

Term Project Report

Team 12

1. Introduction

影像超解析度 (Image Super-Resolution, SR) 目標是從低解析度影像重建高解析度影像。以往是使用卷積神經網路 (CNN)，例如 RCAN, EDSR。自從 Transformer 在自然語言處理中獲得不錯的結果後，也開始被應用於影像領域。Activating More Pixels in Image Super-Resolution Transformer 這篇論文提到現有的 Transformer 在 SR 任務中只利用了有限的輸入資訊範圍，Transformer 的潛力尚未完全發揮。為了讓模型利用更多的輸入像素，作者提出了 Hybrid Attention Transformer (HAT) model。

首先，輸入影像會經過一個淺層特徵提取網路，提取出影像的初步特徵，然後會被送入 RHAG module。RHAG module 由多個 Hybrid Attention Block (HAB) 和 Overlapping Cross-Attention Block (OCAB) 組成。HAB module 結合了 Channel Attention 和 Window-based Self-Attention 機制，可以有效地利用全局統計資訊，並且具有局部擬合能力。OCAB module 則用於增強相鄰窗口特徵之間的交互作用，可以更好地整合跨窗口資訊。RHAG module 會重複許多次，逐漸提取出更深層的特徵。最後再將深層特徵重建成高分辨率的輸出影像。(Fig1.)

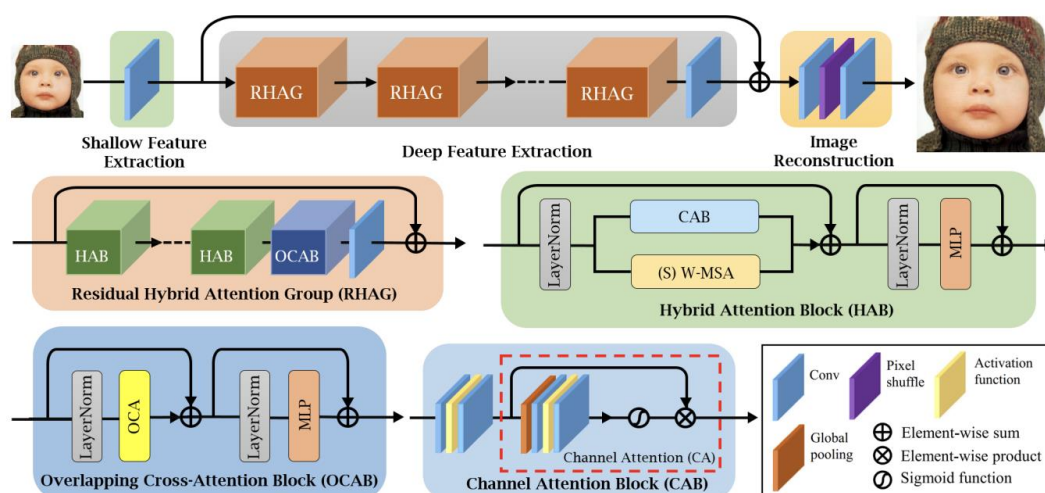


Fig1. HAT model architecture

2. Method

我們使用了論文提供的 pretrained GAN-based HAT model。GAN 的 Generator 目標是生成與真實影像無法分辨的影像，因此 GAN-based model 可以生成更逼真、自然的結果，克服了傳統方法容易產生模糊的問題。而 GAN 的 Discriminator 負責區分真實影像和生成影像，迫使 Generator 生成具有更豐富細節和具真實紋理的影像。GAN 的訓練過程可以有效地學習影像的複雜細節，因此 GAN-based SR model 在重建細節方面表現更為出色。

在 HAT model 的基礎上，我們進一步探討了後處理 (Post-processing) 的影響，並對六種不同的後處理方法進行了評估，包括：Gaussian Blur, Median Blur, Laplacian Sharpening, Unsharp Masking, Histogram Equalization (HEQ), Contrast Limited Adaptive Histogram Equalization (CLAHE)。透過主觀的人眼判斷，我們覺得 Unsharp Masking 的結果最清楚 (Fig2.)。我們還採用了兩種客觀評估指標：NIQE (Natural Image Quality Evaluator) 和 BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator)。NIQE 是一種無參考影像品質評估指標，它根據影像的統計特性來評估影像品質。BRISQUE 也是一種無參考影像品質評估指標，它根據影像的局部紋理和結構來評估影像品質。我們對不同後處理方法的結果進行了客觀評估，結果顯示 Unsharp Masking 的結果在 NIQE 和 BRISQUE 指標上皆取得了最佳的分數 (Fig3.)，因此選擇 Unsharp Masking 作為後處理方法。

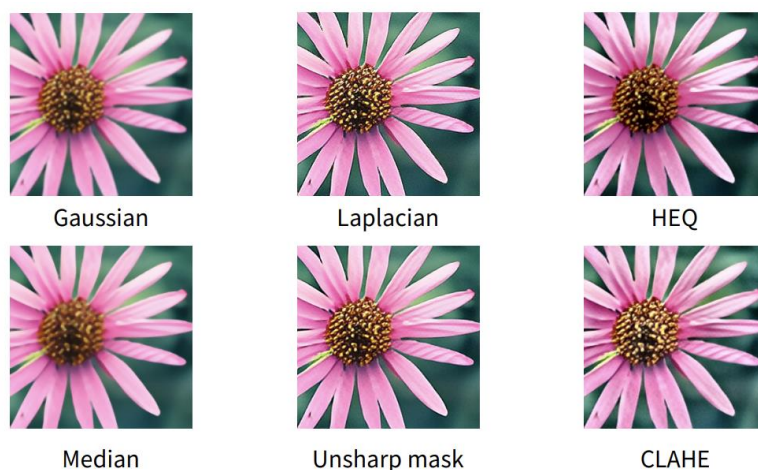


Fig2. Post-processing results

15 images average	None	Gaussian	Median	Laplacian	Unsharp mask	HEQ	CLAHE
NIQE	5.83	7.87	8.65	8.91	4.86	5.09	4.99
BRISQUE	19.71	45.39	43.66	48.83	15.60	22.22	19.48

Fig3. Metrics evaluation

此外，我們也探討了重複使用 HAT model 來放大圖片再縮小是否能夠得到更清晰的影像。研究中我們發現將原始影像經過兩次 HAT model 放大 16 倍，然後使用 Python 內建的 LANCZOS filter 將影像縮小四倍，可以進一步提升影像的清晰度。特別是若先經過 Unsharp masking 處理再縮小影像，其效果比起先縮小再經過 Unsharp masking 處理更為自然，對比度更佳且邊緣處理更為平滑。(Fig. 4)



1xHAT + Filter

2xHAT + downscale + filter

2xHAT + filter + downscale

Fig4. Upscale vs. Upscale*2 then downscale

這樣的方法顯示了在影像處理中，先增強再縮小的流程能顯著改善最終的影像品質，使其細節更豐富且視覺效果更佳。這對於需要高品質圖像的應用場景，如醫學影像分析或數位影像修復等，具有重要的參考價值。

3. Discussion

在我們的研究中，Unsharp Masking 獲得了最好的結果，我們將討論其優於其他方法的原因，以及使用 HAT model 放大影像再縮小的策略能獲得更清晰影像的原因。

在主觀視覺評估和客觀品質指標（NIQE 和 BRISQUE）中，Unsharp Masking 優於其他方法可以歸因於以下幾點：

1. 能有效增強邊緣對比，使影像看起來更加清晰。
2. 能保留圖像中的細節，而不會產生過多的失真或模糊。
3. 能夠靈活調整參數，適應不同的需求。

相比之下，其他後處理方法則存在一些缺點。Gaussian Blur 和 Median Blur 會產生模糊效果，雖然能夠減少噪點，但是會使影像失去銳利程度。Laplacian Sharpening 有時會過度增強邊緣，導致影像看起來不太自然。Histogram Equalization (HEQ) 和 Contrast Limited Adaptive Histogram Equalization (CLAHE) 主要是調整影像的對比度，可能導致影像出現過度飽和或不自然的顏色，而且在細節增強方面效果不如 Unsharp Masking。

此外，將原始圖像經過 HAT model 放大 16 倍，再使用 Python 內建的 LANCZOS filter 將圖像縮小 4 倍，可以顯著提升影像的清晰度，其原理如下：第一次將影像放大 4 倍，能將影像中的細節資訊放大，但同時也會引入一些雜訊。再次將影像放大 4 倍，可以進一步放大影像細節，並利用第一次放大過程中產生的細節資訊來彌補雜訊和模糊。最後將放大後的影像縮小四倍，可以去除放大過程中引入的雜訊，同時保留放大後的細節資訊。雖然經過兩次 HAT model 放大 16 倍再縮小 4 倍看起來像是做白工，但實際上這個過程利用了 HAT model 細節增強能力，以及連續放大縮小操作的雜訊抑制和模糊消除效果，因此提升了影像的清晰度。

4. Conclusion

在此 term project 中，我們探討了使用 Hybrid Attention Transformer (HAT) model 進行影像超解析度重建的方法，並進一步分析了多種後處理技術的效果。通過主觀視覺評估和客觀品質指標 (NIQE 和 BRISQUE) 的測試，我們發現 Unsharp Masking 是最有效的後處理方法。此外，將影像經過 HAT 模型放大再縮小的策略也顯著提升了影像的清晰度。以下是我們的主要結論：

1. HAT model 能夠有效提取和增強影像中的細節，克服了傳統 CNN model 的局限性。透過多層次的 RHAG module，HAT model 充分利用了輸入影像中的資訊，使得重建出的高解析度影像在細節和整體品質上均優於傳統方法。
2. 在各種後處理技術中，Unsharp Masking 能夠有效增強影像邊緣對比和細節，且不會引入過多的失真或模糊。這使 Unsharp Masking 成為提升影像清晰度的最佳選擇。
3. 將影像經過兩次 HAT model 放大 16 倍，再使用 LANCZOS filter 縮小 4 倍的策略，顯著提升了影像的清晰度。這個方法利用了 HAT model 的細節增強能力，並通過多次放大和縮小，有效抑制了雜訊和模糊，進一步提升影像品質。

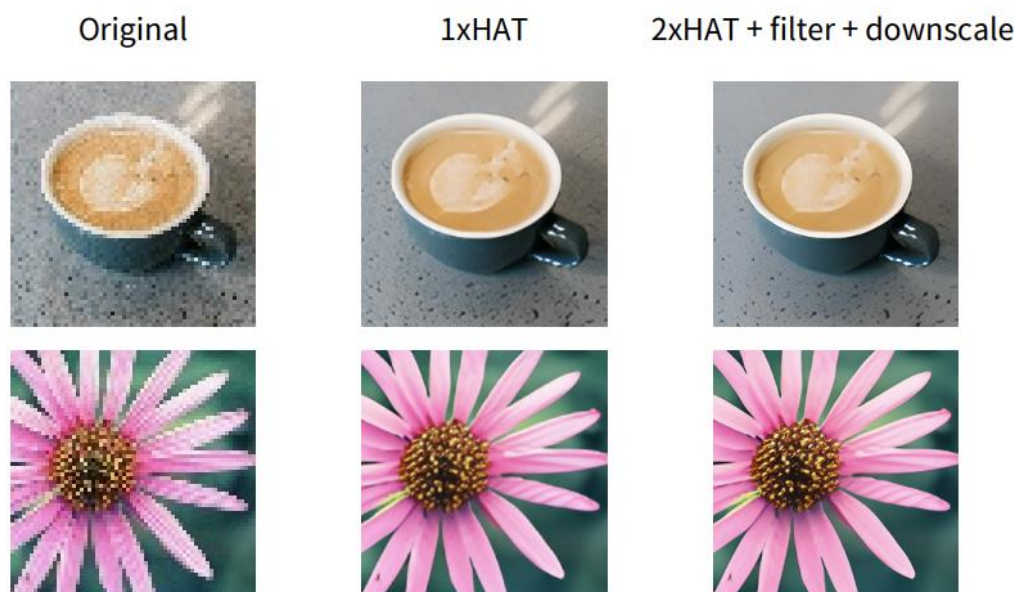


Fig5. Results