110612117 張仲瑜 HW4 report

I. Github

https://github.com/sharkccy/NYCU_CV_2025_Spring/tree/main/hw4

II. Introduction

1. Goal

In this task, we aim to achieve as high Peak Signal-to-Noise Ratio (PSNR) [1] as possible in a blind-type image restoration competition with PromptIR [2] model, and there are two kinds of degraded images in the dataset, 1600 rainy images and 1600 snowy images.

2. Environment

- ✧ Laptop: Windows / RTX4060 * 1
- ✧ Kaggle account * 3: Linux / Tesla T4 * 2

3. Core Idea

I trained the PromptIR model by adding different loss functions including SSIM loss [3], Sobel Gradient loss [4], and Fast Fourier Transform (FFT Frequency) loss [5], and conducted experiments on losses, number of Transformer Blocks [6] in each encoder layer [2], and length of prompt in PGM block [2].

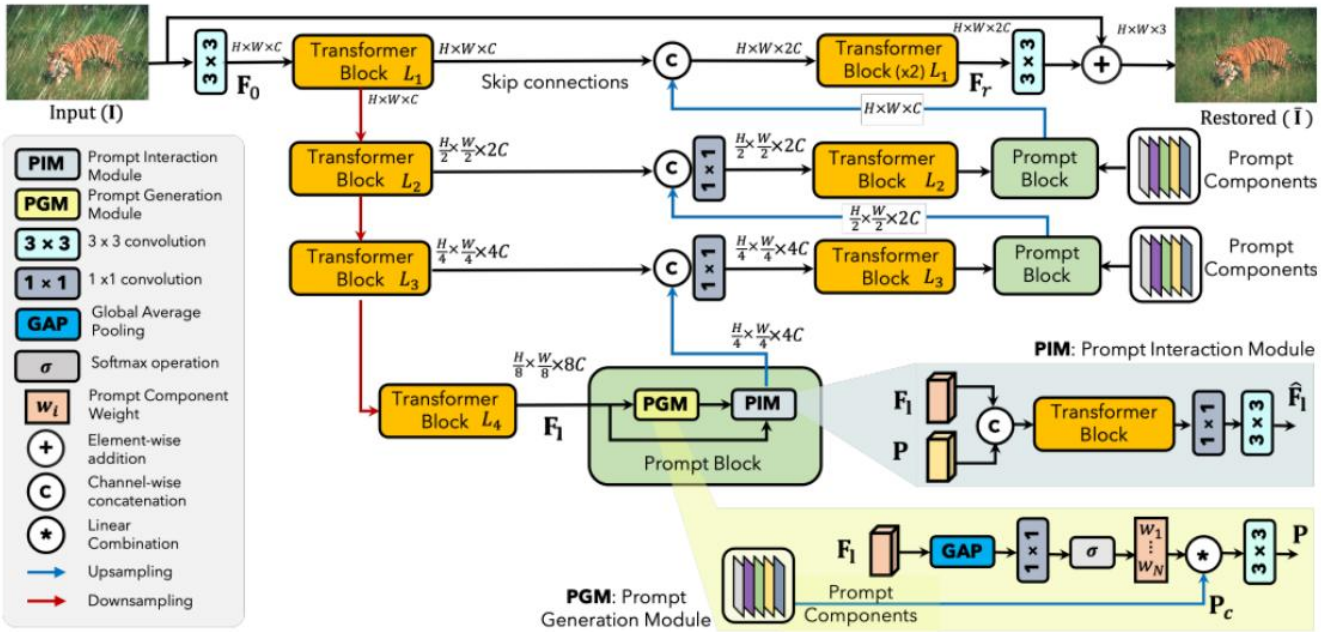
III. Method

1. Data augmentation set

- ✧ Flip, Rotation [7]
  Horizontal, Vertical flip and Rotation are applied to simulate different angles of cells, aiming to increase the robustness of the model.
- ✧ Color Jittering
  ColorJitter [7] (brightness=0.2, contrast=0.2, saturation=0.2,

hue=0.1) is applied to simulate different light conditions of possible rainy and snowy images.

2. Model Architecture
   - ✧ Model name: PromptIR



Total parameters: 41.1 M

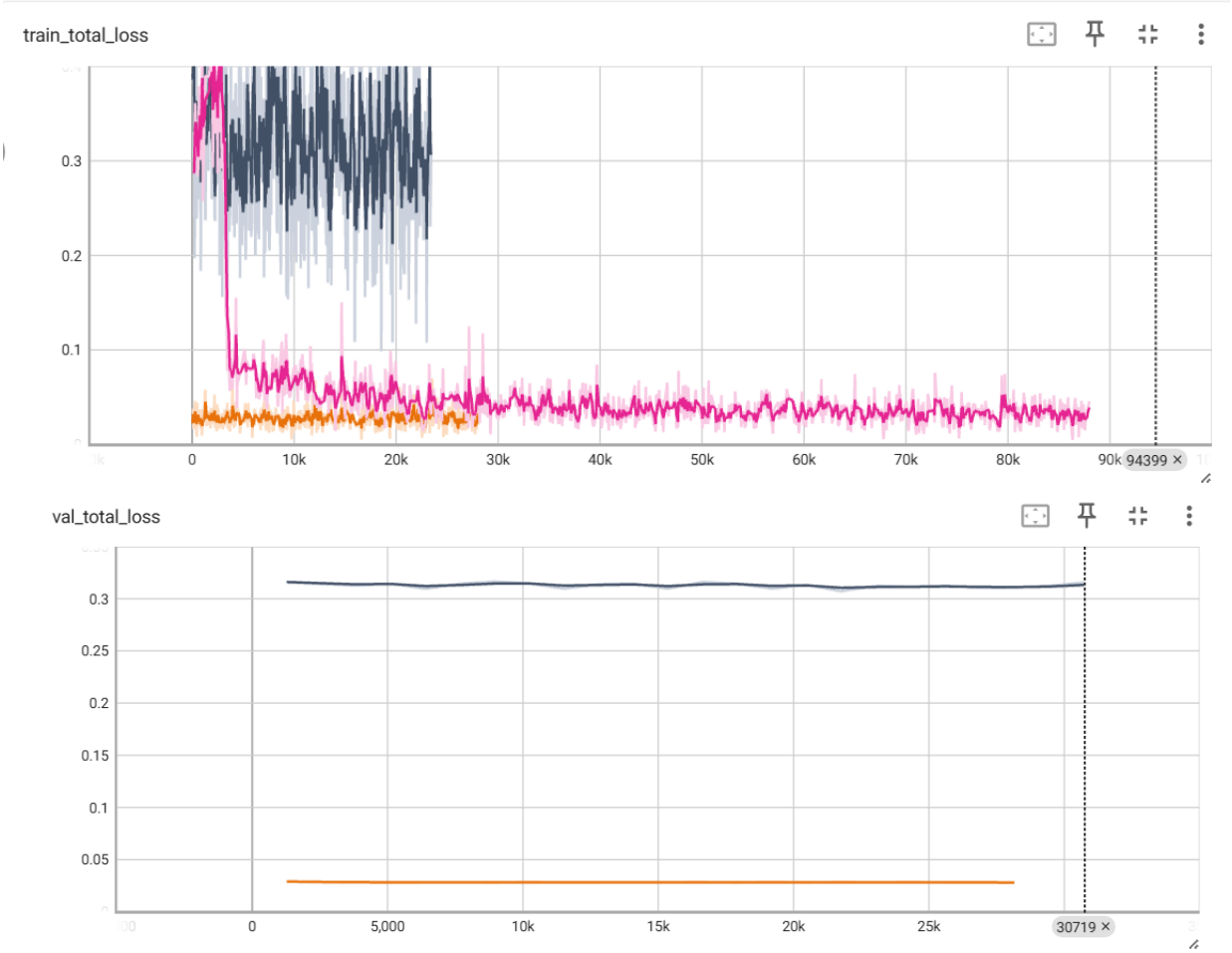
- ✧ Number of blocks in each transformer layer and the number of refinement blocks after the layer:

```
class PromptIR(nn.Module):
    def __init__(self,
        inp_channels=3,
        out_channels=3,
        dim = 48,
        num_blocks = [6,8,8,10],
        num_refinement_blocks = 6,
        heads = [1,2,4,8],
        ffn_expansion_factor = 2.66,
        bias = False,
        LayerNorm_type = 'WithBias',   ## Other option 'BiasFree'
        decoder = False,
    ):
```

3. Hyperparameters (best one)
   - ✧ Trained from scratch
   - ✧ Learning rate: 2e-4 in with Warmup Cosine Annealing in first 62 epochs, and 1e-5 with 0.95 decay factor in 62~100 epochs
   - ✧ Optimizer: AdamW
   - ✧ Batch size: 1
   - ✧ Epoch: 100

IV. Result:

1. Best performance
   - ✧ PSNR: 0.3065 on private testing
   - ✧ Learning curve (epoch vs Training/Val loss)



pink: Initial training

gray and orange: Continue training with validation set applied to detect overfitting in the ending phase.

V. Additional experiment:

1. Hypothesis:
   Incorporating the SSIM loss [3] as an additional loss component can improve the performance of the PromptIR model in terms of PSNR.

   ✧ In the baseline PromptIR design, only L1 loss is used as the pixel-wise loss. However, the SSIM loss, which measures dissimilarity in luminance, contrast, and structural features between the output and ground truth images, has been shown to enhance the performance of image restoration models. By combining SSIM with L1 loss, we aim to capture both pixel-level accuracy and structural similarity, thereby boosting the PSNR.

   ✧ Result

Loss components	PSNR (dB)
L1	27.9
0.8 * L1 + 0.2 * SSIM	29.1

   ✧ Implication
   The incorporation of SSIM loss, which focuses on global luminance, contrast, and structural similarity, enabled the model to better preserve structural details and overall consistency in the restored images. Unlike the L1 loss, which directly minimizes pixel-wise errors, SSIM indirectly enhances PSNR by optimizing for perceptual quality, addressing the limitations of L1 loss in capturing structural information.

2. Hypothesis:
   Adjusting the number of Transformer Blocks in each encoder layer [2] can improve the performance of the PromptIR model in terms of PSNR.

   ✧ The baseline PromptIR architecture defines the number of blocks per layer via the num_blocks parameter which influences the model's capacity to extract hierarchical features.

   ✧ Result

Number of blocks per layer	PSNR (dB)
[4, 6, 6, 8]	29.1
[6, 8, 8, 10]	29.5

   ✧ Implication
   With more trainable parameters due to the increased number of Transformer Blocks, the PromptIR model gained enhanced capacity to capture complex degradation patterns, leading to a PSNR improvement

3. Hypothesis:
   incorporating the Sobel gradient loss [4] and FFT frequency loss [5] as additional loss components can improve the performance of the PromptIR model in terms of Peak Signal-to-Noise Ratio (PSNR)

   ✧ The Sobel gradient loss enhances edge preservation, while the FFT frequency loss improves texture recovery by focusing on high-frequency components. Previous studies have demonstrated that combining such structural and frequency-based losses with traditional pixel losses can enhance image restoration quality, suggesting a potential PSNR improvement in our model.

   ✧ Result

Loss components	PSNR (dB)
L1 + SSIM	29.5
0.7 * L1 + 0.1 * SSIM + 0.1 * gradient + 0.1 *frequency	30.06

   ✧ Implication:
   The incorporation of Sobel gradient loss and FFT frequency loss did boost the PSNR, complementing the L1 and SSIM losses, which focus on pixel-level accuracy and structural similarity.

VI. Thoughts

It's crazy that a model can blind restore the degraded images with the prompt mechanism. During the researching process, I also found another interesting paper using prompt-in-prompt mechanism [8] to further boost performance. However, I struggle to modify it into a blind restoration version due to limited time.

VII. Reference:

1. A. Horé and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM", 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 2010, pp. 2366-2369 https://doi.org/10.1109/ICPR.2010.579
2. Potlapalli, V., Zamir, S. W., Khan, S., & Khan, F. S. (2023). PromptIR: Prompting for All-in-One Blind Image Restoration https://doi.org/10.48550/arXiv.2306.13090
3. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. In Proceedings of the IEEE Transactions on Image Processing, 2004, 13(4), 600-612. https://doi.org/10.1109/TIP.2003.819861
4. Senthil Anandhi, A., Jaiganesh, M. An enhanced image restoration using deep learning and transformer based contextual optimization algorithm. *Sci Rep* **15**, 10324 (2025). https://doi.org/10.1038/s41598-025-94449-5
5. Hu, J., & Wang, Z. (2025). A novel CNN architecture for image restoration with implicit frequency selection. *Connection Science*, *37*(1). https://doi.org/10.1080/09540091.2025.2465448
6. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is All You Need. In Proceedings of the 31st International Conference on

Neural Information Processing Systems (NIPS), 2017, 5998-6008. https://doi.org/10.48550/arXiv.1706.03762

7. Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0
8. Li, Z., Lei, Y., Ma, C., Zhang, J., & Shan, H. (2023). Prompt-In-Prompt Learning for Universal Image Restoration. arXiv preprint arXiv:2312.05038, 1-10. https://doi.org/10.48550/arXiv.2312.05038