2025年臺灣國際科學展覽會
優勝作品專輯

作品編號　190013

參展科別　電腦科學與資訊工程

作品名稱　基於特徵解耦的視覺轉換器之指靜脈辨識模型

得獎獎項　一等獎

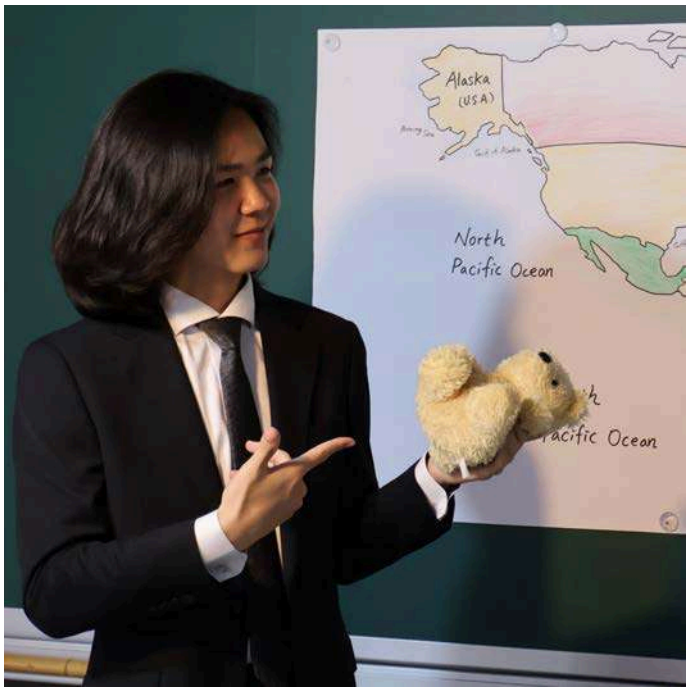
美國國際科技展覽會 ISEF

青少年科學獎

就讀學校　國立宜蘭高級中學
指導教師　夏至賢
柯良穎
作者姓名　林宜辰

關鍵詞　指靜脈辨識、視覺轉換器、特徵解耦

作者簡介



我是林宜辰Emery，就讀宜蘭高中數資班二年級。在國二時我有幸與一位教授做了電腦視覺的科展，增進了我對機器學習領域的了解。在那之後我也不斷地線上自學微積分、資料科學等任何感興趣的領域，這也使我不畏懼在此次研究前期的茫茫文獻海中自學。雖然經歷了無數個漫漫長夜的掙扎，但我始終沉浸於研究帶來的未知與驚喜。"Life is too long not to face any failures, but also too short not to try for yourself."

2025 年臺灣國際科學展覽會

研究報告

區別：

科別：電腦科學與資訊工程

作品名稱：基於特徵解耦的視覺轉換器之指靜脈辨識模型

關鍵詞： 指靜脈辨識 、 視覺轉換器 、 特徵解耦

編號：

摘要

發展安全且可靠的身份辨識技術是當今的重要議題，而指靜脈因其高安全性及難以偽造特性成為我們的主題。本研究首先探討基於 Gabor 濾波器的模型，透過對固定角度 Gabor 濾波器（模型一）及具有可學習 Gabor 參數的後續模型（模型二與三）的實驗結果得以證明，結合高頻靜脈特徵與低頻背景資訊可顯著提高正確辨識率（CIR）。對此，我們提出基於一個特徵解耦模型 SFA-FD（模型四）。透過開發小波解耦模組（WFDM），模型能有效提取並重新組合背景資訊與靜脈特徵。此外，本研究開發的空間特徵注意力模組（SFAM）能提取全域與局部特徵並進行特徵增強，強化模型對靜脈特徵的理解。模型四在 FV-USM、PLUSVein-FV3、MMCBNU-6000、UTFVP、NUPT-FPV、SCUT-FV 資料集中 CIR 達到 100%、99.41%、99.96%、95.17%、99.53% 和 99.90%，凸顯本模型在處理不同年齡層、國籍、影像模糊度的資料下仍能保持高穩定性與泛化能力，顯示其在需要高安全性辨識的應用場景中具備廣泛的實用性。

Abstract

Nowadays, developing secure and reliable identity recognition technology is a critical issue, and finger vein recognition has become our focus because of its high security and resistance to counterfeiting. This research began by investigating Gabor filter-based models to understand feature representation in finger vein recognition. Our initial experiments with fixed-angle Gabor filters (Model I) revealed that 45° orientation achieved optimal feature extraction, while subsequent models with learnable Gabor parameters (Models II and III) demonstrated that combining high-frequency vein features with low-frequency background information significantly enhanced recognition accuracy, achieving 98.06% correct identification rate on challenging datasets.

Building on these insights, we proposed SFA-FD (Model IV), aimed at addressing the limitations of existing technologies in representing and extracting finger vein image features. By developing the Wavelet Feature Decoupling Module (WFDM), the model can effectively discriminate between background information (low-frequancy features) and texture (high-frequancy feature) of finger veins and recombine them to improve recognition accuracy. In addition, with our proposed Spatial Feature Attention Module (SFAM), the model can extract both global and local features, further strengthening the model's understanding of the spatial context of finger vein features.

We conducted experiments on multiple public datasets, including FV-USM, PLUSVein-FV3, MMCBNU-6000, UTFVP, NUPT-FPV, and SCUT-FV, and our SFA-FD model demonstrated excellent correct identification rate (CIR) across these datasets, achieving 100%, 98.47%, 99.75%, 96.11%, and 99.82%, respectively. The exceptional stability and generalization capabilities of the model provide confidence in its performance across diverse databases, including those with low resolution, limited samples, extensive identity categories, age diversity, and nationality diversity. This versatility of the model is a testament to its potential in diverse real-world applications.

壹、研究動機與背景

一、 電子支付工具趨勢

隨著網際網路的普及與資訊及通訊技術（Information and Communication Technology, ICT）的快速發展，加上消費性電子產品的廣泛應用，個人資料安全議題逐漸受到社會大眾的高度關注。與此同時，電子支付工具，如簽帳金融卡、信用卡及電子錢包等支付方式，也逐漸融入人們的日常生活。然而，這些電子支付工具面臨著潛在的資料外洩風險[1]，此安全隱憂不僅影響消費者對電子支付的信心，更可能降低其使用意願。因此，建構完善的線上支付基礎建設，並發展安全、高效且可靠的身份辨識技術，已成為因應全球電子支付產業發展趨勢的關鍵目標。

二、 常見的生物辨識技術

過往以生物特徵為基礎的身份辨識系統逐漸在市面上興起；常見的生物特徵包括指紋、虹膜、人臉，以及指靜脈等特徵。這些生物特徵具有高度的穩定性，不易隨時間變化，因而為身份辨識提供了可靠且穩固的依據。此外，生物特徵可透過感測器快速且方便地獲取，這使得身份辨識的過程更加高效，進一步推動了生物辨識技術的廣泛應用。特別是基於靜脈特徵的身份辨識技術，由於其獨特性和難以偽造的特質，正逐漸受到關注，並被視為未來具有潛力的技術發展方向。

三、 生物辨識技術缺點

相較於內部生物特徵，外部生物特徵（如指紋及人臉）更易受到環境因素干擾，導致辨識技術的準確度不穩定。此外，外部生物特徵因長期暴露於外界，更容易被他人竊取，且由於生物特徵的永久性，使得使用者長期面臨潛在的安全風險。例如，指紋與掌紋辨識過程中，手指上的傷口、油脂及汙漬等因素可

能影響感測器擷取的準確性，進而降低辨識的準確度[2]。同時，指紋與掌紋在日常活動中容易殘留於接觸物品表面，增加其被竊取與複製的風險。

虹膜辨識技術則具有免受口罩或身體接觸影響的優勢，特別是在 COVID-19 疫情期間。然而，其應用需要昂貴的紅外線攝影設備，且在光線亮度控制不當的情況下，可能引起使用者眼部不適。而臉部辨識技術則對性別、遮擋、頭部姿勢、光照及臉部表情等變化極為敏感，這些因素均可能顯著影響模型的辨識準確度[3-5]。此外，與指紋和掌紋相比，臉部特徵由於經常出現在社交平台上，暴露風險更高，進一步增加其被濫用或複製的可能性。

四、 選擇指靜脈辨識的原因

由於外部生物特徵易受環境因素影響，且存在暴露與複製的風險，這使得外部生物特徵在實際應用中可能面臨不穩定性和較低的安全性問題。相對地，靜脈是一種相對較為穩定的內部生物特徵，該特徵隱藏於皮膚表層下導致難以對其進行竊取和偽造，並且不會隨著年齡增長導致靜脈的紋理產生磨損或改變。此外，指靜脈辨識技術還具備非接觸式操作的顯著優勢，這在後疫情時代的公共衛生安全考量下尤為重要。用戶無需直接接觸設備即可完成身份驗證，不僅降低了交叉感染的風險，更提升了使用的便捷性。而在金融交易等對安全性要求極高的應用場景中，指靜脈辨識技術能顯著提高身份驗證的安全性與效率，對於保護用戶的財產安全與隱私具有重要意義。

五、 現今指靜脈辨識技術問題

雖然基於靜脈的生物辨識具有許多顯著優勢，但其仍然存在一些缺點，且主要與靜脈影像取得的方式有關。受限於低成本的近紅外線（near infrared radiation, NIR）攝影機，在擷取靜脈影像時會受到照明環境和使用者行為的影響而導致影像品質受到光照變化、靜脈平移、以及旋轉、等問題影響。

貳、研究目的

本研究建構於一推測：指靜脈的背景資訊亦包含有價值的身份相關，而非如傳統方法將他們視為需要消除的噪點。因此本研究希望能開發一款能將指靜脈的背景資訊納入身分辨識依據，並進一步增強模型對指靜脈紋理特徵間的空間對應關係，使其能夠有效地捕捉全域與局部細節，從而提升指靜脈辨識模型的泛化能力的指靜脈辨識模型。接著使用多種公開指靜脈資料集對模型進行適應性、穩健性、泛化性、辨識正確率等方面的評估測試，確保其能在低解析度影像、少量樣本、大量身分類別以及年齡分布廣泛的情境下，仍能準確地進行指靜脈身份辨識。

參、 文獻探討

根據過往相關的研究文獻顯示，現有的指靜脈辨識技術研究主要聚焦於兩個核心領域：指靜脈的特徵表示（feature representation）技術以及特徵提取（feature extraction）技術。這些研究成果為解決指靜脈辨識中的關鍵技術問題提供了多元且有效的方法論基礎。本研究將從這兩個面向深入探討相關文獻，並分析現有技術的發展現況與挑戰。

一、 指靜脈的特徵表示技術

為使指靜脈辨識模型能有效利用關鍵特徵進行身份辨識，許多過往研究對多種指靜脈特徵表示方法進行探討，並以該特徵作為後續模型進行辨識的基礎。

(一) Al-Tamimi 和 AL-Khafajiy[6]則提出結合限制對比度梯度直方圖（Contrast Limited Adaptive Histogram Equalization, CLAHE）、中值濾波器和主成分分析（Principal Component Analysis, PCA）的影像增強方法，並將增強前後的影像輸入卷積神經網路（Convolutional Neural Network, CNN）進行訓練。但 CLAHE 的應用可能導致指靜脈影像產生光照不一致現象，進而限制模型的辨識能力[7]。

(二) Zhao 等人[9]創新性地提出方向強度向量（Intensity Orientation Vector, IOV）來描述指靜脈背景強度變化，作為身份辨識的輔助特徵。為有效整合 IOV 和方向差值向量（Direction Difference Vectors, DDV），該研究進一步提出語義相似性保留的鑑別二進制特徵學習（Semantic Similarity Preserved Discriminative Binary Feature Learning, SSP-DBFL）方法，以增強特徵間的一致性。然而，該方法仍依賴傳統電腦視覺演算法進行特徵提取，需要針對不同指靜脈資料庫進行耗時的參數調整才能達到理想的準確度。

二、 指靜脈的特徵提取技術

為了充分利用指靜脈特徵，過往的研究對指靜脈辨識模型的架構進行了深入探討，並據此提取出具鑑別度的身份特徵。

(一) Liu 等人[10]設計了雙分支 CNN 架構，包含主幹分支（trunk branch）和平滑遮罩分支（soft mask branch）。主幹分支運用殘差單元（residual units）提取指靜脈特徵，而平滑遮罩分支則採用沙漏網路（hourglass network）來提取全域的指靜脈特徵，並將其結果使用 sigmoid 函數轉換至[0,1]的區間來對主幹分支

的特徵進行相乘來實現注意力機制。但其將特徵圖下採樣（downsample）至4×4 再對其進行上採樣的設計可能導致重要的指靜脈紋理細節丟失。

(二) Zhong 等人[11]採用簡化版 MobileNetV2 作為主幹模型，並結合自動色彩增強（Automatic Color Enhancement, ACE）技術。然而，其特徵填充方式可能因特徵平移而增加模式（pattern）複雜度，影響模型收斂效果。

(三) Liu 等人[12]提出多尺度和多階段的殘差注意力網路（Multiscale and Multistage Residual Attention Network, MMRAN），但其低於 8 的批次量（batch size）訓練策略限制了批次歸一化（Batch Normalization, BN）的效果，更可能導致模型的辨識能力退化[13]。

三、 總結

我們發現過往文獻中提出的特徵表示技術多依賴於傳統電腦視覺算法來提取指靜脈影像的特徵，以捕捉資料中的複雜模式與結構，並作為深度學習模型的輔助資訊以提升辨識準確度。然而，這些傳統算法通常需依據不同資料庫特性進行參數調整，限制了算法的適應性和泛化能力。此外，隨著深度學習的快速發展，CNN 模型已被廣泛應用於指靜脈特徵提取，但其難以捕捉特徵間的長範圍相依性（Long-Range Dependence, LRD），從而限制了模型在處理複雜指靜脈結構時的性能。

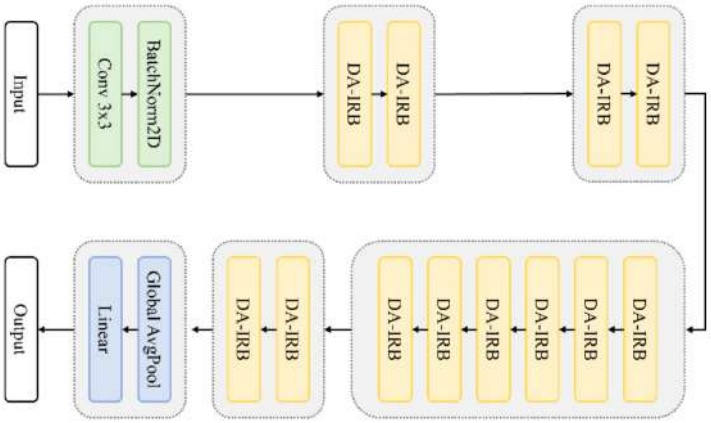
因此，為了結合過往研究成果並利用深度學習技術捕捉更具鑑別力的指靜脈特徵，本研究首先建構 Gabor 濾波器的模型並進行探討。隨後基於視覺轉換器（Vision Transformer, ViT）模型架構，提出了一種能夠解耦並重建手指背景資訊與靜脈紋理特徵的 SFA-FD 模型。該模型透過深度學習技術，自適應地結合紋理特徵與背景資訊，進一步增強了指靜脈特徵的表示能力。

四、 本研究所使用之技術

(一) 輕量化雙注意力機制卷積神經網路（Lightweight Dual-Attention Convolutional Neural Network, LDA-FV）[14]

LDA-FV 通過層層疊加雙注意力反殘差塊（DA-IRB）來構建整個網絡結構。DA-IRB 模塊融合了空間（SAM）和通道（CAM）兩種注意力機制，賦予了模型提取圖像中重要區域和特徵的能力，是 LDA-FV 架構的核心創新點。在 DA-IRB 的加持下，LDA-FV 能夠自動聚焦於對指紋靜脈辨識至關重要的區域和通道特徵，避免了模型資源的浪費，大幅提升了辨識效率和精度。它採用了 1：1：3：1 的模塊比例，通過對模塊數量的合理分配，實現了低階紋理特

徵向高階語義特徵的高效轉換。此外，LDA-FV 還引入了自適應邊際損失函數（AML），使模型能夠根據輸入靜脈圖像的質量動態調整訓練難度，進一步增強了模型的適應性和辨識能力。

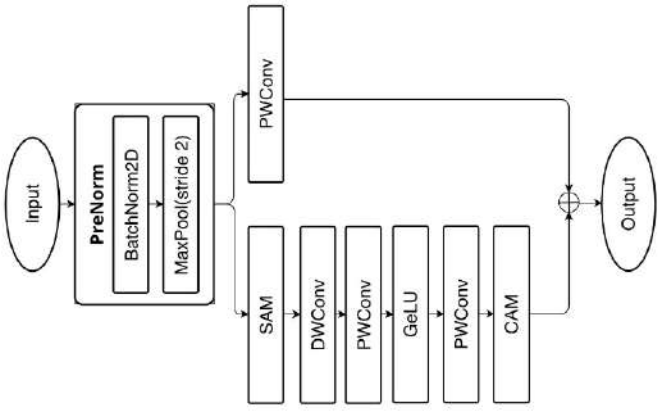


圖一：LDA-FV 之模型架構圖

(二) 雙注意力反殘差模塊（Dual-Attention based Inverted Residual Block, DA-IRB）

DA-IRB 是 LDA-FV 架構中最核心的模塊，其設計靈感來自於 ResNet 中的反殘差結構。在 DA-IRB 的結構中，首先會先進行批次歸一化及使用最大池化（MaxPool）進行下採樣，接著是一個空間注意力機制（Spatial Attention Mechanism, SAM）模塊，它能夠自主感知並聚焦於指紋靜脈在空間維度上的分佈位置，賦予了模型對空間維度特徵的自主感知能力。

之後，DA-IRB 使用一個 7x7 的深度卷積（Depthwise Convolutions, DWconv）對這些區域進行卷積運算，而較大的卷積核能夠有效補償小卷積核在有效感受野（effective receptive field）上的不足，使模型能夠捕捉到更大範圍的語意資訊，進一步提升特徵提取的質量和效率。在 DWconv 特徵提取後，特徵圖會透過兩個 1x1 的逐點卷積（Pointwise Convolutions, PWconv）提升維度並映射特徵圖的通道維度，以防止因縮減維度而丟失特徵資訊。



圖二：DA-IRB 之架構圖

(三) 基於傳統 Gabor 濾波器的卷積層（GaborConv2d）

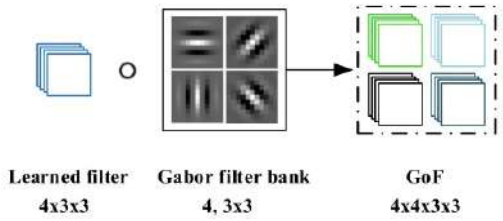
Gabor 濾波器是一種在圖像處理中廣泛使用的線性濾波器，通過調整其參數能夠捕捉圖像的特定方向和尺度的特徵。Gabor 濾波器對邊緣、紋理等圖像特徵特別敏感，使它在模式辨識和電腦視覺領域非常有用。而我們將 Gabor 濾波器的操作融入到標準 2D 卷積中，稱為 GaborConv2d，使得卷積神經網絡能夠更好地捕捉圖像中的紋理和方向信息。傳統的 2D 卷積只能學習較為簡單的低級特徵，而 Gabor 濾波器則擅長於提取複雜的紋理和方向特徵，可以很好地描繪圖像的局部結構。將二者相結合，就能夠賦予卷積神經網絡更強的特徵表示能力。

圖三：各種角度的傳統 Gabor 濾波器（由左至右為 0°、45°、90°、135°）

(四) Gabor 卷積神經網路（Gabor Convolutional Networks, GCNs）[15]

GCNs 將 Gabor 濾波器與卷積神經網路有效結合，使其獲得了 Gabor 濾波器在提取特徵的特性，同時保留了神經網路自身的強大學習能力。GCNs 的關鍵技術是 Gabor Orientation Filters（GoFs）的設計，它是一組三維的濾波器，將 Gabor 濾波器與可學習的二維卷積核結合。

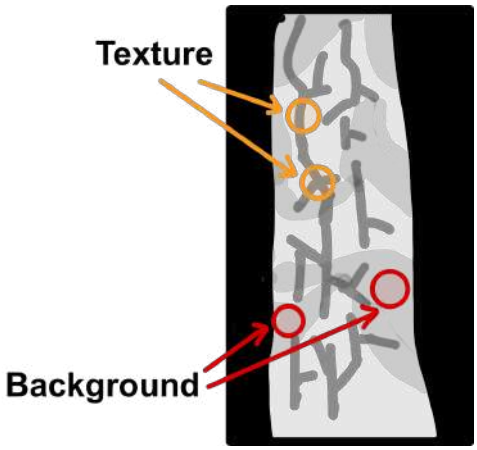
有了 GoFs 之後，GCNs 在卷積運算時就可以產生增強了方向和尺度信息的特徵圖。在優化階段，GCNs 只需要進行對初始可學習卷積核 Ci,o 的梯度更新，而 GoFs 則保持不變，這樣可以減少需要學習的參數量，降低模型計算複雜度。



圖四：GoF 計算示意圖

肆、 研究過程與方法

由於在利用紅外線攝影技術捕捉指靜脈影像時，手指的背景資訊容易受到使用者行為與手指肥厚程度的影響，從而導致背景光源出現變化。然而，除了使用者的行為之外，手指的肥厚程度屬於使用者獨有的生理特徵，因此手指的背景資訊並非毫無價值，它在一定程度上也能揭示使用者的身份，如圖五所示。



圖五、指靜脈影像內的背景資訊（紅色）與紋理特徵（黃色）之說明。

一、 開發基於 Gabor 卷積神經網路之指靜脈辨識模型

在經過大量的指靜脈辨識模型相關的文獻閱讀後，研究者發現 LDA-FV 的模型架構中並沒有特別針對模型提取細微特徵的能力進行提升的設計。對此，本研究開發了四種不同的指靜脈辨識模型，以提高模型在提取語意特徵與細微紋理上的能力。

此外，由於 LDA-FV 模型中的 DA-IRB 採用了 DWConv 和 PWConv 的設計，因此在整合 GaborConv2d 時，需要分別根據這兩種卷積操作的特點對 GaborConv2d 的參數進行調整，使其能夠無縫地嵌入 DA-IRB 中。

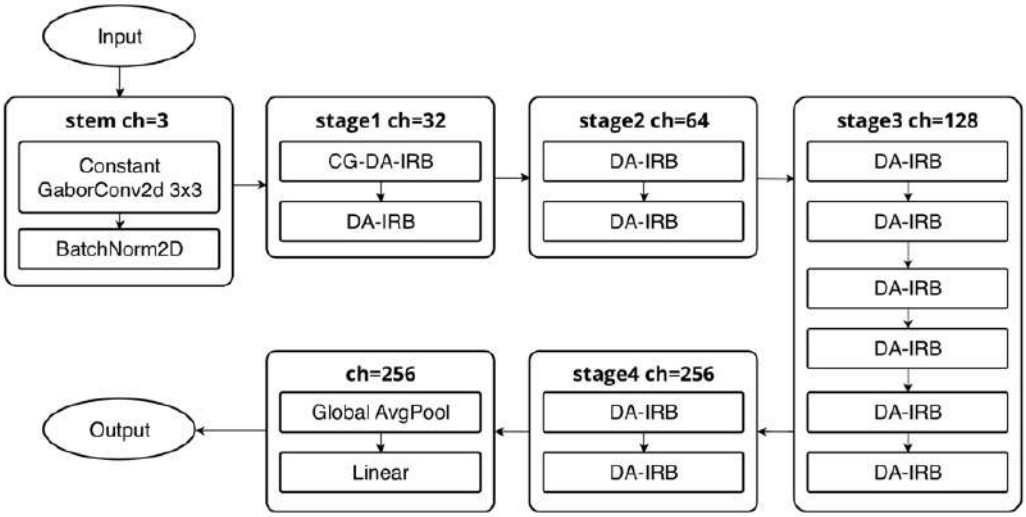
本研究對於 DWConv 使用單個輸出通道的 GaborConv2d 核；而對於 PWConv，由於它本身就是一種暢通道信息的 1x1 卷積操作，因此我們採用標準的 GaborConv2d 核。

(一) GradCAM 模型注意力可視化

為了具體呈現 Gabor 濾波器對於模型提取特徵的效果，除了比較 CIR 外，本研究在實驗中皆利用 GradCAM 將模型在做出預測時所關注的圖像區域，並與原模型進行比較與探討。

(二) 模型一：傳統 Gabor 濾波器整合 - LDA-FV + Constant_Gabor

由於 LDA-FV 模型在提取細微指靜脈特徵方面的能力有待加強，模型一將傳統 Gabor 濾波器整合至 LDA-FV 模型的模型輸入層（stem）以及 stage1 第一個雙注意力反殘差模塊（DA-IRB）中，稱為 Constant GaborConv2d 與 Constant_Gabor-DA-IRB（CG-DA-IRB），模型架構如圖六所示。

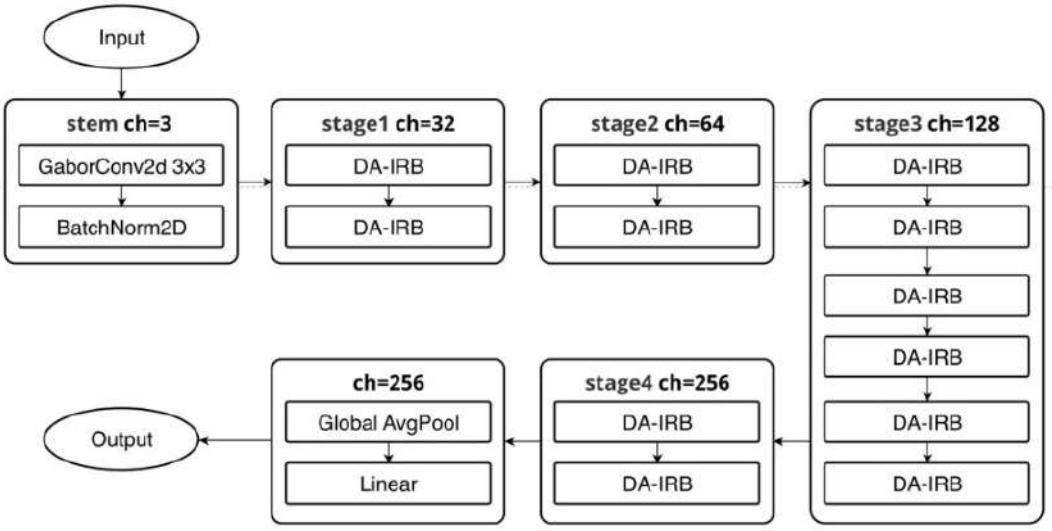


圖六：模型一（LDA-FV + Constant_Gabor）之模型架構圖

(三) 模型二：可學習的 Gabor 卷積層整合 – lPara-Gabor

在模型二中，本研究參考 Gabor Convolutional Networks（GCNs）[15]中提出的可學習的 Gabor 卷積層（GaborConv2d）來取代模型原有的卷積層。它將 Gabor 濾波器核的參數（如核尺寸、方向、寬高比等）設置為可學習的，使得這些參數能夠在模型訓練過程中不斷自我調整優化，因此模型不需要手動測試大量的參數組合，而是讓模型自行學習最適合當前任務和數據的 Gabor 核參數配置。

不過由於 GaborConv2d 的計算量相對較大，為了避免過度加重計算負擔，研究者僅將 GaborConv2d 模塊整合到 LDA-FV 模型的 stem 卷積層中，也就是模型專門提取低階特徵的部分，以此在發揮 Gabor 濾波器對於提取低階特徵的優勢下，模型還保有輕量化的特性，模型架構如圖七所示。

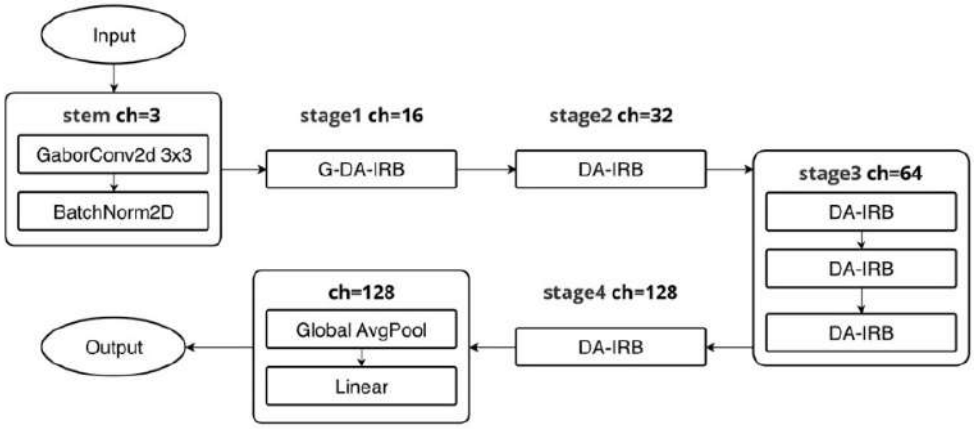


圖七：模型二（lPara-Gabor）之模型架構圖

(四) 模型三：可學習的 Gabor 卷積層整合 – sPara-Gabor

本研究進一步優化了模型架構，並提出了模型三。研究者將 LDA-FV 模型中各階段的通道數從 3 → 32 → 64 → 128 → 256 縮減為 3 → 16 → 32 → 64 → 128。透過這一改動，模型的總參數量從原先的 120 萬大幅降至 21 萬，實現了顯著的參數量減少，為模型的輕量化和實際應用部署奠定基礎。

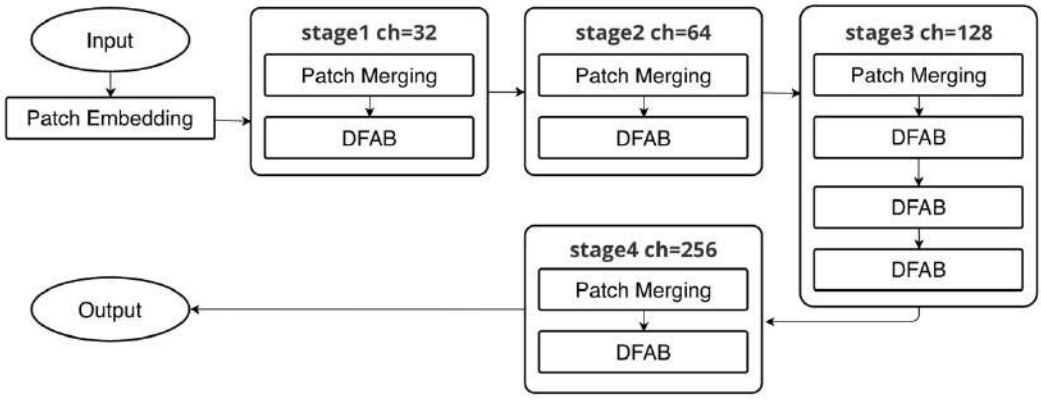
不過由於通道數顯著地減少，可能會影響到模型的辨識成效，因此相較於模型二只替換 stem 中的卷積層，模型三將 stage1 中的卷積層也一併替換，並稱之為 G-DA-IRB，模型架構如圖八所示。



圖八：模型三（sPara-Gabor）之模型架構圖

二、 開發基於視覺轉換器之指靜脈辨識模型 - 模型四：基於特徵解耦之空間特徵注意力模型（Spatial Feature Attention model based on Feature Decoupling, SFA-FD）

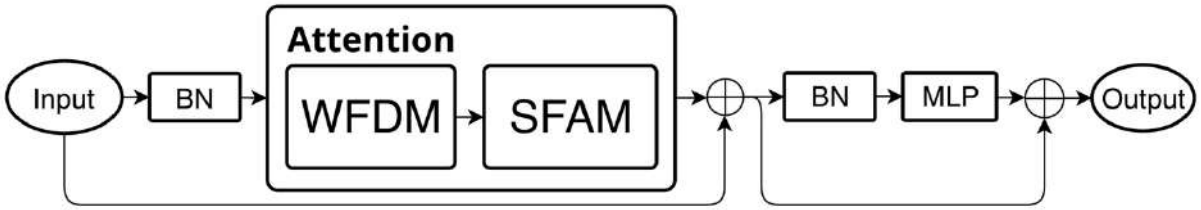
基於模型 I、II、III 的研究結果證實，將指靜脈影像內的低頻背景資訊結合至高頻靜脈特徵中確實能夠有效提升模型辨識的能力。對此，本研究提出了一種基於特徵解耦注意力機制模塊（Decoupled Feature Attention Block, DFAB）建構的 SFA-FD 模型架構來進行指靜脈辨識。我們參考 ConvNeXt [16]的模型設計理念，採用 1:1:3:1 的模塊數量比例進行階段性構建，模型架構如圖九所示。



圖九、模型四（SFA-FD）模型架構圖

在模型的初期，本研究使用重疊圖塊嵌入（overlapping patch embedding）[17]將指靜脈影像轉換為圖塊，並同時增強局部特徵的關聯性。接著，模型在每個階段運算前使用圖塊融合（patch merging）對特徵圖進行降採樣，進一步增強圖塊的感受野。

在模型的主要部分，DFAB 模塊的開發靈感來自經典 Transformer 的注意力區塊，WFDM 和 SFAM 模組會對指靜脈特徵圖中的細微紋理和背景特徵進行解耦，並通過注意力機制進行重建，以增強所提取的指靜脈特徵，如圖十所示。

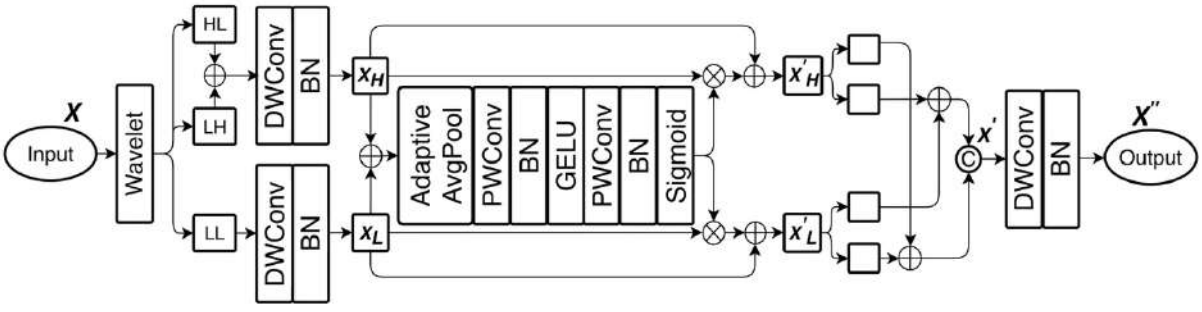


圖十、DFAB 模塊架構圖

DFAB 模塊還會對重建後的指靜脈特徵提取全域與局部特徵，並進行特徵融合，以加強指靜脈紋理特徵之間的關聯性。最終，此模塊會透過一系列卷積操作來對經由注意力機制後的局部特徵進行進一步增強，提升模型的身份辨識能力。

(一) 開發小波解偶模組（Wavelet Feature Decoupling Module, WFDM）

本研究提出一個能夠對指靜脈的細節紋理與背景特徵進行解耦的 WFDM 模組，並通過分別解耦後的特徵重新表示指靜脈特徵，如圖十一所示。



圖十一、WFDM 模組之架構圖

首先，WFDM 模組包含兩個並行處理路徑：

(一) 高頻特徵圖X_H提取

透過離散小波轉換（Discrete Wavelet Transform, DWT）獲取特徵中高低頻（Higher-low）及低高頻（Lower-high）部分，並使用深度卷積（Depthwise Convolution, DWConv）操作提取包含靜脈紋理的高頻特徵圖X_H，如公式(1)所示。

$$X_H = DWConv(Concat(DWT\_HL(X), DWT\_LH(X))) \quad (1)$$

(二) 低頻特徵圖X_L提取

透過使用 DWT 獲得的低低頻（Lower-low）及 DWConv 操作提取包含指靜脈背景資訊的低頻特徵圖，如公式(2)所示。

$$X_L = DWConv(DWT\_LL(X)) \quad (2)$$

接著，WFDM 模組將低頻特徵圖X_L（包含指靜脈背景資訊）與高頻特徵圖X_H(指靜脈紋理)進行特徵融合。融合後的特徵經自適應平均池化層(adaptive average pooling, AdaptiveAvgPool)轉換為特徵向量，以此來透過連續的卷積操作與 Sigmoid 函數將該向量轉換為機率值，如公式(3)所示。

$$A = Sigmoid(Conv(AdaptiveAvgPool(X_H + X_L))) \quad (3)$$

藉由上述操作，WFDM 模組能夠獲取表示低頻特徵圖X_L與高頻特徵圖X_H之間互補關係的機率值，並根據該機率對兩者進行特徵加權融合，從而使 SFA-FD 模型學習到指靜脈細節紋理與背景特徵之間的關聯性，進一步增強模型的特徵表示能力，如公式(4)所示。

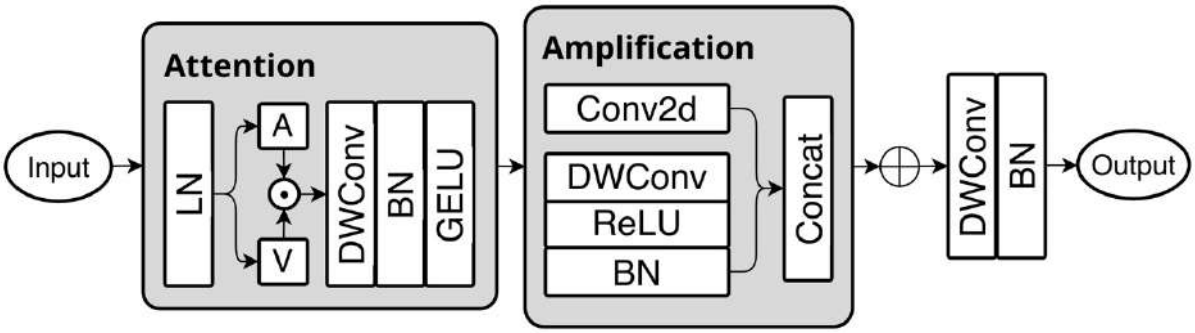
$$X' = ChannelShuffle(X_L \cdot A) + ChannelShuffle(X_H \cdot (1 - A)) \quad (4)$$

而為了有效融合低頻與高頻特徵，本研究將加權後的低頻特徵圖X_L'和高頻特徵圖X_H'拆分成兩部分，並調換順序進行後續的特徵融合與合併，從而獲得初步融合的指靜脈特徵圖X'。最終，融合後的特徵圖將輸入至後續的 DWConv 操作中，進一步增強指靜脈特徵圖X'的特徵，如公式(5)所示。

$$X'' = DWConv(X') \tag{5}$$

(二) 開發空間特徵注意力模組（Spatial Feature Attention Module, SFAM）

過去的指靜脈辨識模型大多基於 CNN 架構，然而由於 CNN 無法有效捕捉長範圍相依性（Long-Range Dependence, LRD），導致模型難以構建紋理特徵間的關聯性，限制了特徵表示能力。為了解決此問題，本研究提出了一種 SFAM 模組，該模組能夠同時提取指靜脈影像中的全域特徵與局部特徵並結合，從而增強模型的特徵表示能力，模組架構如圖十二所示。

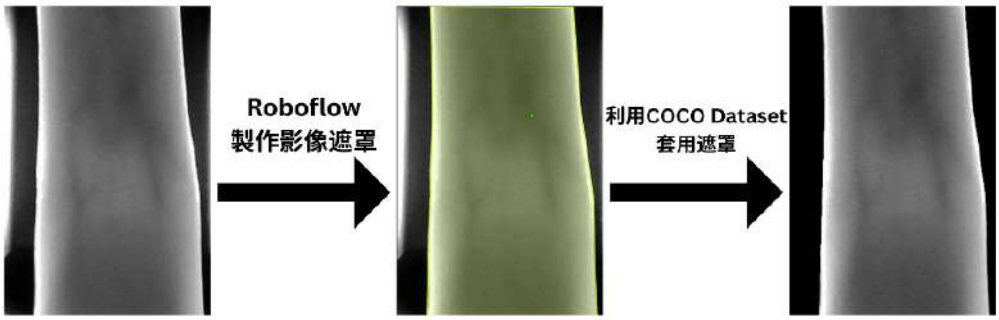


圖十二、SFAM 模組之架構圖

SFAM 模組利用 Conv2Former[18] 與 Amplificaiton Block[19]。在初期階段會先以卷積層提取局部特徵，再透過自注意力機制進行全域特徵建模，爾後將經過多尺度、多層次特徵融合後的特徵圖進一步進行特徵強化，將注意力機制產生的特徵圖進行選擇性放大，突出對辨識任務最有幫助的特徵，抑制雜訊或不重要的資訊。SFAM 模組的設計能同時保留細緻紋理與全域結構資訊，提升模型對複雜影像的理解能力，使其更適合資料量較小或變異性大的場景，能有效提升模型的穩健性，同時也因其使用哈達瑪乘積（hadamard product）而非常見的矩陣相乘，故模型運算需求亦能夠進一步降低。

三、 模型訓練資料集介紹與準備

本研究使用 Roboflow 線上圖片標註網站對指靜脈影像進行遮罩標記，並使用所製作的 COCO Dataset 檔對所有影像套用遮罩，製作出感興趣區域（Region Of Interest, ROI）影像，以避免資料集中的影像背景出現非指靜脈的紋理，確保模型僅會學習到指靜脈的特徵。



圖十一：資料集影像遮罩處理流程圖

(一) FV-USM 資料集[20]

此資料集使用波長 850 nm 的近紅外光 LED 收集 123 名受試者的指靜脈影像，包括 83 名男性和 40 名女性，年齡範圍為 20 至 52 歲。收集的手指包括左手食指、中指，以及右手食指、中指，每根手指視為一個獨立類別，總計 492 個類別。每位受試者在兩次不同的會議中分別蒐集 6 張影像，共 5,904 張解析度為 640×480 的影像。再進行模型訓練前，我們將資料集劃分為訓練集、驗證集和測試集，比例為 4:2。

(二) PLUSVein-FV3 資料集[21]

此資料集透過 NIR LED 感測器和雷射（Laser）感測器收集 60 名受試者的指靜脈影像，包含 35 名男性和 25 名女性，年齡範圍 18 至 79 歲。每位受試者的六根手指（左、右手的食指、中指、無名指）的正面和反面影像均被收集，並且每個手指會被視為不同的類別。因此，該資料集具有 360 個身份類別和 7,200 張解析度分別為 1280×1024 的原始指靜脈影像和 736×192 的 ROI 指靜脈影像。為進行公平的模型評估，本研究僅使用正面的指靜脈影像，因為 FV-USM 僅包含正面的指靜脈影像。我們將 LED 和 Laser 的正面指靜脈的 ROI 影像資料集分別切分為訓練集、驗證集和測試集，比例為 3:2。

(三) MMCBNU-6000 資料集[22]

此資料集使用波長 850 nm 的近紅外光 LED 收集來自 20 個國家的 100 名受試者的指靜脈影像，受試者包含 83 名男性和 17 名女性，年齡範圍 16 至 72 歲。每位受試者的雙手食指、中指和無名指影像被重複收集 10 次，總計 6,000 張影像和 600 個類別。在進行模型訓練前，我們將資料集劃分為訓練集、驗證集和測試集，比例為 3:2。

(四) UTFVP 資料集[23]

此資料集使用波長 850 nm 的近紅外光 LED 收集 60 名受試者的指靜脈影像，包含 44 名男性和 16 名女性，年齡範圍主要為 19 至 30 歲。影像來自雙手的食指、中指和無名指，總計 360 個類別和 1,440 張解析度為 672×380 的影像。在進行模型訓練前，我們將其切分為訓練集和測試集，比例為 5:5。

(五) NUPT-FPV 資料集[24]

此資料集使用波長 850 nm 的近紅外光 LED 收集 140 名受試者的指靜脈影像，包含 108 名男性和 32 名女性，年齡範圍 16 至 29 歲。每位受試者的雙手食指、中指和無名指的影像被重複收集 10 次，並且不同的手指會被視為相異的類別。其中，每個手指在兩次採樣中都會被各別蒐集 2 張指靜脈影像，並且採樣的間隔時間為至少一週。因此，此資料集共有 16,800 張影像和 840 個類別。在進行模型訓練前，此資料集被劃分為訓練集和測試集，比例為 5:5。

(六) SCUT-FV 資料集[25]

此資料集使用 850nm 的近紅外光 LED 收集 568 根手指的 61,344 張影像。對於每根手指，前 6 次採集是在正常手勢下進行，另外 12 次則分別在順時針與逆時針旋轉的狀態下採集，每張影像的旋轉角度均小於 20°。每次採集會在 6 種不同的近紅外光（NIR）強度下分別拍攝 6 張影像。與其他指靜脈資料集相比，SCUT-FV 資料集的影像變化更大，這使得在訓練過程中要達到良好性能更具挑戰性。在進行模型訓練前，此資料集被劃分為訓練集和測試集，比例為 3:3。

表一：模型訓練資料集比較

	FV-USM	PLUSVein-FV3	MMCBNU-6000
收集方法	NIR(850nm)	NIR(850nm、950nm) Laser	NIR(850nm)
受試者資訊	共 123 名 83 位男性 40 位女性 20 ~ 52 歲 馬來西亞	共 60 名 35 位男性 25 位女性 18 ~ 79 歲 奧地利	共 100 名 83 位男性 17 位女性 16 ~ 72 歲 20 個國家
手指種類	雙手 食指與中指	雙手 食指、中指、無名指	雙手 食指、中指、無名指

類別	492 個	360 個(正面)	600 個
影像	5904 張	7200 張	6000 張
原始解析度	640 × 480	1280 × 1024	640 × 480
	UTFVP	NUPT-FPV	SCUT-FV
收集方法	NIR(850nm)	NIR(850nm)	NIR(850nm)
受試者資訊	共 60 名 44 位男性 16 位女性 19～30 歲 尼德蘭	共 140 名 108 位男性 32 位女性 16～29 歲 中國	正常姿勢 6 個 順/逆時鐘旋轉 （角度< 20°）12 個 中國
手指種類	雙手 食指、中指、無名指	雙手 食指、中指、無名指	共 568 根手指
類別	360 個	840 個	10224 個
影像	1440 張	16800 張	61344 張
原始解析度	672 × 380	300 × 450	640 × 288

伍、研究結果與討論

一、 評估指標

為了對本研究所提出的 SFA-FD 模型架構的可行性進行評估，我們選用了 FV-USM、PLUSVein-FV3、MMCBNU-6000、UTFVP 及 NUPT-FPV 公用資料集進行整體性能測試，並採用正確辨識率（Correct Identification Rate, CIR）指標來衡量模型的安全性，CIR 的計算方式如公式(6)所示。

$$CIR = \frac{Correct\ case}{Number\ of\ total\ case} \tag{6}$$

其中，正確辨識次數（Correction case）表示模型正確辨識的次數，總辨識次數（Number of total case）表示模型總共辨識過的次數。當 CIR 指標數值越高時，模型的安全性與系統辨識性能越好；相反，CIR 值越低表示模型的辨識性能與安全性較差。

二、 模型實驗環境設置

在指靜脈辨識模型的超參數（hyperparameter）設置方面，本研究將初始學習率（learning rate）設置為 0.001，影像尺寸（image size）設為 112，批次量（batch size）設為 32，循環次數（epochs）設為 40，並選用 AdamW 作為模型優化器；餘弦退火（cosine annealing）作為學習率調節器（scheduler）來進行訓練，以確保模型在訓練過程中不會過早收斂或陷入局部最優解。

模型訓練在搭載 12th Gen Intel® Core™ i7-12800HX CPU、24GB RAM 和 Nvidia RTX 3070Ti GPU 的 Windows 11 作業系統上，並使用 Python 與 PyTorch 進行。

三、 不同的 Gabor 濾波器角度對模型一（LDA-FV + Constant_Gabor）的影響

由於 Gabor 濾波器的參數設置會直接影響其對特徵的提取效果，因此本研究分別測試了 0°、45°、90° 和 135° 四種不同角度設置下，模型在多種訓練資料集上的正確辨識率（CIR）。同時，為了直觀地分析模型在不同 Gabor 濾波器角度下對指靜脈區域的關注程度，本研究使用 GradCAM 模型可視化技術，繪製模型的注意力熱力圖現，並與本研究其他模型進行比較。

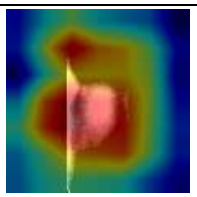
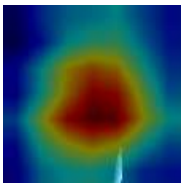
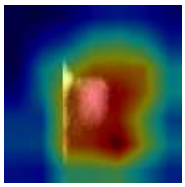
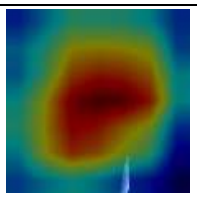
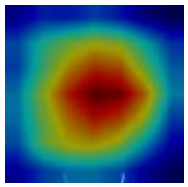
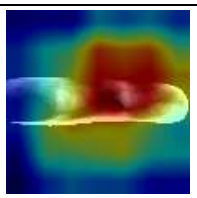
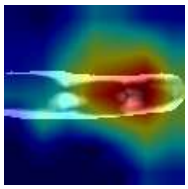
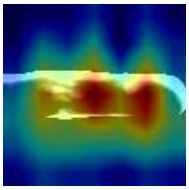
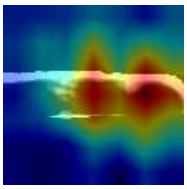
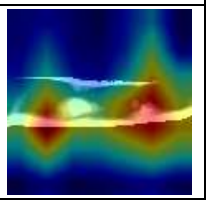
表二：不同的 Gabor 濾波器角度對 LDA-FV + Constant_Gabor 模型的 CIR 比較

角度（°）	FV-USM	PLUSVein LED	PLUSVein Laser
0	99.90	**96.94**	96.11
45	99.90	**96.94**	95.00
90	99.90	**96.94**	**96.67**
135	99.90	96.67	95.00

根據表二中的數據顯示，在 FV-USM 數據集上，不同的 Gabor 濾波器角度（0°、45°、90°、135°）對模型性能影響甚微，CIR 均保持在 99.90%。因此研究者認為對於圖像質量較高、一致性好的資料集，Gabor 濾波器的角度參數並不是影響模型性能的關鍵因素。然而，在 PLUSVein LED 和 PLUSVein Laser 兩個資料集上，不同角度的 Gabor 濾波器導致模型 CIR 值出現明顯波動。以 PLUSVein LED 為例，CIR 在 0°、45°、90° 時達到 96.94%，但在 135°時卻下降到 96.67%。這種現象可能是因為 PLUSVein 數據集包含更多質量參差不齊的指靜脈圖像樣本。這導致模型在該數據集上對 Gabor 濾波器角度的選擇更加敏感，因此當面對複雜多變的實際應用場景時，預設的 Gabor 角度很難始終保證

最佳性能。不恰當的角度選擇可能導致一些關鍵特徵被忽略或干擾，進而影響了模型的 CIR。

表三：不同的 Gabor 濾波器角度對模型一（LDA-FV + Constant_Gabor）的注意力比較

資料集	LDA-FV	theta=0°	theta=45°	theta=90°	theta=135°
FV-USM					
PLUSVein LED					

根據表三的比較可以看出，在大多數情況下，使用 Gabor 濾波器的模型能夠更準確、更集中地關注指靜脈的紋理區域。相比之下，原始 LDA-FV 模型的注意力可能較為分散，有時會關注到一些非關鍵區域，如 PLUSVein LED 的圖像。而在 FV-USM 資料集上，尤其是當角度設為 45° 時，使用 Gabor 濾波器明顯表現出更為集中的特徵定位能力。

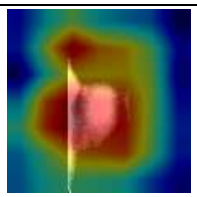
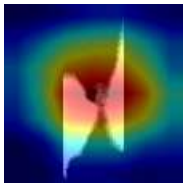
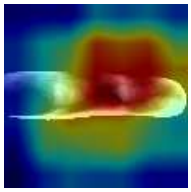
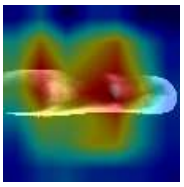
四、 模型二（lPara-Gabor）與原始模型的比較

表四：模型二（lPara-Gabor）與原始模型的 CIR 比較

模型	FV-USM	PLUSVein LED	PLUSVein Laser
LDA-FV	99.90	97.50	97.22
模型二	**100.00**	**97.78**	**98.06**

從表四的 CIR 比較來看，模型二在三個資料集上的 CIR 都有顯著提升。PLUSVein LED 和 PLUSVein Laser 資料集的 CIR 也分別從 97.50% 和 97.22% 提高到了 97.78% 和 98.06%。因此研究者認為引入可學習的 GaborConv2d 模塊對於提升模型的辨識性能是有效的，也證實 GaborConv2d 能夠透過自適應地優化 Gabor 濾波器參數，更好地提取指靜脈的細微紋理特徵，適應不同資料集的特點。

表五：模型二（lPara-Gabor）與原始模型的注意力比較

資料集	LDA-FV	模型二
FV-USM		
PLUSVein LED		

從 GradCAM 可視化結果來看，引入 GaborConv2d 後，模型能夠更加準確地關注指靜脈的紋理區域，相較於原始 LDA-FV 模型和模型一都有明顯改善。

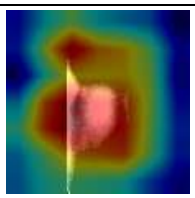
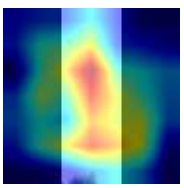
五、 模型三（sPara-Gabor）與原始模型的比較

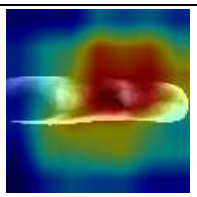
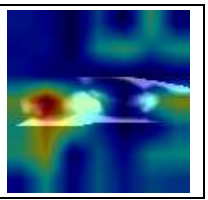
表六：模型三（sPara-Gabor）與原始模型的 CIR 比較

模型	FV-USM	PLUSVein LED	PLUSVein Laser
LDA-FV	99.90	97.50	97.22
模型三	**100.00**	**97.78**	**98.06**

從表六的結果來看，模型三在三個資料集上都取得優於 LDA-FV 模型的 CIR。此結果證實了可學習的 Gabor 卷積層能夠提取到更具辨識力的靜脈特徵，並驗證了本研究提出方法的有效性。此外，即便 sPara-Gabor 模型的參數量從原始的約 120 萬減少至 21 萬，但其在 FV-USM 和 PLUSVein 資料集上仍達到了 100% 與 98.06%，較參數量更多的 LDA-FV 模型、模型一、模型二有更加突出的表現。這說明了該模型能在大幅降低參數量的同時保持甚至提升其辨識能力，這對於日後將模型應用於金融領域的嵌入式系統有顯著的優勢。

表七：模型三（sPara-Gabor）與原始模型的注意力比較

資料集	LDA-FV	模型三
FV-USM		

PLUSVein LED		

在 FV-USM 資料集上，模型三的注意力熱力圖相較於原始 LDA-FV 模型更加集中和準確。雖然在 PLUSVein 資料集上仍然在關鍵區域仍有一定的關注，但相較於原始 LDA-FV 模型的熱力圖並無明顯地提升，研究者認為可能的原因為 GradCAM 可視化方法是通過梯度加權計算來生成注意力熱力圖，但其結果可能受到模型架構、梯度流等因素的影響，並不一定能完全準確地反映模型的實際關注點。而在只有單一對象的圖像中，Grad-CAM 有時可能無法定位對象的整個區域。此外，Grad-CAM 通過產生熱力圖來可視化哪些部分的輸入影像對模型的預測貢獻最大，提高深度學習模型的「可解釋性」，然而這並不直接增加模型的「解釋性」，因此即使它的決策可以透過熱力圖等方式被可視化，但卻未必能追蹤其內部計算過程與運作原理。因此，研究者認為未來除了 GradCAM 的可視化方法外，還應結合其他分析手段來評估模型的性能和行為。

六、 模型四中 WFDM 與 SFAM 模組效能評估

為了評估 WFDM 模組和 SFAM 模組在 SFA-FD 模型中的貢獻，本研究進行了模組效能的對比測試，通過對 SFA-FD 模型的不同配置進行實驗，以分析這兩個模組對模型性能的影響。在測試中，我們依次移除模型中的 WFDM 模組和 SFAM 模組。實驗結果顯示，WFDM 和 SFAM 模組在各個資料及上皆對模型的性能提升具有顯著貢獻，如表八所示。

(一) WFDM 模組效能分析效能分析

藉由對 WFDM 模組的效能測試結果可知，指靜脈影像中的背景資訊並不完全是環境光源的干擾，事實上，這些背景資訊中隱含了部分與使用者身份相關的特徵。而 WFDM 模組通過解耦低頻與高頻特徵，成功地將這些背景特徵與指靜脈的紋理特徵結合起來，使模型能夠更全面地進行身份辨識。因此，WFDM 模組不僅增強了紋理特徵的提取能力，還利用了背景資訊中的隱含身份信息來輔助辨識，顯著提升了模型的辨識性能，如表九所示。

(二) SFAM 模組效能分析效能分析

藉由對 SFAM 模組的效能評估結果，我們發現雖然僅依靠 Conv2Former 提取全域特徵能夠有效提升模型的辨識能力，但因為 SFA-FD 是用於指靜脈

辨識，因此當 SFA-FD 模型提取全域特徵來感知指靜脈整體的結構關係時，容易會因為局部特徵缺乏而導致模型難以更進一步利用全域特徵來進行提升模型的辨識性能。SFAM 模組的加入不僅彌補了這一缺陷，還使得模型能夠同時提取全域和局部特徵，而後的特徵強化機制也進一步強化了 FSA-FD 模型的整體辨識能力。尤其是在處理具有複雜結構的指靜脈影像時，SFAM 模組透過平衡全域與局部特徵，顯著提升了模型的準確性與穩定性，如表十所示。

表八、在各資料集進行模型四各模組效能（CIR）評估

WFDM	SFAM	FV-USM	PLUSVein		NUPT-FPV	MMCBNU-6000	UTFVP	SCUT-FV
			LED	Laser				
✓	✓	**100.00**	**99.65**	**99.17**	**99.53**	**99.96**	**95.17**	**99.90**
✗	✓	99.95	99.48	99.10	99.48	99.92	94.72	99.89
✓	✗	99.95	99.58	98.99	99.41	99.95	94.91	99.90
✗	✗	99.95	99.41	98.96	99.42	99.88	94.24	99.89

表十、在各資料集進行模型四 WFDM 模組效能（CIR）評估

高頻	低頻	FV-USM	PLUSVein		NUPT-FPV	MMCBNU-6000	UTFVP	SCUT-FV
			LED	Laser				
✓	✓	**100.00**	**99.65**	**99.17**	**99.53**	**99.96**	**95.17**	**99.90**
✗	✓	100.00	99.51	99.10	99.49	99.93	94.51	99.90
✓	✗	99.95	99.48	98.89	99.51	99.92	94.52	99.89
✗	✗	99.95	99.48	99.10	99.48	99.92	94.72	99.89

表十一、在各資料集進行模型四 SFAM 模組效能（CIR）評估

Amplification	FV-USM	PLUSVein		NUPT-FPV	MMCBNU-6000	UTFVP	SCUT-FV
		LED	Laser				
After Attention	**100.00**	**99.65**	**99.17**	**99.53**	**99.96**	**95.17**	**99.90**
Before Attention	99.95	99.58	99.03	99.44	99.89	94.94	99.89
None	99.95	99.24	99.06	99.05	99.88	94.35	99.89

七、 本研究開發之模型於各公開資料集的定量評估

(一) 具模糊指靜脈影像之資料集 - 模型一（45°）、模型二、模型三、模型四

在實驗中，模型二、模型三、模型四在具有大量模糊指靜脈影像的 FV-USM 資料庫上的 CIR 皆達到了 100.00%，顯示了本研究開發之模型在面對模糊影像

時依然能夠精確提取具有鑑別力的特徵。這表明該模型在模糊的條件下，依然能維持高度準確性。其中，就模型四而言，我們推測這一結果得益於本研究提出之 WFDM 和 SFAM 模組，因其能夠有效解耦指靜脈的細微紋理特徵與背景特徵，並通過注意力機制重建這些特徵，從而提升辨識性能。此外，與其他研究所提出的方法相比，本研究開發之模型架構具有較高的準確度，如表十一所示

表十一：與過往研究在 FV-USM 資料庫之 CIR 數值比較

Methods	Years	CIR (%)
DenseNet-121 [26]	2017	99.71
EfficientNet-B0 [27]	2019	99.78
Semi-PFVN [28]	2022	99.80
PVTv2-B0 [17]	2022	97.45
FastViT-T8 [29]	2023	99.61
LDA-FV [14]	2024	99.71
模型一 (45°) (ours)	2024	99.90
模型二 (ours)	2024	**100.00**
模型三 (ours)	2024	**100.00**
模型四 (ours)	2024	**100.00**

(二) 具年齡與光源多樣性之資料集 - 模型一（45°）、模型二、模型三、模型四

此資料庫的受試者年齡範圍較廣，涵蓋了不同年齡層，為驗證模型的適應性（adaptability）和穩健性（robustness）提供了理想的測試環境。實驗結果表明，模型四在 PLUSVein-FV3 資料庫中，針對由 LED 和 Laser 感測器拍攝的影像，分別達到了 99.65% 和 99.17% 的 CIR 指標，如表十二所示。這表示該模型在年齡分布較廣的資料庫中也能保持較高的辨識正確率，顯示出對多樣化數據來源的良好適應性。

值得注意的是，我們還將各模型與同樣能夠捕捉 LRD（Long-Range Dependence）的 FastViT 模型及其他現有先進的指靜脈辨識技術進行了比較。儘管 FastViT 模型具有強大的長距離依賴捕捉能力，經過微調訓練後，其在 PLUSVein-FV3 資料庫上的平均 CIR 仍僅為 95.19%。

表十二、與過往研究在 PLUSVein-FV3 資料庫之 CIR 數值比較

Methods	Years	CIR (%)		
		LED	Laser	Average
DenseNet-121 [26]	2017	99.34	98.61	98.98
EfficientNet-B0 [27]	2019	99.17	98.44	98.81
Semi-PFVN [28]	2022	99.38	98.72	99.05
PVTv2-B0 [17]	2022	96.81	96.53	96.67
FastViT-T8 [29]	2023	97.05	93.33	95.19
LDA-FV [14]	2024	98.61	98.30	98.46
模型一 (45°) (ours)	2024	96.94	96.67	96.81
模型二 (ours)	2024	97.78	98.06	97.92
模型三 (ours)	2024	97.78	98.06	97.92
模型四 (ours)	2024	**99.65**	**99.17**	**99.41**

(三) 大量身分類別資料庫 - 模型三、模型四

為了進一步驗證本研究開發之模型在大量身份類別資料庫中的辨識能力，本研究使用了包含 840 個身份類別的 NUPT-FPV 資料庫進行評估。該資料庫具有較大規模身份數據，對模型的特徵表示和辨識能力提出了更高的要求。如表十三所示，模型四在該資料庫上取得了 99.53%的 CIR。

表十三：與過往研究在 NUPT-FPV 資料庫之 CIR 數值比較

Methods	Years	CIR (%)
DenseNet-121 [26]	2017	97.60
EfficientNet-B0 [27]	2019	99.51
Semi-PFVN [28]	2022	98.93
PVTv2-B0 [17]	2022	94.01
FastViT-T8 [29]	2023	97.05
LDA-FV [14]	2024	98.18
模型三 (ours)	2024	99.76
模型四 (ours)	2024	**99.53**

(四) 具年齡及國籍多樣性之資料集 - 模型四

在 MMCBNU-6000 資料庫中，模型四的 CIR 指標達到了 99.96%，如表十四所示。該資料庫包含來自 20 個不同國家的受試者，其數據亦覆蓋了廣泛的

年齡範圍，這進一步證明了模型四的多元適應性與泛化能力，並顯示其在面對不同年齡層和國籍的受試者時，依然能夠保持高準確度的身份辨識能力。

表十四、與過往研究在 MMCBNU-6000 資料庫之 CIR 數值比較

Methods	Years	CIR (%)
DenseNet-121 [26]	2017	99.78
EfficientNet-B0 [27]	2019	99.92
Semi-PFVN [28]	2022	99.86
PVTv2-B0 [17]	2022	97.45
FastViT-T8 [29]	2023	99.71
LDA-FV [14]	2024	99.72
模型四 (ours)	2024	**99.96**

(五) 影像數量較少之資料集 - 模型四

為了驗證模型四在少量指靜脈影像資料庫中的有效性，本研究選擇了 UTFVP 資料庫對模型進行性能評估。該資料庫中的樣本量相對較少，但能夠提供針對模型泛化能力的挑戰。如表十五所示，其在 UTFVP 資料庫上的 CIR 達到了 95.17%。此結果說明，儘管數據量較少，模型四依然能夠提取出具有高度鑑別力的指靜脈特徵，展現其出色的特徵表示能力。

此外，在本研究之模型與其他當前先進的辨識模型的比較中，尤其包括基於 ViT（Vision Transformer）架構的模型。結果顯示，本研究提出之模型四在 CIR 上比 PVTv2-B0 高出 11.68%，顯示出顯著的性能優勢。

表十五：與過往研究在 UTFVP 資料庫之 CIR 數值比較

Methods	Years	CIR (%)
DenseNet-121 [26]	2017	90.23
EfficientNet-B0 [27]	2019	91.13
Semi-PFVN [28]	2022	93.13
PVTv2-B0 [17]	2022	83.49
FastViT-T8 [29]	2023	82.19
LDA-FV [14]	2024	87.70
模型四 (ours)	2024	**95.17**

(六) 影像數量較多之資料集 - 模型四

在包含 61,344 張影像的最大規模 SCUT-FV 資料集上，SFA-FD 仍保持了 99.90%的 CIR，如表十六所示。雖然相比小規模資料集性能略有下降，但仍保持在實用水準之上，展現了模型在複雜真實世界條件下的良好可擴展性。大規模資料集通常包含更多的個體差異、姿勢變化和環境干擾。SFA-FD 通過 DFAB 設計中的強化學習機制和自適應特徵選擇能力，能夠在大規模部署中保持穩定的性能表現。

表十六：與過往研究在 SCUT-FV 資料庫之 CIR 數值比較

Methods	Years	CIR (%)
DenseNet-121 [26]	2017	99.71
EfficientNet-B0 [27]	2019	99.19
Semi-PFVN [28]	2022	99.58
PVTv2-B0 [17]	2022	99.71
FastViT-T8 [29]	2023	93.43
LDA-FV [14]	2024	99.66
模型四 (ours)	2024	**99.90**

陸、結論

本研究首先開發基於 Gabor 濾波器的模型（模型一、模型二、模型三）並透過實驗證明高頻靜脈特徵與低頻背景特徵的結合對提升正確辨識率(CIR)至關重要。在 FV-USM 上，模型二與三皆達到 100% 的正確識別率，在 PLUSVein LED 和 PLUSVein Laser 數據集上分別達到 97.78% 和 98.06%。此外，這些結果亦促成了模型四的系統性特徵解耦方法。

本研究基於 Vision Transformer 模型架構提出了一種針對指靜脈特徵進行解耦與重建的模型四（SFA-FD）來進行指靜脈辨識。此模型透過本研究提出的 WFDM 模組實現對手指影像中的背景資訊與靜脈紋理的解耦，並在解耦後對特徵進行重建，以提取更加具鑑別力的指靜脈特徵。接著，模型四透過 SFAM 模組進一步強化對背景資訊與紋理特徵之間身份辨識相關性的學習，使得模型能夠更深入地理解這些特徵間的相互作用，從而增強指靜脈模型對指靜脈特徵的感知能力。

為了全面評估本研究開發之模型的穩健性與泛化性，本研究利用了多個各具特色的公用指靜脈資料集，包括 FV-USM、PLUSVein-FV3、MMCBNU-6000、UTFVP、NUPT-FPV 和 SCUT-FV 進行實驗。這些資料集涵蓋了多種不同的受試者群體、感

測器類型、影像模糊度、影像數量及類別數量，以驗證模型在多樣環境下的適應能力。

研究結果顯示，模型四模型在 FV-USM、PLUSVein-FV3、MMCBNU-6000、UTFVP、NUPT-FPV 及 SCUT-FV 資料集中，分別達到了 100%、99.41%、99.96%、95.17%、99.53% 和 99.90% 的 CIR。這樣的結果充分證實了其在不同資料集下的穩健性與泛化能力，展現其在指靜脈辨識領域中的技術優勢。

從基於背景信息與靜脈特徵結合的 Gabor 特徵處理相關模型（模型一、二、三），到具備卓越準確性與適應性的模型四（SFA-FD），本研究皆展示了技術的系統性進步。其中，模型四更是構建了一個全面的影像辨識框架，適用於多樣化的現實應用場景，具有極高的實用潛力。這也進一步表明，本研究所提出的指靜脈辨識模型在多樣且複雜環境中的電子支付領域具備廣泛的應用潛力，能夠提供高效且精準的身份驗證解決方案，從而顯著提升交易的安全性與可靠性。

柒、參考資料

[1] D. Dayanikli and A. Lehmann, “Password-based credentials with security against server compromise,” European Symposium on Research in Computer Security, pp. 147-167, 2024.

[2] Y.-Y. Chen, S.-Y. Jhong, C.-H. Hsia, and K.-L. Hua, “Explainable AI: A multispectral palm-vein identification system with new augmentation features,” ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 17, no. 111, pp. 1-21, 2021.

[3] V. Albiero, K.S. Krishnapriya, K. Vangara, K. Zhang, M. C. King, and K. W. Bowyer, “Analysis of gender inequality in face recognition accuracy,” IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, pp. 81-89, 2020.

[4] C. Bisogni, M. Nappi, C. Pero, and S. Ricciardi, “PIFS scheme for head pose estimation aimed at faster face recognition,” IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 4, no. 2, pp. 173-184, 2022.

[5] Peña, A. Morales, I. Serna, J. Fierrez, and A. Lapedriza, “Facial expressions as a vulnerability in face recognition,” IEEE International Conference on Image Processing, pp. 2988-2992, 2021.

[6] M. S. H. Al-Tamimi and R. S. S. AL-Khafaji, “Finger vein recognition based on PCA and fusion convolutional neural network,” International Journal of Nonlinear Analysis and Applications, vol. 13, no. 1, pp. 3667-3681, 2022.

[7] Y. Zhong, J. Li, T. Chai, S. Prasad, and Z. Zhang, “Different dimension issues in deep feature space for finger-vein recognition,” Chinese Conference on Biometric Recognition, pp. 295-303, 2021.

[8] L. Zhang, W. Li, X. Ning, L. Sun, and X. Dong, “A local descriptor with physiological characteristic for finger vein recognition,” International Conference on Pattern Recognition, pp. 4873-4878, 2021.

[9] P. Zhao, S. Zhao, J.-H. Xue, W. Yang, and Q. Liao, “The neglected background cues can facilitate finger vein recognition,” Pattern Recognition, vol. 136, pp. 109199, 2023.

[10] W. Liu, H. Lu, Y. Li, Y. Wang, and Y. Dang, “An improved finger vein recognition model with a residual attention mechanism,” Chinese Conference on Biometric Recognition, pp. 231-239, 2021.

[11] Y. Zhong, J. Li, T. Chai, S. Prasad, and Z. Zhang, “Different dimension issues in deep feature space for finger-vein recognition,” Chinese Conference on Biometric Recognition, pp. 295-303, 2021.

[12] W. Liu, H. Lu, Y. Wang, Y. Li, Z. Qu, and Y. Li, “MMRAN: A novel model for finger vein recognition based on a residual attention mechanism,” Applied Intelligence, vol. 53, pp. 3273-3290, 2023.

[13] P. Luo, R. Zhang, J. Ren, and Z. Peng, J. Li, “Switchable normalization for learning-to-normalize deep representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 2, pp. 712-728, 2021.

[14] L.-Y. Ke and C.-H. Hsia, “Finger-vein secure access based on lightweight dual-attention convolutional neural network for quality distance education,” Sensors and Materials, vol. 36, p. 945, 2024.

[15] S. Luan, C. Chen, B. Zhang, J. Han, and J. Liu, “Gabor convolutional networks,” 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, pp. 1254–1262, 2018.

[16] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, “A ConvNet for the 2020s,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11976-11986, 2022.

[17] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, “PVT v2: Improved baselines with pyramid vision transformer,” Computational Visual Media, vol. 8, pp. 415-424, 2022.

[18] Hou, Q., Lu, C.-Z., Cheng, M.-M., & Feng, J. (2024). Conv2Former: A simple transformer-style ConvNet for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 8274–8283.

[19] Huang, Y., Ma, H., & Wang, M. (2023). Axially enhanced local attention network for finger vein recognition. IEEE Transactions on Instrumentation and Measurement, 72, 1–10.

[20] M. S. M. Asaari, S. A. Suandi, and B. A. Rosdi, "Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics," Expert Systems with Applications, vol. 41, no. 7, pp. 3367-3382, 2014.

[21] C. Kauba, B. Prommegger, and A. Uhl, "Focussing the beam - a new laser illumination based data set providing insights to finger-vein recognition," IEEE International Conference on Biometrics Theory, Applications and Systems, pp. 1-9, 2018.

[22] Y. Lu, S. J. Xie, S. Yoon, Z. Wang, and D. S. Park, "An available database for the research of finger vein recognition," International Congress on Image and Signal Processing, vol. 1, pp. 410-415, 2013.

[23] B. T. Ton and R. N. J. Veldhuis, "A high quality finger vascular pattern dataset collected using a custom designed capturing device," International Conference on Biometrics, pp. 1-5, 2013.

[24] H. Ren, L. Sun, J. Guo, and C. Han, "A dataset and benchmark for multimodal biometric recognition based on fingerprint and finger vein," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 2030-2043, 2022.

[25] Tang, S., Zhou, S., Kang, W., Wu, Q., & Deng, F. (2019). Finger vein verification using a Siamese convolu4onal neural network. IET Biometrics, 8, 306–315.

[26] G. Huang, Z. Liu, L.V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700-4708, 2017.

[27] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," International Conference on Machine Learning, pp. 6105-6114, 2019.

[28] T. Chai, J. Li, S. Prasad, Q. Lu, Z. Zhang, "Shape-driven lightweight cnn for finger-vein biometrics," Journal of Information Security and Applications, vol. 67, pp. 103211, 2022.

[29] P. K. A. Vasu, J. Gabriel, J. Zhu, O. Tuzel, and A. Ranjan, "FastViT: A fast hybrid vision transformer using structural reparameterization," IEEE/CVF International Conference on Computer Vision, pp. 5785-5795, 2023.

【評語】190013

1. 此作品旨在建構 GLA-FD 模型分離手指影像的背景與靜脈資訊，並用類似 transformer 模型技術得自我注意力機制來偵測局部與全域特徵相依性。以達到指靜脈身份辨識的目的。

2. 整體而言，資料收集豐富並且實驗結果亦驗證所提方法之潛力。

3. 建議可以探討其它人體(生物)特徵，例如體溫等的使用是否對於辨識效能有幫助。