航測及遙測學刊　第三十一卷　第 1 期　第 37- 45 頁　民國 115 年 3 月
J Journal of Photogrammetry and Remote Sensing
Volume 31, No.1, 2026, pp. 37 – 45
DOI：10.6574/JPRS.202603_31(1).0003

使用深度學習進行快速無參考標的無人機影像品質評估

林亞立 [1*]　蘇冠秦 [1]　鄒來翰 [1]　林昭宏 [2]　饒見有 [2]
賴威伸 [3]　胡智超 [3]

摘要

隨著無人飛行載具(UAV)應用於基礎設施監測，影像品質穩定性成為影響深度學習與攝影測量精度的關鍵。然而 UAV 影像常受環境干擾影響，現行品質篩選多仰賴人工檢視，且傳統結構相似度指標(SSIM)需參考影像並受限於空間對齊，難以實務應用。為此，本研究提出一套基於 Swin-Unet 的快速無參考影像品質評估方法。首先設計改良型 CLIP-SSIM 結合 Swin-Transformer，建立高精度影像品質圖(RMSE = 0.0193)，再以該品質圖作為標註資料訓練 Swin-Unet 模型，使單張影像推論時間降至 0.3 秒，並維持良好準確度(RMSE = 0.04)。結果顯示，本方法可有效取代人工檢視流程，滿足高頻 UAV 影像應用需求。

關鍵詞：影像品質評估、深度學習、無人機影像、影像結構相似度指標

1. 前言

無人機因具備高效率、低成本、高機動性與低風險等優勢，已廣泛應用於基礎建設監測、地形測繪、災後勘查與環境監控等多元任務(Rakha & Gorodetsky, 2018)。尤其在橋梁、道路、水壩與電塔等難以親近或具潛在危險性的設施上，無人機可大幅降低作業人力需求與風險負擔，提升檢測作業的時效性與空間覆蓋率。伴隨影像感測與控制技術之進步，無人機得以快速擷取高解析度圖像，為智慧基礎設施管理提供豐富的視覺資料來源。然而，無人機所拍攝之影像品質常受到外部環境因素干擾，例如強風造成飛行震動、光照變化導致曝光不均，以及拍攝角度變化引發畫面偏移等，皆可能造成影像模糊、變形與對比不足，進而影響後續結構健康監測、三維重建與劣化偵測等任務之精度與可靠性(Sieberth *et al.*, 2015)。特別是在進行攝影測量與深度學習推論等任務時，影像品質的微小變異即可能對成果產生顯著影響。因此，建立一套客觀、快速且可自動化執行之影像品質評估(Image Quality Assessment, IQA)機制，作為影像前處理與篩選之依據，已成為實務應用中亟需解決的重要課題。

傳統影像品質評估方法如峰值訊雜比(Peak Signal-to-Noise Ratio, PSNR)(Horé & Ziou, 2010)與結構相似性指數(Structural Similarity Index Measure, SSIM)(Wang *et al.*, 2004)雖被廣泛應用於影像品質評估研究中，但相關文獻亦指出其在實務應用上仍存在明顯限制。PSNR 主要基於像素強度差異進行計算，對影像結構變化與人眼主觀感知之關聯性有限，已被證實在多種影像失真情境下與人類視覺評價之相關性不足。相較之下，SSIM 雖進一步考量亮度、對比與結構資訊，在模糊與幾何失真評估上具備較佳表現，然而該方法仍需仰賴參考影像進行比較，且對影像尺寸、裁切位置與像素對齊條件高

[1] 國立成功大學測量及空間資訊學系 博士生
[2] 國立成功大學測量及空間資訊學系 教授
[3] 交通部運輸研究所運輸工程組 研究員
[*] 通訊作者, E-mail: alecfree2@gmail.com

收到日期：民國 114 年 11 月 19 日
修改日期：民國 114 年 12 月 18 日
接受日期：民國 115 年 01 月 30 日

度敏感。此一特性使得 SSIM 難以直接應用於無人機巡檢等缺乏對應參考影像，且拍攝視角與尺度變化頻繁之實務情境，進而限制其實際可行性。為解決此問題，近年來興起的深度學習技術提供了無需參考影像的評估方案。透過卷積神經網路(Convolutional Neural Network, CNN)、殘差網路(Residual Network, ResNet)(Kang *et al.*, 2014、Ma *et al.*, 2017)，以及具備全局特徵建模能力之 Transformer 架構(Vaswani *et al.*, 2017)，影像品質評估得以轉化為具備語意理解能力的分類或回歸任務，具備更高的判別力與自適應能力。然而，深度學習模型之訓練效果高度仰賴大量且高品質之標註資料，而無人機影像常處於環境多變、光線與構圖條件差異大的情境下，造成標註資料取得困難，進而限制模型於實務領域的泛化能力與應用潛力。

為解決上述挑戰，本研究提出一個二階段影像品質評估架構。首先以基於 Swin－Transformer (Liu *et al.*, 2021)的機率加權模型產生高精度影像品質圖，作為訓練標註；再透過 Swin-Unet 架構(Cao *et al.*, 2021)，進行全影像推論取代逐像素預測，兼顧推論速度與準確性，使本架構能有效支援無人機快速影像品質評估任務，具有實務應用潛力。本研究聚焦於橋梁檢測任務中無人機影像品質評估的實務挑戰。橋梁檢測作業往往需於高風速、背光或遮蔽等複雜環境下執行飛行任務，致使所拍攝影像品質參差不齊。由於橋梁結構細節眾多，裂縫、剝落等劣化特徵常需仰賴高品質影像才能準確辨識，影像品質的不穩定將直接影響後續深度學習模型之偵測準確性與三維建模結果的可靠性。因此，本研究旨在發展一套具效率、無需參考影像且適用於橋梁檢測環境的自動化影像品質評估架構，期能有效支援橋梁劣化檢測與結構安全監測工作，提升無人機應用於基礎建設巡檢作業的整體效能。

2. 研究方法

人機在每次任務中往往會拍攝大量影像，尤其是在橋梁巡檢任務中，為求完整覆蓋橋體結構，通常會以多角度、多視角方式進行連續拍攝，動輒產生數百至上千張高解析度影像。在此情境下，若仍仰賴傳統人工方式逐張進行影像品質評估與篩選，不僅耗費大量人力與時間，更無法滿足效率與高頻率的實務需求，進而可能延誤檢測進度或降低整體資料處理效率。為解決此問題，本研究建立一套自動化的影像品質評估流程，使用實地無人機拍攝的橋梁影像資料集，並訓練三種不同目的之模型以應對實務需求，如圖 1 所示。

首先，從資料集中選取 600 張高品質影像進行直方圖均化(histogram equalization) (Pizer *et al.*, 1987)，統一亮度與對比度，並將這些高品質影像評分為 1.0。接著，以演算法對這些高品質影像製作各種程度的退化影像，用以模擬真實無人機環境中的干擾條件，如：亮度變化、對比變化、高斯模糊、水平與垂直移動模糊。接著，則透過 CSSIM 指標對退化影像進行評分，再以 CSSIM 指標預測分數作為標註，並取對應的退化影像作為基於 Swin-Transformer 架構之機率加權模型之訓練資料。該模型會針對每像素周圍 500x500 區域進行分析，逐像元偵測 CSSIM 數值，生成高解析度的影像品質圖，成功實現無參考影像品質評估。 然而，因該模型採逐像素推論方式，計算時間成本過高，不利於即時應用。為提升實用性，本研究進一步以該模型所生成的影像品質圖作為標註影像，訓練一個基於 Swin-Unet 架構之模型，讓模型可以直接推論整張影像的影像品質圖，大幅提升推論速度與實務應用彈性。

2.1 CSSIM 影像品質指標

SSIM 為傳統常用之客觀影像品質指標，主要根據亮度、對比與結構三項元素進行評分，其中結構分數依賴區域內兩張影像的共變異數計算。因此，SSIM 需在影像對齊的前提下進行評估，對深度學習模型因影像裁切所導致的像素錯位極為敏感，進而影響評估準確性。為解決此問題，本研究提出 CSSIM 指標，將原本依賴兩影像共變異數之結構分數，替換為透過 CLIP 模型(Radford *et al.*, 2021)之圖像編碼器(image encoder)所提取之高階特徵向量間的餘弦相似度，如式(1)所示。於本研究中，CLIP

僅作為影像特徵編碼工具使用，而非重新訓練之深度學習模型，其可將輸入之影像轉換為固定維度之特徵向量表示。由於影像在輸入編碼器前會經由模型內部之標準化處理，不同尺寸之影像皆可被轉換為相同維度之特徵向量，進而進行一致之相似度計算。

CSSIM 指標之計算流程如圖 2 所示，其係以高品質影像(品質分數為 1.0)與對應之退化影像作為比較對象，透過影像特徵相似度量測方式，評估兩影像間之結構差異程度。具體而言，影像首先經由 CLIP 之圖像編碼器轉換為特徵向量表示，其中高品質影像所對應之特徵向量記為 $CLIP_x$，退化影像所對應之特徵向量記為 $CLIP_y$，並以兩者之餘弦相似度作為結構相似性量測基礎。該結構相似性結果再與亮度與對比度等影像品質相關因素進行綜合計算，以得到最終之 CSSIM 指標，並作為後續影像品質圖生成模型之訓練標註。

$$Cosine\ Similarity\left(CLIP_x, CLIP_y\right) = \frac{CLIP_x \cdot CLIP_y}{\|CLIP_x\| \cdot \|CLIP_y\|} \quad \text{.....(1)}$$

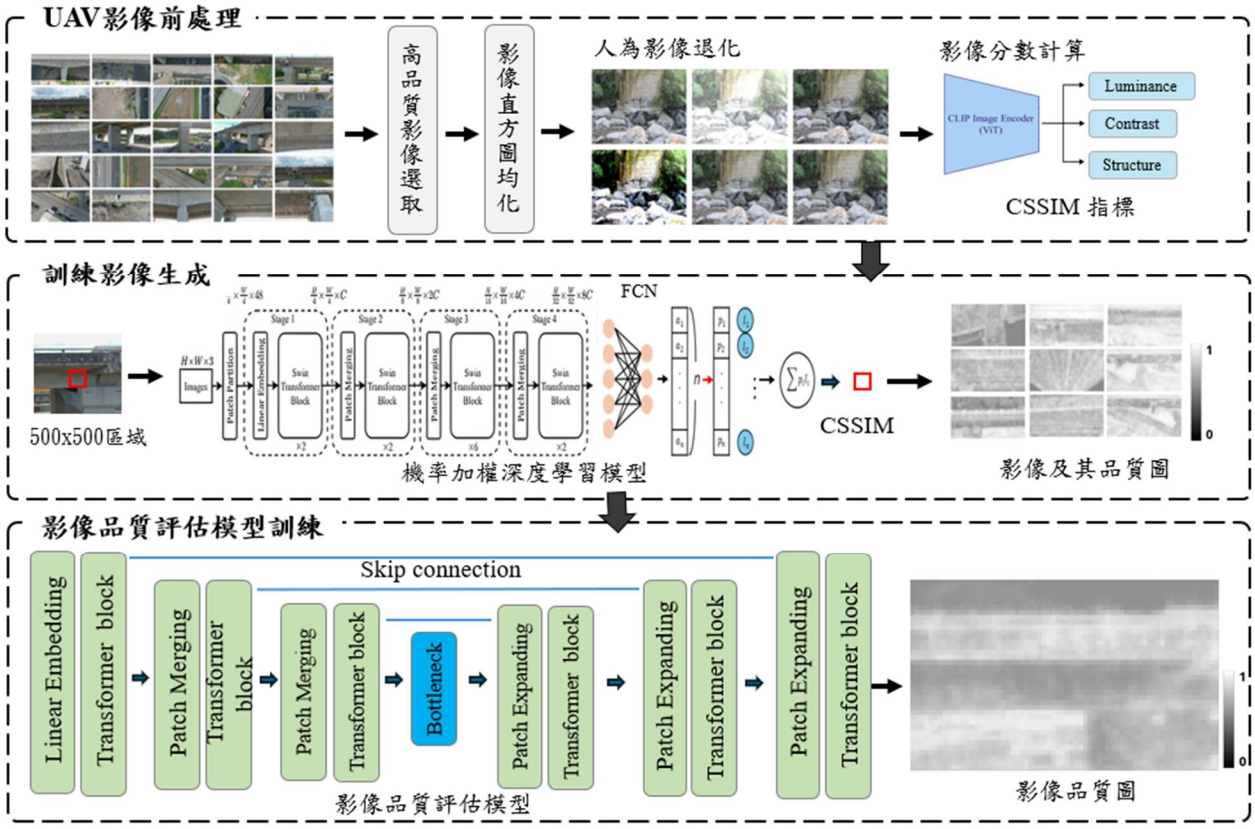


圖 1 快速無參考影像評估架構圖

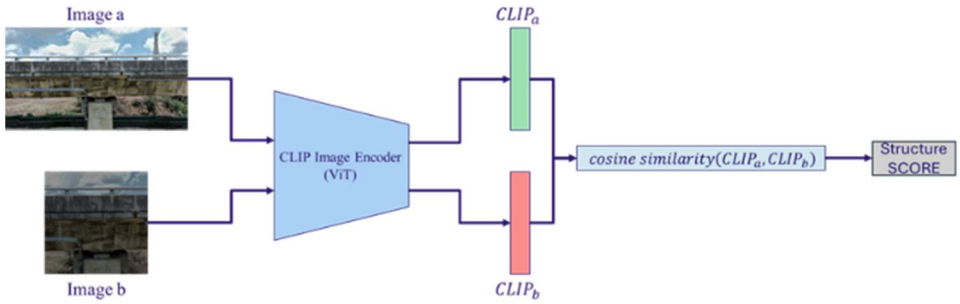


圖 2 CSSIM 結構評分計算架構圖

2.2 基於 Swin Transformer 影像品質圖生成模型

本研究建立一個基於 Swin-Transformer 架構的機率加權模型，以生成高品質的影像品質圖，如圖 3 所示。模型以影像中裁切之區域影像作為輸入，首先透過 Swin Transformer 骨架進行階層式特徵萃取，隨後經由全連接網路(FCN)輸出對應於不同品質區間之預測機率分佈。各品質區間代表 CSSIM 值域中之一段連續範圍，模型透過學習輸入影像特徵與品質區間之關聯性，輸出該區域落於各品質區間之機率。最終再利用各區間之代表數值與其對應機率進行加權計算，取得單一之 CSSIM 預測數值，作為該區域之影像品質評估結果。在骨架方面，Swin Transformer 模型具備自注意力機制、滑動視窗與階層式結構，可有效捕捉全局語意與多尺度特徵，提升品質預測之準確性。在解碼器設計上，本研究將影像品質預測任務由傳統回歸問題轉換為分類形式，藉由學習品質區間之機率分佈來進行數值推估。相較於直接回歸連續數值，此一機率加權策略可降低單一數值預測不穩定所造成之影響，並有助於改善回歸模型中常見之訓練不穩定與梯度消失問題。

由於 CSSIM 分數之計算需考量目標像元周圍之結構資訊，無法僅依賴單一像元之像素值進行預測，因此本研究採用區域式預測策略。模型會以目標像元為中心，擷取其周圍 500×500 影像區域作為輸入，並以該區域中心像元所對應之 CSSIM 數值作為預測目標，使模型能透過周圍結構特徵推估中心位置之影像品質。透過此設計，模型可在無需參考影像之情況下，逐一對影像中各像元進行 CSSIM 數值預測，進而生成具備空間連續性之影像品質圖。然而，此逐像元區域推論方式需對影像中大量像元重複進行預測，導致推論時間與運算資源消耗顯著增加。因此，本研究進一步利用上述模型所產出之影像品質圖作為訓練標註資料，建立一套可直接對整張影像進行品質圖生成之高效率模型，以因應實務應用情境下對即時性與運算效率之需求。

2.3 即時無參考影像品質評估模型— Swin Unet 模型

為提升影像品質圖生成效率，本研究採用 Swin-Unet 作為最終影像品質評估模型架構，如圖 4 所示。模型整體採用編碼器－解碼器架構，輸入為整張無人機影像，經由編碼器逐步萃取階層式影像特徵，並透過跳接(skip connection)將編碼階段之局部結構資訊傳遞至解碼器，以於解碼階段回復空間解析度，最終輸出與輸入影像同尺度之影像品質圖。

在此架構下，模型所輸出之每一像素值係對應於該位置之影像品質評估結果，其數值範圍與 CSSIM 指標一致。此一對應關係係透過訓練階段以退化影像及其對應之影像品質圖(由前述逐像元 CSSIM 預測模型產生)作為監督訊號所建立，使模型得以學習 CSSIM 在影像空間中的分佈特性，而非重新定義評估指標本身。因此，本研究將影像品質圖生成問題視為類語意分割任務，透過像素層級的品質值回歸，直接推估整張影像中各區域之品質分佈情形，並可進一步透過空間整合方式反算整張影像之 CSSIM 指標。此策略不僅顯著降低推論時間，亦維持良好預測穩定性，有效支援無人機應用場域中對於快速影像品質評估之實際需求。

3. 成果與討論

本研究所使用之影像資料為橋梁檢測任務中所取得之橋梁表面影像，皆由研究團隊實地操作無人機進行拍攝。影像皆採用高解析度相機取得，原始解析度為 5472 × 3076 像素，具備足夠細節以支援後續劣化辨識與三維建模等應用。資料內容涵蓋多種橋梁構件表面材質與拍攝角度，並反映實務巡檢作業中常見之光照變化、視角偏移與局部遮蔽等情境，如圖 5 所示。

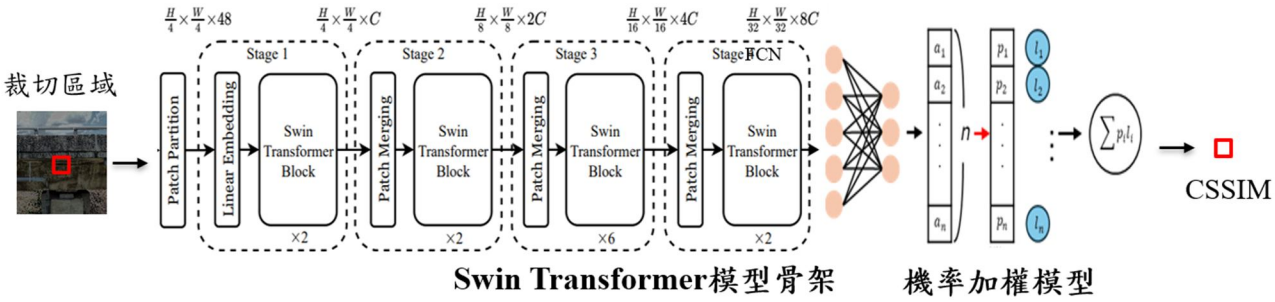


圖 3 基於 Swin Transformer 模型架構圖

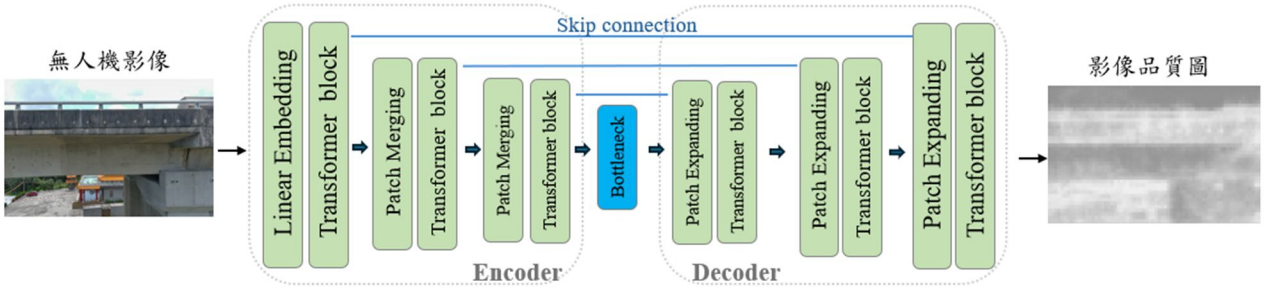


圖 4 Swin-Unet 影像品質評估模型架構圖

圖 5 實地無人機拍攝的橋梁影像資料集

在資料前處理階段，首先由原始影像中篩選出 600 張影像結構清晰、無明顯模糊或過曝現象之影像，作為高品質影像樣本。高品質影像為依據橋梁構件可清楚辨識且不影響後續缺失判釋之工程判準，而非主觀品質評分結果。為了降低初始影像亮度與對比差異對後續品質退化分析之影響，本研究對所選取之高品質影像進行直方圖均化處理，以建立一致之影像品質基準。在完成基準對齊後，透過演算法方式對上述高品質影像施加不同程度之模擬退化處理，以產生對應之退化影像，模擬無人機實際巡檢環境中可能遭遇之光照變化與影像干擾情形。透過此方式，可在受控條件下建立具參考對應關係之影像品質資料，進一步用於影像品質評估模型之訓練與驗證。所有模型訓練與實驗皆於配備 NVIDIA GeForce RTX 4060 Ti GPU 與 64 GB 記憶體之本地工作站上進行，作業系統為 Windows 平台。

3.1 基於 Swin Transformer 的機率加權模型

基於 Swin Transformer 的機率加權模型之訓練參數設定如下：訓練圈數(epoch)為 50，學習率為 10e-5，優化器為 ADAM，損失函數為 RMSE。該模型由骨架與解碼器部分所構成，為評估模型效能，本研究比較多種不同的骨架，如：CNN、ResNet-18、Vision Transformer 與 Swin Transformer，並針對本

研究所設計之四種不同預測策略進行比較，包括傳統數值回歸模型，以及三種機率加權分類模型。後者係將連續之影像品質分數離散為不同解析度之品質區間，分別對應 11、21 與 41 個分類區間，用以評估不同品質分級解析度設定下之預測表現，其成果詳列於表 1。實驗結果顯示，Swin Transformer 架構在所有預測策略中皆優於其他骨架模型，且機率加權分類方法明顯優於傳統回歸模型，能更靈活捕捉品質變化細節，提升整體預測準確度。分類區間數由 11 提升至 21 時，預測效果明顯提高；惟進一步提升至 41 時，誤差僅略有降低(約 0.004)，顯示模型已擷取主要資訊，再增加分類數效益有限。因此，本研究選用使用 21 類分類區間之 Swin Transformer 模型作為產出高精度影像品質圖之最佳模型。

3.2 Swin Unet 影像品質評估模型

為降低逐像素推論所需之運算成本，本研究以高精度影像品質圖作為訓練資料，並採用 Swin-Unet 架構作為影像品質圖預測模型。訓練參數設定如下：訓練圈數為 150，學習率為 10e-5，優化器採用隨機梯度下降法(Stochastic Gradient Descent, SGD)，以確保在模型訓練過程中具備穩定且可控之收斂行為，損失函數同樣採用 RMSE。該模型與基於 Swin Transformer 的機率加權模型的模型效能比較如表 2。結果顯示基於 Swin Transformer 的機率加權模型雖可達最高預測精度(RMSE = 0.019)，但平均每張影像推論時間為 532.3 秒，實務應用受限；相較之下，Swin-Unet 影像品質評估模型雖預測精度略降(RMSE = 0.044)，但推論速度大幅提升，每張影像僅需 0.3 秒，展現極佳即時性，適用於無人機巡檢等快速應用情境。

此外 Swin-Unet 影像品質評估模型的預測成果如圖 6 所示。此外，Swin-Unet 影像品質評估模型之預測成果如圖 6 所示。圖中之顏色條表示模型所預測之影像品質分數分佈，其中顏色由淡色至深色分別對應由低至高之影像品質數值，較低品質區域代表影像中可能存在亮度不足、過曝或模糊等失真情形。從品質圖中可以具體呈現畫面中品質下降的區域，並能對應原始影像中的失真類型。例如，案例 1 右下角柱子亮度過高，存在過曝現象；案例 2 在橋梁底下拍攝整體亮度較暗；案例 3 因背光造成橋側亮度過低； 案例 4 與案例 3 為同一張，但加上移動模糊，導致影像的品質較低；而案例 5 與案例 6 整體無明顯問題，可排除為低品質影像。

表 1 不同模型骨架與預測策略下的影像品質圖精度比較

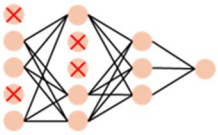
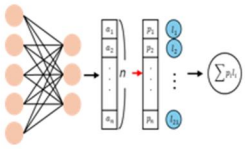
預測策略	骨架	RMSE
 回歸模型	CNN	0.1648
	ResNet-18	0.1571
	Vision Transformer	0.0751
	Swin Transformer	**0.0523**
機率加權模型 (11個分類區間)	CNN	0.0855
	ResNet-18	0.0794
	Vision Transformer	0.0444
	Swin Transformer	**0.0362**
 機率加權模型 (21個分類區間)	CNN	0.0704
	ResNet-18	0.0578
	Vision Transformer	0.0337
	Swin Transformer	**0.0193**
機率加權模型 (41個分類區間)	CNN	0.0695
	ResNet-18	0.0561
	Vision Transformer	0.0314
	Swin Transformer	**0.0189**

表 2 兩模型影預測精度與平均推論時間比較

模型	RMSE	平均推論時間
基於 Swin-Transformer 的機率加權模型	0.0019	532.3 秒
Swin-Unet 即時無參考影像品質評估模型	0.0044	0.3 秒

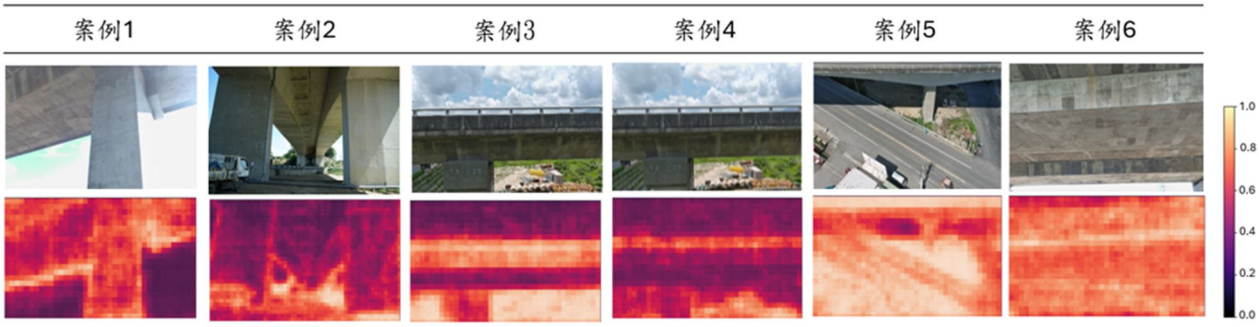


圖 6 Swin-Unet 影像品質評估模型架構圖

4. 結論

本研究提出一套基於 Swin-Unet 架構之即時無參考影像品質評估模型，能於 0.3 秒內完成單張影像品質圖之生成，並達到 RMSE＝0.044 之預測精度，展現良好之效率與準確性。為解決傳統 SSIM 指標於影像裁切後易產生像素錯位之問題，本研究進一步設計結合 CLIP 圖像編碼器與 SSIM 概念之 CSSIM 指標，並搭配基於 Swin Transformer 之機率加權模型，有效克服無參考條件下品質評估精度與標註資料不足的雙重挑戰。實驗結果顯示，Swin Transformer 機率加權模型(21 分類區間設計)可達最佳預測準確度(RMSE＝0.0193)，惟其逐像元推論機制導致單張影像推論時間高達 532.3 秒，限制其於即時應用情境中之可行性。為提升實務應用性，本研究進一步將上述高精度模型所產生之影像品質圖作為標註資料，訓練一套基於 Swin-Unet 架構之整張影像品質評估模型，以在合理精度下大幅縮短推論時間，成功實現一套可應用於無人機即時影像篩選之高效率、低成本、無參考影像品質評估方法。於實務構造物巡檢流程中，該模型可作為影像品質即時判斷與篩選之決策支援工具，協助辨識因模糊、過曝或光照不足而不適合後續分析之影像，並提供即時重拍或補充拍攝之依據，以避免低品質影像進入後續缺失判釋或建模流程。

未來研究將進一步整合影像品質評估結果與影像補救或後處理機制，例如針對低品質影像進行自動化亮度調整、去模糊或雜訊抑制處理，或作為觸發重拍與任務調度之依據，以強化系統於實際巡檢作業中之完整性與實用性。同時亦將擴充資料集所涵蓋之影像失真類型，如色偏、雜訊與壓縮失真，以提升模型在多樣現場環境下之泛化能力，並持續優化系統效率與可部署性，推動自動化 UAV 檢測任務之品質管控邁向實務應用。

參考文獻

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., and Wang, M., 2021. Swin-Unet: Unet-like pure transformer for medical image segmentation, in Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Strasbourg, France, DOI: 10.48550/arXiv.2105.05537.

Horé, A., and Ziou, D., 2010. Image quality metrics: PSNR vs. SSIM, in Proceedings of the 20th International Conference on Pattern Recognition,

Istanbul, Turkey, pp. 2366–2369, IEEE, DOI: 10.1109/ICPR.2010.579.

Kang, L., Ye, P., Li, Y., and Doermann, D., 2014. Convolutional neural networks for no-reference image quality assessment, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, pp. 1733-1740, IEEE, DOI: 10.1109/CVPR.2014.224.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows, in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, pp. 10012–10022.

Ma, K., Liu, W., Zhang, K., Duanmu, Z., Wang, Z., and Zuo, W., 2017. End-to-end blind image quality assessment using deep neural networks, IEEE Transactions on Image Processing, 27(3): 1202-1213, DOI: 10.1109/TIP.2017.2774045.

Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., Terhaarromeny, B., Zimmerman, J.B., and Zuiderveld, K., 1987. Adaptive histogram equalization and its variations, Computer Vision, Graphics, and Image Processing, 39(3): 355–368, DOI: 10.1016/S0734-189X(87)80186-X.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I., 2021. Learning transferable visual models from natural language supervision, In Proceedings of the 38th International Conference on Machine Learning (ICML), pp. 8748–8763.

Rakha, T., and Gorodetsky, A., 2018. Review of Unmanned Aerial System applications in the built environment: Towards automated building inspection procedures using drones, Automation in Construction, 93: 252-264, DOI: 10.1016/j.autcon.2018.05.002.

Sieberth, T., Wackrow, R., and Chandler, J. H., 2015. UAV image blur—Its influence and ways to correct it, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-1/W4: 33–39, DOI: 10.5194/isprsarchives-XL-1-W4-33-2015.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I., 2017. Attention is all you need, In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, Vol. 30, pp. 5998–6008.

Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P., 2004. Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing, 13(4): 600-612, DOI: 10.1109/TIP.2003.819861.

Real-time and Reference-free UAV Image Quality Assessment using Deep Learning

Ya-Li Lin [1*] Guan-Chin Su [1] Lai-Han Zou [1] Chao-Hung Lin [2]
Jiann-Yeou Rau [2] Wei-Shen Lai [3] Chih-Chao Hu [3]

Abstract

With the increasing use of unmanned aerial vehicles (UAVs) in infrastructure monitoring and environmental inspection, stable image quality has become critical for deep learning and photogrammetry applications. However, UAV images are often degraded by environmental disturbances, while existing quality filtering still relies on manual inspection, making it unsuitable for high-frequency or real-time deployment. This study proposes a real-time, reference-free image quality assessment (IQA) framework based on a Swin-Unet architecture to improve screening efficiency and ensure data quality stability, while simultaneously generating image quality maps (IQMs) for downstream applications. To overcome limitations of traditional SSIM-based methods, including the requirement for reference images and sensitivity to pixel misalignment, an improved metric, termed CLIP-SSIM (CSSIM), is introduced to construct an image scoring model. A probability-weighted Swin-Transformer is first employed to generate high-accuracy IQMs (RMSE = 0.0193); however, its pixel-wise inference is computationally expensive (532 seconds per image). Therefore, the generated IQMs are used as supervisory labels to train a Swin-Unet model, enabling real-time inference (0.3 s per image) with acceptable accuracy (RMSE = 0.04). The proposed approach provides an efficient, accurate, and scalable solution for UAV image screening, effectively replacing manual inspection in high-frequency UAV applications.

Keywords: Image Quality Assessment, Deep Learning, UAV Imagery, Structural Similarity

[1] Ph.D. Student, Department of Geomatics, National Cheng-Kung University
[2] Professor, Department of Geomatics, National Cheng-Kung University
[3] Researcher, Transportation Engineering Division, Institute of Transportation
[*] Corresponding Author, E-mail: alecfree2@gmail.com

Received Date: Nov. 19, 2025
Revised Date: Dec. 18, 2025
Accepted Date: Jan. 30, 2026