

# Generative Models for Visual Signals – Assignment

*Prof. Chih-Chung Hsu*

*TA: Chia-Ming Lee (zuw408421476@gmail.com)*

## I. Introduction:

This homework assignment is focused on providing a deep understanding of two advanced image processing models: Denoising Diffusion Probabilistic Models (DDPM) and Deep Image Prior (DIP). This document will elaborate on the core concepts, theoretical underpinnings, and practical examples to better grasp how these models operate and can be applied in real-world scenarios.

### 1.1. Denoising Diffusion Probabilistic Models (DDPM)

DDPMs are generative models that transform Gaussian noise into a complex data distribution through a process akin to reversing thermodynamic diffusion. This is achieved by a sequence of learnable reverse diffusion steps, each of which is conditioned on its predecessor. The models are trained to predict the noise added at each timestep, enabling the recovery of the original data from pure noise. For instance, one practical application is generating detailed and diverse human faces from randomized noise inputs.

A simplified equation representing this process is:  $x_t = \sqrt{\alpha_t} * x_0 + \sqrt{1-\alpha_t} * \epsilon$ , where  $\epsilon$  is noise, and  $\alpha_t$  represents the variance schedule over timesteps.

### 1.2. Deep Image Prior (DIP)

Deep Image Prior utilizes the inherent structure of convolutional neural networks (CNNs) to encode priors for natural images. By fitting a CNN with randomly initialized weights directly to a target image, DIP leverages the bias of CNNs towards natural image statistics without training on large datasets. This approach is particularly effective for tasks like image denoising, super-resolution, and inpainting, providing high-quality restorations based on the network's innate biases alone.

The optimization process can be described as minimizing the difference between the CNN output and the target image, which typically leads to underfitting to noise and emphasizes significant image structures.

## II. Task: Technical Integration of DDPM and DIP

For this assignment, you are tasked to explore and propose methods to integrate the principles of DDPM and DIP. Consider how the probabilistic modeling of DDPM can enhance the image-specific priors of DIP, or how the architectural bias of DIP can be used to initiate the reverse diffusion process in DDPM. Provide theoretical formulations and possible practical implementations. Here are two examples, and your solution could be one of them or a novel one.

### Example 1: Accelerating DDPM with DIP-based Initial Priors

In this example, you are asked to explore the possibility of using Deep Image Prior (DIP) to provide a quick initial prior for the Denoising Diffusion Probabilistic Model (DDPM) training process. The motivation behind this approach is to address the slow backward learning process in DDPM by leveraging the image-specific prior automatically learned by the CNN architecture in DIP.

To implement this idea, you can start by training a DIP model on the target image for a relatively short period. The goal is to capture the high-level structures and patterns present in the image without overfitting to the noise. The trained DIP model can then be used to generate an initial prior for the DDPM training process.

Next, you should modify the DDPM training algorithm to incorporate the DIP-based initial prior. Instead of starting from pure noise, the DDPM model can be initialized with the output of the DIP model. This initialization can provide a more informative starting point, potentially reducing the number of diffusion steps required for the DDPM model to converge.

You should experiment with different DIP training durations and architectures to find the optimal balance between capturing meaningful image priors and computational efficiency. They should also investigate the impact of the DIP-based initialization on the quality and diversity of the generated samples from the DDPM model.

To evaluate the effectiveness of this approach, you can compare the convergence speed and sample quality of the DDPM model with and without the DIP-based initial prior. They should provide quantitative metrics, such as the number of diffusion steps required to reach a certain level of sample quality, as well as qualitative comparisons of the generated samples. Furthermore, you can explore variations of this idea, such as using DIP to provide intermediate priors at different stages of the DDPM training process. They can also investigate the potential of using DIP to guide the noise estimation and denoising steps in DDPM, leveraging the learned image prior to improve the accuracy of these steps.

## **Example 2: Guiding DIP Early Stopping with DDPM-inspired Supervision**

In this example, you are challenged to develop a more principled approach to determine the optimal early stopping point in Deep Image Prior (DIP) training by incorporating supervision information inspired by the Denoising Diffusion Probabilistic Model (DDPM).

The main idea is to introduce gradual denoising steps during the DIP training process, similar to the denoising steps in DDPM. By adding noise to the target image at different levels and using these noisy versions as intermediate targets, you can guide the DIP model to learn a hierarchical representation of the image.

To implement this approach, you should modify the DIP training algorithm to include multiple denoising stages. At each stage, the target image is corrupted with noise of varying levels, creating a sequence of noisy images. The DIP model is then trained to reconstruct these noisy images in a progressive manner, starting from the most heavily corrupted image and gradually moving towards the clean target image.

During training, you should monitor the reconstruction quality of the DIP model at each denoising stage. They can use metrics such as peak signal-to-noise ratio (PSNR) or structural similarity index (SSIM) to quantify the similarity between the reconstructed images and the corresponding noisy targets. By analyzing the improvement in reconstruction quality across the denoising stages, you can develop a criterion for determining the optimal stopping point for DIP training.

You should experiment with different noise levels, denoising schedules, and DIP architectures to find the most effective configuration. They should also investigate the impact of this approach on the final reconstructed image quality and compare it with traditional early stopping methods used in DIP.

Furthermore, you can explore additional techniques to enhance the denoising guidance in DIP training. For example, they can incorporate perceptual loss functions or adversarial training objectives to improve the perceptual quality of the reconstructed images. They can also investigate the use of learned denoising priors or conditional denoising models to guide the DIP training process.

To evaluate the effectiveness of this approach, you should provide quantitative comparisons of the reconstruction quality and early stopping accuracy compared to baseline DIP methods. They should also present qualitative results showcasing the visual quality of the reconstructed images at different denoising stages and the final output.

### III. Evaluation Criteria

- Theoretical Justification (30%):
  - Provide a clear and coherent explanation of the proposed solution, highlighting how it combines the strengths of DDPM and DIP. Justify the design choices and assumptions made in the proposed approach. Discuss the potential benefits and limitations of the proposed solution compared to using DDPM or DIP alone.
- Experimental Verification (40%):
  - Implement the proposed solution and conduct experiments to validate its effectiveness. Compare the performance of the proposed approach with standalone DDPM and DIP methods in terms of either image quality, generation speed, or both. Provide quantitative metrics to support the claims, such as PSNR, SSIM, FID, or generation time. Present qualitative results showcasing the visual quality of the generated or reconstructed images. Analyze the experimental results and discuss the observed improvements or trade-offs compared to the baseline methods.
- Ablation Studies and Analysis (30%):
  - Conduct ablation studies to investigate the impact of different components or hyperparameters in the proposed solution. Vary the key parameters, such as noise levels, denoising schedules, or architectures, and evaluate their influence on the performance. Provide insights and interpretations based on the ablation studies, justifying the chosen configurations.

Note: The focus of this assignment is on demonstrating the effectiveness of combining DDPM and DIP techniques, rather than achieving state-of-the-art performance. The proposed solution should show improvements in either image quality, generation speed, or both, compared to using DDPM or DIP individually. The claims made in the solution should be supported by theoretical justifications and experimental verification.

The evaluation criteria emphasize the importance of providing a clear theoretical justification for the proposed approach, conducting thorough experiments to validate its effectiveness, and presenting the work in a well-organized and understandable manner. You are expected to provide quantitative and qualitative results, perform ablation studies to analyze the impact of different components, and discuss the observed improvements or trade-offs compared to the baseline methods.

By meeting these evaluation criteria, you can demonstrate their understanding of the DDPM and DIP techniques, their ability to combine them effectively, and their skills in conducting rigorous experiments and analysis.

### IV. Submission Requirements

#### 4.1. GitHub Repository (50%):

- Create a GitHub repository to host your implementation and related files.
- Include well-documented and organized code for your proposed solution.
- Provide clear instructions in the README.md file on how to run the code and reproduce the experiments.
- Use appropriate git commit messages and branches to track the development progress.
- Ensure that the repository is accessible to the instructor and teaching assistants.

#### **4.2. Report (50%):**

- Write a comprehensive report describing your proposed solution, experiments, and findings.
- The report should be in a format of your choice (e.g., PDF, Markdown, LaTeX) and can be written in any preferred language.

#### **4.3. Submission Deadline:**

- The GitHub repository link and the report should be submitted via Moodle by 2024.6.11.
- Late submissions will be subject to the course's late submission policy.