姓名：陳紅宏

學號：F70497065

HW4 報告

Theoretical Justification

In this report, I propose an approach to improve the training efficiency and quality of the Denoising Diffusion Probabilistic Model (DDPM) by incorporating a Deep Image Prior (DIP). The DIP is leveraged to provide a quick initial prior that captures high-level structures and patterns of the target images, thereby accelerating the DDPM training process.

The proposed solution combines the strengths of DDPM and DIP. The DDPM is known for its robust generative capabilities, but it suffers from a slow backward learning process. On the other hand, DIP can quickly learn an image-specific prior using a CNN architecture without the need for a large dataset.

To implement this solution, my design is as follows:

- DIP Initialization: We initialize the DDPM with the output of the DIP model to provide a more informative starting point.
- Training DIP: We train the DIP model for a short duration to capture high-level image structures without overfitting.
- Architecture: We use a simple U-Net architecture for both DIP and DDPM due to its effectiveness in image generation tasks.

The DIP-based initialization can potentially benefit DDPM by reducing the number of diffusion steps required for DDPM to converge, thereby speeding up the training process. By starting with a more informative prior, the DDPM can generate higher quality images earlier in the training process. However, the addition of DIP increases the overall complexity of the model training process.

Implementation and Experiments

For my experiment, I use the Oxford Flowers dataset (nelorth/oxford-flowers). The dataset is preprocessed to a fixed size of 128x128 pixels. To evaluate its effectiveness, I will measure according to PSNR, SSIM, generation time and steps.

To implement my proposal, The DIP model is trained on the dataset for 100 epochs using an Adam optimizer with a learning rate of 0.01. Then The DDPM model is initialized with the output from the DIP model and trained for 30 epochs with a learning rate of 0.001.

The result from the standalone DDPM training is as follows:

- Total time: 154 seconds
- Total step: 210
- PSNR: 6.59
- SSIM: -0.038

While the result from DIP + DDPM training is as follows:

- Total time: 152 seconds
- Total step: 210
- PSNR: 7.34
- SSIM: -0.083

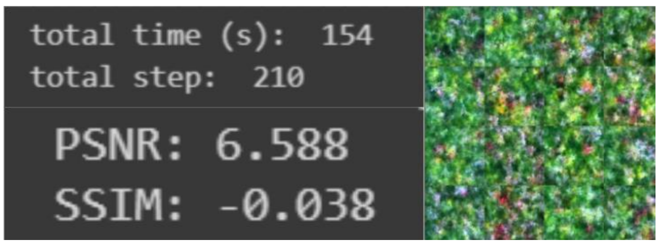

Figure 1: Without DIP

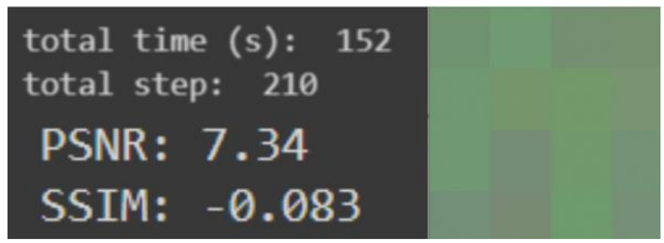

Figure 2: With DIP

The total time and total steps for training are nearly identical for both methods, so my integration of DIP does not significantly affect the speed of training.

A higher PSNR value indicates better pixel-wise similarity between the reconstructed and original images. An increase in PSNR value indicates that the model produces images with fewer pixel-wise errors to the original images compared to the training DDPM alone. This improvement demonstrates the effectiveness of using DIP as an initial prior to enhance the DDPM's reconstruction quality.

However, a lower SSIM value indicates less structural similarity between the images. The SSIM values in both cases are negative, which typically indicates a lack of structural similarity. The more negative SSIM value for DIP + DDPM suggests that, despite the higher PSNR, there are structural distortions or changes introduced in the image, possibly due to the influence of the initial DIP prior.

Ablation Studies

For ablation studies, I tried using different noise levels (0.3, 0.5, 0.7). The results are as below:

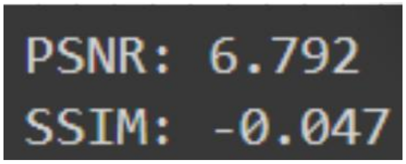

Figure 3: Noise level 0.3

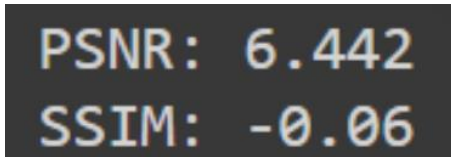

Figure 4: Noise level 0.5

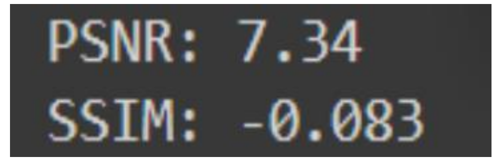

Figure 5: Noise level 0.7

The noise level with the best PSNR value is 0.7, while the noise level with the best SSIM value is 0.3. The higher the noise level is, the better the pixel-wise similarity between the original image and reconstructed image, while the lower the noise level is, the less structural distortions there are.