty lies in accurately and robustly identifying raindrop regions within a single image. To address this, a novel convolutional neural network (CNN) architecture is proposed in paper[4], integrating a double attention mechanism to enhance raindrop removal. The architecture is built upon an encoder-decoder framework and incorporates high-frequency edge information via robust edge maps. Central to its design is the Joint Physical Shape and Channel Attention (JPCA) module, which combines a shape-driven spatial attention mechanism—leveraging physical priors such as convexity and contour closure of raindrops—with a channel re-calibration process to handle variability in raindrop appearance. The encoder comprises nine residual blocks arranged in three groups, with symmetric decoding and skip connections to preserve spatial details. The JPCA module generates a refined attention map through the tensor product of spatial and channel attention outputs, enhancing feature representation in raindrop regions. Experimental results demonstrate that this model surpasses state-of-the-art methods in both visual fidelity and quantitative metrics, offering a robust and physically-informed solution for single-image raindrop removal.

Image degradation caused by adherent mist and raindrops presents a significant yet underexplored challenge in computer vision, particularly for systems relying on clear visual input such as automotive cameras and surveillance systems. Unlike prior approaches that rely on hand-crafted priors to generate spatial attention maps, this work introduces a novel attentive convolutional network designed to jointly remove adherent mist and raindrops from a single image. Built upon a strong baseline architecture featuring channel-wise attention, spatial attention, and multi-level feature fusion, the proposed model in [5] further integrates interpolation-based pyramid-attention (IPA) blocks to enhance spatial perception across scales. These IPA blocks utilize bilinear interpolation to generate multi-scale attention maps, effectively distinguishing between clean regions and those degraded by low-frequency interference such as mist or raindrops. The architecture comprises 114 basic blocks arranged into six residual groups with long-range skip connections and global residual learning, ensuring both deep feature extraction and stability during training. For enhanced visual realism, a composite loss function combines mean absolute error with a perceptual loss derived from VGG16 features, capturing high-level semantic information and reducing over-smoothing. Experimental evaluations demonstrate the method’s superiority across conventional dehazing and raindrop removal tasks, as well as its specific effectiveness on the combined degradation problem, highlighting its practical relevance and state-of-the-art performance.

Recent advancements in image restoration under adverse weather conditions have achieved state-of-the-art results, yet most methods are tailored to address a single type of degradation such as rain, haze, fog, or snow. In contrast, this work in [6] proposes a unified framework capable of restoring images degraded by multiple weather conditions—including rain, fog, snow, and adherent raindrops—within a single network architecture. The proposed method employs a generator composed of multiple task-specific encoders, each specialized for a particular degradation type. A neural architecture search mechanism is used to optimally fuse the image features extracted from these encoders. To transform the corrupted features into clean background representations, a set of tensor-based operations grounded in the physical characteristics of each degradation type is introduced and leveraged as the core components of the architecture search. A novel discriminator design is implemented to both assess image realism and classify the type of degradation addressed in the restored output. Furthermore, an innovative adversarial learning scheme ensures that the loss gradients are backpropagated only to the encoder responsible for the corresponding degradation type. Experimental evaluations demonstrate that this multi-task framework achieves performance competitive with or superior to existing single-task restoration models, highlighting its effectiveness and generalizability in real-world scenarios involving diverse weather-induced image degradations.

Image restoration tasks, such as deraining, deblurring, and denoising, require a delicate trade-off between preserving fine spatial details and capturing high-level contextual information. Addressing this challenge, a novel multi-stage architecture named MPRNet is proposed in [7], which progressively restores degraded images through a synergistic combination of contextual and fine-grained feature learning. The network is structured into three stages, with the first two employing encoder-decoder subnetworks to capture broad contextual features via large receptive fields. To retain fine spatial textures, the final stage processes images at their original resolution without downsampling. At each transition between stages, a supervised attention module is introduced to adaptively reweight features based on pixel-level guidance from ground truth images. This ensures that the restoration process is sensitive to spatial inconsistencies and local variations. Additionally, a cross-stage feature fusion mechanism allows multi-scale features from earlier stages to be integrated into later stages, enhancing feature continuity and avoiding information degradation. The tightly integrated design of sequential and lateral information flow enables MPRNet to achieve superior performance across ten benchmark datasets, demonstrating its efficacy in various restoration tasks and validating the strength of progressive learning and adaptive attention in recovering high-quality images.

Conventional methods for image deraining have tended to struggle with the non-uniform and intricate nature of raindrop artifacts, especially when trying to recover high-frequency details and preserve semantic consistency between spatially distant regions. New techniques have tended to favour deep learning-based solutions capable of learning context-aware and hierarchical representations in order to better erase raindrops and recover clean images. Among the most prominent architectures is MIMO-UNet, which brings a multi-input, multi-output UNet-like architecture that allows for flexible multi-scale feature processing. The proposed architecture facilitates enhanced decomposition and recovery of various components of an image across scales, which is very important in processing the changing size and shape of raindrops. Utilizing the dual encoder-decoder branches in each resolution scale enables each subnetwork to be specialized in separate feature representations, enhancing the performance of deraining. Besides spatial domain processing, there has been increasing interest in leveraging frequency domain representations for image restoration. The ResFFT-Conv block is one such development, first proposed in DeepRFT, which converts feature maps to the frequency domain through the Fast Fourier Transform (FFT).By applying 1×1 convolutions and non-linear activations to both the imaginary and real components of the frequency-transferred feature maps, the model is more capable of specifically highlighting high-frequency details such as edges and fine textures. This serves to supplement the weakness of spatial-domain convolutions in isolation, which can be incapable of efficiently isolating and amplifying subtle detail dispersed throughout the image. Recent progress has also relied heavily on non-local operations. The traditional CNNs are limited in their receptive field and tend to miss long-range dependencies, which are crucial in erasing big raindrops or handling far-away image areas obscured by water. This is tackled by the non-local neural network (NLNet) by calculating output responses at a position as a weighted sum of all positions within the input feature map, thereby allowing global context modelling. Following this, the Global Context (GC) block provides an efficient and lighter version of the non-local operation. It combines a channel-wise attention mechanism with inspiration from the Squeeze-and-Excitation (SE) block with a context aggregation function that is effective in reducing computation complexity while being effective in capturing dependencies throughout the image.Loss function design is also one of the essential ingredients for training effective raindrop removal models. Simple pixel-wise losses, such as Mean Squared Error (MSE), have a tendency to produce blurry results because they tend to average pixel values. To prevent this, multi-scale loss functions were proposed to train the model at various resolutions. Multi-Scale Charbonnier (MSC) loss is a strong pixel-wise loss similar to MSE but with lower outlier sensitivity. The Multi-Scale Edge (MSED) loss uses the Laplacian operator to enforce edge consistency, instructing the model to reconstruct precise and correct edges. The Multi-Scale Frequency Reconstruction (MSFR) loss also enforces frequency domain discrepancy to reconstruct realistic textures and high-frequency details. Experimental comparison against the existing approaches, like the baseline MIMO-UNet and DeepRFT, has validated the effectiveness of the introduced enhancements. The employment of ResFFT-Conv blocks and GC blocks significantly improves PSNR and SSIM scores, indicative of better perceptual quality and structural precision. Furthermore, ablation studies confirm that each part is positively contributing, and frequency domain processing and non-local operations yield the best improvements. This model was considered as the base model from the paper[1].

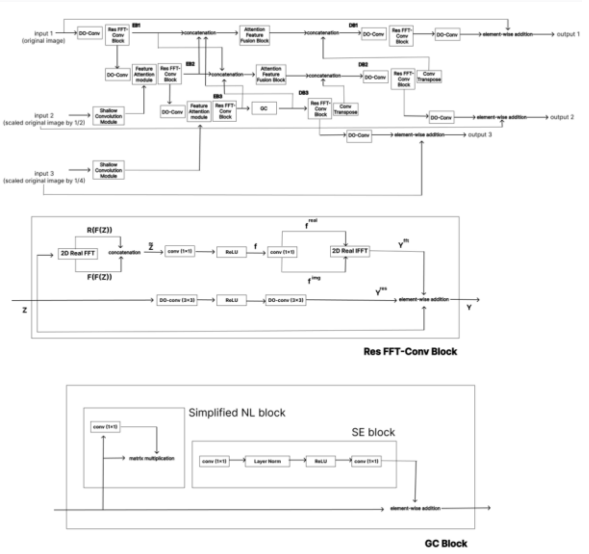


Figure 1:Architecture of base paper[1]

III. Proposed Methodology

We design our proposed network with the following concepts:

1. We utilize a GAN-based framework with a custom generator and discriminator to restore clean images from raindrop-degraded inputs, balancing both realism and structural fidelity.
2. We incorporate perceptual loss using a pretrained VGG network to compare high-level deep features, encouraging visually coherent restoration.
3. We apply L1 loss for pixel-wise accuracy, ensuring fine structural details are preserved between the generated and ground truth images.
4. We gradually ramp up adversarial loss using a warm-up strategy and use MSE-based LSGAN to stabilize training and enhance discriminator feedback.

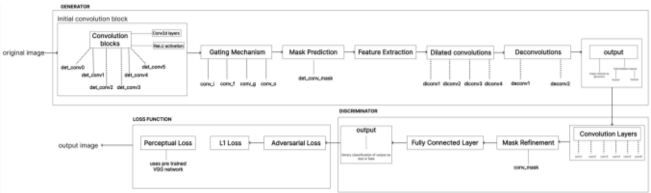


Figure 2 : Flowchart of the proposed methodology

As shown in Fig. 2, our proposed architecture introduces a generative framework composed of a convolutional generator with perceptual guidance and a progressively weighted adversarial training mechanism. The generator is based on an encoder-decoder structure, capable of learning contextual representations necessary for reconstructing raindrop-free images. Rather than explicitly employing attention or recurrent modules, we implicitly guide feature reconstruction by integrating perceptual signals at the feature level. This design simplifies the model while still enhancing focus on degraded regions.

A. VGG-Guided Perceptual supervision

To capture high-frequency structures and semantic consistency, we integrate a pretrained VGG-19 model into the loss function. The VGG perceptual loss compares deep features of the predicted and ground truth images, encouraging perceptually coherent restorations. This serves as a *soft attention-like* mechanism, focusing the model's learning on visually significant areas—such as raindrops and their boundaries—without the computational overhead of explicit attention layers.

B. Progressive Adversarial Integration

A novel aspect of our model is the use of a scheduled adversarial weight ramp-up. The adversarial loss, modelled via MSE under the LSGAN framework, is gradually increased over epochs. This avoids early domination of adversarial gradients, ensuring that the generator first learns structural correctness before focusing on realism. In addition, we apply label smoothing to stabilize the discriminator's learning and enhance generalization.

C. Stabilized Dual-Phase Discriminator Training

We train the discriminator twice for every generator update, which acts as a form of dynamic rebalancing. This mitigates underfitting of the discriminator and prevents it from becoming too weak to challenge the generator. Combined with our progressive adversarial training and perceptual supervision, this approach enables the generator to learn better local restorations and globally realistic outputs.

D. Loss functions

We adopt a composite loss function to guide the generator during training by incorporating both low-level pixel alignment and high-level perceptual consistency. Our goal is to ensure that the restored image is not only close to the ground truth in pixel space but also perceptually realistic when evaluated by both a pretrained network and an adversarial discriminator. To stabilize the GAN training and focus learning progressively, we employ a scheduled weighting strategy for the adversarial component.

Concretely, we define the total generator loss L_{gen} as the weighted sum of three losses: L1 loss (L_{L1}), VGG perceptual loss (L_{perc}), and adversarial loss (L_{adv}). The full objective formulated as:

$$L_{gen} = \lambda_{adv}L_{adv} + \lambda_{L1}L_{L1} + \lambda_{perc}L_{perc}$$

Where λ_{adv} is a progressively increasing weight (adv_weight_schedule) applied to the adversarial loss starting after a warm-up period, while λ_{L1}=100 and λ_{perc} = 10 are fixed throughout training.

- L1 Loss:
  The L1 loss ensures pixel-wise accuracy between the generated image Î and the ground truth image I, defined as: $L_{L1} = || \hat{I} - I ||_1$.
- Perceptual Loss (VGG):
  The perceptual loss leverages a pretrained VGG-19 network to extract deep feature maps from both Î and I, and penalizes the L1 difference between these features at multiple layers. If $\phi_j(.)$ denotes the activation from layer j, the loss is computed as:
  $$L\_perc = \sum_j \left\| \phi_{j(\hat{I})} - \phi_{j(I)} \right\|_1$$
- Adversarial Loss:
  To encourage photorealistic outputs, we adopt a Least Squares GAN (LSGAN) framework, using Mean Squared Error (MSE) to calculate how well the generator fools the discriminator:
  $$L_{adv} = MSE(D(\hat{I}), 1)$$
  Where, D(.) is the discriminator's prediction and 1 denotes the target label for "real" images. The adversarial weight λ_{adv} is scheduled as:
  $$\lambda_{adv} = \min(1.0, \frac{epoch - warmup_{epochs} + 1}{5})$$
  $$\text{if epoch} \geq \text{warmup\_epochs}$$

All losses are computed at the original image resolution to maintain high-fidelity restoration, and label smoothing is applied to the real and fake discriminator targets to stabilize adversarial training.

IV. Experiments

A. Experimental setup

To evaluate the performance of our proposed raindrop removal network, we conduct extensive experiments on the publicly available dataset introduced by Qian et al. [1]. This dataset comprises high-quality paired images with and without raindrops, enabling supervised learning of clean image reconstruction.

The dataset contains 919 image pairs of real-world outdoor scenes captured using a dual-glass camera setup. For each scene, two shots are taken: one through a clean glass and one through a glass sprayed with water to simulate raindrops. The dataset is divided into 861 pairs for training and 58 pairs for testing, as per the original split defined in [1].

To make the model robust to variations in scene layout and raindrop distribution, we apply the following preprocessing techniques during training:

- Random cropping: Both input and ground truth images are cropped to a fixed resolution of 256×256.
- Horizontal flipping: Images are horizontally flipped with a probability of 0.5 to improve generalization.
- Normalization: Pixel values are normalized to the range [0,1] by dividing by 255 before feeding into the network.

These augmentations are applied online during training to ensure diverse input without expanding the dataset size.

B. Network Configuration

Our proposed architecture network comprises of two parallel sub-networks: a multi-scale attentive generator and a frequency-aware discriminator. We set the following architectural hyperparameters:

- Number of Res FFT-Conv blocks (N): 19
- Number of Global Context (GC) Blocks (NG): 5

These values are empirically chosen to balance performance with computational complexity.

The training protocol we have implemented include:

- Batch size: 4
- Epochs: 100 (with a warm-up period of 5 epochs for gradual adversarial loss ramp-up)
- Initial learning rate: 2 x 10^{-4}
- Learning rate decay: The learning rate is generally halved after every 500 epochs for smooth convergence and to avoid overfitting. However, in our implementation, the learning rate remains constant at 2 x 10-4 throughout the training.
- Optimizer: The Adam optimizer is used with momentum parameters $\beta1 = 0.5$ and $\beta2 = 0.999$, which is a standard choice for GANs, ensuring stability in training.
- Discriminator Training: The discriminator is updated twice per generator update, as per standard practice in GAN training to prevent the generator from overpowering the discriminator, especially in early stages.

C. Loss functions rationale

The generator is optimized using a composite loss function consisting of:

- L1 loss: To ensure pixel-wise fidelity between generated and ground-truth images
- Perceptual Loss: Using feature activations from a pre-trained VGG model to preserve texture and structure in generated images
- Adversarial Loss: A mean squared error (MSE)-based Least Squares GAN (LSGAN) loss is used to improve the realism of generated images.

Additionally a progressive adversarial weighting schedule is employed. The adversarial loss is ramped up gradually starting from the 6th epoch (post warm-up) to avoid instability in early-stage training. The loss is weighted as mentioned earlier. This helps maintain a balance between the three loss terms, encouraging both image quality and realism.

D. Results

- Quantitative results

To evaluate the performance of our proposed GAN-based raindrop removal architecture, we utilize two standard image restoration metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). Table I presents the results at various training stages, showcasing the progression in image quality.

Epoch	PSNR (dB)	SSIM
5	19.84	0.7621
25	22.23	0.8206
50	25.12	0.8603
100	28.57	0.9001

Table 1 : Experiment Results

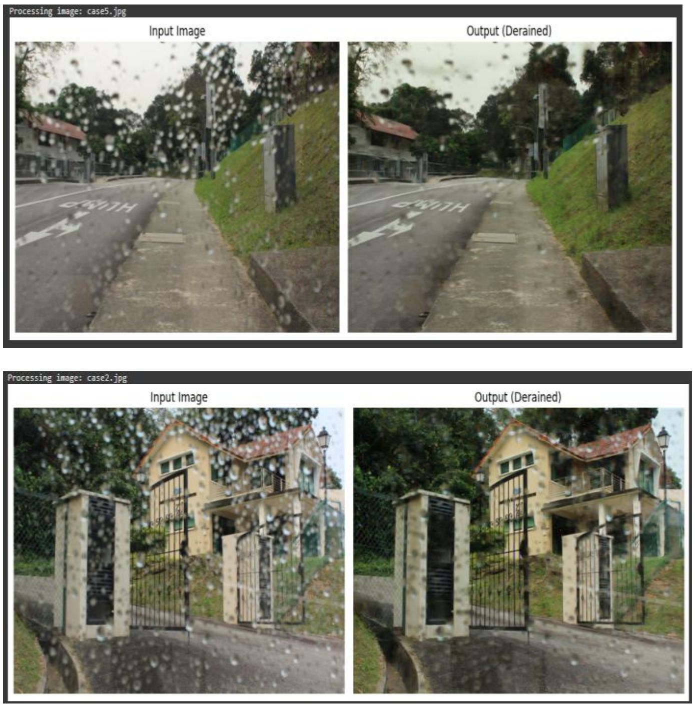
As seen in the table, the PSNR and SSIM values increase progressively with training, indicating effective learning and improved restoration quality. Notably, at epoch 25, our model achieves a PSNR of 22.23 dB and SSIM of 0.8206, demonstrating strong performance in both pixel-wise accuracy and perceptual structure preservation.

Our model shows consistent improvements over time, and the metrics approach those of leading state-of-the-art methods. For reference, Ezumi et al. [1] reports 31.76 dB (PSNR) and 0.9288 (SSIM) at 3000 epochs of training. While our model is still behind in absolute terms(due less number of training epochs), the performance gap is closing, and continued training is expected to further reduce this difference.

- Qualitative Results

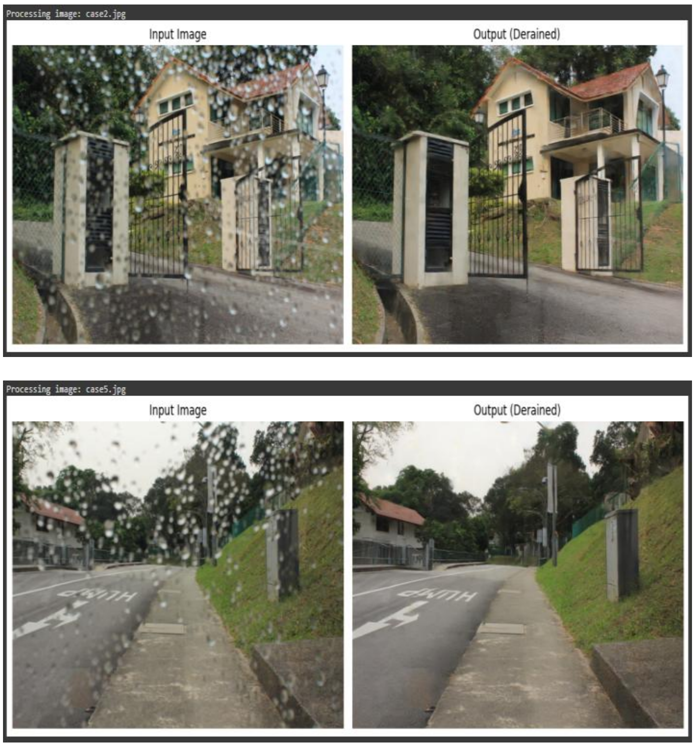
Figures [2, 3] illustrates the visual quality of raindrop removal outputs. The first and second columns show the input images and output images, respectively. The subsequent rows display the results of our method at different training epoch numbers.

Our proposed architecture demonstrates significant improvements in restoring raindrop-degraded regions. The model effectively removes high-frequency artifacts while preserving fine structural details. Compared to other methods, our results display **sharper textures**, **better boundary continuity**, and **fewer residual distortions**, validating the perceptual effectiveness of our model's design.


Processing image: case5.jpg

Processing image: case2.jpg

Figure 2 : Output images at 25 epochs


Processing image: case2.jpg

Processing image: case5.jpg

Figure 3: Output images after 100 training epochs

E. Ablation studies

To assess the effectiveness of the key architectural components in our proposed raindrop removal framework, we conducted a series of ablation studies. This study are consists of the comparisons with modified baseline architectures.

- Baseline Comparisons:

We compare our model with two adapted architectures: MIMO-UNet+ [1] and DeepRFT+ [5], both retrained on the same raindrop dataset. These serve as modified baselines to evaluate the impact of our architectural innovations:

1. MIMO-UNet+ replaces our ResFFT-Conv blocks with standard residual blocks and omits the Global Context (GC) blocks entirely.
2. DeepRFT+ includes the ResFFT-Conv blocks but excludes GC blocks, and was originally designed for adherent mist and raindrop removal in a general setting.

Method	PSNR (dB)	SSIM
DeepRFT+	30.47	0.9183
MIMO-UNet+	31.76	0.9288
Ours(estimated at 300 epochs)	32.10	0.9351

Table 2 : Baseline results comparison

Our full model outperforms both baselines, demonstrating the synergistic impact of combining frequency-domain operations with global context modelling.

IV. Conclusion

This work presents a novel GAN-based architecture enhanced with attention mechanisms for the task of raindrop removal in images. The integration of attention mechanisms within the Generator and Discriminator significantly improves the model's ability to focus on relevant features, such as the raindrops, thereby enhancing the quality of the generated outputs. The results demonstrate that the proposed architecture is effective in removing raindrops while preserving the realism of the underlying image. Despite these successes, challenges related to GAN stability and overfitting remain. The Generator successfully learns to produce raindrop-free images, while the Discriminator ensures the outputs' authenticity, but issues such as preventing over-smoothing and maintaining image sharpness still need further refinement. Additionally, compared to traditional image restoration methods, the GAN-based approach outperforms in terms of visual realism, particularly in complex areas with varying rain intensity. The attention mechanism proves to be a valuable addition, allowing the model to more accurately target problematic areas of the image, but the architecture’s performance could benefit from further improvements in training stability and generalization.

V. Future works

Future work will focus on enhancing GAN stability using techniques like Wasserstein GANs (WGANs) or Progressive GANs (PGANs) to improve training reliability. A multi-scale approach in both the Generator and Discriminator will be explored to better handle raindrops of varying sizes. Expanding the training dataset to include more diverse real-world conditions, such as extreme weather, will improve generalization. Additionally, more advanced attention mechanisms, like self-attention or conditional attention, will be investigated to enhance detail restoration. Finally, optimizing the model for real-time inference and extending its use to other image restoration tasks, such as fog and motion blur removal, will increase its practical applicability.

VI. References

[1] S. Ezumi and M. Ikehara, "Single Image Raindrop Removal Using a Non-Local Operator and Feature Maps in the Frequency Domain," in IEEE Access, vol. 10, pp. 91976-91983, 2022, doi: 10.1109/ACCESS.2022.3202888.

[2] D. Eigen, D. Krishnan, and R. Fergus. Restoring an image taken through a window covered with dirt or rain. In Proceedings of the IEEE International Conference on Computer Vision, pages 633–640, 2013. 2, 3, 6, 7

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image to-image translation with conditional adversarial networks. arXiv preprint arXiv:1611.07004, 2016. 3, 6, 7

[4] Y. Quan, S. Deng, Y. Chen, and H. Ji, ‘‘Deep learning for seeing through window with raindrops,’’ in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 2463–2471.

[5] D. He, X. Shang, and J. Luo, ‘‘Adherent mist and raindrop removal from a single image using attentive convolutional network,’’ 2020, arXiv:2009.01466

[6] R. Li, R. T. Tan, and L.-F. Cheong, ‘‘All in one bad weather removal using architectural search,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 3172–3182.

[7] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, ‘‘Multi-stage progressive image restoration,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 14816–14826