Rain drop removal using GAN

TOC

Problem Statement

Outdoor image captures in uneven weather conditions or other external factors may allow water droplets to collect on the camera lens. This can degrade the captured images by introducing artifacts into the image or by a blurring effect due to refraction of light inside the water droplets. These degradations can obscure the details in the image and can hinder its usefulness in machine vision applications by reducing the processing downstream for machine vision tasks.

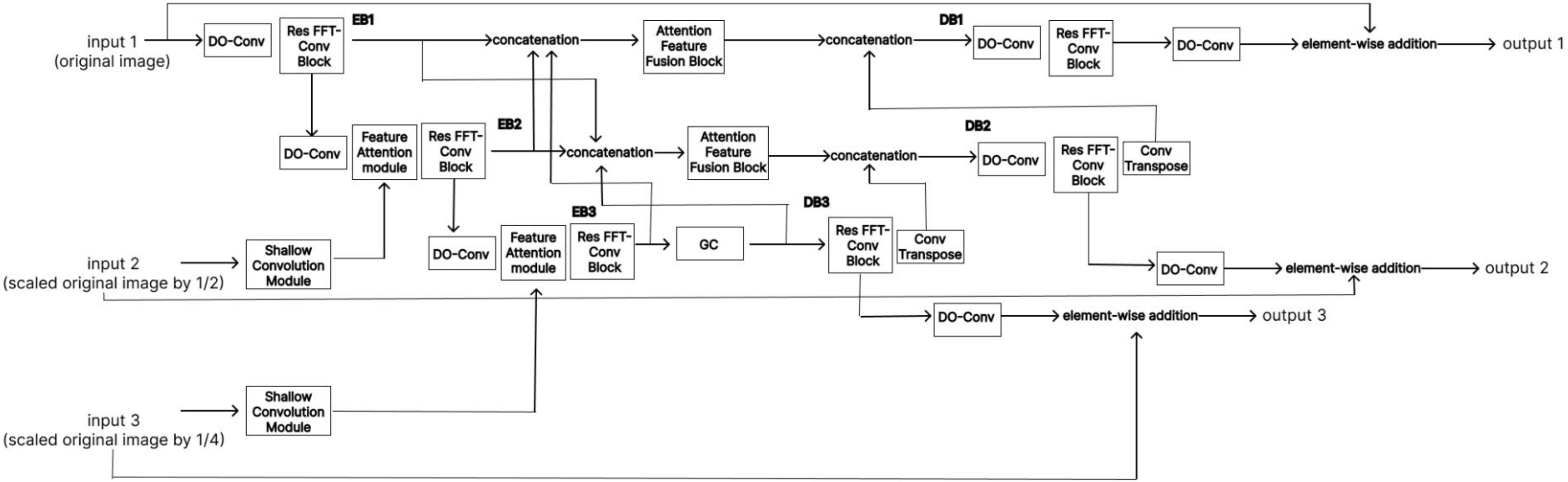
Problems to solve

1. Raindrops exhibit diverse shapes, sizes, and transparency levels, making them difficult to model
2. It must be ensured that the removal process maintains fine-grained details and textures of the underlying scene
3. Rain-removal algorithm must analyze pixel associations to remove blurring while preserving underlying scene information
4. Droplets of varying sizes can produce different degrees of blurring effect

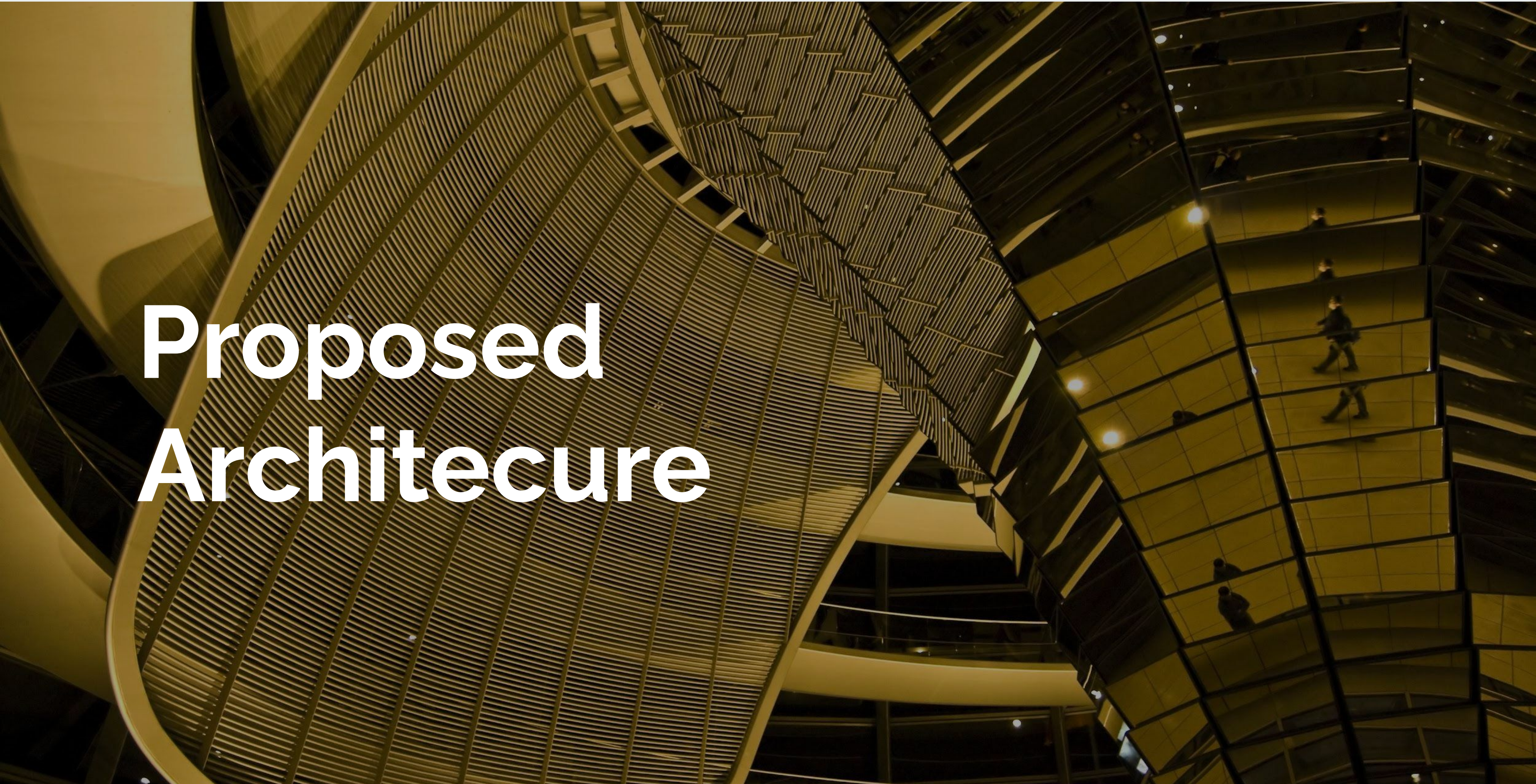
Project objective

The project proposes a GAN based model with attentive masks implemented in Generator and Discriminator to remove water droplets from images

Architecture of the Base Paper



Proposed Architecure



Architecture overview

01

Input Image

Serves as the raw input for the model. Used to predict both the clean background and a rain mask.

02

Generator

Takes the rainy image + initial mask (default: 0.5 values). Iterative mask refinement (LSTM-inspired updates). Multi-stage outputs: Frame1, Frame2, Final clean image. Uses encoder-dilated-decoder structure with skip connections.

03

Discriminator

Receives real or generated (fake) clean image. Uses multiple convolutional layers to extract features. Applies a learned mask to focus on important regions.

04

Loss functions

For generator, $\mathcal{L}$rec: L1 loss between output & ground truth. $\mathcal{L}$adv: Adversarial loss from discriminator. For discriminator, Binary cross-entropy to distinguish real vs. generated

05

Output

Final clean background image with rain streaks removed. High detail preserved. Consistency and realism enforced by adversarial training

Generator

The generator is designed to remove rain streaks from the input image by predicting a clean background and an accurate rain mask. It utilizes a multi-stage refinement mechanism to progressively improve the output quality.

- Takes a rainy image and an initial uniform mask as input.
- Employs an encoder–dilated convolution–decoder structure for rich feature extraction.
- Incorporates skip connections for better spatial information preservation.
- Refines output in three stages: Frame1, Frame2, and Final output.
- Learns and updates the rain mask iteratively using a memory-inspired feedback mechanism.

Discriminator

The discriminator is a convolutional neural network trained to differentiate between real clean images and generator outputs. It plays a critical adversarial role by guiding the generator to produce more realistic, rain-free images.

- Takes both real and generated images as input.
- Employs multiple convolutional layers with ReLU activation for deep feature extraction.
- Outputs two components: a learned attention mask and a classification score (real or fake).
- The mask highlights regions where rain artifacts are likely to remain, aiding the generator’s refinement.
- Includes downsampling and fully connected layers to ensure global and contextual discrimination.

Loss Functions

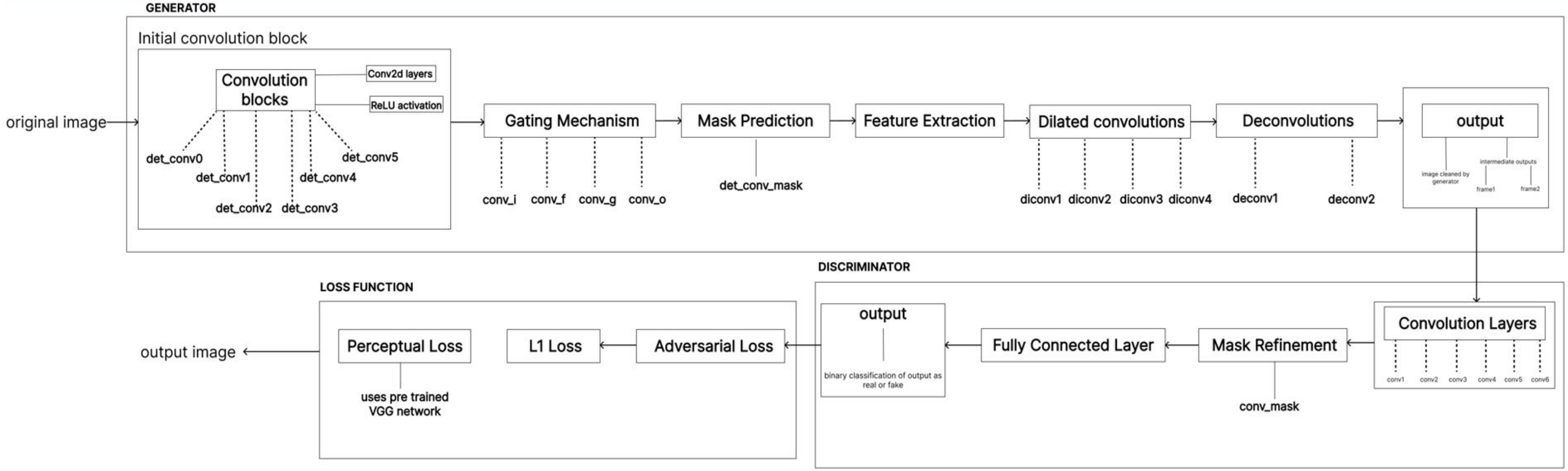
The training process employs a hybrid loss function that ensures both visual fidelity and adversarial robustness. This includes:

- **Adversarial Loss**: Encourages the generator to produce outputs indistinguishable from real clean images.
- **Content Loss (L1 Loss)**: Measures pixel-wise difference between generated and ground-truth images for accurate restoration.
- **Perceptual Loss**: Uses feature maps from a pretrained VGG network to ensure high-level similarity.
- **Mask-Guided Loss**: Focuses training on rain-affected regions as highlighted by the discriminator's attention mask.

These combined losses ensure the generator produces sharp, realistic, and rain-free images without artifacts.

Flowchart

This flowchart highlights the various components comprising the generator and the discriminator along with the three loss functions and how their are sequenced during a single loop



Results

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.

Output images after 100 training epochs

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).

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

Comparisons

Our proposed model outperforms both DeepRFT+ and MIMO-UNet+ baselines. The combination of frequency-domain operations with global context modeling significantly enhances image restoration quality.

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

Conclusion

This work introduces a GAN-based architecture with integrated attention mechanisms for effective raindrop removal. The attention modules enhance focus on rain-affected regions, leading to more realistic and visually consistent outputs. Compared to traditional methods, the model shows superior performance, particularly in complex scenes. However, challenges like GAN stability, overfitting, and occasional over-smoothing remain.

Future Works

- Integration of WGAN or PGAN for more reliable training.
- Enhancing Generator/Discriminator to better handle varied raindrop sizes.
- Including diverse real-world scenarios (e.g., extreme weather).
- Exploring self-attention and conditional attention for finer detail recovery.
- Optimizing for real-time inference and extend to fog/motion blur removal.

Thank you.

Varshaa Puvvada Srinath
Bhuvana Prabha B
Vatsal Ojha