

# HW4: Generative Models for Visual Signals

F74094067 楊育勝

## 1. 方法發想

在傳統的深度影像先驗(Deep Image Prior, DIP)框架下，我們通過一個隨機初始化的神經網絡(U-net、ResNet、CNN 等)，直接對單張目標影像進行訓練。這種方法旨在利用深度神經網絡在訓練過程中所學到的自然影像統計先驗知識，將其內化為模型參數，實現對目標影像的高質量重建。然而，DIP 面臨的一個關鍵問題是：我們要如何判斷訓練的最佳停止點？過早停止會導致嚴重欠擬合，無法捕捉足夠的細節；而過度訓練則可能使網絡開始擬合雜訊，導致圖像品質下降。傳統的做法是通過監測訓練損失的變化趨勢來決定提早停止或是固定一個訓練次數來決定中止的時機，但這種啟發式方法往往效果是有限。

作業的指示提出了一種新穎的方法，通過借鑒去雜訊擴散概率模型(Denoising Diffusion Probabilistic Model, DDPM)的思想。DDPM 通過學習逐步去除雜訊的條件分佈，實現從純雜訊到複雜數據分佈(如自然影像)的高質量生成。將 DDPM 這種階段式去雜訊的訓練過程引入 DIP 框架，理論上有助於深度網絡捕捉影像的層次結構資訊，為訓練過程提供更合理的停止時機判據。

我們會將 DIP 的單階段訓練模式更改為多階段逐步去雜訊的訓練模式。在這一新框架下，神經網絡的訓練目標將不再是單一的乾淨影像，而是一系列由高到低、雜訊水平逐漸遞減的帶雜訊影像序列，類似於 DDPM 模型中的逆向去除雜訊的過程(Reverse Process)。

這個方法的關鍵在於：

- 通過分階段訓練，模型能更高效建立影像的層次結構
- 雜訊強度與各階段模型訓練次數的調節，可以避免過度擬合雜訊
- 最後低雜訊階段訓練完即可終止訓練，實現合理收斂時機的判斷

這種融合了 DDPM 思想的新 DIP 訓練框架，透過引入分階段去雜訊以及各階段模型訓練次數自主調節的機制，實現了對模型訓練收斂狀態的合理控制。理論上，在最終低雜訊階段，模型將達到適度欠擬合的最佳狀態，從而有效克服了傳統 DIP 方法的欠擬合或過擬合困境，獲得更優秀的重建性能。

## 2. 方法實現

為了將 DDPM 啟發的逐步去除雜訊，透過監督式學習的方式融入 DIP 訓練流程中，我們採取了以下具體步驟：

1. 基於 DDPM 前向過程，生成一系列具有不同雜訊強度的雜訊影像序列，每張影像的噪聲強度是由 `beta_scheduler` 控制。在這次的作業中我實作了三種常見的 DDPM `beta_scheduler`：

- Linear Beta Scheduler
- Scaled Linear Beta Scheduler
- Sigmoid Beta Scheduler

2. 我設計了一種動態訓練次數的調度機制來調節各階段模型的訓練次數，根據每個訓練階段對應的 `beta` 值(雜訊強度)，分配不同的訓練次數，使模型在處理低雜訊影像時獲得更多訓練機會。也就是說當 `beta` 值較大(雜訊較高)時，對應的最大訓練次數較少；隨著 `beta` 值下降，逐步去除噪音，該階段可以訓練的次數會逐漸增加。具體而言，我定義了一個指數衰減函數，這個函數會根據當前階段雜訊影像的 `beta` 值( $\beta(i)$ )、最大訓練次數上限( $E_{max}$ )、敏感度( $k$ )還有縮放係數( $\alpha$ )來決定當前階段要訓練多少個 epoch。

$$E(i) = E_{max} * (1 - \alpha * e^{(-k * \beta(i))})$$

3. 在訓練過程中，我們使用雜訊影像序列中的影像逐步作為訓練目標，引導模型從最初雜訊非常嚴重的影像開始，逐漸重建出清晰的原始影像。每個階段結束後，訓練目標都會動態更新為下一個雜訊強度的影像。我們使用均方誤差(MSE)損失和基於預訓練 VGG 網絡的感知損失(Perceptual Loss)作為訓練目標，兩者以一個權重因子結合( $\lambda$ )，旨在同時最小化像素水平和特徵層次的差異，獲得理想的重建質量。損失函數公式如下：

$$L(x, x') = MSE(x, x') + \lambda * PercepLoss(x, x')$$

4. 為了評估模型的表現，我們持續記錄了每個訓練回合(epoch)中，輸出影像與當前訓練目標影像(雜訊影像)以及原始乾淨影像之間的峰值信噪比(PSNR)。最終，我們將這些 PSNR 值繪制為訓練曲線，直觀展示了模型在逐步去噪監督下的學習進展。

### 3. 實驗結果

這次的實驗會使用到一張 F16 的圖片(圖一)，還有一張有雜訊的 F16 圖片當作要去除雜訊的目標(圖二)，長和寬皆為 512。

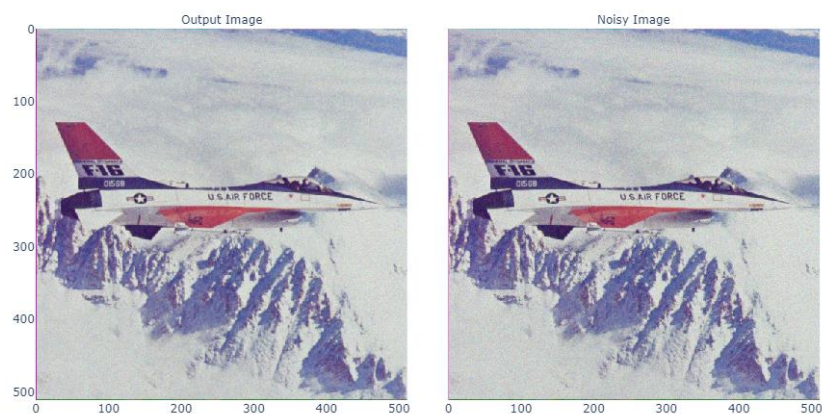


圖一、 原始 F16 圖片

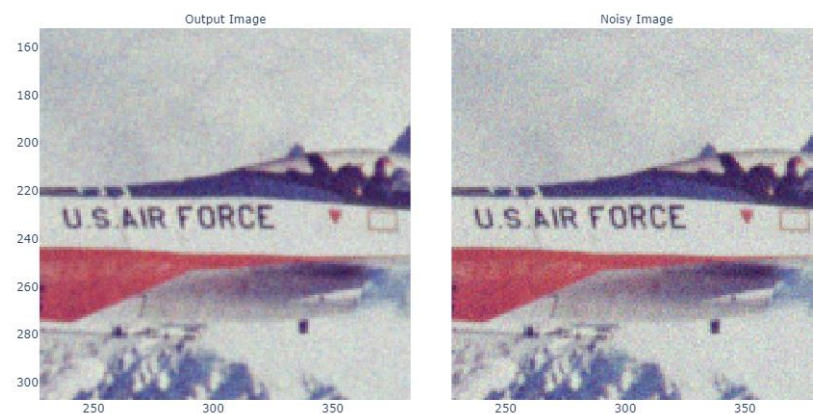


圖二、 有雜訊的 F16 圖片

我使用原始的 DIP 對有些有雜訊的 F16 圖片(圖二)去雜訊，並根據去完雜訊的結果作為基準，下方是訓練結果(圖三)(圖四)、訓練過程(圖五)以及訓練的 PSNR 曲線(圖六)。



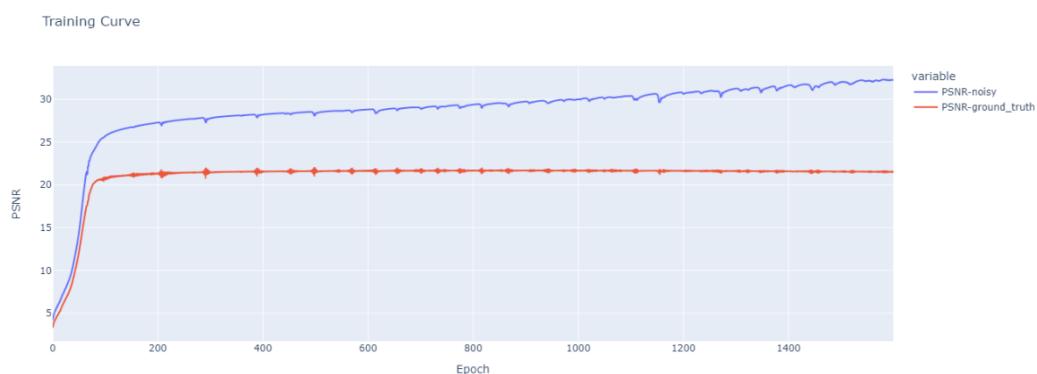
圖三、 原始 DIP 去雜訊對比圖



圖四、 原始 DIP 去雜訊對比圖(放大)



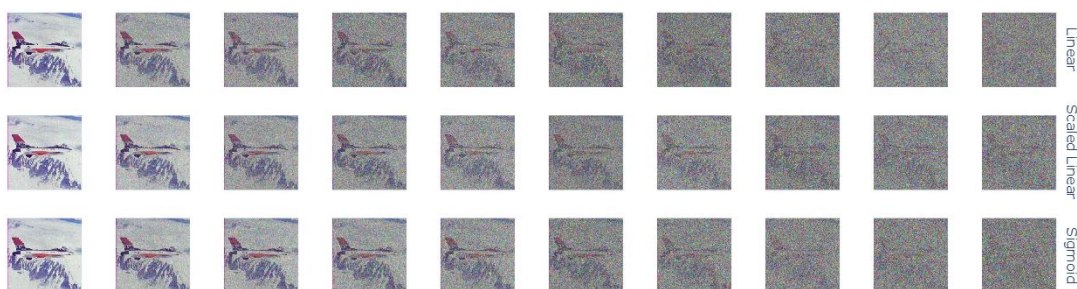
圖五、 原始 DIP 訓練過程



圖六、 原始 DIP 去雜訊訓練 PSNR 曲線

為了可以全面評估 DDPM 逐步去噪監督融入 DIP 訓練框架的新方法，我設計了一系列的控制變量實驗，探索雜訊強度、調度策略、DIP 網路架構以及是否使用感知損失等因素對方法性能上的影響。

下方是使用不同調度策略對圖片加雜訊的圖片(圖七)。

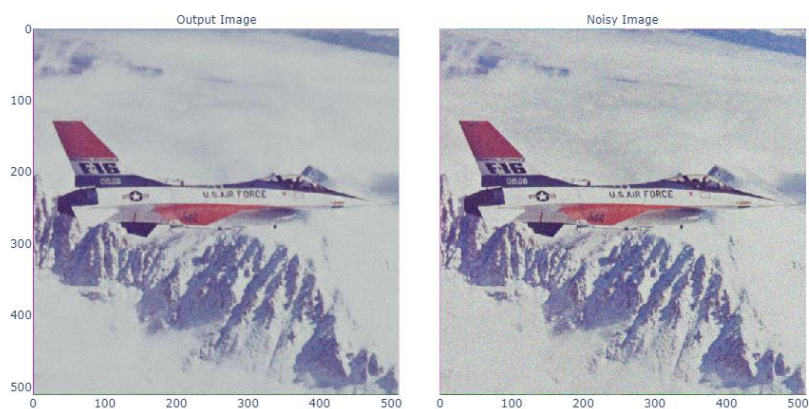


圖七、 不同 Scheduler 對圖片加雜訊的過程比較

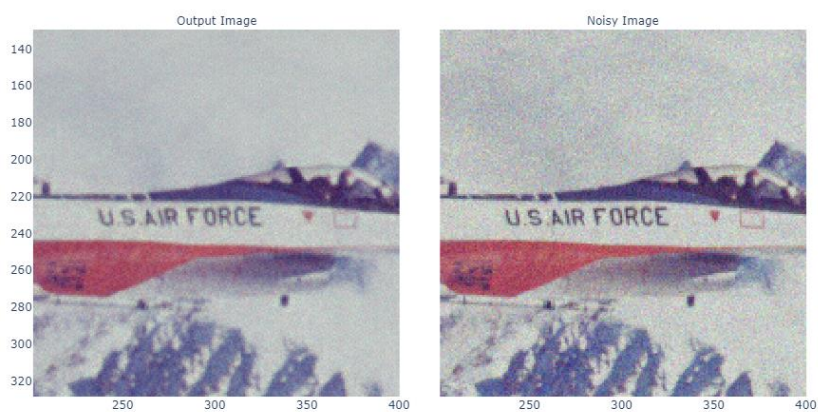


下方是使用不同的網路架構訓練的結果與 PSNR 曲線：

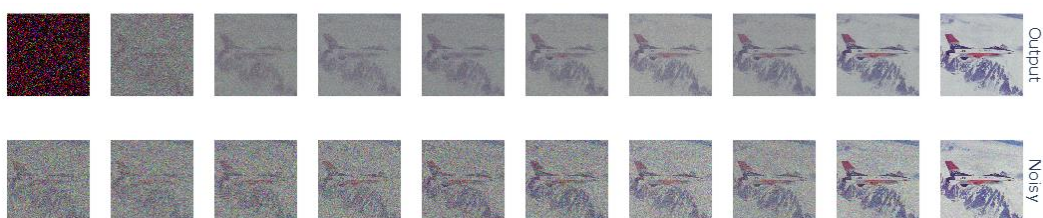
- Unet(圖八-圖十一)



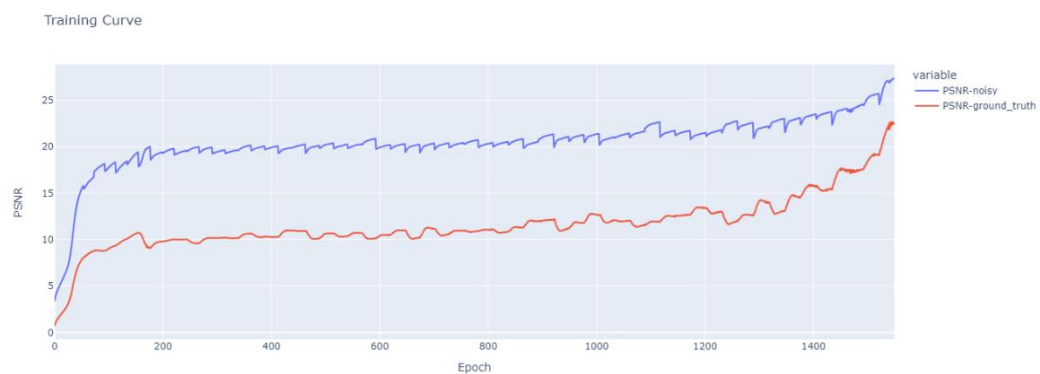
圖八、DDPM+DIP(unet)去雜訊對比圖



圖九、DDPM+DIP(unet)去雜訊對比圖(放大)

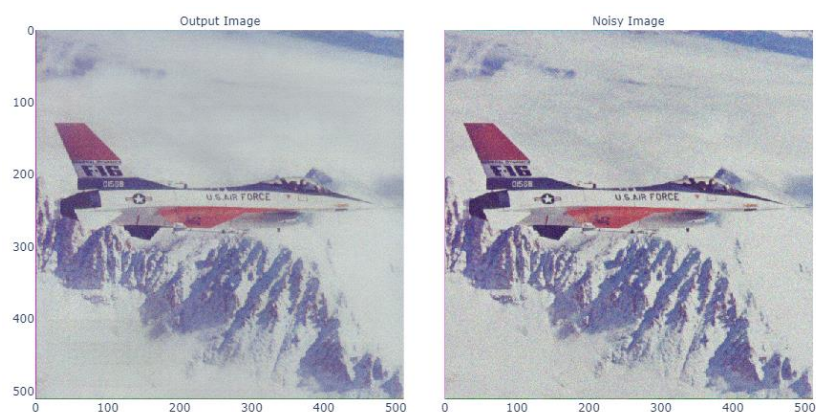


圖十、 DDPM+DIP(unet)訓練過程

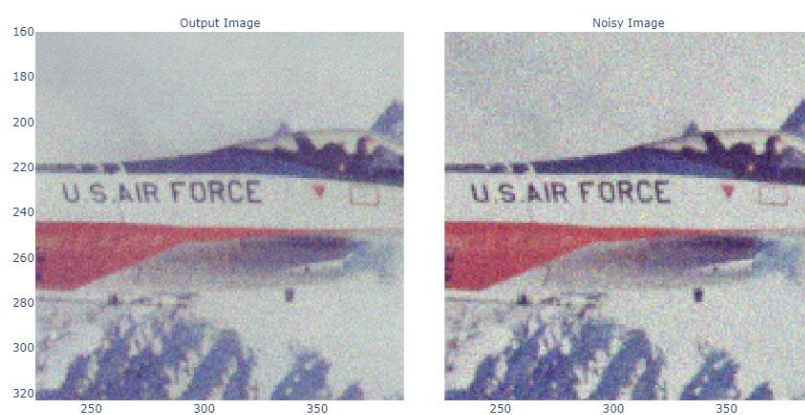


圖十一、 DDPM+DIP(unet)訓練 PSNR 曲線

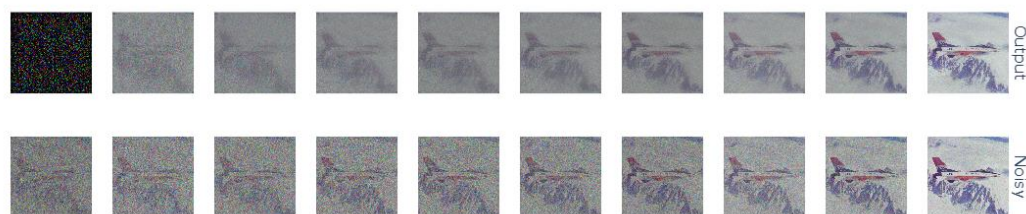
- ResNet18(圖十二-圖十五)



圖十二、DDPM+DIP(ResNet)去雜訊對比圖



圖十三、DDPM+DIP(ResNet)去雜訊對比圖(放大)



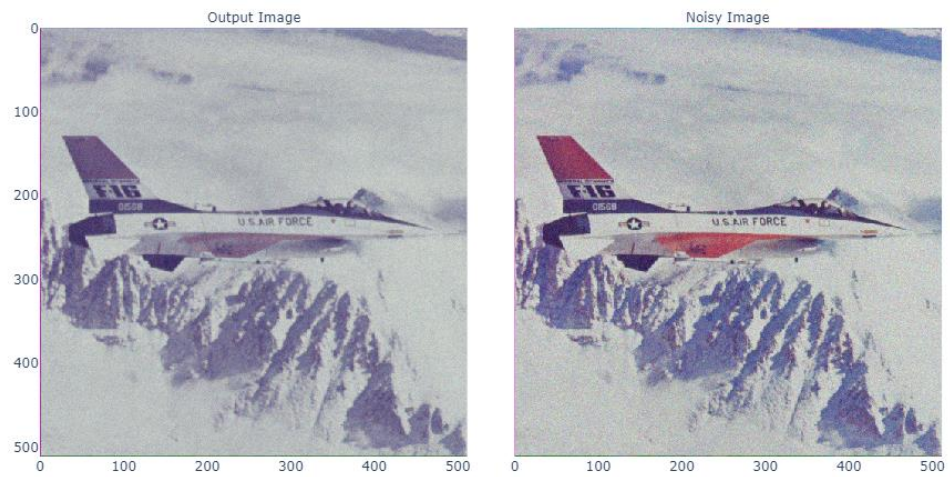
圖十四、DDPM+DIP(ResNet)訓練過程



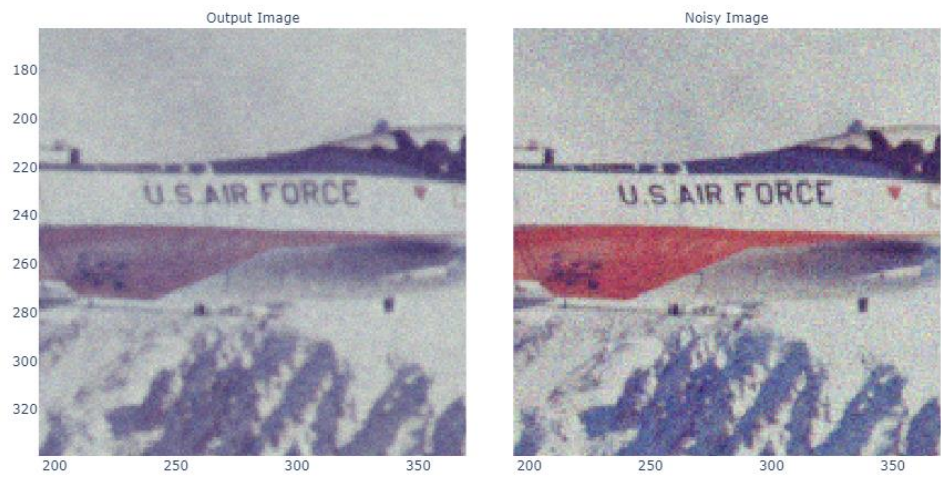
圖十五、DDPM+DIP(ResNet)訓練 PSNR 曲線



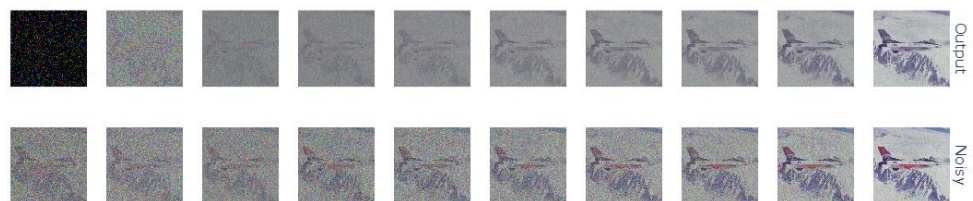
- DnCNN(圖十六-圖十九)



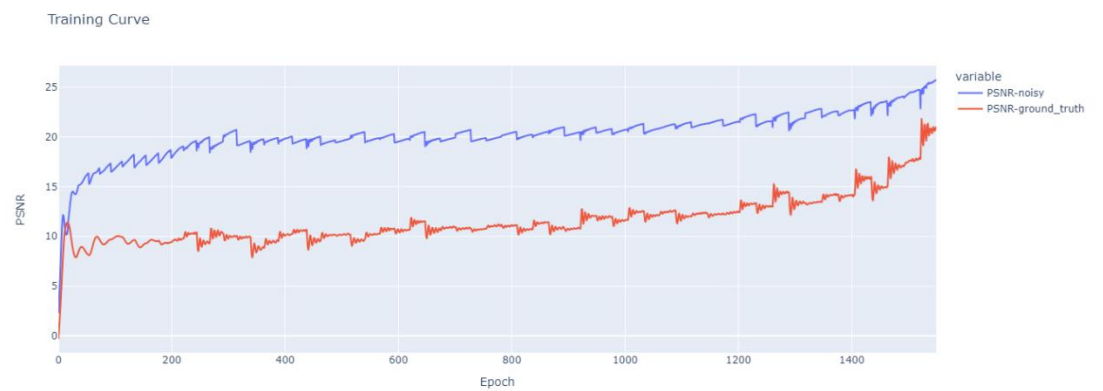
圖十六、DDPM+DIP(DnCNN)去雜訊對比圖



圖十七、DDPM+DIP(DnCNN)去雜訊對比圖(放大)



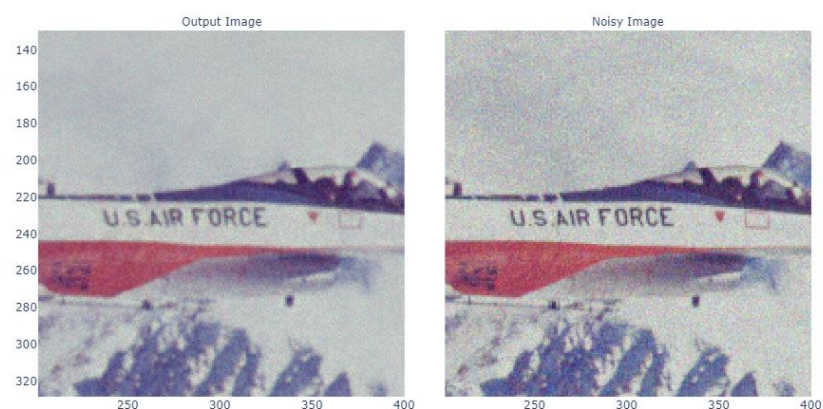
圖十八、DDPM+DIP(DnCNN)訓練過程



圖十九、DDPM+DIP(DnCNN)訓練 PSNR 曲線

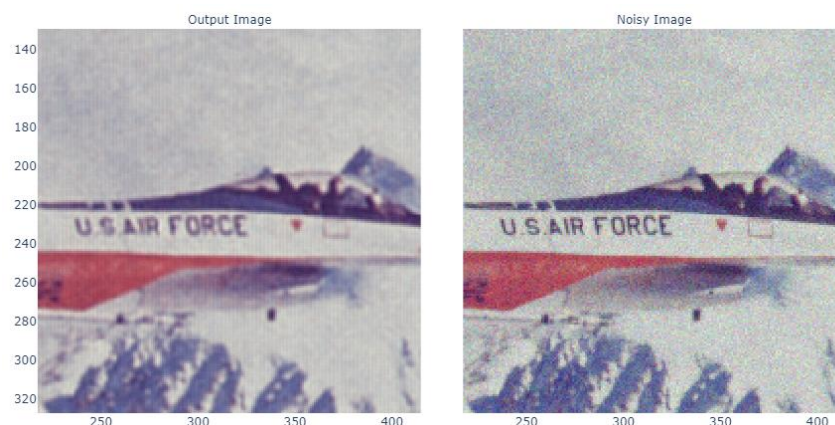
下方分別是使用感知損失與不使用感知損失訓練出來的結果

- 使用感知損失的 unet(圖二十)



圖二十、使用感知損失的結果

- 不使用感知損失的 unet(圖二十一)



圖二十一、沒有使用使用感知損失的結果

由於實驗結果過多，因此我整理部分實驗內容於下方表格中，可以提供參考。

編號	雜訊強度(beta)		scheduler類型	DIP網路架構	有無使用感知損失	PSNR Noisy 分數	PSNR Clear 分數
	beta-min	beta-max					
1	0.001	0.01	linear	unet	無	28.157	24.083
2	0.001	0.01	linear	unet	有	28.4932	24.0118
3	0.001	0.01	linear	resnet	無	31.4185	26.2007
4	0.001	0.01	linear	resnet	有	31.1696	26.6652
5	0.001	0.01	linear	dncnn	無	29.6394	25.1487
6	0.001	0.01	linear	dncnn	有	29.3284	25.777
7	0.001	0.01	scaled linear	unet	無	28.8386	24.3055
8	0.001	0.01	scaled linear	unet	有	28.3089	23.1985
9	0.001	0.01	scaled linear	resnet	無	31.7383	25.208
10	0.001	0.01	scaled linear	resnet	有	31.275	26.2238
11	0.001	0.01	scaled linear	dncnn	無	30.0483	23.8576
12	0.001	0.01	scaled linear	dncnn	有	29.6844	26.8152
13	0.001	0.005	linear	unet	無	27.6523	21.7758
14	0.001	0.005	linear	unet	有	28.7964	25.4856
15	0.001	0.005	linear	resnet	無	30.8027	26.0081
16	0.001	0.005	linear	resnet	有	31.2264	25.0907
17	0.001	0.005	linear	dncnn	無	28.9348	25.4697
18	0.001	0.005	linear	dncnn	有	28.9571	26.5411
19	0.001	0.005	scaled linear	unet	無	28.9473	25.3
20	0.001	0.005	scaled linear	unet	有	27.5057	22.2945
21	0.001	0.005	scaled linear	resnet	無	31.6003	25.6256
22	0.001	0.005	scaled linear	resnet	有	30.9409	24.2465
23	0.001	0.005	scaled linear	dncnn	無	30.0582	25.9969
24	0.001	0.005	scaled linear	dncnn	有	29.8755	25.6349



## 4. 實驗分析

雜訊強度是影響模型訓練和影像重建質量的關鍵因素。在解決影像重建問題時，模型需要學會從一個雜訊分佈中恢復出原始的影像分佈，而雜訊的強度直接決定了這個恢復過程的難度。較高的雜訊意味著模型需要處理更多的干擾因素，需要具備更強的去噪能力，才能逆向重建出高質量的影像。反之，較低的雜訊強度則讓恢復過程變得相對簡單，但也增加了模型過度擬合雜訊細節的風險。因此我認為調節雜訊強度對於控制模型學習的難度和評估其極限性能至關重要。

為了系統地探索雜訊強度對模型表現的影響，我設計了三種不同的添加雜訊的調度策略，即 Linear Beta Scheduler、Scaled Linear Beta Scheduler 和 Sigmoid Beta Scheduler。這些策略通過不同的 beta 值分配方式，生成了具有不同雜訊強度的影像序列。

在訓練過程中，我密切關注了不同雜訊強度下模型的收斂行為。實驗發現，當訓練目標為高雜訊影像時，由於影像中的細節和紋理已被掩蓋，模型的主要任務是還原影像的整體輪廓和邊緣結構，相對來說較為簡單，模型能夠相對迅速收斂並產生理想的重建結果。然而，隨著雜訊強度的逐步降低，過度擬合的問題開始凸顯，驗證了我們最初的假設：DIP 方法在後期訓練時容易陷入擬合噪聲的困境，導致細節損失和影像質量下降。這是因為在低噪聲條件下，模型不僅需要捕捉影像的主要結構，還必須精確重建細節和紋理信息。如果訓練過度，模型就有可能將雜訊也學習為影像的一部分，從而引入失真。

為了解決這個問題，我設計了一種基於雜訊強度動態調節訓練次數的機制。該機制通過一個指數衰減函數，根據當前雜訊強度、預設的最大訓練次數上限、敏感度參數和縮放係數來決定該階段的實際訓練次數。當雜訊較高時，對應的最大訓練次數會被適當減小，避免過度擬合；隨著雜訊逐步減小，該階段可以適度地獲得更多的訓練機會，使模型在低噪聲條件下進一步精細化細節捕捉。

這種策略的關鍵優勢在於，它使得模型能夠在前期重點學習影像的主要結構，後期則集中捕獲細節，最終達到一個適度擬合的最佳狀態。通過調整敏感度參數和縮放係數，可以權衡模型在不同雜訊強度下的訓練時間分配，獲得最佳的重建性能。我也發現透過這樣可以在訓練結束時獲得相對最佳的影像品質並且有很好的去雜訊結果，並且不需要像原始 DIP 一樣訓練這麼多 epoch，達到 Early Stopping 的效果。

在網絡架構選擇方面，我評估了多種常見 DIP 網絡，包括 U-Net、ResNet 和標準卷積網絡。不同架構在捕獲影像階層特徵方面存在偏好差異：U-Net 編碼器-解碼器結構有利於捕獲全局信息，細節保留相對遜色，這樣對於我們要避免擬合雜訊很有幫助。ResNet 和 DnCNN 生成的圖片品質相對 U-Net 就差一些，雖然在去雜訊上表現仍然不錯，但是影像相對來說會比較模糊。除此之外，在訓練上 U-Net 的訓練會穩定一些(可以參考上方 PSNR 的訓練曲線)。

此外，我探索了感知損失在訓練中的作用，與僅使用像素級別的 MSE 損失相比，引入基於 VGG 網絡的感知損失能夠更好地捕捉人眼對影像質量的主觀感知，有效減少了一些結構性和紋理性的失真。不過，過度依賴感知損失也可能導致細節流失，因此我通過調整 MSE 和感知損失的權重比例，權衡細節保留和整體感知質量。

在超參數調整方面，我也嘗試了不同的學習率策略、權重和優化器來獲得最佳的結果。但是我發現更改這些並不會明顯影響生成圖片的結果，基本上對於重建圖片，像是去噪任務來說，影響結果最大的還是噪音強度、DIP 模型選擇。

我綜合評估了生成影像的質量，包含客觀指標 PSNR 和人工視覺評估。在 PSNR 分數指標上，融合了 DDPM 思想的新方法相比於傳統 DIP 雖然沒有顯著提升影像重建質量，甚至沒有提升，但是在我人工視覺的評估下，我認為新方法的去雜訊結果使有明顯比較好的。除此之外，要達到相同水準的生成品質條件下，新方法所需要的訓練時間也會略短於傳統的 DIP 方法。我認為雖然在結果上沒有顯著的提升與進步，但是我認為這個方法的潛能還沒有被完整地開發出來，仍然還有進度的空間。