

GAI Project 4: Generative Models for Visual Signals

資訊 113 F74096027 吳瀚惟

1. Theoretical Justification

DIP (Deep Image Prior) 是一種基於深度學習的圖像修復技術，他利用隨機初始化的深度卷積網路來進行圖像去雜訊、修復和超分辨率等逆向工程，僅需使用受損的圖像進行訓練。DIP 技術被廣泛應用於圖像品質提升中，他能有效去除圖像中的 noise、修復破損部分並且提升清晰度。實作時，我們需要根據具體目標去選擇適合的 DIP 網路結構(U-Net, DnCNN, ResNet...) 和超參數組合，也必須關注迭代的次數和 overfitting 的問題。雖然 DIP 技術已經算是成熟，在學界也有顯著成果，但是他仍存在一些潛在的問題，如作業中提到的：**在 training 過程中難以確定最佳 early-stopping 時機的問題。**

DIP 雖然能利用卷積神經網路的結構先驗來重建圖像，但他容易 overfitting noise，並且很難去確定何時應該停止訓練，這會導致訓練過程中的影像品質不穩定，並且可能在重建圖像時引入更多的 noise。

我選擇實作作業文件中的 **Example 2: Guiding DIP Early Stopping with DDPM-inspired Supervision** 來嘗試解決上述問題。DDPM 在實作上，通過學習逐步去除 noise 的過程，能夠反向的從 noise 中去還原高品質的圖像。我借鑑這個想法，將 DIP 的訓練過程改為多階段逐步去噪的方式，等於是結合了 DDPM(引入他逐步去雜訊的過程來更好的捕捉圖片的層次結構以避免

overfitting)與 DIP 的優勢。實作步驟大致如下：

(1) **生成多階段的雜訊影像**：使用不同程度的雜訊來污染目標圖像，生成一系列不同雜訊水平的影像序列。這些雜訊影像將作為 DIP 模型的中間目標，用於指導模型學習圖像的層次結構表示。

(2) **逐步去噪訓練**：修改 DIP 的訓練演算法，讓它包含多個去雜訊階段。每個階段的目標圖像是不同雜訊水平的影像序列。接著，DIP 模型從雜訊等級最高的影像開始訓練，逐步移向較乾淨的目標圖像。

(3) **監控重建品質**：在每個去雜訊階段，我使用 PSNR 和 SSIM 兩種指標來量化重建圖片與相應雜訊目標之間的相似性。通過分析各個去雜訊階段的重建品質改進情況，作為制定 DIP 訓練的最佳 early-stopping point 依據。

在實際操作前，我假設多階段去雜訊能夠幫助模型逐步學習圖像的層次結構，從而提高重建品質，在實驗中對照傳統 DIP 與新方法的生成圖片品質、PSNR 以及 SSIM 值可以幫助我們較為全面的判斷此方法是否真的可以改善傳統 DIP 方法現存的問題。

當然，這套方法也有潛在的侷限性，例如多階段訓練可能會增加訓練的時間和複雜度，我也會在實驗中觀察這點。

2. Experimental Verification + Ablation Studies and Analysis

實驗目標是找出結合 DDPM 和 DIP 的新方法是否合理，且去雜訊的結果有

沒有比較好。因此我先實作了傳統 DIP(用於稍後去雜訊結果比較)與傳統 DDPM 方法(讓我更好的熟悉逐步去噪過程)，再來才是結合兩者的新方法。

在新方法中，通過多個階段訓練 DIP 模型，每個階段使用不同噪聲水平的圖像進行訓練，而每個階段的輸出作為下一階段的輸入，模擬逐步去噪過程。

我設計了一種動態調整訓練次數的方式，如下圖所示：

```
# 動態調整訓練次數的指數衰減函數
def calculate_epochs(beta, Emax, k, alpha):
    return int(Emax * (1 - alpha * np.exp(-k * beta)))
```

這個方法可以讓每個 stage 分配不同的訓練 epochs，做動態調整，讓模型分配較多訓練 epochs 在處理低雜訊的圖像，也就是說在訓練的過程，當 beta 值(雜訊)下降，訓練的次數也會隨之增加。我設計了以下幾種變因，來嘗試找出最佳配置，並驗證是否在解決難以確定最佳 early-stopping 時機問題的同時也比傳統 DIP 方法有著更好的 PSNR 和 SSIM 數據：

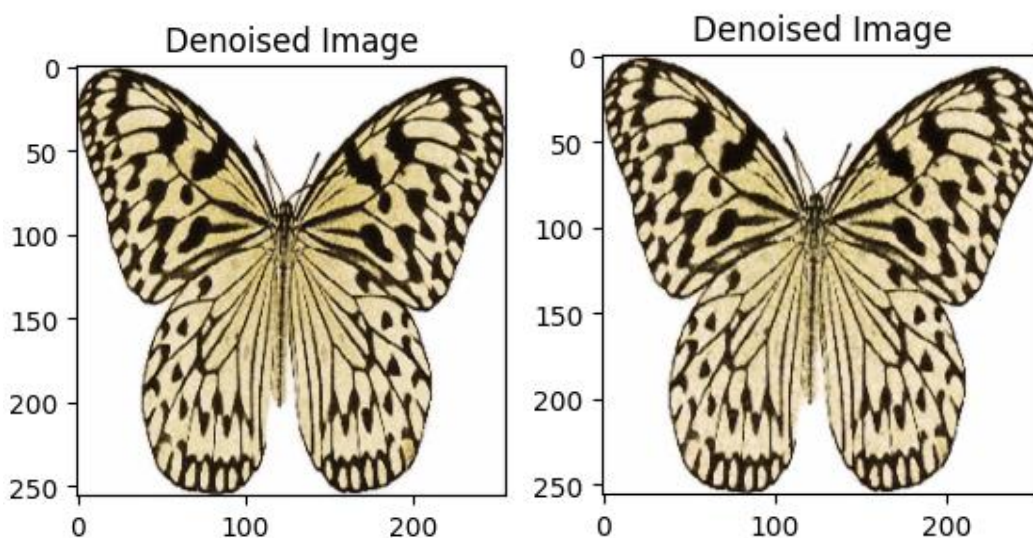
- (1) 三種 DIP 網路結構：U-Net, DnCNN, ResNet。
- (2) 感知損失，用於衡量預測圖像和真實圖像在特徵空間的差異。可透過調整 `use_perceptual_loss` 的值來選擇是否要使用(我設定的權重是 0.1)，或是單純使用 MSE loss。
- (3) Linear 和 Scaled-Linear Scheduler，兩種調度策略，用來控制每張圖片的雜訊強度。透過 `generate_beta_scheduler` 函數生成不同的 beta scheduler，控制每個階段加入的噪聲水平。

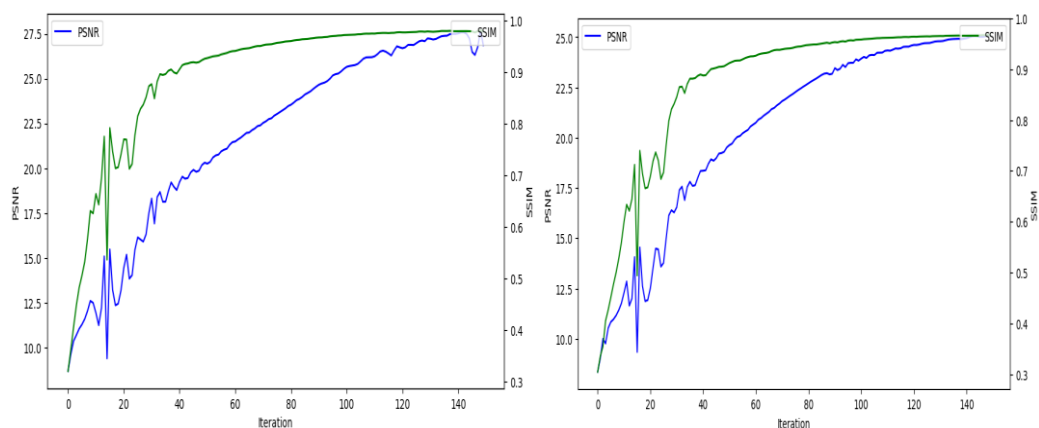
(4) 不同的 β_{\max} 值：0.01 和 0.02。

我通過調整以上配置來做消融實驗，並與傳統 DIP 比對結果。我使用的是 huggan/smithsonian_butterflies_subset 資料集中的第一張蝴蝶圖片(如下圖)，來觀察模型的去雜訊效果。



在傳統 DIP 框架下，我使用 U-Net 結構，設定 noise level 為 0.01(左)和 0.02(右)，分別訓練出來的去噪圖片以及指標變化曲線如下：





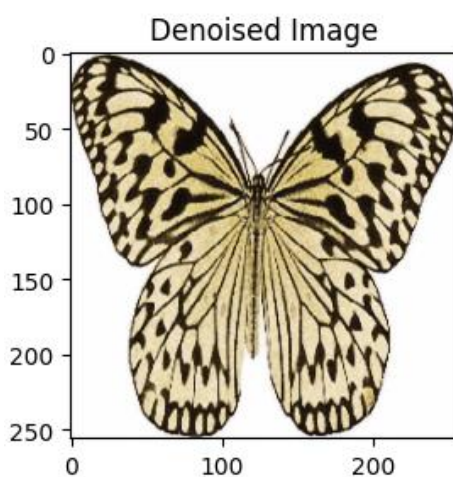
其 PSNR 與 SSIM 值如下所表所示：

DIP: 用U-NET	noise=0.01	noise=0.02
PSNR:	26.8612	25.1043
SSIM:	0.9789	0.967

接下來，我根據前文列出的變因，在新方法上進行消融實驗。由於實驗筆數過多，我僅統計了部分的實驗結果，如下表所示：

	noise level (beta_max)	DIP 結構	Scheduler	有無使用感知損失	PSNR value	SSIM value
1	0.01	U-Net	linear	X	27.7683	0.9827
2				V	28.0903	0.9868
3			scaled linear	X	28.4366	0.9882
4				V	28.9212	0.9905
5	0.02		linear	X	27.0062	0.9784
6				V	27.5224	0.9782
7			scaled linear	X	27.5031	0.9817
8				V	27.8355	0.9832

首先，我將展示第一項實驗的去噪圖片結果：



雖然用肉眼難以看出與使用傳統 DIP 方法去噪明顯的差距，但是根據 PSNR 和 SSIM 的評量結果，我們可以發現新方法生成的去噪圖片，其 PSNR 值和 SSIM 值皆略高於使用傳統 DIP 方法跑出來的結果，這證明了作業 example 2 所提的方法之有效性與實作可能性，而我也在訓練的過程中觀察到新方法中的動態調整訓練次數機制可以讓模型達到適度欠擬合的最佳狀態，也實現了在傳統 DIP 框架上希望解決的問題。另外，我們可以發現在其他條件皆相同的情況下，使用感知損失可以達到較高的 PSNR 和 SSIM 值(相較於未使用)；而使用 linear scheduler 會相較於使用 sccaled linear scheduler 的結果來的稍差。隨著 beta_max 值的設定增大，模型是必須要處理更多的圖象干擾因素，也因此 PSNR 與 SSIM 的值會相對較小。另外，在我實測的三種 DIP 網路結構中，U-Net 的效果是相較於使用 DnCNN 或是 ResNet 來的較佳。最後，回到第一點中所提到的錢在侷限性，雖然在模型設計上確實較為複雜，但我並未觀察到使用新方法的訓練過程和使用傳統 DIP 方法的訓練過程有付出明顯更大的時間成本。總結而言，作業中提到的這個方法確實合理，也是未來值得更進一步研究並改良的方法。