Visual Recognition using Deep Learning HW4

110550085 房天越

GitHub Repo Link

https://github.com/TianYueh/NYCU-Visual-Recognition-2025-HW4

Introduction

In this homework, we aim to implement image restoration by removing snow and rain artifacts from photographs. Our core idea is to integrate noise-type prompts (labels) into a U-Net backbone [2] using Feature-wise Linear Modulation (FiLM) [1]. By conditioning the network on the noise category, we aim to adaptively scale and shift feature maps to improve denoising performance.

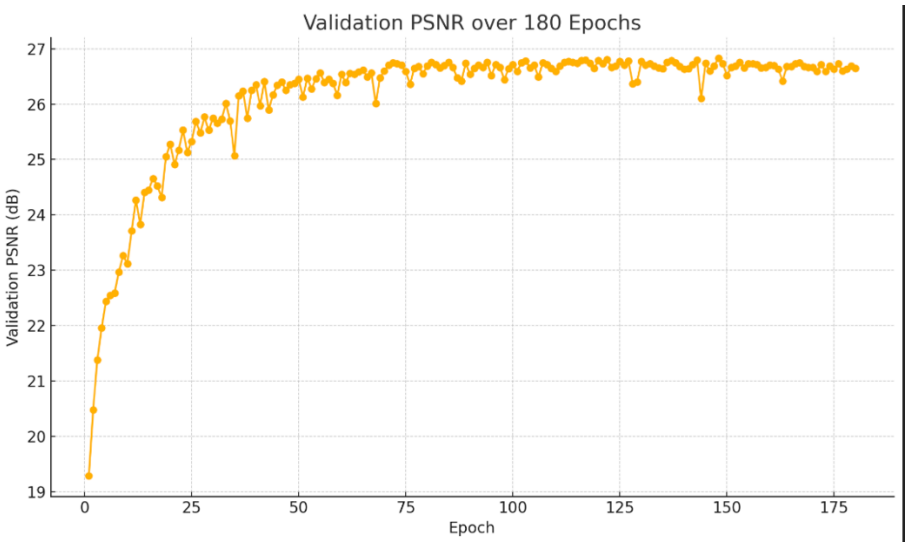
Method

1. Data Pre-processing
   All input images are converted to RGB and normalized to the [0,1] range.
   The file weather.csv is used to assign a binary label to each test image, the images mentioned in the csv file is for rain. This file is handwritten because we have only 100 images.
   We use torchvision.transforms.ToTensor() without further augmentation.Model Architecture
2. Model Architecture
   Base Model: Four-level encoder–decoder U-Net [2].
   FiLM Blocks: Inserted after each encoder block to modulate features with a learned prompt embedding [1].
   Prompt Embedding: Learned via nn.Embedding(2, 16).
   Loss Function: Mixed loss: L = 0.5·L1 + 0.5·MSE.
3. Hyperparameters
   Prompt dimension: 16
   Optimizer: Adam
   Learning rate: 1e-4
   Batch size: 4
   Training epochs: 180
4. Architectural Modifications
   Hypothesis: A generic U-Net lacks contextual understanding of noise types.
   Modification: FiLM blocks provide type-specific modulation of feature maps.

Expected Benefit: Enables the network to apply differentiated denoising strategies for snow vs. rain.

Results

1. Training Curves



From this curve, we can see that the model converges at epoch about 80.

The best result is at epoch 148, which has PSNR = 26.82 on the validation set, and 27.25 in the submission file.

2. Example Visualization of the Pictures

Snow, Before vs. After (0.png)

Rain, Before vs. After (1.png)

Here we can observe that the result for snow is quite good.
However, for the rain pictures, it has a feeling like oil painting, the result PSNR is also worse than snow (Discussed in Additional Experiments)

References

[1] Perez, E., Strub, F., de Vries, H., Dumoulin, V., & Courville, A. (2018). *FiLM: Visual Reasoning with a General Conditioning Layer*. CVPR.
[2] Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. MICCAI.

Additional Experiments

1. Training Separately for Snow and Rain
   In our dataset, the pictures are given with prefix “snow” or “rain”, so I tried to train two models separately, and use the best models on both sides.
   The code for this part is also given in the submission files.
   The result is that, the PSNR for Snow goes up to 27.81, while that for Rain is at most 26.14 (validation set).
   The Overall performance on the submission file is 27.40 (public), which is better than the model by training one model.
2. Improvement by Depth of UNet
   The depth of UNet significantly impacts the performance.
   Model with only two layers has the highest PSNR of 24.32 (On epoch 142), which is significantly lower than the current performance