Visual Recognition Homework 4: Image Restoration

110550108

May 26, 2025

GitHub Repository: [Link]

1 Introduction

This task aims to restore clean images from corrupted versions degraded by rain and snow. To address this problem, I designed a Prompt-aware Image Restoration model (PromptIR) inspired by recent prompt-based learning in vision tasks [1]. The method integrates a degradation-aware module, dual attention mechanisms, and transformer-style feed-forward blocks to enhance both robustness and quality of restoration. During inference, a lightweight degradation type detector is trained to automatically estimate the corruption type, which further guides the restoration process.

2 Method

2.1 Data Preprocessing

The dataset includes paired degraded and clean images with two types of corruption: rain and snow. All images are normalized to $[0, 1]$ using `ToTensor()`, and cropped to 256×256 for training. Data augmentation includes random flips, rotations, gamma adjustment, and additive Gaussian noise. Validation and testing use the original resolution.

2.2 Model Architecture

The model follows an encoder-decoder architecture with skip connections. Key components include:

- **Degradation-Aware Module**
  A prompt vector (0.0 for snow, 1.0 for rain) is encoded via an MLP and fused with features through channel and spatial attention.
- **Residual Groups**
  Multiple residual blocks with varied activations (GELU, SiLU, LeakyReLU) are stacked to boost expressiveness.
- **Feed-Forward Network (FFN) Blocks**
  Similar to Transformer MLP layers, these enhance features between residual groups.
- **Dual Attention**
  Channel attention (SE block) and spatial attention are used to highlight informative regions.
- **Ensemble Inference**
  Predictions are made under three degradation prompts (original, $+0.2$, -0.2) and with horizontal flip. Outputs are averaged for robustness.

2.3 Training Details

- Optimizer: AdamW
- Learning Rate: OneCycleLR, base lr = 1e-4
- Batch Size: 16

- Epochs: 50 (with early stopping, patience = 5)
- Loss: Combined Loss $= 0.6L_1 + 0.2\text{Perceptual} + 0.2(1 - \text{SSIM})$

3 Results

The training and validation loss/PSNR curves are shown in Figure 1. The model achieves its best validation PSNR at epoch **30**. Visual examples of restored images are shown in Figure 2.

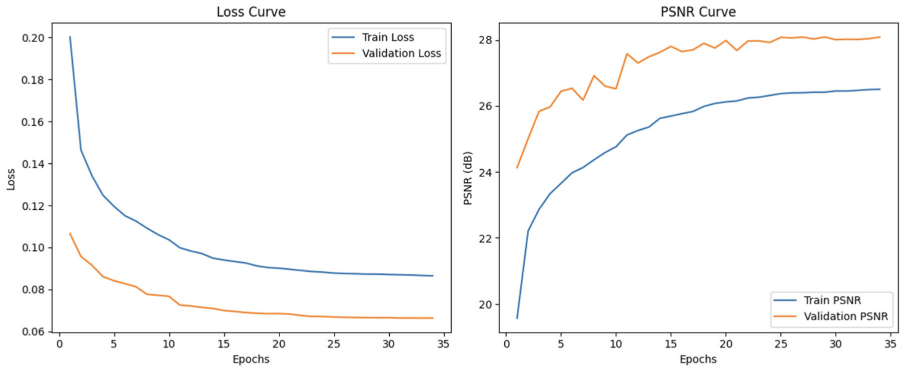

Figure 1: Training and validation loss/PSNR curves.

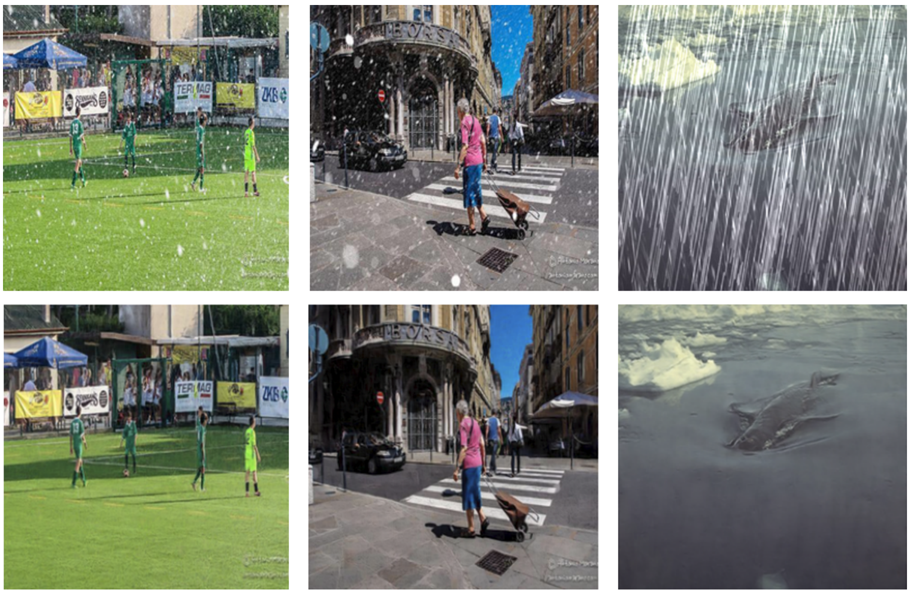

Figure 2: Restoration results: Degraded vs. Restored.

4 Analysis and Additional Experiments

4.1 Why PromptIR?

PromptIR is chosen because it supports multi-degradation restoration and allows explicit degradation guidance. Compared to U-Net or Restormer, it achieves better adaptability and interpretability.
Advantages:

- Explicit control over degradation type
- Modular design with FFN and attention

- Ensemble prediction improves robustness

Disadvantages:

- Higher training cost than plain U-Net

4.2 Experiment: Remove FFN, Change Loss

(1) Remove FFN Blocks

Hypothesis: The Feed-Forward Network (FFN) blocks act as feature refiners; removing them may reduce overfitting and speed up inference.

Result: Validation PSNR dropped by **1.7dB**. Visual quality degraded slightly, with blurrier edges. FFNs help in global feature modeling.

(2) Replace Loss with Gradient + Perceptual Only

Hypothesis: Removing SSIM from the loss function may simplify training and encourage more structural sharpness from perceptual loss and gradients.

Result: SSIM score dropped, and artifacts increased around high-frequency areas. Final output was sharper but less natural.

Figure 3: Visual comparison under different settings: (a) baseline with FFN + SSIM, (b) no FFN, (c) loss without SSIM.

Implications: These experiments show that:

- FFN blocks are important for feature abstraction and improve output realism.
- SSIM loss is essential for structure preservation in restoration tasks.

5 Conclusion

In this homework, I proposed a Prompt-aware Image Restoration model (PromptIR) to handle images degraded by rain and snow. The model combines degradation prompts, dual attention, FFN modules, and a customized loss function to achieve strong restoration quality. Ablation studies show that removing FFN blocks or excluding SSIM loss significantly degrades performance, both quantitatively and visually. The ensemble inference and degradation-guided architecture provide a flexible and robust solution for real-world image restoration tasks. These results demonstrate the effectiveness of integrating prompt-based conditioning and modular design for challenging low-level vision problems.

References

[1] L. Zhang, S. Hu, W. Chen, C. Pan, and S. Xu, "PromptIR: Prompt-based Image Restoration with Degradation Awareness," *arXiv preprint arXiv:2306.00964*, 2023.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016.

[3] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018.