Synthesizing and Restoring Weather-corrupted Images with Conditional Diffusion Models

Young-Ho Go and Sung-Hak Lee
Kyungpook National University, Korea
E-mail: gyh7454@knu.ac.kr and shak2@ee.knu.ac.kr

***Abstract*— AI-based image processing plays a pivotal role in sensor fusion and image transformation for future mobility systems. However, real-world images captured by autonomous vehicles and surveillance cameras are highly susceptible to weather-induced degradation. In particular, raindrops adhering to camera lenses significantly impair object recognition and pose safety concerns.**

This paper proposes a diffusion-based framework for raindrop removal. While denoising diffusion probabilistic models (DDPM) enable stable training and high-resolution restoration, they require large-scale paired datasets, which are difficult to obtain in raindrop scenarios. To overcome this challenge, the proposed method leverages the conditional generation capability of the Palette diffusion model to synthesize a diverse dataset of paired clean–raindrop images.

Raindrop masks are extracted from input images and used to train an inpainting-based generator. During inference, synthetic masks simulate various raindrop patterns. The restoration network repurposes the Palette architecture for image-to-image translation, with customized loss functions and normalization strategies to enhance performance. A post-processing module further improves visual consistency through tone adjustment and chroma compensation.

Experimental results demonstrate that the proposed framework effectively preserves fine structural details and achieves superior visual quality compared to existing approaches. Moreover, the method improves the robustness of vision systems in adverse weather conditions.

I. INTRODUCTION

AI-based vision systems are critical to future mobility applications but remain highly susceptible to visual degradation under adverse weather conditions. As shown in Fig. 1, raindrops on camera lenses can distort visual input and hinder reliable scene understanding. Traditional methods for raindrop removal struggle to reconstruct the complex distortions introduced by water droplets. Recently, the denoising diffusion probabilistic model (DDPM) [1] has emerged as a more stable and accurate alternative to the generative adversarial network (GAN) [2]. DDPM achieves high-quality reconstructions by learning to reverse a noise-adding process, often using a U-Net-based architecture [3], and has shown promising results in tasks such as inpainting and image translation.

However, its application to raindrop removal is limited by the scarcity of large-scale paired datasets, which are difficult to collect due to scene misalignment and lighting variations. GAN-based methods also suffer from mode collapse and artifact generation.

To overcome these limitations, this study proposes a two-stage framework consisting of the Make-Raindrop Diffusion (MRD) module and the Removal-Raindrop Diffusion (RRD) module. The MRD module is designed to generate diverse synthetic clean–raindrop image pairs through conditional inpainting based on the Palette diffusion model [4]. The RRD module is then trained using both synthetic and real-world data to restore clean images from raindrop-contaminated inputs.

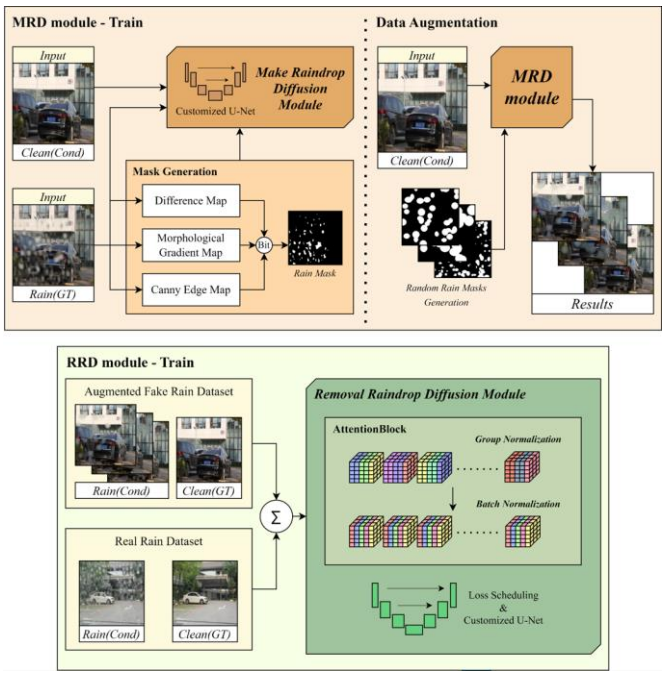
To further improve performance, the framework incorporates Smooth L1 loss, Batch Normalization, and DDIM-based sampling [5], resulting in enhanced training stability and faster inference. This structure enables accurate data generation and high-quality image restoration under challenging weather conditions.

A post-processing module further refines luminance using a Mutual Image Translation Module (MITM) [6], followed by luminance blending and chroma compensation to improve tone balance and visual coherence. Experimental results demonstrate that the proposed method effectively removes raindrop artifacts and restores high-fidelity structural details under real-world conditions.

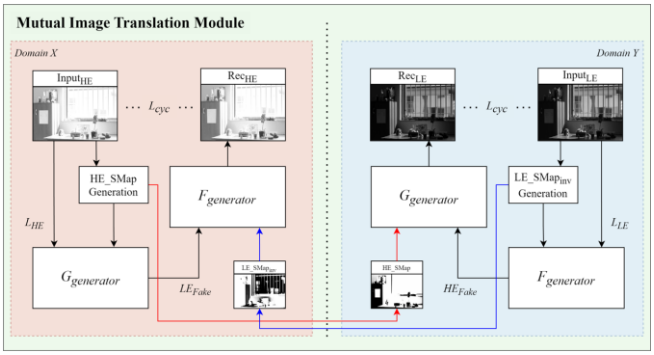
(a) (b)

Fig. 1. Comparison of image degradation caused by raindrops: (a) Raindrop image, (b) Clean Image

979-8-3315-7206-8/25/$31.00 ©2025 IEEE



(a)



(b)

Fig. 2. Overall flowchart of the proposed system: (a) MRD–RRD training, (b) MITM training.

II. METHODOLOGY

This study proposes a two-stage framework to address visual degradation caused by raindrops adhering to camera lenses in adverse weather conditions. As illustrated in Fig. 2(a), the framework consists of two main modules: the Make-Raindrop Diffusion (MRD) module and the Removal-Raindrop Diffusion (RRD) module. In addition, Fig. 2(b) depicts the training pipeline for the Mutual Image Translation Module (MITM), which serves as a post-processing component.

The MRD module is responsible for generating accurate raindrop masks during training. These masks are derived by analyzing localized degradation patterns in ground truth raindrop images using difference maps and edge detection techniques. The resulting masks are used as conditional inputs to simulate diverse raindrop patterns during inference.

The RRD module performs the core task of raindrop removal. It receives a raindrop-contaminated image as input and reconstructs a clean version using a dedicated restoration network trained to suppress occlusion-induced artifacts.

The MITM module is trained on paired high- and low-exposure images to learn exposure-dependent feature alignment. It receives a switching map as an additional input to a CycleGAN-based generator [7], enabling the correction of luminance imbalances during inference by injecting restored content into underexposed regions.

This overall architecture enables the framework to both generate synthetic raindrop data for augmentation and perform accurate restoration with high perceptual fidelity.

A. *Palette Diffusion-Based Conditional Synthesis*

To support both raindrop synthesis and removal, the proposed framework adopts a dual-module structure based on the Palette diffusion model. This unified design enables flexible handling of two complementary tasks: generating realistic raindrop artifacts and restoring clean images.

While the original DDPM excels at generating high-quality images from noise, it lacks task specificity without additional input. To address this, the Palette model is trained to reconstruct a ground truth (GT) image using both a conditional input and its corresponding GT, allowing explicit control over the generation process.

The model employs a U-Net–based encoder–decoder architecture that integrates multi-scale features, supporting stable and accurate reconstruction across varying spatial resolutions.

B. *Make-Raindrop Diffusion Module*

The MRD module is designed to generate high-precision raindrop masks by analyzing differences between clean and raindrop-contaminated images. The mask generation process consists of three main stages: initial mask estimation, boundary-aware refinement, and morphological post-processing.

In the first stage, a pixel-wise difference map is computed between the clean (Cond) and raindrop (GT) images. Since raindrop-affected regions exhibit large intensity deviations, this difference map serves as the initial indicator of degradation. However, it often contains irrelevant details such as fine object textures or edge noise, making it unsuitable for direct mask construction.

To address this issue, an auxiliary edge map is generated to capture prominent boundaries in the image. A bitwise AND operation between the edge map and the initial difference map suppresses false positives along object boundaries, enabling more accurate localization of raindrop regions.

Fig. 3. Generation of synthetic rain images conditioned on input.

Next, morphological gradient operations are applied to extract structural edge features. Both the gradient edge map and the Canny edge map are inverted using a bitwise NOT operation and then fused with the difference map using another bitwise AND. This step removes residual object contours and isolates components specific to raindrops.

In the final stage, morphological closing (dilation followed by erosion) and opening (erosion followed by dilation) are applied sequentially to eliminate small noise artifacts and fill holes within the raindrop regions. A Gaussian blur is then applied to smooth the mask boundaries and compensate for potential loss of size caused by earlier processing.

The resulting refined mask is used to selectively inject noise into masked regions of clean images during training. The DDPM is then trained to predict this injected noise and reconstruct the corresponding raindrop appearance, effectively learning to synthesize realistic degradations in a conditional inpainting setting.

C. Data Augmentation

During the inference phase of the MRD module, data augmentation is performed by generating synthetic raindrop images using only clean (Cond) inputs. Unlike the training phase, which requires GT images, this process is designed to produce diverse clean–raindrop pairs to support effective dataset expansion.

Random coordinates are sampled across the image, and at each location, synthetic raindrop instances are generated with randomly sampled radii and shapes.

Although circular droplets are used by default, geometric distortions such as elliptical and irregular forms are also introduced to simulate realistic variations in raindrop morphology.

To ensure visual diversity, the generated masks are uniformly distributed across the image plane. The total number and overall area of raindrops are stochastically varied to reflect natural randomness. As illustrated in Fig. 3, this strategy allows a single clean image to be augmented into multiple synthetic variants with different droplet configurations.

This augmentation approach addresses the limitations of real-world datasets, particularly the difficulty of capturing spatially aligned clean–raindrop image pairs. It also enables the inclusion of challenging cases, such as large or deformed droplets, that are often underrepresented in real-world data. By synthesizing a wide variety of raindrop patterns, the proposed framework significantly enhances the generalization capability and robustness of the downstream raindrop removal model.

D. Removal-Raindrop Diffusion Module

The RRD module is constructed on top of the Palette diffusion model and reconfigured to perform image-to-image translation for raindrop removal. Several key architectural modifications are introduced to enhance reconstruction quality, color consistency, and inference efficiency.

First, the original loss function is replaced with the Smooth L1 loss to improve stability during training. As illustrated in Fig. 4, the Smooth L1 loss combines the advantages of MSE and mean absolute error (MAE), exhibiting reduced sensitivity to large outliers while ensuring stable convergence. This property is particularly beneficial for recovering high-frequency structural details in regions affected by occlusions.

Second, the normalization strategy is updated from Group Normalization to Batch Normalization. While Group Normalization is well suited for gray-to-color image translation tasks due to its independence from batch size, it normalizes features within grouped channels, which may limit inter-channel interactions. In contrast, Batch Normalization shares channel-wise statistics across the entire batch, thus better preserving RGB consistency and improving feature coherence.

Third, to accelerate inference, the RRD module adopts DDIM as its sampling strategy in place of standard DDPM sampling. The DDIM formulation reduces the number of iterative steps required during sampling while maintaining comparable reconstruction fidelity. DDIM uses a deterministic, non-Markovian reverse path that removes stochastic noise and directly estimates the previous state by moving toward a clean image estimate, enabling larger timestep strides (about 20 to 50 steps) with minimal quality loss. This modification is particularly important for real-time or resource-constrained applications.

The RRD module is trained using a large set of synthetic clean–raindrop image pairs generated by the MRD module. With the above architectural enhancements, the proposed framework achieves superior performance in edge preservation, color fidelity, and training robustness compared to previous raindrop removal approaches.

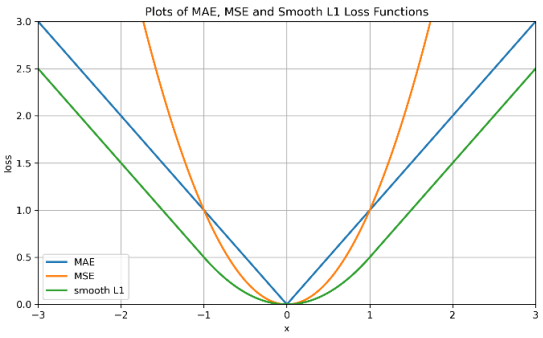


Fig. 4. Graphs of MAE, MSE, Smooth L1 Loss

E. Image Blending

To enhance the perceptual quality of restored images, the luminance channel (L) of the RRD output undergoes tone correction using MTIM. This module is trained to recover details in underexposed regions and improves object visibility under challenging lighting conditions. However, this tone enhancement process may lead to over-saturation in bright areas, resulting in loss of visual balance. To mitigate this effect, a luminance-domain blending strategy is employed:

$$L_{fused} = L_{surr} \cdot L_{RR} + (1 - L_{surr}) \cdot L_{HDR} \quad (1)$$

In this equation, L_{RR} denotes the luminance channel of the image restored by the RRD module, while L_{HDR} represents the tone-enhanced luminance obtained via MITM.

The term L_{surr} refers to a surround map derived from L_{HDR}. This blending formulation maintains a balanced representation of both dark and bright regions, thereby enhancing the overall perceptual quality.

As blending is applied only to the luminance channel, directly recombining it with the original a and b chroma channels can cause luminance–chroma mismatch, often resulting in desaturation artifacts. To address this, a chroma compensation step is introduced to restore color consistency while preserving hue.

A naive scaling of the a and b channels based on the luminance ratio can exaggerate chroma in darker regions and introduce unnatural color shifts. Instead, a hue-preserving chroma scaling method is applied. The chroma magnitude is first computed as:

$$C = \sqrt{(a - 128)^2 + (b - 128)^2} \quad (2)$$

$$L_{ratio} = \frac{L_{fused}}{L_{RR}}, \qquad C_{new} = C * L_{ratio} \quad (3)$$

Here, C is the chroma magnitude of the original image, and L_{ratio} quantifies the luminance change introduced by blending. This ratio is then used to selectively scale the chroma magnitude while preserving the original hue, thereby preventing color distortion and desaturation artifacts.

(a) (b) (c) (d) (e)

Fig. 5. Qualitative comparison of raindrop removal results: (a) Input raindrop image, (b) Ground truth clean image, (c) YUK, (d) TUM, (e) Proposed final output.

(a) (b) (c) (d) (e)

Fig. 6. Qualitative comparison of raindrop removal results: (a) Input raindrop image, (b) Ground truth clean image, (c) YUK, (d) TUM, (e) Proposed final output.

The compensated a and b channels are then computed by applying the ratio C_{new} to the original chroma vectors:

$$a_{comp} = a * C_{new} + 128, \quad b_{comp} = b * C_{new} + 128 \quad (4)$$

$$I_{Blend} = Lab2BGR(L_{blend}, a_{comp}, b_{comp}) \quad (5)$$

This transformation selectively amplifies or suppresses chroma based on luminance change while maintaining the original hue orientation. Finally, I_{Blend} is reconstructed by converting the modified Lab channels back into RGB space. Equation (5) yields the final output image with enhanced tonal balance and preserved chromatic integrity, completing the post-processing stage.

III. Simulation Results

The proposed framework was trained using approximately 10,000 clean–raindrop image pairs. Among them, about 1,000 samples were collected from real-world scenes to construct a real raindrop dataset, while the remaining 9,000 pairs were synthetically generated using the MRD module. Both real and synthetic data were jointly utilized to train the RRD module for robust performance across diverse conditions.

To evaluate raindrop removal performance, qualitative comparisons were conducted against existing baseline methods. As illustrated in Figs. 5 and 6, each example includes: (a) a raindrop-contaminated input image, (b) the GT clean image, (c) the output of YUK [8], (d) the output of TUM [9], and (e) the final output generated by the proposed framework incorporating both DDIM sampling and the blending-based post-processing module.

In Fig. 5, the proposed method achieves enhanced visual clarity and demonstrates robustness to variations in raindrop size, position, and transparency. The framework effectively eliminates raindrop artifacts while preserving overall structural consistency across the image.

Fig. 6 further highlights the model's ability to maintain fine texture details and produce natural color representation. Notably, improvements in darker regions contribute significantly to the overall perceptual quality. Compared to prior methods, the proposed approach delivers superior restoration accuracy and aesthetic quality under both subtle and severe degradation scenarios.

IV. Conclusions

This paper presents a diffusion-based framework for synthesizing and removing raindrop artifacts in images captured under adverse weather conditions. The proposed method consists of two key components: the Make-Raindrop Diffusion (MRD) module, which generates diverse synthetic clean–raindrop image pairs, and the Removal-Raindrop Diffusion (RRD) module, which restores clean images using both real and synthetic training data.

To enhance both reconstruction quality and inference efficiency, the RRD module incorporates Smooth L1 loss, Batch Normalization, attention reconfiguration, and a DDIM-based sampling strategy. In addition, a luminance-based image blending and chroma compensation module is introduced to improve perceptual quality and color consistency in the final output.

Extensive experiments confirm that the proposed framework outperforms existing raindrop removal methods in terms of structural preservation, visual fidelity, and generalization across varied raindrop patterns. The model demonstrates robustness against changes in shape, density, and transparency, producing visually pleasing and accurate restorations under real-world conditions.

Future work will focus on designing lightweight architectures for real-time video applications and extending the framework to handle additional weather-related degradations, such as fog, snow, and dust. Further research will also explore multi-domain learning and physics-based data generation techniques to enhance realism and scalability in adverse environmental conditions.

V. ACKNOWLEDGMENT

This research was supported by Korea Creative Content Agency(KOCCA) grant funded by the Ministry of Culture, Sports and Tourism(MCST) in 2024(Project Name: Development of optical technology and sharing platform technology to acquire digital cultural heritage for high quality restoration of composite materials cultural heritage, Project Number: RS-2024-00442410, Contribution Rate: 50%) and the Institute of Information & Communications Technology Planning & Evaluation(IITP)-Innovative Human Resource Development for Local Intellectualization program grant funded by the Korea government(MSIT)(IITP-2025-RS-2022-00156389, 50%).

REFERENCES

[1] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 6840–6851, Jun. 2020, doi: https://doi.org/10.48550/arXiv.2006.11239.

[2] I. J. Goodfellow *et al.*, "Generative Adversarial Networks," *Adv. Neural Inf. Process. Syst.*, vol. 27, Jun. 2014, doi: 10.5555/2969033.2969125.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer, 2015, pp. 234–241.

[4] C. Saharia *et al.*, "Palette: Image-to-Image Diffusion Models," in *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings*, New York, NY, USA: ACM, Aug. 2022, pp. 1–10. doi: 10.1145/3528233.3530757.

[5] J. Song, C. Meng, and S. Ermon, "Denoising Diffusion Implicit Models," *arXiv Prepr. arXiv2010.02502*, Oct. 2020, [Online]. Available: http://arxiv.org/abs/2010.02502

[6] Y.-H. Go, S.-H. Lee, and S.-H. Lee, "Multiexposed Image-Fusion Strategy Using Mutual Image Translation Learning with Multiscale Surround Switching Maps," *Mathematics*, vol. 12, no. 20, p. 3244, Oct. 2024, doi: 10.3390/math12203244.

[7] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2017, pp. 2242–2251. doi: 10.1109/ICCV.2017.244.

[8] Y.-K. Han, S.-W. Jung, H.-J. Kwon, and S.-H. Lee, "Rainwater-Removal Image Conversion Learning with Training Pair Augmentation," *Entropy*, vol. 25, no. 1, p. 118, Jan. 2023, doi: 10.3390/e25010118.

[9] W.-T. Chen, Z.-K. Huang, C.-C. Tsai, H.-H. Yang, J.-J. Ding, and S.-Y. Kuo, "Learning Multiple Adverse Weather Removal via Two-stage Knowledge Learning and Multi-contrastive Regularization: Toward a Unified Model," in *2022 IEEE/CVF Conference on Computer Vision and Pattern*