NYCU Visual Recognition using Deep Learning
HW4 report

109612019, 林伯偉

Github Link : https://github.com/Albert5865/NYCU-Computer-Vision-2025-Spring/tree/main/HW4

1. Introduction:

This report details the approach and results for Homework 4 of the Visual Recognition using DeepLearning course at NYCU (Spring 2025). The task involves training a single model to restore images degraded by rain and snow, using a dataset of 1600 training images per degradation type and 50 test images per type. The evaluation metric is Peak Signal-to-Noise Ratio (PSNR), with constraints including no external data, a pure vision-based approach, and training from scratch without pre-trained weights. The chosen approach leverages the PromptIR framework, which uses tunable prompts to encode degradation-specific information, enabling a single model to handle both rain and snow efficiently. This report outlines data pre-processing, model modifications, results, and additional experiments, focusing on enhancements made to the original PromptIR implementation.

2. Method:

To achieve the desired segmentation, I used the following approach:

Data Pre-processing :

The dataset was pre-processed for consistency and training efficiency:

- **Normalization:** Pixel values normalized to [0, 1] by dividing by 255.
- **Resizing :** Images resized to 256×256 for batch processing.
- **Augmentation:** Applied random horizontal flips, rotations (±10◦), and color jitter (brightness and contrast adjustments of 0.1) to enhance dataset diversity and reduce overfitting.

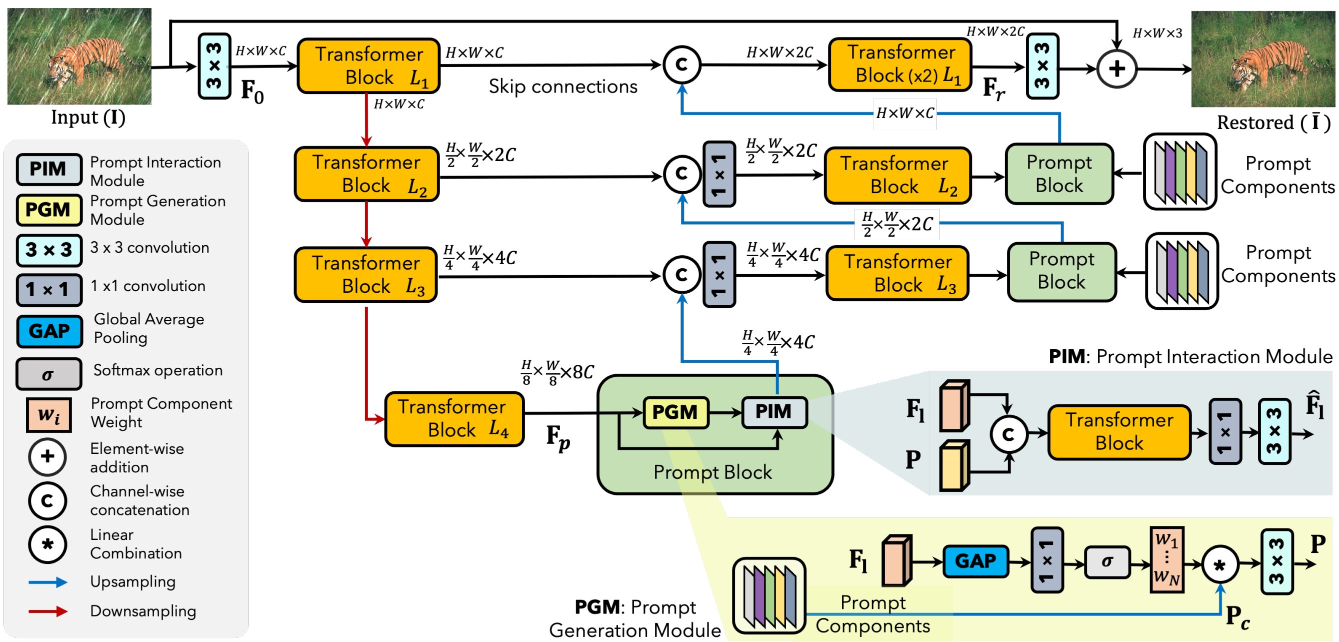
Model Architecture :

The base PromptIR model was modified with the following updates:

- **Scheduler Change:** Replaced the default LinearWarmupCosineAnnealingLR scheduler with ReduceLROnPlateau, configured with a factor of 0.7, patience of 0, threshold of 1e-4, and minimum learning rate of 1e-6, to adaptively adjust the learning rate based on training loss plateaus.

- **Loss Function Change:** Modified from a pure L1 loss to a weighted combination: 0.7 × L1 loss + 0.3 × SSIM loss (using SSIM with a window size of 11), balancing pixel-wise accuracy with structural similarity

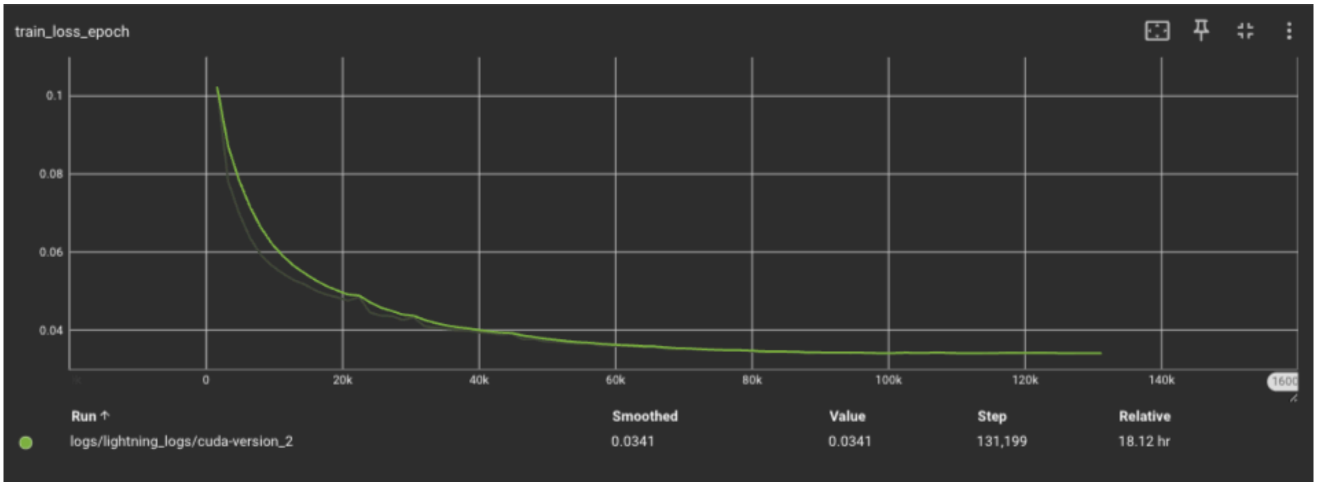
- **Oringinal PromptIR Architecture:** The model is modified from this original PromptIR



Hyperparameters :

- Model : Modified PromptIR
- Learning rate : 0.0002
- Batch size : 2 (constrained by GPU memory)
- Patch size: 128
- Weight decay : 1e-5
- Optimizer : AdamW
- Epoch : 150

3. Results:



{"private_psnr": 28.788398663197913, "public_psnr": 29.51551621141678}

The training loss curve starts at around 0.1, drops sharply, and stabilizes between 0.02 and 0.03, ending at 0.0341 after 131,199 steps. The smooth curve indicates stable training, while final PSNR values of 28.79 (private) and 29.52 (public) confirm the model's success in image restoration and effective convergence.

Result Visualizations :

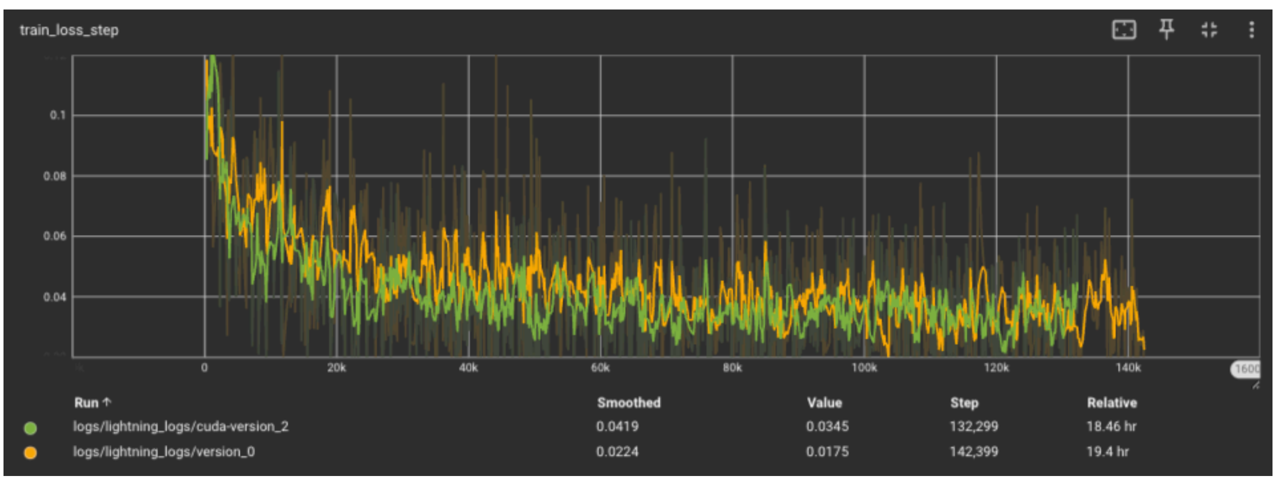
4. Additional experiments:

1. Scheduler Comparison:

- **Hypothesis**: ReduceLROnPlateau improves convergence over LinearWarmupCosineAnnealingLR and reduces oscillation .

- **How it works**: LinearWarmupCosineAnnealingLR increases the learning rate linearly during a warmupphase, then decreases it following a cosine annealing schedule, which can lead to oscillations if thedecay is too aggressive or misaligned with the loss landscape. In contrast, ReduceLROnPlateau mon-itors the training loss and reduces the learning rate by a factor (0.7) if the loss does not improve by athreshold (1e-4) within a patience period (0 epochs), down to a minimum (1e-6). This adaptive approachminimizes oscillations by ensuring the learning rate adjusts to the models convergence needs, prevent-ing premature or excessive reductions. We chose ReduceLROnPlateau to stabilize training and improvefinal performance, especially given the complex degradation patterns of rain and snow.

- **Implication**: The adaptive scheduler better aligns with the training dynamics, reducing oscillations and enhancing convergence.



Green: ReduceLROnPlateau **Yellow: LinearWarmupCosineAnnealingLR**

2. Loss Function Tuning:

- **Hypothesis**: A weighted loss combining L1 and SSIM enhances both pixel accuracy and structural quality.

- **How it works**: The original L1 loss computes the mean absolute error between predicted and groundtruth pixels, focusing on pixel-wise accuracy but often neglecting structural coherence.

SSIM (StructuralSimilarity Index), with a window size of 11, measures perceptual similarity by comparing luminance,contrast, and structure between patches, emphasizing visual quality. The new loss, 0.7×L1+0.3×SSIM,balances these objectives: the L1 component ensures pixel fidelity, while SSIM improves texture and edge preservation.

- **Implication**: The adaptive scheduler better aligns with the training dynamics, reducing oscillations and enhancing convergence.

Performance of the original PromptIR:

{"private_psnr": 27.021484994643743, "public_psnr": 27.767700126742575}

Performance of the modified PromptIR with experiment changesperformance:

{"private_psnr": 28.788398663197913, "public_psnr": 29.51551621141678}

Experiment results:

PromptIR was selected for its prompt-based adaptability, offering efficiency over task-specific models. Its advantages include flexibility and reduced computational overhead, though prompt tuning adds complexity. Visualizations with the modified loss function showed improved texture preservation. Experiments confirm the efficacy of adaptive scheduling, data augmentation, and hybrid loss functions, suggesting further exploration of these enhancements.

References:

1. **PromptIR: Prompting for All-in-One Blind Image Restoration (NeurIPS'23)**Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan (https://arxiv.org/abs/2306.13090)
2. **PromptIR: Prompting for All-in-One Blind Image Restoration (NeurIPS'23)** (https://github.com/va1shn9v/PromptIR?tab=readme-ov-file)
3. **Understanding SSIM** Jim Nilsson, Tomas Akenine-Möller (https://arxiv.org/abs/2006.13846)