

國立虎尾科技大學資訊工程系
專題製作報告

畫派辨識與風格轉換系統

Art Style Recognition and Style Transfer System

參與成員：陳蓉霞 林元傑

指導教授簽名：黃建宏

中華民國 114 年 6 月 3 日

摘要

近年來，基於深度學習的圖像處理技術（如卷積神經網路）在藝術應用中取得重大進展。例如，神經風格轉移（Neural Style Transfer, NST）已廣泛應用於數位藝術創作，能夠將畫派風格進行轉換。此外，圖像分類模型已成功運用於藝術品的自動分類與分析，幫助博物館進行藝術品數位化管理。本研究旨在結合人工智慧與藝術創作，在先前研究的畫派辨識模型訓練方法上做調整，並融入神經風格轉移，設計成一套具教育性與互動性的畫派辨識與風格轉換系統，以達到畫派分類自動化，及實現藝術創作普及化。

本研究主要專注於七個畫派的處理，分別是巴洛克、立體派、盛世文藝復興、印象派、極簡主義、普普藝術及浮世繪。最終畫派辨識模型的整體辨識準確率約有85%，而風格轉換的部分，平均 Gram Matrix 差異為 1.55，SSIM 約為 0.30，整體呈現良好的風格轉換與內容結構保留。系統亦具備操作介面與畫派知識介紹，適用於藝術教育與創作輔助。

關鍵字：電腦視覺、圖像處理、深度學習、神經風格轉移

Abstract

In recent years, deep learning-based image processing techniques, such as Convolutional Neural Networks (CNNs), have made significant progress in artistic applications. For example, Neural Style Transfer (NST) has been widely used in digital art creation, enabling the transformation of artistic styles. In addition, image classification models have been successfully applied to the automatic classification and analysis of artworks, assisting museums in the digital management of their collections. This study aims to integrate artificial intelligence with artistic creation by refining previous training methods for art style recognition models and incorporating NST. The goal is to develop an educational and interactive system for art style recognition and style transfer, achieving automated art style classification and promoting the popularization of artistic creation.

This study primarily focuses on seven art styles: Baroque, Cubism, High Renaissance, Impressionism, Minimalism, Pop Art, and Ukiyo-e. The final model achieved an overall classification accuracy of approximately 85%. For the style transfer component, the system achieved an average Gram Matrix difference of 1.55 and a Structural Similarity Index (SSIM) of approximately 0.30, indicating effective style transformation while preserving the structural integrity of the original content. In addition, the system incorporates a user-friendly interface and provides informative content on each identified art style, making it suitable for applications in art education and creative assistance.

Keywords: Computer Vision, Image Processing, Deep Learning, Neural Style Transfer

目錄

研究動機與目的	4
系統設計	4
研究方法	4
研究結果	11
結論與未來研究方向	13
誌謝	13
團隊分工	14
參考文獻	15

研究動機與目的

隨著美術館的數位化，更多觀眾更容易接觸到各家美術館的藏品，以奇美博物館的數位典藏為例，任何人都可以輕易地線上欣賞其雕塑、繪畫、樂器等收藏品。然而這衍伸出了一個問題，如何有效分類作品以便於檢索，以繪畫為例，常以藝術家或畫派作為分類依據，若由人工來分類，往往費工也費時，因此如何達到繪畫分類自動化勢在必得。卷積神經網路（Convolutional Neural Network, CNN）因其在影像辨識上的優異表現，已被多項研究應用於畫派分類，並展現出不俗的準確率。基於此，本研究參考相關文獻中的模型訓練方法，並依據本研究選擇的七種畫派進行調整與優化。同時，我們亦結合神經風格轉移技術，建立一套整合畫派分類與藝術風格轉換的完整系統，期望在提升分類效能的同時，促進數位藝術創作的普及與應用。

系統設計

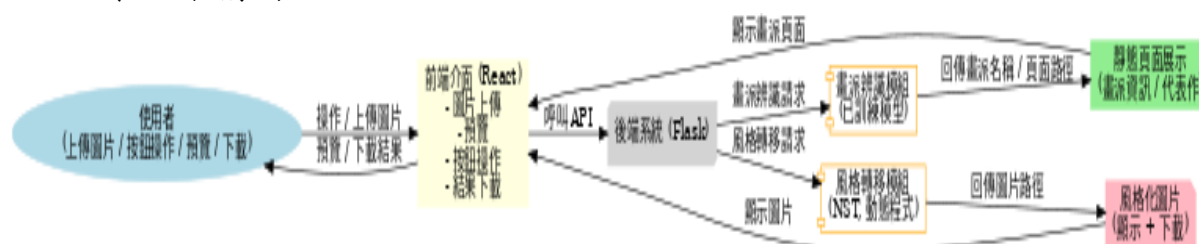
● 系統環境

軟體：Python 3.8、HTML、JavaScript、Flask、PyTorch

硬體：12th Gen Intel(R) Core(TM) i5-12500H、16.0 GB RAM、NVIDIA GeForce RTX 3050 Laptop GPU

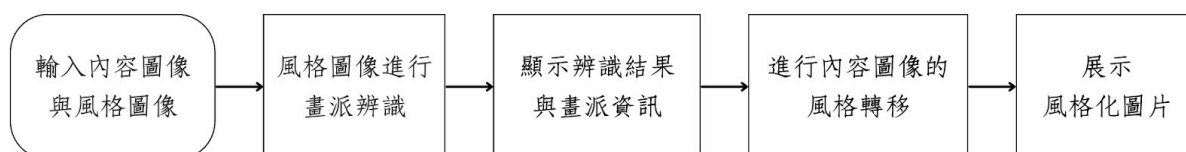
通訊設備/協定：HTTP、JSON

● 系統架構圖



圖一、系統架構圖

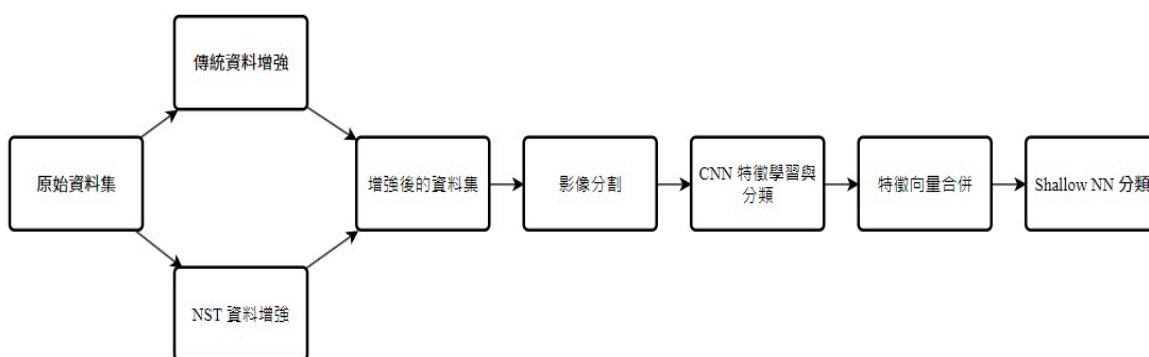
研究方法



圖二、系統流程圖

畫派辨識

本研究在畫派辨識的部分提出一種新的方法，結合神經風格轉移作為資料增強技術，並採用兩階段深度學習分類框架，以提升藝術風格分類的準確性。整體方法包含兩個核心部分：（1）透過 NST 與傳統資料增強技術進行訓練資料的強化，（2）運用兩階段深度學習模型，對影像區塊進行特徵學習，再整合結果進行最終分類。



圖三、畫派辨識模型訓練流程圖

● 資料集

本研究採用的資料集為 WikiArt dataset，是一個來自 WikiArt.org 的大型藝術畫作圖像資料集，收錄超過 80,000 幅來自全球知名藝術家的作品，涵蓋中世紀至當代的多種藝術風格。資料集中每張圖像通常附有風格、藝術家、創作年份等標籤，被廣泛應用於藝術風格分類、藝術家辨識、風格轉換等深度學習與計算藝術研究領域。由於其豐富多樣的圖像與標註資訊，WikiArt 成為藝術類圖像辨識任務中的標準資料集之一，但也存在風格界線模糊與類別不均等挑戰。

本研究目前僅採用該資料集中的 7 個畫派，分別是巴洛克、立體派、盛世文藝復興、印象派、極簡主義、普普藝術及浮世繪，目的是為了降低分類難度，及加速訓練與實驗的進行。

● 資料前處理

1. 資料平衡

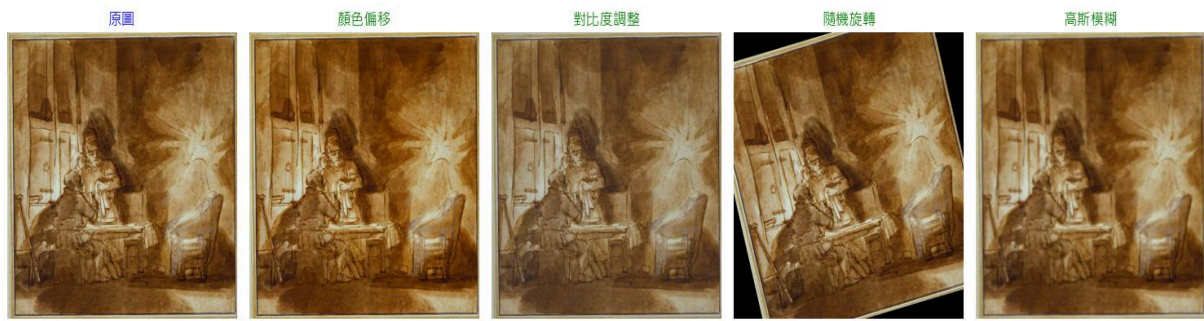
資料集中該 7 種畫派圖片數不平衡，分別為巴洛克：4240，立體派：2235，盛世文藝復興：1343，印象派：13060，極簡主義：1337，普普藝術：1483，及浮世繪：1167。為了達到資料平衡，本研究採取了以下作法，如果畫派的圖片數達到 2000，則隨機不重複複製圖片到相對應的資料夾；如果畫派的圖片數不足 2000，則先在目標資料夾放入約 50 張 NST 處理過後的畫派圖片，並將全部該畫派的圖片複製到相對應的資料夾後，再隨機選擇該畫派的圖片，隨機做四種資料增強方式的其中之一，直到圖片數達到 2000。這四種資料增強方式分別是顏色偏移、對比度調整、隨機旋轉及高斯模糊，都是圖像處理中常見的資料增強方法。

顏色偏移：改變影像的色相與亮度，以模擬不同光源環境。

對比度增強：強調細節，使畫作中的筆觸與紋理特徵更加明顯。

隨機旋轉：增加影像的角度變化，使模型對圖像方位具有更強的適應能力。

高斯模糊：模擬不同畫作的柔和度，增強模型對紋理的適應性。



圖四、資料增強對比圖

2. 訓練集及測試集分割

平衡完後的資料集會被分為訓練集及測試集，比例分別為 80% 及 20%。也就是說每個畫派 2000 張圖片中，會有 1600 張作為訓練集，400 張作為測試集。

3. 動態分割

本研究參考了「Two-Stage Deep Learning Approach to the Classification of Fine-Art Paintings」的作法，在模型訓練前對每張圖片進行動態分割，切割為上左、上右、下左、下右及中間，中間部分與其他四個部分重疊。這五小塊加上原圖會再調整大小為 224x224，以符合後續模型的輸入規格。



圖五、動態分割示意圖

● 模型訓練

我們參考的文獻中，是採用兩階段深度學習模型來進行畫派的分類，以提升準確率，第一階段使用深度卷積神經網路，第二階段則使用淺層神經網路。第一階段為了節省

訓練時間與資源，及提升模型泛化能力，作者使用了遷移學習，這部分我們採用了結構簡單、效果穩定的 VGG16。以下將完整說明訓練流程。

1. 第一階段 - VGG16

為了適應本研究的分類任務，我們移除了 VGG16 預設的全連接層，並加上自訂的分類層，包括一層具備 L2 正則化的全連接層、Dropout 層以及最終的 softmax 分類輸出層。整體模型使用 Adam 優化器進行訓練，並加入 ReduceLROnPlateau 與 EarlyStopping 等策略以動態調整學習率與避免過度訓練。訓練資料為每張原始圖片與其五個切割 patches 所組成的影像群組，共六張圖對應同一標籤，使模型能同時學習圖像的整體與局部風格特徵。

2. Patch 加權平均最佳化

在完成第一階段模型訓練後，每張輸入圖像及其五個對應的局部區域（patches）皆可透過 VGG16 模型獲得對應的畫派預測機率分佈。為了整合這六張圖像的預測資訊，本研究採用加權平均的方式進行機率融合，以指定的權重組合進行加總，形成單一的整合預測結果。目前我們採用等權重平均，意即每張圖像與其 patches 對最終預測具有同等貢獻。

3. 第二階段 - Shallow NN

在完成加權平均機率整合後，第二階段模型使用淺層神經網路（Shallow Neural Network）對這些機率向量進行進一步分類判斷。Shallow NN 的輸入為加權平均後的機率向量分佈。透過簡單的全連接層結構與 Dropout、Batch Normalization 等設計，Shallow NN 學習如何更有效地綜合這些預測結果，進行更精確的畫派分類。模型訓練過程中使用了早停機制（EarlyStopping）避免過擬合。最後模型輸出七個畫派的機率向量分佈，本研究取各畫派預測機率的最高值作為最終預測的畫派。

● 模型評估

1. 準確率（Accuracy）：衡量模型預測正確的比例。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

TP（True Positive）：模型正確分類為某畫派的樣本數。

TN（True Negative）：模型正確分類為非該畫派的樣本數。

FP（False Positive）：模型錯誤分類為某畫派的樣本數。

FN（False Negative）：模型錯誤分類為非該畫派的樣本數。

2. F1 分數（F1-score）：綜合精確率（Precision）和召回率（Recall），用來衡量類別不均衡時的分類效能。

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

精確率（Precision）：衡量模型預測為該畫派的樣本中，真正屬於該畫派的比例。

$$Precision = \frac{TP}{TP + FP}$$

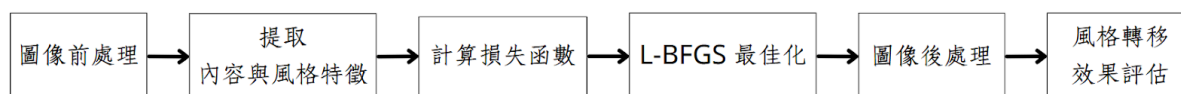
召回率（Recall）：衡量所有實際屬於該畫派的樣本中，被正確辨識出的比例。

$$Recall = \frac{TP}{TP + FN}$$

3. 混淆矩陣（Confusion Matrix）：用來顯示模型在各個畫派分類上的正確與錯誤情況。

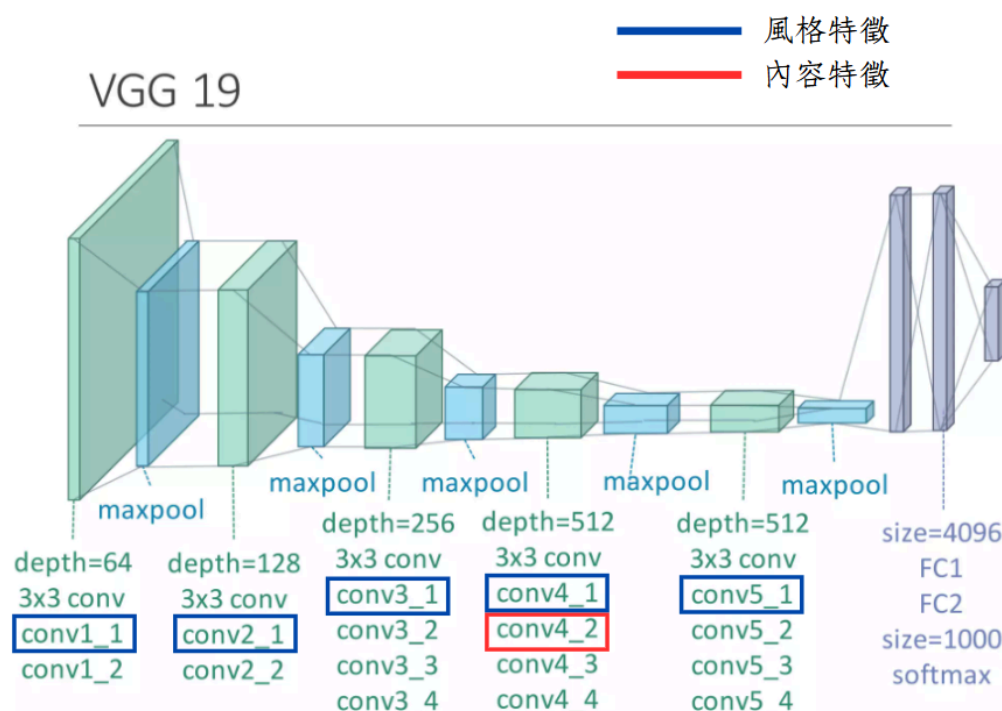
風格轉移

本研究運用神經風格轉移進行畫風轉換，透過 CNN 進行學習風格特徵並生成風格化圖像。我們採用 VGG-19 模型作為特徵提取器，綜合內容損失、風格損失與總變分（Total Variation, TV）損失進行最佳化，並透過 L-BFGS 演算法進行圖像更新，以獲得高品質的風格轉移結果。



圖六、風格轉移之流程圖

- 影像前處理
在進行風格轉移之前，需先對輸入影像（內容與風格影像）進行標準化處理。主要步驟如下：
 1. 圖像縮放至統一尺寸（如 512x512 像素），確保計算效率與視覺效果的平衡。
 2. 轉換為浮點數格式並正規化，使像素值介於 $[0, 1]$ 或 $[-1, 1]$ 之間。
 3. 依據 VGG-19 模型的要求，進行均值標準化，使其符合模型預訓練數據分佈。
- 提取內容與風格特徵
採用 VGG-19 模型作為特徵提取網路，該模型為 ImageNet 預訓練的深度卷積神經網路。我們從不同層提取內容與風格特徵，並計算相應的損失函數。
 1. 內容特徵：從 conv4_2 層提取內容影像與合成影像的特徵圖，並透過均方誤差（MSE）計算內容損失。
 2. 風格特徵：從 conv1_1、conv2_1、conv3_1、conv4_1、conv5_1 層提取風格影像與合成影像的 Gram 矩陣，計算風格損失。



圖七、內容特徵與風格特徵提取

- 損失函數設計（Loss Function）

為了有效進行風格轉移，本研究使用三種損失函數來引導圖像更新。內容損失（Content Loss）確保生成圖像保留內容圖像的主要結構，風格損失（Style Loss）使生成圖像的紋理與風格圖像一致，而總變分損失（TV Loss）平滑生成圖像，使其更自然具連續性。

1. 內容損失（Content Loss）

$$L_{content} = \frac{1}{2N_l H_l W_l} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2$$

其中， F^l 為生成影像的特徵圖， P^l 為內容影像的特徵圖， N_l 、 H_l 、 W_l 分別為該層的通道數、高度與寬度。

2. 風格損失（Style Loss）

$$L_{style} = \sum_l w_l \cdot \frac{1}{4N_l^2 H_l^2 W_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2$$

其中， G^l 與 A^l 分別為合成影像與風格影像在第 l 層的 Gram 矩陣， w_l 為該層的權重。

3. 總變分損失（Total Variation Loss, TV Loss）

$$L_{TV} = \sum_{i,j} (I_{i,j+1} - I_{i,j})^2 + (I_{i+1,j} - I_{i,j})^2$$

該損失函數用於平滑影像，減少像素間的劇烈變化，避免雜訊與人工紋理。

4. 總損失函數（Total Loss）

$$L_{total} = \alpha L_{content} + \beta L_{style} + \gamma L_{TV}$$

其中， α 、 β 、 γ 為權重超參數，決定內容、風格與平滑損失的相對影響力。

● 最佳化（Optimization）

在圖像生成過程中，本研究使用 L-BFGS 方法來更新影像，使其收斂至最小損失。該方法能夠有效利用歷史資訊進行二階近似，以提升收斂速度。

1. 設定初始圖像為內容圖像。
2. 在每次迭代中，根據損失函數計算梯度，並使用 L-BFGS 進行圖像更新。
3. 當損失變化趨於穩定或達到最大迭代次數，則停止最佳化。

● 圖像後處理（Post-processing）

1. 進行像素值正規化（Rescale），確保像素範圍介於 $[0, 255]$ 。
2. 轉換為整數格式（uint8）以符合圖像輸出需求。
3. 展示最終風格化影像。

● 風格轉移效果評估（Evaluation）

為了量化本研究所實現之風格轉移系統的轉換品質，我們採用 Gram 矩陣相似度（Gram Matrix Similarity）與結構相似性指數（Structural Similarity Index, SSIM）作為評估指標，分別衡量風格一致性與內容保留程度。

1. Gram 矩陣相似度：透過計算生成圖像與風格圖像之 Gram 矩陣的均方誤差（MSE），衡量風格相似程度，誤差越小表示風格轉移效果越佳。

$$L_{Gram} = \frac{1}{N} \sum_{i,j} (G_{ij}^{gen} - G_{ij}^{style})^2$$

其中， G_{ij}^{gen} 與 G_{ij}^{style} 分別代表合成影像與風格影像在第 i, j 通道之 Gram 矩陣元素， N 為 Gram 矩陣的總元素數量。

2. 結構相似性指數 (SSIM)：用來衡量內容圖像與生成圖像的結構相似度，數值介於 0 到 1 之間，當越接近 1，代表內容保留效果越佳。

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

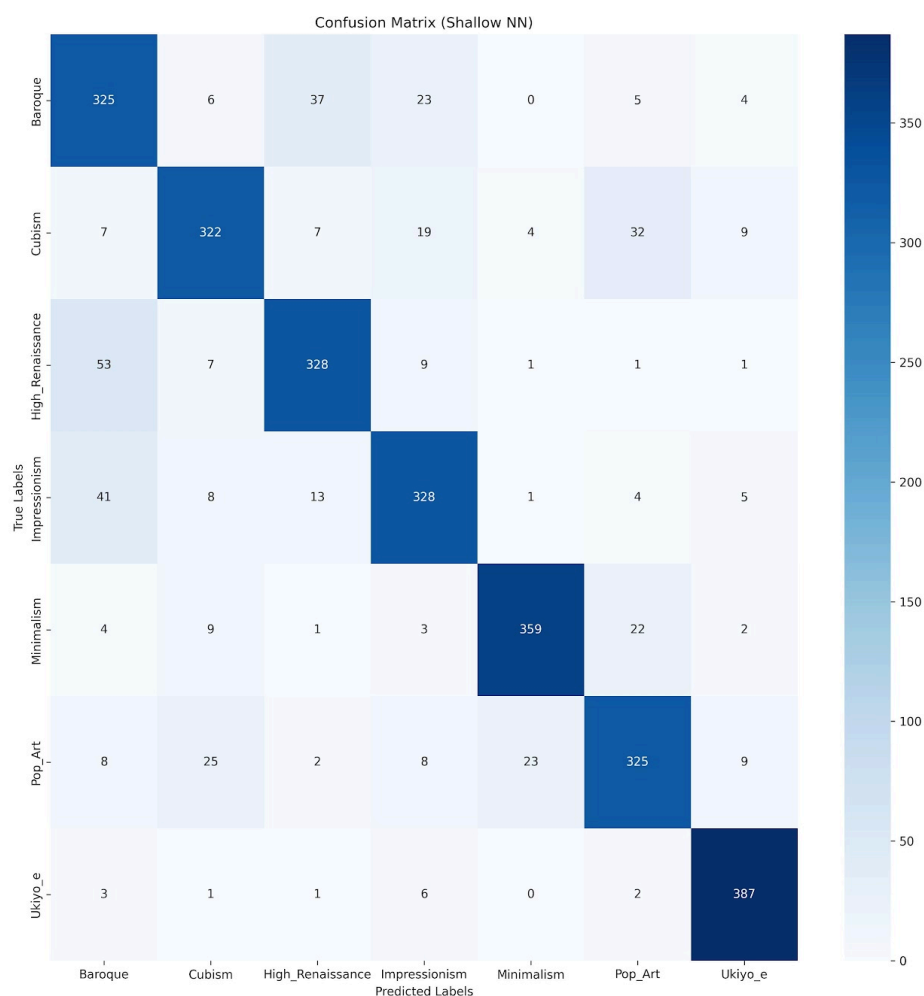
其中， x 與 y 分別代表內容影像與合成影像的像素值矩陣， μ_x 、 μ_y 為兩者的平均值， σ_x 、 σ_y 為標準差， σ_{xy} 為兩影像之間的共變異數， C_1 與 C_2 為避免分母為零的穩定因子。

系統整合

前端介面使用 React 設計，提供圖像上傳、預覽與結果下載功能；後端則以 Flask 框架整合模型推論與風格轉換操作，並提供 API 介面與圖片路徑回傳。

研究結果

最終七個畫派的混淆矩陣如圖八，而各畫派的其他分類評估指標（精確率、召回率、F1分數及準確率）則詳見圖九，整體的辨識準確率約有85%。由圖可看出，某幾個畫派是彼此易混淆的類別，如（巴洛克，盛世文藝復興）、（巴洛克，印象派）及（立體派，普普藝術），這也是後續我們希望改善的部分。



圖八、Shallow NN 混淆矩陣

Art_Style	Precision	Recall	F1-score	Support
Baroque	0.74	0.81	0.77	400
Cubism	0.85	0.81	0.83	400
High_Renaissance	0.84	0.82	0.83	400
Impressionism	0.83	0.82	0.82	400
Minimalism	0.93	0.90	0.91	400
Pop_Art	0.83	0.81	0.82	400
Ukiyo_e	0.93	0.97	0.95	400

圖九、各畫派分類評估指標

神經風格轉移中，評估指標包含 Gram Matrix 差異（Gram Diff）與結構相似度指標（SSIM），用以評估風格強度與內容結構保留程度。每組畫派皆搭配相同的內容圖與風格圖進行三次實驗，固定迭代次數為 750 次，平均後作為該畫派之指標參考依據。最終各畫派風格轉換評估指標的結果彙整為圖十。

Art_Style	Gram_Diff	SSIM
Baroque	0.5333	0.27
Cubism	3.1867	0.2867
High_Renaissance	0.7467	0.2833
Impressionism	0.14	0.41
Minimalism	0.64	0.39
Pop_Art	2.9967	0.2633
Ukiyo_e	2.62	0.1733
Average	1.5519	0.2967

圖十、各畫派風格轉換評估指標

整體平均 Gram Diff 和 SSIM 分別為 1.55 及 0.297，顯示在保有基本的內容結構下，皆可成功擷取風格特徵，達到風格轉換的效果。而部分實驗結果中出現 Gram Diff 過高或過低的情況，可能為轉換幅度不足或風格過度強化，需進一步調整損失權重以平衡內容與風格之間的取捨。

而文藝復興（High Renaissance）與極簡主義（Minimalism）為轉換效果較不理想之畫派。畫面常呈現不自然的平鋪或輪廓模糊現象，說明這類風格較難透過目前模型架構成功轉換。文藝復興畫派多以寫實細節為主，極簡主義則過於抽象、色塊單純，皆可能導致風格特徵難以充分學習並作用於內容圖。

結論與未來研究方向

本研究成功將畫派辨識與風格轉移整合成系統，並提供各畫派的背景知識，為藝術學習者和創作者提供即時分析工具，提升對該畫派的認識，也幫助他們快速實現藝術創作。畫派辨識模型的部分整合了 NST 資料增強與傳統資料增強，加上兩階段深度學習分類框架，進一步提升分類的準確性。風格轉移方面，使用 L-BFGS 演算法進行圖像更新，使輸出結果在風格化與原始內容保留之間達成良好平衡。經實驗評估，整體 Gram Diff 平均值為 1.55，代表風格特徵成功轉移；SSIM 平均為 0.297，顯示轉換後的圖像能有效保留內容結構。此外，系統亦提供進度條、結果預覽與圖片下載等功能，提升使用體驗。

未來的研究除了可以專注於畫派的擴充，也可以專注於解決畫派間混淆的問題，如加入創作年代、藝術家等資訊作為模型輔助輸入；或結合多種模型，利用綜合判斷來降低混淆；也可以改用多階段分類架構，降低相近畫派間的混淆。在風格轉移方面，未來可進一步導入多風格混合或風格強度控制參數，讓使用者自由調整轉換效果，提升創作彈性；亦可考慮加入主觀評分機制，結合使用者回饋，進而提升風格轉移品質與使用者滿意度。

誌謝

本研究為國科會大專學生研究計畫之相關成果，國科會計畫編號為：114-2813-C-150-011-E。

年度	學生姓名	執行機關	內容
114	陳晉霖	國立虎尾科技大學資訊工程系	計畫名稱：基於深度學習之畫派辨識與風格轉移系統 計畫編號：114-2813-C-150-011-E 成果報告：無電子檔 執行起迄：2025/07/01~2026/02/28 指導教授：黃建宏 核定金額：58,000元

圖十一、大專生計畫補助

團隊分工

工作項目 / 組員	陳蓉霞	林元傑
畫派資料集收集與統整		✓
畫派資料集前處理		✓
畫派分類相關文獻統整		✓
畫派辨識模型訓練與調整		✓
畫派辨識模型評估		✓
NST 相關文獻收集與統整	✓	
NST 圖像處理	✓	
NST 相關程式碼撰寫	✓	
NST 效果評估	✓	
畫派背景資訊收集與統整		✓
前端靜態網頁設計		✓
後端系統設計與整合	✓	
前後端整合	✓	
報告書撰寫	✓	✓

圖十二、工作分配表

參考文獻

- [1] Sandoval, C., Hyon, Y., & Kim, J. (2019). "Two-stage deep learning approach to the classification of fine-art paintings." *IEEE Access*, 7, 41770–41781. <https://doi.org/10.1109/access.2019.2907986>
- [2] Zheng, X., Chalasani, T., Ghosal, K., Lutz, S., & Smolic, A. (2019). "STaDA: Style transfer as data augmentation." *arXiv preprint arXiv:1909.01056*.
- [3] Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). "A neural algorithm of artistic style." *arXiv preprint arXiv:1508.06576*.
- [4] Simonyan, K., & Zisserman, A. (2014). "Very deep convolutional networks for large scale image recognition." *arXiv preprint arXiv:1409.1556*.
- [5] Liu, C., Liu, J., & Belongie, S. (2019). "Improving style transfer with wavelet transforms." *arXiv preprint arXiv:1906.01283*.
- [6] LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." *Nature*, 521(7553), 436–444.
- [7] Chang, M. (2016). "Applied deep learning 11/03 convolutional neural networks." Lecture slides, National Taiwan University. [https://www.csie.ntu.edu.tw/~yvchen/f105 adl/](https://www.csie.ntu.edu.tw/~yvchen/f105%20adl/)
- [8] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). "How transferable are features in deep neural networks?" *Advances in Neural Information Processing Systems*, 27 (NeurIPS 2014), 3320–3328.
- [9] Nocedal, J. (1980). "Updating quasi-Newton matrices with limited storage." *Mathematics of Computation*, 35(151), 773–782.
- [10] Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). "A limited memory algorithm for bound constrained optimization." *SIAM Journal on Scientific Computing*, 16(5), 1190–1208.
- [11] Kingma, D. P., & Ba, J. (2015). "Adam: A method for stochastic optimization." *International Conference on Learning Representations (ICLR) 2015*.

