Visual Recognition using Deep Learning HW4 Report

110550093 蔡師睿

1 Introduction

This task addresses image restoration under adverse weather conditions, specifically rain and snow. The training dataset comprises 1,600 degraded images for each condition (3,200 images in total), and the test set includes 100 degraded samples. The constraint is that only the PromptIR[1] model may be used, and all training must be trained from scratch. Our goal is to produce high-quality restorations, with Peak Signal-to-Noise Ratio (PSNR) serving as the evaluation criterion. The implementation can be found at *https://github.com/Ray-1026/Visual-Recognition/tree/main/HW4*.

2 Method

2.1 PromptIR: Prompting for All-in-One Blind Image Restoration

PromptIR[1] is a unified image-resotration framework that treats each corruption as a ”task prompt,” which borrows the prompting idea from LLM. As shown in the Fig. 1, it uses a single U-Net-like network[2] with transformer blocks[3] in the encoding and decoding stages. This framework contains two primary modules, Prompt Generation Module (PGM) and the Prompt Interaction Module (PIM).

Prompt Generation Module (PGM) transforms static prompt components into input-conditioned prompts by generating attention-based weights from the input features. These weights are then used to reweight the prompt components. This design enables the model to adaptively encode degradation-specific information while maintaining flexibility and parameter efficiency.

Prompt Interaction Module (PIM) concatenates the prompts with input features along the channel dimension and processes them through a Transformer block[3] to inject degradation-specific guidance. This plug-and-play block, composed of a Multi-Dconv Head Transposed Attention (MDTA) and a Gated-Dconv Feedforward Network (GDFN), enables efficient channel-wise attention and selective feature refinement.

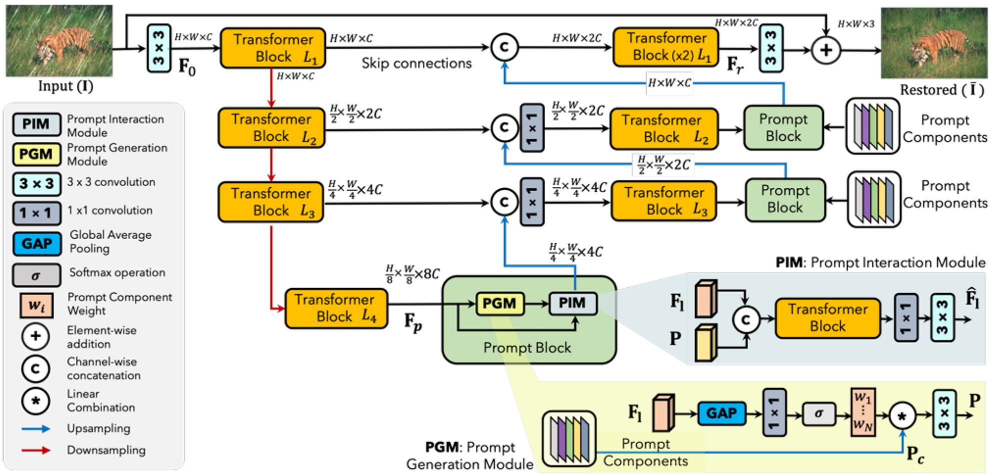


Figure 1: Overview of the PromptIR approach.

2.2 Modification

Charbonnier loss[4] is a differentiable variant of the L1 loss, often described as a smooth L1 loss. The loss is designed to be robust to outliers because it behaves like L1 loss for large errors and like L2 near zero. Its advantages include robustness to outliers, smooth gradient behavior, and widespread use in image restoration tasks. Given a prediction $\hat{y}$ and ground truth y, the Charbonnier loss is defined as:

$$\mathcal{L}_{\text{Charbonnier}}(\hat{y}, y) = \sqrt{(\hat{y} - y)^2 + \epsilon^2}$$

, where ϵ is a small constant (set to 10^{-6} here) to ensure numerical stability and to make the function smooth at zero.

SSIM loss[5] is derived from the Structural Similarity Index Measure (SSIM) [6], a perceptual metric that evaluates the similarity between two images. In training deep learning models, SSIM loss is used to maximize structural similarity between the predicted and ground-truth images by minimizing:

$$\mathcal{L}_{\text{SSIM}} = 1 - \text{SSIM}(\hat{y}, y)$$

, where SSIM($\cdot$) values range from -1 to 1, with 1 indicating perfect similarity. By minimizing SSIM loss, the model is guided to produce outputs that are perceptually closer to the ground truth, particularly in terms of structural fidelity.

Total loss. The original PromptIR[1] framework employs only L1 loss. To improve PSNR on deraining and desnowing tasks, I modified the loss function to combine Charbonnier loss and SSIM loss. As a result, the total loss function are defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Charbonnier}} + \lambda_s \mathcal{L}_{\text{SSIM}}$$

, where λ_s denotes the weighting factor for the SSIM loss, which is set to 0.84 in this work following the recommendation in [5].

Train validation split. The dataset is only divided into training and testing sets, and thus I split 10% of the training dataset as a validation set to monitor the training process. After tuning all hyperparameters and determining that 150 epochs yielded satisfactory results, the validation set is merged back into the training set, and the model is retrained using the full training data.

2.3 Hyperparameters

- epochs: 150
- batch size: 4
- AdamW:
  - learning rate: $2e{-}4$
  - weight_decay: $1e{-}2$
- Linear Warmup Cosine Annealing Schedule:
  - warmup_epochs: 15
  - max_epochs: 150
  - eta_min: 0

3 Results

3.1 Training curve

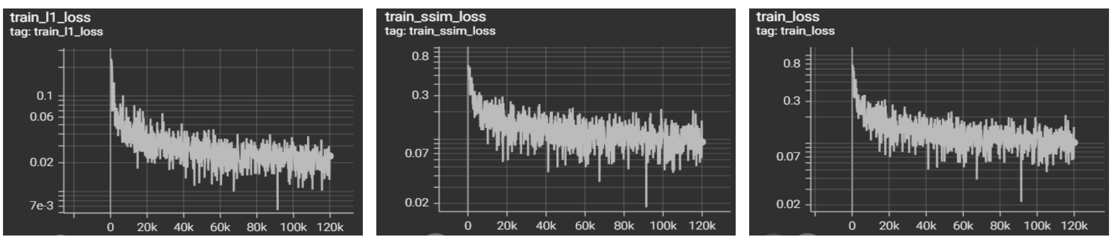


Figure 2: Charbonnier loss (left), SSIM loss (mid), total loss (right).

3.2 Visual results

Figure 3: Visual results in derain and desnow tasks.

3.3 Public score

110550093	1	2025-05-28 00:50	299410	110550093	30.77

Figure 4: Public score on Codabench

4 Additional Experiments - Ablation Studies

4.1 Ours v.s. default settings

We compare our modified approach with the default PromptIR[1] without any modifications.

Methods	PSNR ↑
Default settings	30.32
Ours	**30.77**

4.2 Different weighting factors on SSIM loss

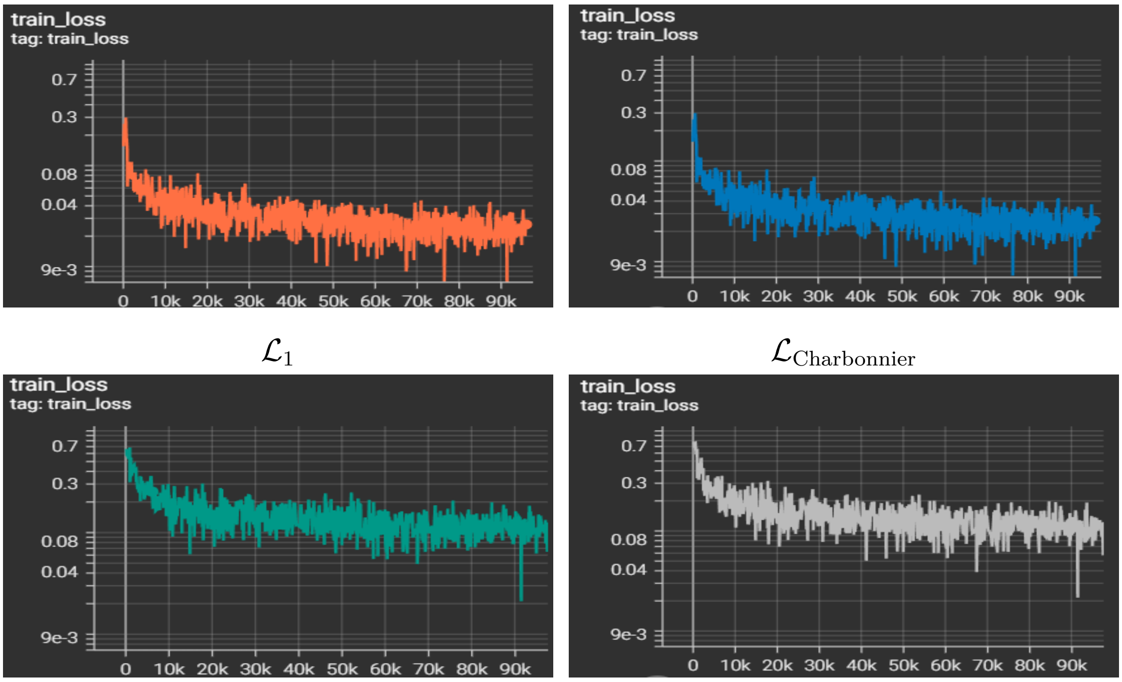
I applied different weighting factors on SSIM loss to demonstrate that 0.84 is a satisfactory choice as recommendation in in [5].

λ_s	PSNR ↑
0.10	30.58
0.84	**30.77**

4.3 Loss functions

As the table and figure show below, while using Charbonnier loss alone does not outperform the default L1 loss, combining Charbonnier loss with SSIM loss yields the highest PSNR results.

Loss functions	PSNR ↑
$\mathcal{L}_1$	30.32
$\mathcal{L}_{\text{Charbonnier}}$	30.28
$\mathcal{L}_1 + \mathcal{L}_{\text{SSIM}}$	30.75
$\mathcal{L}_{\text{Charbonnier}} + \mathcal{L}_{\text{SSIM}}$	**30.77**



$\mathcal{L}_1$ $\mathcal{L}_{\text{Charbonnier}}$

$\mathcal{L}_1 + \mathcal{L}_{\text{SSIM}}$ $\mathcal{L}_{\text{Charbonnier}} + \mathcal{L}_{\text{SSIM}}$

Figure 5: Comparisons of loss curve.

References

[1] V. Potlapalli, S. W. Zamir, S. H. Khan, and F. Shahbaz Khan, "Promptir: Prompting for all-in-one image restoration," *Advances in Neural Information Processing Systems*, vol. 36, pp. 71 275–71 293, 2023.

[2] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739.

[3] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[4] J. T. Barron, "A general and adaptive robust loss function," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4331–4339.

[5] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017. DOI: `10.1109/TCI.2016.2644865`.

[6] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. DOI: `10.1109/TIP.2003.819861`.