

國立中央大學

資訊工程學系
碩士論文

使用特徵增強策略在 MLP-Mixer 影像分類器
Applying Feature Enhancement Strategies in
MLP-Mixer Image Classifier

研究生：童子祐
指導教授：陳慶瀚博士

中華民國 111 年 6 月

摘要

由於卷積神經網路模型擁有龐大的演算法，運算過程被視為黑盒子(black box)無法對其提出合理的解釋與分析，因此本研究提出透過增強影像特徵的方式並結合 MLP-Mixer 分類器，增加整個辨識系統的可解釋性與準確度，該辨識系統架構應用於魚類、種子和中歐森林生物辨識資料集。首先針對影像先進行形狀、紋理與顏色的特徵增強，再將特徵增強過後的影像(Feature-Enhanced Image, FEI)作為 MLP-Mixer 分類器的輸入，分別輸出三個特徵增強方式的 Top-5，作為此三個 Top-5 作為決策融合的輸入，透過多類別羅吉斯回歸(Multinomial Logistic Regression)輸出最終決策結果。本篇研究在 40 種魚類資料集上達到 99%的辨識率，優於未使用特徵增強的 MLP-Mixer 分類器的 96%辨識率；在 560 類種子資料集上達到 90.65%的辨識率，優於混合式神經網路(ResNet-50+Siamese)的 70.23%辨識率；在中歐森林資料集 153 類上達到 97.91%的辨識率，優於採用單個卷積神經網路架構的 93.4%辨識率。

Abstract

Since the convolutional neural network model has a huge algorithm, the operation process is regarded as a black box and cannot provide a reasonable explanation and analysis. Therefore, this study proposes to enhance the image features and combine the MLP-Mixer classifier to increase the overall Interpretability and accuracy of the identification system architecture applied to the fish, seed and central European forest biometric datasets. Firstly, the features of shape, texture and color are enhanced for the image, and then the image after feature enhancement is used as the input of the MLP-Mixer classifier, and the Top-5 of the three feature enhancement methods are output respectively, as the three Top-5 as the input of the MLP-Mixer classifier. The input of the decision fusion, the final decision result is output through multi-class Logis regression. This study achieves a recognition rate of 99% on 40 fish datasets, which is better than the 96% recognition rate of the MLP-Mixer classifier without feature enhancement; Achieving a recognition rate of 90.65% on the 560-category seed dataset, which is better than the 70.23% recognition rate of the hybrid neural network; It achieves a recognition rate of 97.91% on 153 categories of the Central European Forest dataset, which is better than the 93.4% recognition rate using a single convolutional neural network architecture.

誌 謝

感謝兩年前陳慶瀚老師能夠收我為指導學生，才有幸進入 MIAT 實驗室大家庭，在這兩年內，學習了許多新的知識與程式能力，尤其是在嵌入式系統的能力上提升許多，也增強舊有的技術能力。老師總是能在研究上指點迷津，給予許多建議與想法，對於害怕進入陌生領域的我，因為有老師的協助，才能一步一步的完成所有研究，也能夠在其中去發掘新知識並自我學習，並且能在完成最後的畢業論文研究主題，因此要再次感謝老師，非常榮幸能夠當老師您的學生。

也非常感謝陳慶逸教授與許佳興教授能夠撥空前來擔任口試委員，並且給予論文上的建議與想法，能夠讓論文的章節架構更加完善。

非常感謝我的家人，爸爸、媽媽、阿公和阿嬤，當我最堅強的後盾，在學業期間給予我許多關愛與鼓勵，讓我能夠專心完成學業，未來必定好好行孝報恩。

感謝實驗室的博班學長，建鈞學長和冠維學長，能夠在研究領域與論文上給予適時的協助，並提點相關建議，讓我能好好的利用所學。

感謝同儕們，庭萱、文義、耀德、文韜、聖文和昀劭，因為有你們一起同甘共苦面對各種挑戰，互相扶持，才能順利地過關斬將，很開心能夠與你們相識，希望未來大家能夠在職場上一起加油繼續奮鬥。

感謝學弟妹們，垣溯、聖哲、浩源、俐秀、羅捷、晨瑋和峻葦，謝謝你們給予實驗室帶來許多歡樂，對於研究議題也能夠互相討論與學習，預祝大家能夠準時寫出最優良的論文。

童子祐 謹致

中華民國 111 年 7 月 13 日

目錄

摘要	I
Abstract	II
誌謝	III
目錄	IV
圖目錄	VI
表目錄	VIII
第一章、緒論	1
1.1 研究背景	1
1.2 研究目標	3
1.3 論文架構	3
第二章、影像分類	4
2.1 Canny 邊緣檢測	4
2.2 方向梯度直方圖(HOG)	6
2.3 局部二值模式(LBP)	9
2.4 單尺度視網膜增強算法(SSR)	11
2.5 深度學習	12
2.6 MLP-Mixer	14
第三章、影像辨識分類系統	18
3.1 分類系統架構	18
3.2 分類器系統離散事件建模	22
第四章、系統整合與驗證	28
4.1 實驗開發環境介紹	28
4.2 實驗資料集介紹	29
4.3 特徵增強策略在 MLP-Mixer 影像分類驗證	31

第五章、	結論與未來展望	38
5.1	結論	38
5.2	未來展望	38
參考文獻	40

圖目錄

圖 2.1 最大值示意圖	5
圖 2.2 雙閥值檢測與邊緣連線	6
圖 2.3Canny 邊緣增強影像	6
圖 2.4(a)為 non-overlap 區塊劃分，(b)為 overlap 區塊劃分	8
圖 2.5 統計梯度方向與大小之直方圖	8
圖 2.6 經由 HOG 特徵增強後的影像	9
圖 2.7 基本 LBP 流程圖	10
圖 2.8LBP 旋轉不變最終輸出最小值作為中心像素點	10
圖 2.9 經由等價 LBP 特徵增強後的影像	11
圖 2.10 經由 SSR 特徵增強後的影像	12
圖 2.11 卷積神經網路架構	13
圖 2.12 常見的 Pooling 方式	14
圖 2.13MLP-Mixer 應用於影像分類	16
圖 2.14Mixer Layer 架構	17
圖 3.1 特徵增強策略分類器系統主要架構	19
圖 3.2 影像特徵增強架構圖	20
圖 3.3 影像特徵增強系統架構	21
圖 3.4 特徵增強策略 MLP-Mixer 分類器離散事件建模	22
圖 3.5 影像特徵增強離散事件建模	23
圖 3.6 形狀特徵增強離散事件建模	24
圖 3.7 紋理特徵增強離散事件建模	25
圖 3.8 顏色特徵增強離散事件建模	26
圖 3.9 決策融合離散事件建模	27
圖 4.1 魚類資料集範例	30

圖 4.2 種子資料集範例	30
圖 4.3 中歐森林資料集影像範例	31

表目錄

表 4.1 訓練環境	28
表 4.2MLP-Mixer_L/16_224 規格表	28
表 4.3 測試環境	29
表 4.4 魚類資料集於本系統訓練參數與結果	31
表 4.5MIAT 魚類資料集在有無採用特徵增強於 MLP-Mixer 比較	32
表 4.6 土魷魚特徵增強分類結果	32
表 4.7 種子資料集於本系統訓練參數與結果	33
表 4.8MIAT 種子資料集在本系統與混合式神經網路之比較	33
表 4.9 韭蔥特徵增強分類結果	34
表 4.10 越瓜特徵增強分類結果	34
表 4.11 油菜特徵增強分類結果	35
表 4.12 甜椒特徵增強分類結果	35
表 4.13 中歐森林資料集於本系統訓練參數與結果	36
表 4.14 開源中歐森林資料集在本系統與單個神經網路之比較	36
表 4.15Hedera helixSTERILE 特徵增強分類結果	37
表 4.16Sophora japonica 特徵增強分類結果	37

第一章、緒論

1.1 研究背景

在現今蓬勃發展的深度學習架構中[1, 2]，影像前處理皆採自動化特徵擷取的方式，經由大量數據透過演算法計算取得重要特徵來進行特徵擷取，例如 K. Simonyan 和 A. Zisserman 所提出的 VGGNet[3]與 K. He 等人提出的深度殘差網路(Deep residual network, ResNet)[4]皆為卷積神經網路(Convolution Neural Network, CNN)的架構，是現今的主流方法。卷積層以滑動卷積核並以特定的特徵檢測器(Feature Detector)做卷積運算[5]，透過不同的權重組合來做不同卷積運算，最終獲得各種圖片特徵表示，如邊、角、顏色或去除圖片中的噪點等效果，以此擷取出不同的特徵，但此權重來自於學習演算法的運算結果，自動找出最好的特徵來擷取，固萃取出的特徵我們無法理解其為形狀、紋理或顏色等特徵，進而導致分類錯誤時，難以解釋錯誤原因。

由於 CNN 模型擁有多層的卷積層，如 ResNet-50 有 50 層卷積層，又於卷積層上使用大量的矩陣乘法運算[6]，故使用 CNN 自動擷取特徵十分耗費計算資源[7]，而且需要依靠大量的資料集來進行模型權重更新，所以在訓練樣本充分的情況下才能得到具有良好鑑別性的特徵。為了在有限的訓練樣本情況下能夠以較低的計算資源需求達到神經網路分類效能，我們將藉由增強影像特徵表示，讓神經網路後續提取特徵更為容易，進而提高辨識率[8, 9]並加快訓練和辨識速度。

根據不同的資料集和應用目標，其代表的特徵皆不一樣，根據 Ritter 等人研究[10]，人類較常根據事物的形狀進行分類，而不是紋理、顏色或其他等特徵來作為分類依據，但卷積神經網路則是傾向以紋理來進行分類。在 R. Geirhos 等人研究中[11]，將一張帶有大象紋理的貓咪圖片經由人類與 CNN 進行比較，此實驗結果為對於人類來說此張圖片仍然為貓咪，但對於 CNN 來說此張圖片為大象，故此實驗證

實了 CNN 具有紋理偏差，因此會忽略形狀特徵，但對於某些應用目標而言，形狀特徵則為重要特徵，且根據 H. Li 等人研究指出[12]，VGG 在特徵擷取中遺失了許多有用的特徵[13]。

卷積神經網路(CNN)與視覺轉換(Vision Transformers , ViT)[14]為現今影像辨識中最為人知的深度學習模型，於 2021 年 Google 團隊提出一個新的模型架構-MLP(Multi layer perceptron)-Mixer[15]，不需要卷積層與注意力層也能提取特徵，並獲得優良的辨識結果且能與 CNN、ViT 相提並論。

在深度學習中擁有龐大的參數量且具有不透明性，整個運算過程被視為黑盒子(Black box)[16]，從輸入資料至輸出結果的過程中，我們無法瞭解其內部的整體運作，或是理解其判別結果從何而來。當人類越來越依賴深度學習後，尤其是應用於較影響人類的生活上如：醫學、交通等，就必須對此模型做進一步的解釋，使得大眾對此模型產生信任感，因此衍伸出可解釋人工智慧(Explainable AI , XAI)。於深度學習中，可解釋人工智慧方法已有多種，如：Perturbation-Based、Gradient-Based、Propagation-Based、CAM(Class Activation Map)-Based[21]等方法。

Perturbation-Based 是基於遮擋影像部分區域的方法[17]，此方法最早由 M. D. Zeiler 等人提出[22]，透過指定大小的方塊在原始圖片上滑動，可以計算出影像區塊的重要性。Gradient-Based 是基於梯度的方法[18, 19]，第一個用於 Gradient-Based 上的方法為 Sensitivity analysis[23]。Layer-Wise Relevance Propagation 是經由泰勒分解式(Taylor decomposition)做反向傳遞[20]，進而得出像素的重要性。Grad-CAM(Gradient- Class Activation Map)[24]是將權重值替換成梯度，可加強細節表現且 Grad-CAM 比 CAM 更好應用於各種模型上。

從上述可以發現深度學習中的特徵擷取雖已將傳統手動的特徵擷取改為自動化，並將單一特徵提取增強為泛用的特徵提取，但在深度學習的特徵擷取中，我們仍然無法清楚理解其提取出的特徵作用，本研究欲改善此特徵擷取的問題，增加神經網路辨識的可解釋性並提高其效能與辨識準確度。

1.2 研究目標

本論文致力於提出一個結合影像特徵增強的神經網路系統，並使用形狀、紋理與顏色等傳統影像特徵擷取方法來增強影像特徵表現，進而提升辨識效能。形狀特徵增強我們採用方向梯度直方圖(Histogram of Oriented Gradients , HOG)[25]，紋理特徵增強採用局部二值模式(Local Binary Patterns , LBP)[26]，顏色特徵增強則採用單尺度視網膜增強算法 (Single-Scale Retinex , SSR)[27]，分別輸出形狀、紋理、顏色特徵強化影像，以此作為神經網路分類器的輸入。

本論文設計一個廣泛應用於魚類、種子與中歐森林資料集的影像辨識系統，將魚類、種子與中歐森林影像分別轉換為形狀、紋理和顏色特徵增強影像(Feature-Enhanced Image, FEI)，再經由 MLP-Mixer 分類，將此分類輸出的 Top-5 作為決策融合 stacking 的輸入，並經由多類別羅吉斯回歸(Multinomial Logistic Regression)分類，最終輸出一個綜合三種特徵增強之分類結果，並根據分類結果探討分類的解釋性。

1.3 論文架構

本論文分為五個章節，第一章敘述深度學習中特徵擷取的方法，並探討現今神經網路模型對於特徵擷取的特性、並對 MLP-Mixer 做描述，以及提出本研究目標與研究方法。第二章影像分類技術，介紹影像特徵增強技術與 MLP-Mixer 模型架構及特性。第三章提出透過影像特徵增強，並結合神經網路 MLP-Mixer 之分類系統，並透過 MIAT 方法論進行階層式模組化的設計和離散事件建模。第四章系統整合實驗與效能評估，透過比較不同的神經網路模型來進行驗證與分析。第五章總結本篇論文的研究結果並提出未來展望。

第二章、 影像分類

由於透過 CNN 做影像特徵擷取後較難以進一步分析擷取出的特徵代表什麼，固採用影像特徵增強並結合 MLP-Mixer，使神經網路提取特徵時更加容易。形狀、紋理和顏色特徵具有廣泛且直觀的適用性為較常使用的影像特徵如：圖像檢索系統 (Query By Image Content , QBIC) 基於顏色、紋理和形狀進行圖像比較[28]。本節將針對影像之形狀、紋理與顏色做特徵增強回顧。

2.1 Canny 邊緣檢測

Canny 邊緣檢測由 John Canny 所提出[29]，影像邊緣指局部區域內亮度變化明顯的部分。一張原始的彩色圖像由 Red、Green、Blue 所組成，先將彩色圖像轉換為灰階圖像，常用的灰階轉換公式如下(2.1)：

$$Gray = 0.299 * R + 0.587 * G + 0.114 * B \quad (2.1)$$

接著採用高斯濾波器(Gaussian Filter)去除影像雜訊，可避免後續誤將雜訊判斷成邊緣資訊，公式(2.2)為生成大小 $(2k+1)*(2k+1)$ 的高斯濾波器方程式，k 表示遮罩大小。

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp \left[-\frac{(i - (k + 1))^2 + (j - (k + 1))^2}{2\sigma^2} \right] \quad (2.2)$$

$$, where \ 1 \leq i, j \leq (2k + 1)$$

邊緣方向並無固定可以為任意方向，透過邊緣檢測的運算元計算出水平與垂直方向，再計算出每一像素點的梯度大小與梯度方向，而邊緣檢測常用的運算元有：Sobel、Roberts 和 Prewitt，水平與垂直方向之 Sobel 運算元如下(2.3)(2.4)：

$$S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (2.3)$$

$$S_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (2.4)$$

再由以下公式計算出邊緣的梯度大小(2.5)與方向(2.6)，其中 G_x 、 G_y 分別為像素點在 x 與 y 方向的梯度值。

$$G(x,y) = \sqrt{G_x^2 + G_y^2} \quad (2.5)$$

$$\theta(x,y) = \tan^{-1}\left(\frac{G_y}{G_x}\right) \quad (2.6)$$

由於獲得的梯度邊緣像素寬度為多個像素，固需將非最大梯度值刪除，保留局部區域的最大梯度值，以達到邊緣細線化之效果，找尋最大梯度方向分為四個：0°、45°、90°、180°，圖 2.1 為例，若此梯度方向為45°，則尋找45°方向裡的最大值將其保留，反之則為 0。

$$\begin{bmatrix} 0.3 & 0.7 & 1 \\ 0.1 & 0.5 & 0.7 \\ 0.8 & 0.2 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0.3 & 0.7 & 1 \\ 0 & 0 & 0.7 \\ 0 & 0 & 0 \end{bmatrix}$$

圖 2.1 最大值示意圖

設定兩個閾值-高邊界與低邊界，高於高邊界的像素點幾乎是邊緣，但由於閾值較高，圖像邊緣即有可能為非封閉曲線，故須多設一個低邊界閾值。高於高邊界閾值的像素點視為強邊緣，低於低閾值的像素點為非邊緣，介於高閾值與低閾值之中的像素點為弱邊緣點，若旁邊有強邊緣點，則此像素點被視為邊緣點，圖 2.2 為例：點 A 與點 C 皆界於高閾值之上，故皆為強邊緣點，而點 B 雖為弱邊緣點，但因旁邊的點 A 與點 C 皆為強邊緣點，故點 B 也為強邊緣點；點 E 介於低閾值之下，故點 E 為非邊緣；點 D 雖為弱邊緣點，但由於旁邊沒有強邊緣點，故點 D 為非邊緣。將是邊緣點的像素值設為 255，反之非邊緣點的設為 0，並與原圖做合併，所產生出來的圖即為 Canny 邊緣檢測後的增強影像，如圖 2.3。

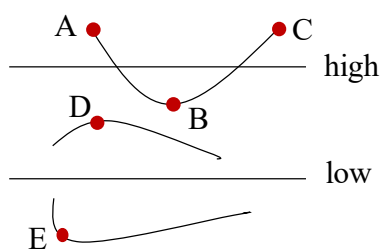


圖 2.2 雙閥值檢測與邊緣連線

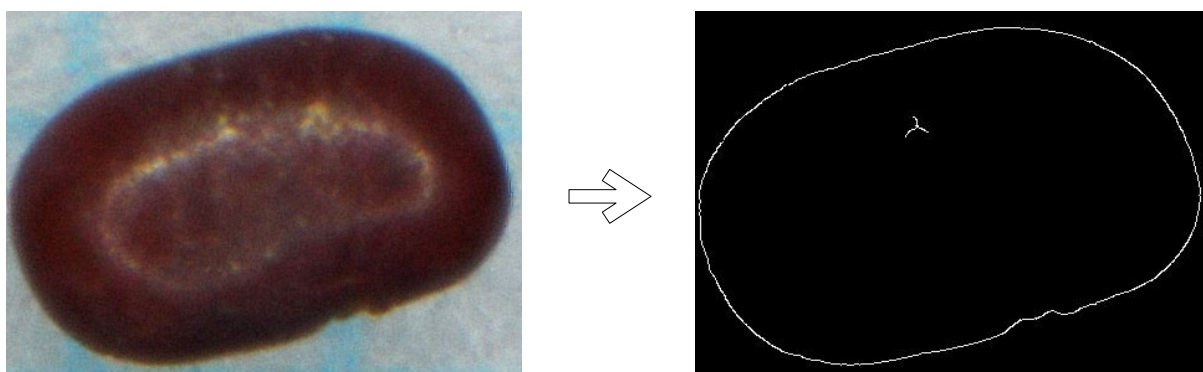


圖 2.3Canny 邊緣增強影像

2.2 方向梯度直方圖(HOG)

方向梯度直方圖(Histogram of Oriented Gradients , HOG)由 N. Dalal 和 B. Triggs 所提出[25]，應用於局部特徵描述上，經由計算影像中局部區域的梯度方向後統計每個區域的梯度直方圖來構成特徵，常用於行人檢測或靜態圖像上，降低光照與環境變化的影響，並盡可能地減少圖像的雜訊干擾。首先對原始影像進行預處理，因色彩訊息對此演算法無太大影響，固先將影像轉換為灰階圖像後再對影像進行 Gamma 校正，提升陰影與光照變化的強健性、減少圖像中的局部陰影、曝光等與增加圖像的對比度，平方根法的 Gamma 標準化公式如下(2.7)：

$$I(x,y) = I(x,y)^{\gamma} \quad (2.7)$$

Gamma 校正需依照以下步驟來執行：正規化，將所有像素點的值轉換至 0~1 之間，其公式如下(2.8)：

$$N(p) = (p + 0.5)/256 \quad (2.8)$$

p 表示其像素點的值。指數調整，給定一個 gamma 值後，將正規化後的像素值與 gamma 值進行指數運算，其公式如下(2.9)：

$$CPS(p, g) = N(p)^{1/g} \quad (2.9)$$

g 代表 gamma 值。還原正規化，將進行指數調整後的值還原至 0~255 之間，其公式如下(2.10)：

$$uN(p) = CPS(p, g) * 256 - 0.5 \quad (2.10)$$

預處理圖像過後，接著分別計算圖像的水平與垂直的方向梯度，水平梯度運算元為： $[-1, 0, 1]$ ，垂直梯度運算元為： $[-1, 0, 1]^T$ ，其水平(2.11)、垂直(2.12)梯度公式如下：

$$G_x(x, y) = I(x + 1, y) - I(x - 1, y) \quad (2.11)$$

$$G_y(x, y) = I(x, y + 1) - I(x, y - 1) \quad (2.12)$$

分別求得水平與垂直的方向梯度後，再計算圖像的梯度大小(2.13)和梯度方向(2.14)，公式如下：

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (2.13)$$

$$\theta(x, y) = \tan^{-1}\left(\frac{G_y(x, y)}{G_x(x, y)}\right) \quad (2.14)$$

由於 HOG 是用來描述局部特徵的，固需將圖像做切割，使 HOG 對於圖像的細節更為敏感。首先將圖像切割成大小相同的小方格(cell)，每個 cell 由 $n \times n$ 個像素所構成；一個區塊(block)由 n 個 cell 所組成；一張原始圖像由 n 個 block 組成，區塊分割為重疊(overlap)與不重疊(non-overlap)，以圖 2.4 為例，一張原始圖分割成 4×4 個 cell，每個 cell 為 3×3 個像素，每 2×2 個 cell 為一個 block。

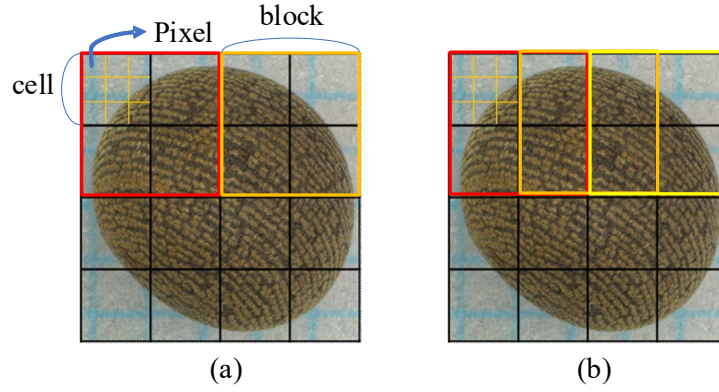


圖 2.4(a)為 non-overlap 區塊劃分，(b)為 overlap 區塊劃分

而 overlap 區塊所選取的 block 有重疊的部分，使得每個 block 間的連貫與相關性較高，再將每個 cell 的梯度方向做直方圖累計，將梯度方向依 $0^\circ \sim 180^\circ$ 或是 $0^\circ \sim 360^\circ$ 做轉換，以圖 2.5 為例，將 $0^\circ \sim 180^\circ$ 分割為 9 塊，即每 20° 為一區塊，而梯度方向旋轉 180 度後仍保持不變，將每一區塊的梯度大小累加後並用直方圖來做呈現。

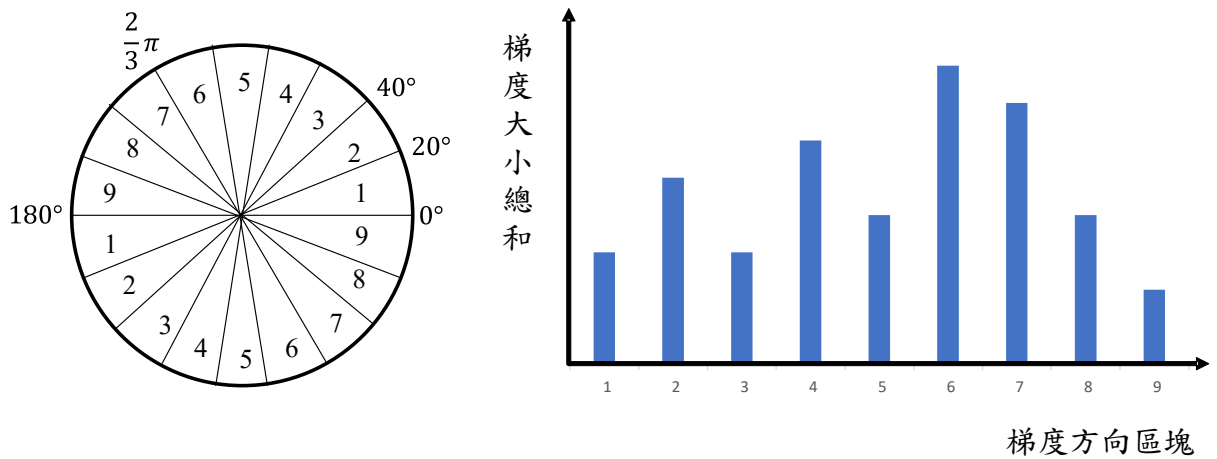


圖 2.5 統計梯度方向與大小之直方圖

局部因素的變化會影響梯度強度，使其變化範圍增加，固需在 block 內進行正規化後能有效將資訊壓縮，最終將 block 內的每個 cell 之特徵向量串接。常用的正規化方法有四種： $L_2 \text{ norm}$ (2.15)、 $L_2 \text{ hys}$ 、 $L_1 \text{ norm}$ (2.16)、 $L_1 \text{ sqrt}$ (2.17)

$$f = \frac{v}{\sqrt{\|v\|_2^2 + \varepsilon^2}} \quad (2.15)$$

$$f = \frac{v}{\|v\|_1 + \varepsilon} \quad (2.16)$$

$$f = \frac{v}{\sqrt{\|v\|_1 + \varepsilon}} \quad (2.17)$$

v 為一向量， ε 為一極小的常數，而 L_2 *hys* 則是藉由 L_2 *norm* 的結果將其截短後再進行一次正規化。學者研究發現，這四種正規化的方式皆有效改善數據，採用 L_2 *norm*、 L_2 *hys*、 L_1 *sqrt* 的效果大同小異，且皆優於 L_1 *norm*。

一張原始影像的每個像素點將由 HOG 的運算產出一個新的值並取代原像素值，此新的影像即為 HOG 增強後的影像，如圖 2.6。

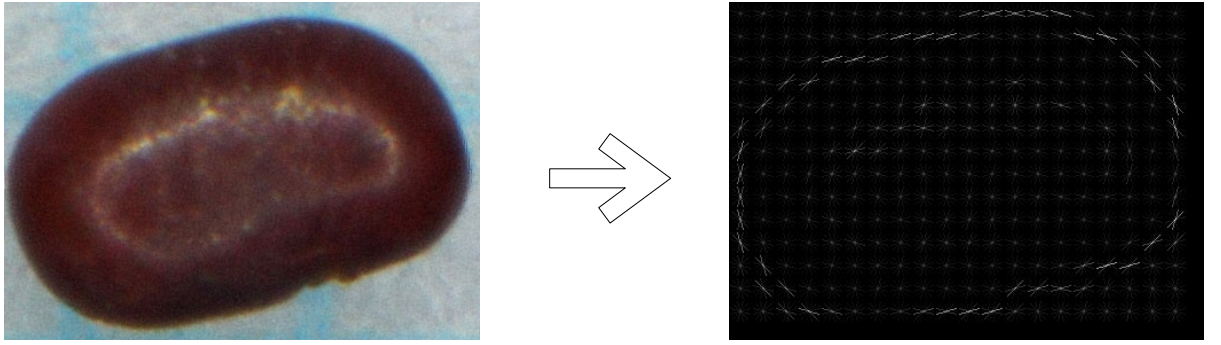


圖 2.6 經由 HOG 特徵增強後的影像

2.3 局部二值模式(LBP)

局部二值模式(Local Binary Patterns, LBP)由 T. Ojala 等人所提出[26]，應用於局部紋理特徵分類上，透過各個像素與其鄰近像素之間做比較，將其結果轉換為二進位數並保存。首先將彩色圖像轉換成灰階圖像，基本 LBP 採遮罩方式，經由每個灰階像素與鄰近的 8 個像素值進行大小比較，若大於或等於中心像素值，則給予二值化數值 1，反之則給予數值 0，並依照同一方向將其轉為二進位值，公式如(2.18) 和(2.19)， P 代表鄰近中心之數量， x_c 與 y_c 為中心像素座標， v_c 為中心像素值， v_p 為鄰近像素值。圖 2.7 呈現基本 LBP 擷取特徵之流程。

$$\text{LBP}(x_c, y_c) = \sum_{p=0}^{P-1} s(v_c - v_p) 2^p \quad (2.18)$$

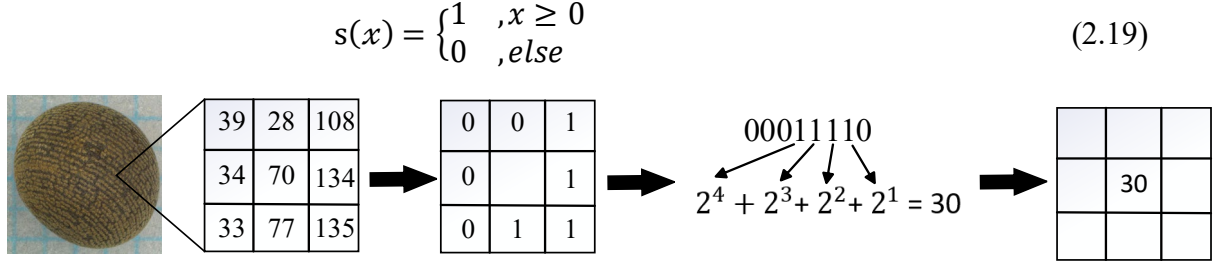


圖 2.7 基本 LBP 流程圖

由於基本的 LBP 僅能在固定範圍內的區域，為了達到灰度不變與旋轉不變，後續學者繼續對 LBP 進行改良，採用了圓形鄰近區域，半徑為 R 的圓形區域內，有 P 個像素點，可根據不同的半徑大小做出不同的選擇，中心座標為 (x_c, y_c) ，利用以下公式(2.20)和(2.21)計算出鄰近像素點之座標 (x_p, y_p) ：

$$\begin{aligned} x_p &= x_c + R \cos \frac{2\pi p}{P} \\ y_p &= y_c - R \sin \frac{2\pi p}{P} \end{aligned} \quad (2.20)$$

$$f(x, y) \approx [1 - x \quad x] \begin{bmatrix} f(0,0) & f(0,1) \\ f(1,0) & f(1,1) \end{bmatrix} \begin{bmatrix} 1 - y \\ y \end{bmatrix} \quad (2.21)$$

透過半徑 R 與鄰近點之個數 P 做運算，由於計算結果有可能為非整數情況，固利用雙線性差值公式將其轉換為整數。依據不同的 R 、 P 組合，會有不同的特徵值，且當同一張圖片作旋轉後，也會有不同的特徵值出現，為了使旋轉能不影響其結果，固學者將旋轉過後的二進位值皆視為同樣的，定義為旋轉結果中的最小值，圖 2.8 為 LBP 旋轉不變之示意圖：

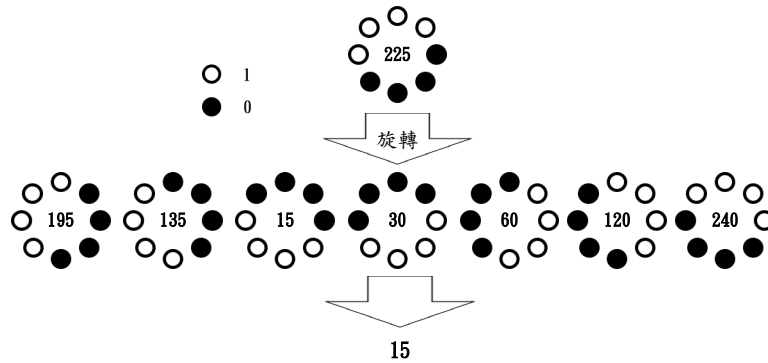


圖 2.8 LBP 旋轉不變最終輸出最小值作為中心像素點

對於半徑為 R 的範圍內有 P 個鄰近點，則會產生出 2^P 種情況，除去旋轉不變的情況，

大幅降低變異的可能性，但由於 P 的個數增加，則會產生許多情況，導致數據量過於龐大，且用特徵值所繪出的直方圖會過於稀疏，固後續提出等價(Uniform)LBP[30]降低了特徵資料的維度。T. Ojala 等人發現大多數的圖像皆能以特定幾種的特徵值來表示完整的圖像特徵資訊，且此種圖像通常都擁有較少的空間轉換，由下列公式(2.22)(2.23)來描述 LBP 等價模式：

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if } U(LBP_{P,R}) \leq 2 \\ P + 1 & \text{otherwise,} \end{cases} \quad (2.22)$$

$$U(LBP_{P,R}) = |s(g_{p-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (2.23)$$

U()函數計算空間轉換的次數-累加特徵值二進位數當中 0 到 1 或 1 到 0 的轉變次數，在 P 個像素的圓形鄰近點內，轉變次數若大於兩次的視為 P+1 類，反之稱為 uniform pattern，等價模式的 LBP 將大幅減少特徵值二進位數的種類。

一張原始影像的每個像素點將由等價 LBP 的運算產出一個新的值以取代原像素值，此新的影像即為等價 LBP 增強後的影像，如圖 2.9。

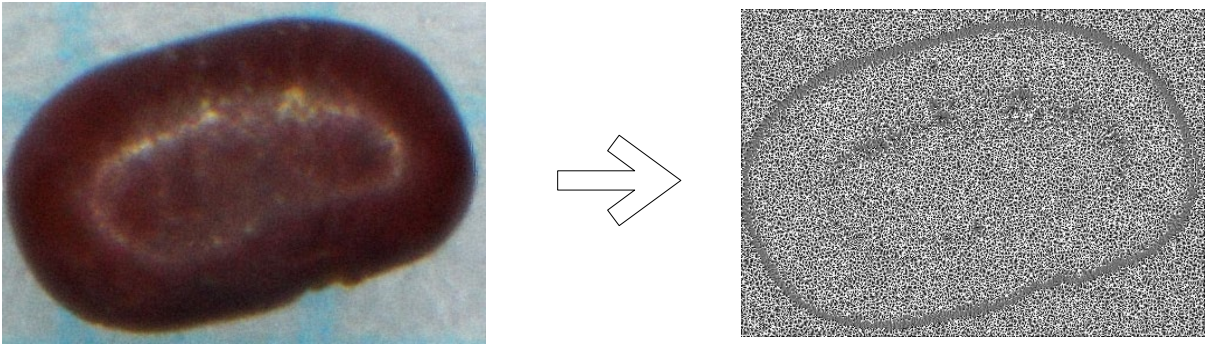


圖 2.9 經由等價 LBP 特徵增強後的影像

2.4 單尺度視網膜增強算法(SSR)

單尺度視網膜增強算法(Single-Scale Retinex , SSR)由 D. J. Jobson 等人所提出[27]，應用於影像色彩增強上，取得影像資訊的反射影像，可在色彩恆常性與動態

壓縮上達到平衡。

影像可視為由入射影像與反射影像所組合而成的，首先將彩色圖像分解成 Red、Green、Blue 三個通道，分別對三個通道進行後續處理，Jobson 等人發現採用高斯函數與原始圖像進行卷積運算能與入射影像相似，如公式(2.24)，其中 i 為 R、G、B 三個顏色通道的其中一個， $*$ 為卷積運算符號， $I_i(x, y)$ 為第 i 個通道的像素值， $G(x, y, c)$ 為高斯函數(2.25)， K 則由(2.26)推導而出，其中 c 為高斯核，控制高斯函數之鄰近大小，影響影像色彩與細節的保留。

$$R_i^{SSR}(x, y) = \log I_i(x, y) - \log(I_i(x, y) * G(x, y, c)) \quad (2.24)$$

$$G(x, y, c) = Ke^{-(x^2+y^2)/c^2} \quad (2.25)$$

$$\iint G(x, y, c) dx dy = 1 \quad (2.26)$$

一張原始影像拆分成 R、G、B 三個通道，每個通道的每個像素點將由 SSR 的運算產出一個新的值以取代原像素值，再將新的三通道合併，即為 SSR 增強後的影像，如圖 2.10。

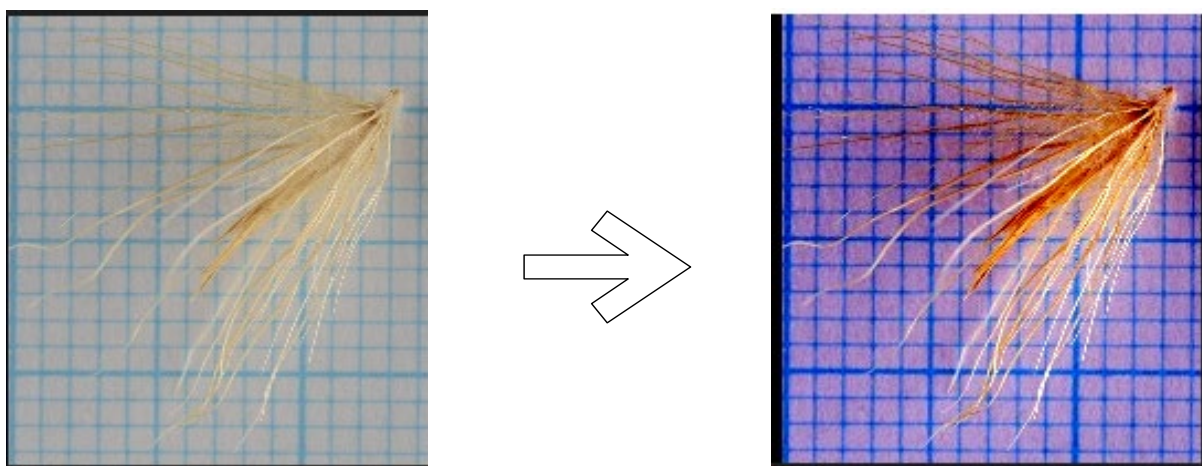


圖 2.10 經由 SSR 特徵增強後的影像

2.5 深度學習

深度學習[31]是由機器學習中延伸而出的，模擬人類的神經系統，架構採用堆疊式多層人工神經網路，對資料進行特徵學習與特徵擷取以取代傳統的手動特徵擷

取。深度學習採用了分層的概念，輸入與輸出之間藉由多層的神經元相互連結，而每個神經元中皆有激勵函數(Activation Function)，激勵函數將神經元的輸入映射至輸出。

2.5.1 卷積神經網路(CNN)

卷積神經網路(CNN)是現今影像分類的主流方法，在傳統的機器學習中，需要手動提取特徵後再由專門分類的網路進行分類，而 CNN 將手動提取特徵改為自動提取特徵，進而提取出最適合的特徵，並將此特徵交給後續的全連接層做分類，減少傳統特徵擷取的人力耗費與提升整體的準確率。CNN 由卷積層(Convolution layer)、池化層(Pooling layer)、全連接層(Fully Connected layer)所組合而成，圖 2.11 為卷積神經網路架構示意圖。

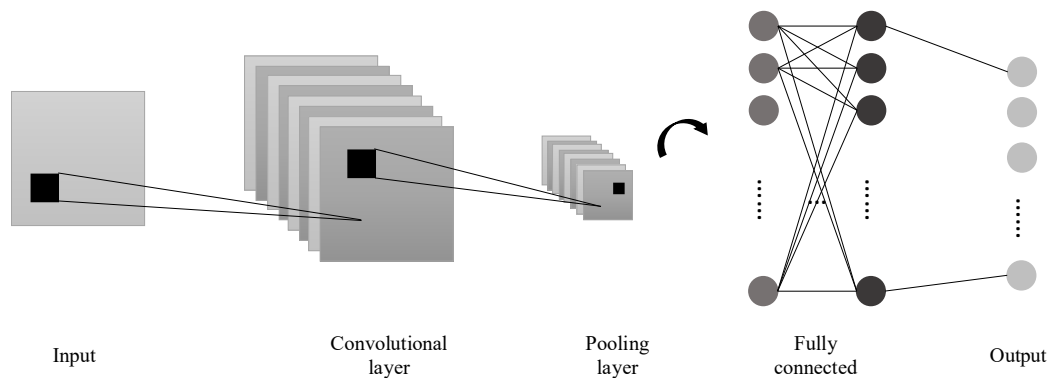


圖 2.11 卷積神經網路架構

卷積層以滑動卷積核並與特定的特徵檢測器(Feature Detector)做卷積運算[5]，並利用共享權重的方式來降低參數量，進而達到特徵擷取與特徵映射的效果，由於卷積層為線性運算，固需加上非線性的激勵函數。

池化層針對特定區域提取特徵，具有了圖像的平移、旋轉與尺度不變性，進而再次降低神經網路的參數量，並防止過擬合(overfitting)。常用的 pooling 方式為 Max Pooling 與 Average Pooling，Max Pooling 取出固定範圍內的最大值，將獲得更明顯的特徵，如邊緣特徵，而 Average Pooling 是輸出固定範圍內的平均值，能提取到較為平滑的特徵，圖 2.12 為 Max Pooling 與 Average Pooling 示意圖。

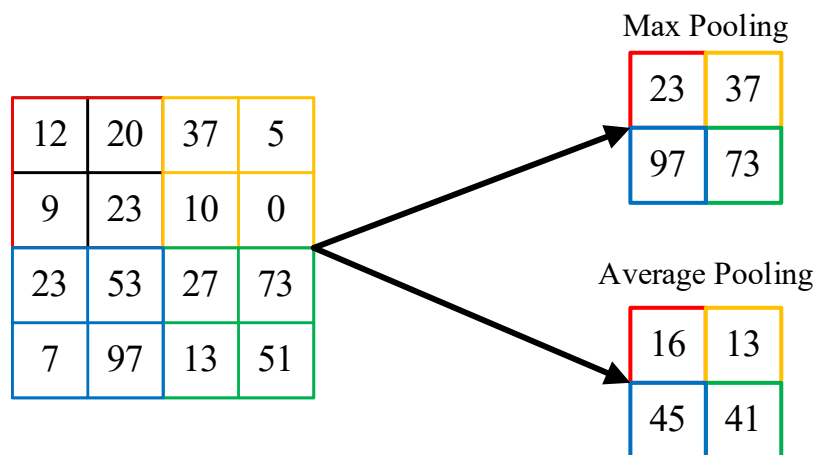


圖 2.12 常見的 Pooling 方式

全連接層類似於機器學習中的分類器，將先前取出的特徵輸入至全連接層內，藉由調整權重與偏差取得結果。

Y. Lecun 等人提出了 LeNet-5 網路架構[1]，基於梯度的反向傳播對模型進行訓練，至此奠定了 CNN 模型的架構。A.Krizhevsky 所提出的 AlexNet[32]於 2012 年的 ImageNet LSVRC 比賽上榮獲冠軍，至此 CNN 備受眾人關注，AlexNet 採用了 ReLU 非線性激勵函數，而 LeNet 則是使用 tanh，ReLU 解決了 tanh 容易發生的梯度消失問題且計算量較小和收斂速度快，固成為最常使用的激勵函數。後續也越來越多新穎的 CNN 網路模型出現，如 VGGNet[3]、GoogleNet[33]與殘差神經網路(ResNet)[4]，VGGNet 與 GoogleNet 分別加深了網路的深度與寬度，皆提升了辨識率；後續經過實驗發現，並非加深層數即可獲得較好的辨識率，當層數到達一定數量後，網路會變得難以訓練，ResNet 使用殘差學習(Residual learning)中的恆等映射(Identity mapping)，解決模型退化問題，ResNet 也在 2015 年的 ImageNet 榮獲冠軍。但由於 CNN 網路的層數較多，較耗費計算資源，且於分類錯誤時無法清楚理解其擷取出的特徵作用，造成無法解釋。

2.6 MLP-Mixer

在電腦視覺領域中，網路模型架構起初是由多層感知器(multi-layer perception,MLP)開始，後續卷積的崛起取代了 MLP 架構，接著自注意力機制(Self-

attention)的出現與卷積並駕齊驅,直到 2021 年 Google 團隊提出新的網路模型 MLP-Mixer[15]後,網路模型架構又回到了 MLP,在 MLP-Mixer 尚未被提出前,電腦視覺領域中最受討論的網路模型非 CNN 與 ViT 莫屬,而 I. Tolstikhin 等人主張卷積與自注意力雖皆能獲取優異的辨識性能,但並非必要。

MLP-Mixer 僅依賴矩陣乘法與非線性層,使其架構簡單與計算速度較快,並且成功地在電腦視覺上與 CNN、ViT 不相上下。MLP-Mixer 主要由每個小塊全連接層(per-patch fully-connected)、Mixer layer 與全連接層(Fully-connected)所組合而成,如圖 2.13。透過 Fully-connected 進行分類,無法獲得局部區域之間的訊息,固使用 per-patch fully-connected,將維度為 H, W, C 的輸入影像,切割成數個大小為 $P \times P \times 3$ 的不重疊小塊(patch),總共可分割成 S 塊,其公式如(2.27), H, W 分別代表輸入影像的高與寬, C 則為輸入影像的通道數(channel),若輸入影像為彩色的話,則 $C=3$,若為灰階圖的話 $C=1$, p 為切割影像的長、寬。

$$S = \frac{WH}{P^2} \quad (2.27)$$

再將每塊 patch 平坦化為一維向量,其長度為 $3P^2$,即獲得一張量且維度為 $(S, 3P^2)$,再經由線性投影將 $3P^2$ 轉換成大小為 C 的隱藏維度,其張量維度為 $S \times C = \text{patches} \times \text{channels}$,共有 S 個名為令牌(token)的 patch 向量,其大小為 $1 \times C$ 。per-patch fully-connected 內實現了將輸入影像經過分割與線性映射後轉為序列,以利於後續進行局部區域之訊息融合。

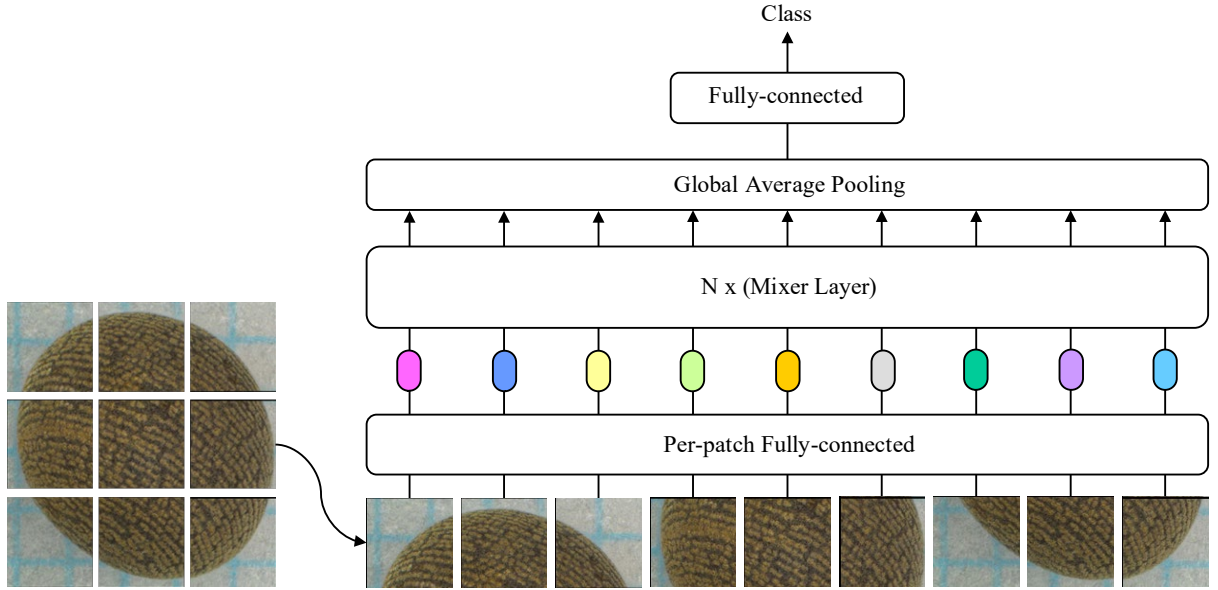


圖 2.13 MLP-Mixer 應用於影像分類

Mixer 層包含兩種的 MLP，一種為通道混合(channel-mixing)MLP 應用於局部特徵上，結合不同通道(channels)之間的特徵，另一種為令牌混合(token-mixing) MLP 應用於空間上，融合不同空間位置的訊息，而 MLP 由全連接與激勵函數所組成，為了提升性能，Mixer 層內添加了層的正規化(Layer norm)和殘差連結(skip-connection)，如圖 2.14，圖中的 MLP1 即為 token-mixing MLP，MLP2 為 channel-mixing MLP。透過 per-patch fully-connected 輸出的二維矩陣 $m \times n$ 作為 Mixer 層的輸入，先將矩陣進行正規化後並對其轉置成 $n \times m$ 矩陣後傳入 token-mixing MLP 進行計算，再將 $n \times m$ 矩陣轉回 $m \times n$ 矩陣並正規化以此作為 channel-mixing MLP 的輸入。 D_s 為 token-mixing MLP 的可調寬度與 S 成正比，而 D_c 是 channel-mixing MLP 的可調寬度與 C 成正比。

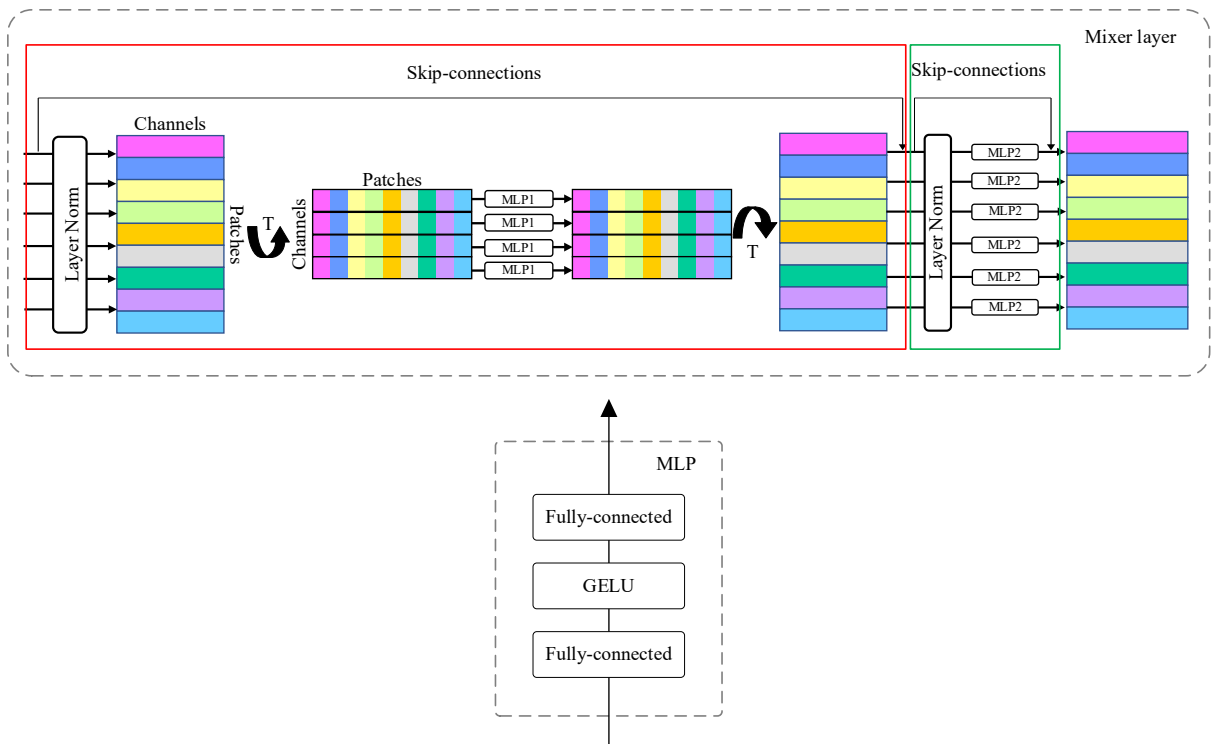


圖 2.14 Mixer Layer 架構

第三章、 影像辨識分類系統

CNN 十分耗費計算資源且其自動擷取出的特徵難以分析特徵，後續較難針對分類結果進行解釋分析，由於卷積層的權重難以修改，固本研究提出一個增強影像特徵之分類系統，先將輸入的原始影像做特徵增強，使用形狀、紋理與顏色三種多模組影像增強的方式，來提升特徵的明顯度，再結合 MLP-Mixer 分類器，分別輸出形狀、紋理與顏色之分類結果再透過決策融合來進行最終的分類。

本章節使用 MIAT 方法論[34]來呈現本研究之系統架構與設計，首先由 Integrated Computer Aided Manufacturing(ICAM) Definition for Function Modeling(IDEF)中的 IDEF0[35]來創建系統之功能模組，將系統切分成多個獨立模組，產生階層式的模組設計，再由 Grafcet[36]進行離散事件建模，針對每個獨立模組創建動作(action)與轉移條件，最後使用高階合成來整合軟硬體之程式碼完成驗證，第一小節介紹特徵增強策略於 MLP-Mixer 影像分類器之階層式架構，第二小節進行分類器的離散事件建模。

3.1 分類系統架構

特徵增強策略分類器系統架構如圖 3.1，第一層的 A0 模組即為分類系統的主要模組，將 A0 向下切分成兩個子模組 A1 與 A2，A1 為影像增強的 MLP-Mixer 分類器，而 A2 為決策融合 stacking 輸出，原始資料輸入後，依序經過多個功能模組運算後，即可獲得最終影像辨識之種類結果。

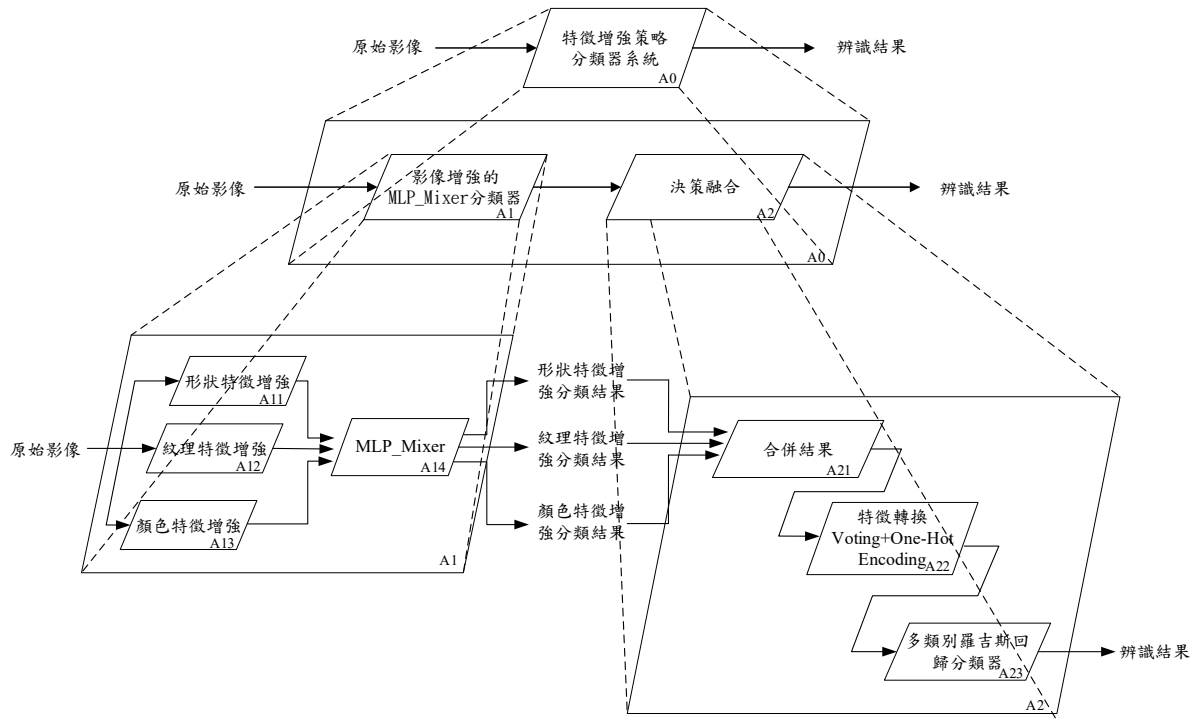


圖 3.1 特徵增強策略分類器系統主要架構

首先進入 A1 子模組，先對影像分別進行特徵增強，本研究針對形狀、紋理、顏色進行特徵增強，採用平行運算模式，同時進行不同的特徵增強，A11 為形狀特徵增強模組，A12 為紋理特徵增強模組，A13 為顏色特徵增強模組，將增強過後的影像分別作為 A14 子模組 MLP-Mixer 的輸入，而後分別輸出形狀、紋理與顏色特徵增強之分類結果，將其結果再作為 A2 決策融合的輸入，進入到 A21 將三個結果進行合併後，再到 A22 做 Voting+One-Hot Encoding 特徵轉換，將轉換過後的特徵輸入到 A23 做多類別羅吉斯回歸(Multinomial Logistic Regression)分類，最終輸出一個綜合三種特徵增強分類之結果。

特徵增強影像(Feature-Enhanced Image, FEI)為特徵擷取運算過後的特徵圖(feature map)與原始影像進行疊圖，其架構流程如圖 3.2。本研究採三種特徵增強方式，將以 S_N^{3+n} 、 T_N^{3+n} 、 C_N^{3+n} 分別代表形狀特徵增強、紋理特徵增強與顏色特徵增強之輸出結果，其中 n 為所有特徵擷取方法之 feature map 的總通道數， $n \in \mathbb{Z}^+$ ，3 為 RGB 的通道數， N 為一集合，其集合元素為特徵擷取運算方法，如紋理特徵增強輸出為 T_N^4 , where $n = 1$ and $N = \{\text{LBP}\}$ 。

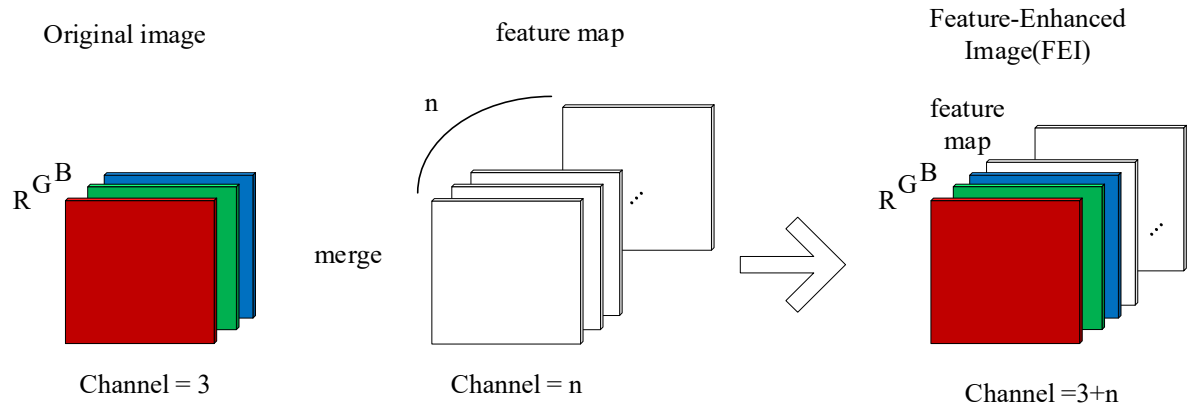


圖 3.2 影像特徵增強架構圖

而影像特徵增強又可分為多個子模組，如圖 3.3，A11 形狀特徵增強再向下切分為兩個子模組，首先為 A111HOG 影像增強，採用 HOG 運算擷取出影像邊緣資訊，再將擷取出的 feature map 輸入至 A112 與原始影像疊圖，輸出 S_N^4 , where $n = 1$ and $N = \{HOG\}$ 。A12 紋理特徵增強，首先先進入 A121LBP 影像增強模組，採用等價 LBP 運算出紋理特徵後，將其 feature map 輸入至 A122 與原始影像進行疊圖，輸出為 T_N^4 , where $n = 1$ and $N = \{LBP\}$ 。A13 顏色特徵增強模組向下分割為兩個子模組，分別為 A131SSR 影像增強模組，採用 SSR 運算出顏色特徵後，再將 feature map 輸入至 A132 與原始影像疊圖，其輸出為 C_N^6 , where $n = 3$ and $N = \{SSR\}$ 。

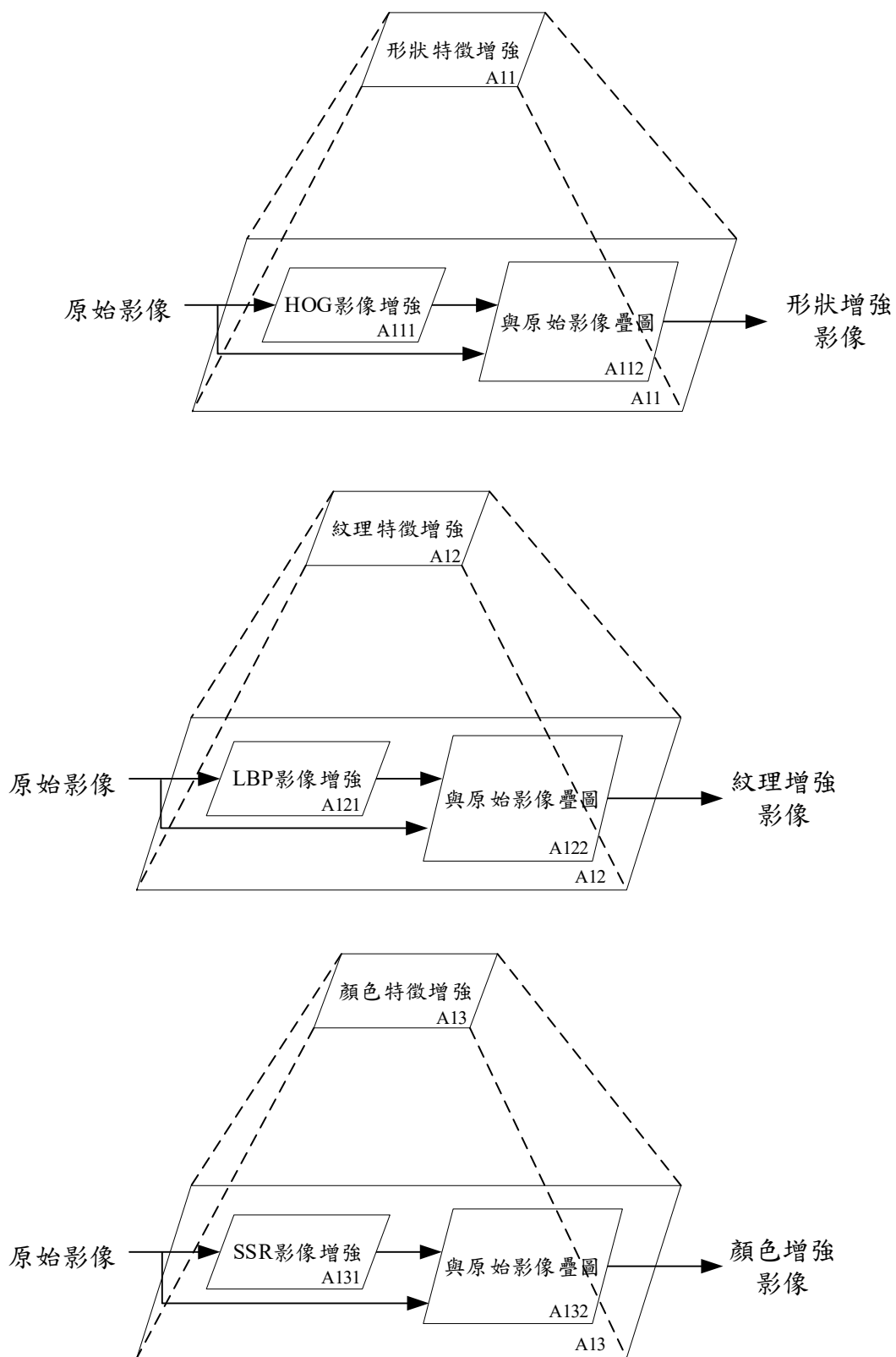


圖 3.3 影像特徵增強系統架構

3.2 分類器系統離散事件建模

此小節將以離散事件建模來介紹分類器的運作過程與其轉移條件，圖 3.4 為特徵增強策略在 MLP-Mixer 分類器的 Grafcet，0 為初始狀態，直到第一個轉移條件 Image 成立，代表原始影像輸入後，系統開始辨識，將進入狀態 1 影像增強的 MLP-Mixer 分類器階段，在此階段會先進行影像增強，再將增強後的影像輸入至 MLP-Mixer 進行分類後，取得各別的輸出 Top-5 結果，便觸發下個轉移條件，即狀態 16 完成且 output_rank5 成立後，將從狀態 1 進入至狀態 2 決策融合，綜合 3 個模型的結果進行決策，輸出一個最終結果，當狀態 24 完成且 done 條件成立，即整個系統執行結束，將會回到初始狀態 0 等待，並將各個資料與轉移條件等調整回初始設定，等待下次觸發 Image 條件。

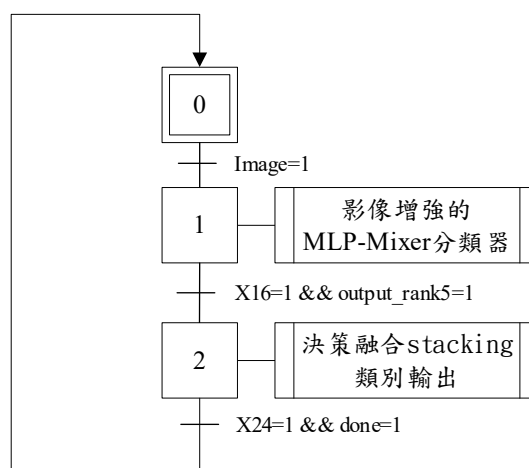


圖 3.4 特徵增強策略 MLP-Mixer 分類器離散事件建模

3.2.1 影像特徵增強離散事件建模

圖 3.5 為影像特徵增強之離散事件建模，其為平行架構，同時進行 3 種特徵增強，當 Image 成立後，才進入到狀態 1，其初始狀態為 10，轉移條件成立後，轉移至狀態 11，讀取影像，當 ImageRGB 成立時，同時轉移至狀態 12、13、14，分別進行形狀特徵增強、紋理特徵增強及顏色特徵增強，當各別的子 Grafcet 皆完成後，才可進入至下個狀態 15，分別儲存特徵增強後的影像，當 Enhanced_image 成立時，

轉移至狀態 16MLP-Mixer 分類器，進行影像分類並各別輸出 Top-5 結果，即完成狀態 1 的流程，並回到狀態 10 等待。

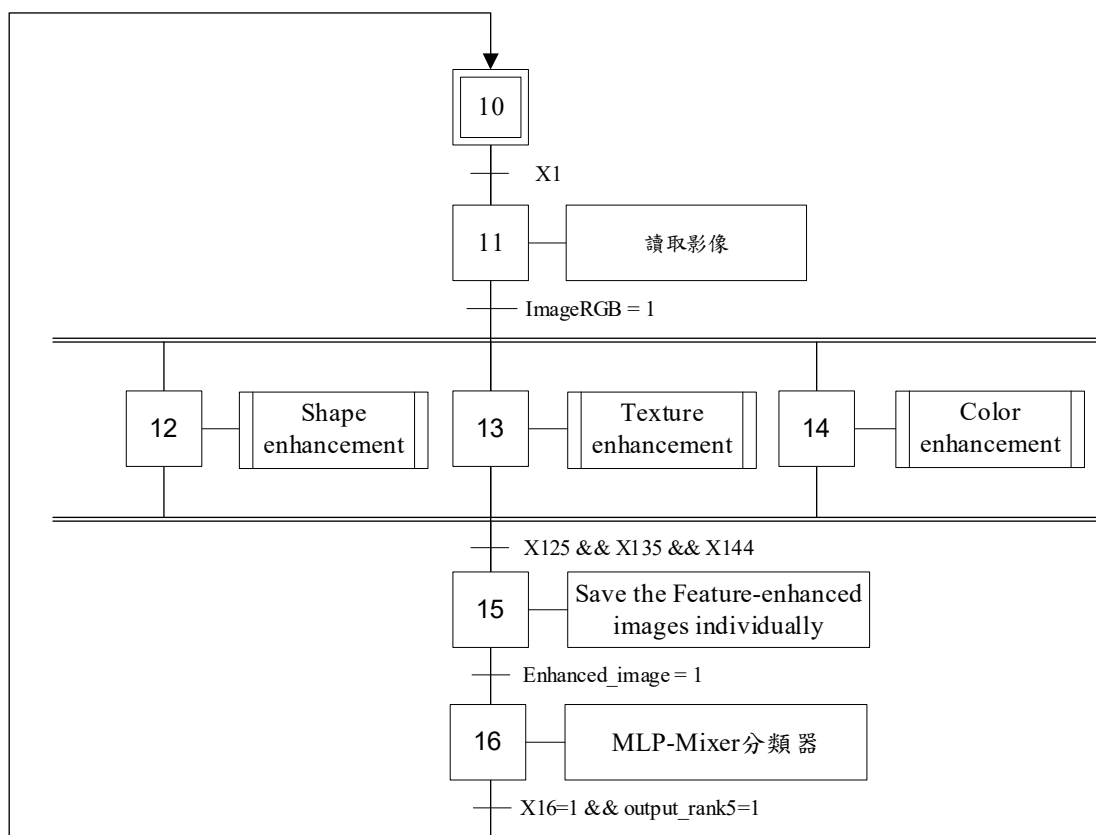


圖 3.5 影像特徵增強離散事件建模

圖 3.6 為形狀特徵增強之離散事件建模，狀態 12 的初始狀態為狀態 120，當轉移條件成立後，轉移至狀態 121 將影像轉成灰階圖像，當 Gray 成立，轉移至狀態 122 進行 Gamma 校正，當 Gamma_done 成立後，轉移至狀態 123，計算影像之梯度大小與方向，並用角度區分進行統計，Gradient 成立後，轉移至狀態 124，進行 L2 正規化，當 Feature_hog 成立後，會進入至下個狀態 125，將提取出的 feature map 與原圖做疊圖，完成後回到狀態 120 等待。

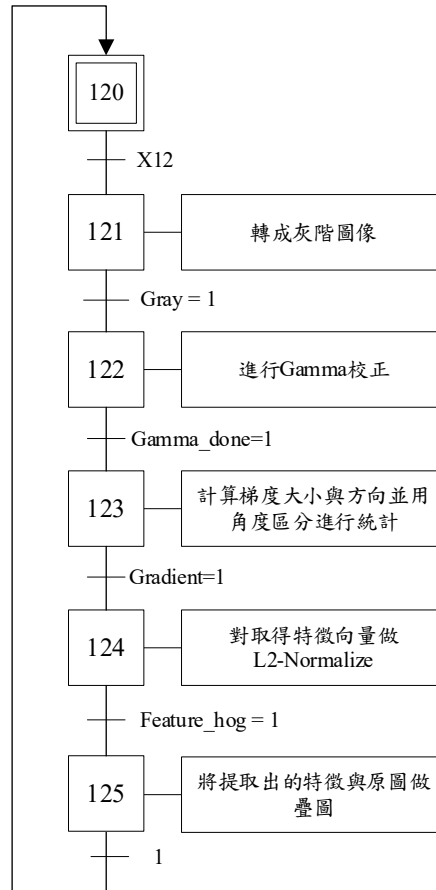


圖 3.6 形狀特徵增強離散事件建模

圖 3.7 為紋理特徵增強之離散事件建模，狀態 130 轉移至狀態 131，將影像轉成灰階圖像，當 Gray 成立，狀態轉移至 132，進行等價局部二值化，當 Binary_count 成立時，轉移至狀態 133，計算 0 與 1 之間的轉換次數，當 transfer 成立時，轉移至狀態 134，依照轉換次數給予相對應的類別，當 Feature_lbp 成立後，會進入至下個狀態 135，將提取出的 feature map 與原圖做疊圖，完成後將回到初始狀態 130 等待下次的觸發。

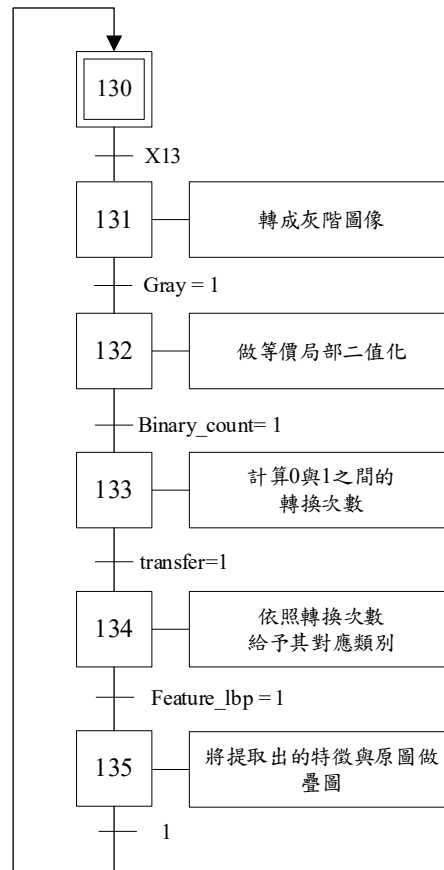


圖 3.7 紋理特徵增強離散事件建模

圖 3.8 為顏色特徵增強之離散事件建模，狀態 140 轉移至狀態 141，讀取彩色影像並將 R、G、B 三通道分離，當 Split_rgb 成立時，轉移至狀態 142，經由公式 (2.24) 運算後獲取新的像素值，當 New_pixel 成立時，轉移至狀態 143，將新的三通道合併，當 Feature_ssr 成立後，會進入至下個狀態 144，將提取出的 feature map 與原圖做疊圖，完成後將回到初始狀態 140 等待下次的觸發。

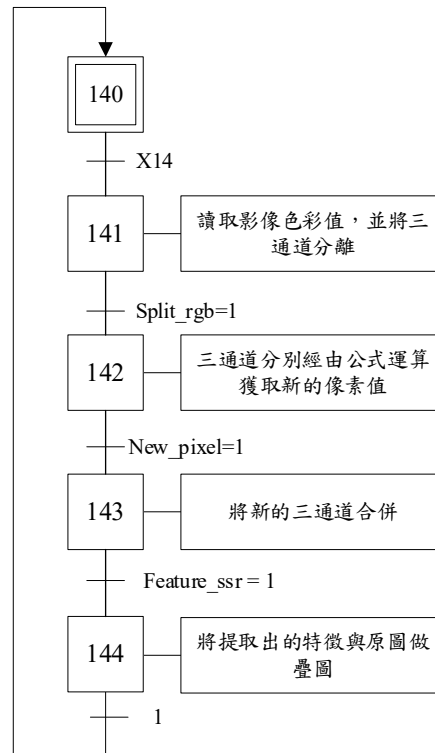


圖 3.8 顏色特徵增強離散事件建模

3.2.2 決策融合離散事件建模

圖 3.9 為決策融合 Stacking 之離散事件建模，當轉移條件成立時，初始狀態 20 轉移至狀態 21，將三個模型輸出的 rank5 進行合併，當 merge_rank5 成立時，轉移至狀態 22，將合併的輸出進行 Voting+One-Hot Encoding 後輸出新的特徵向量，當 encoding 成立時，轉移至狀態 23，將新的特徵向量輸入至 Meta-model 進行決策融合，decision_fusion 成立時，轉移至狀態 24，輸出最終相似類別前 5 名，完成後回到初始狀態 20 等待下次觸發。

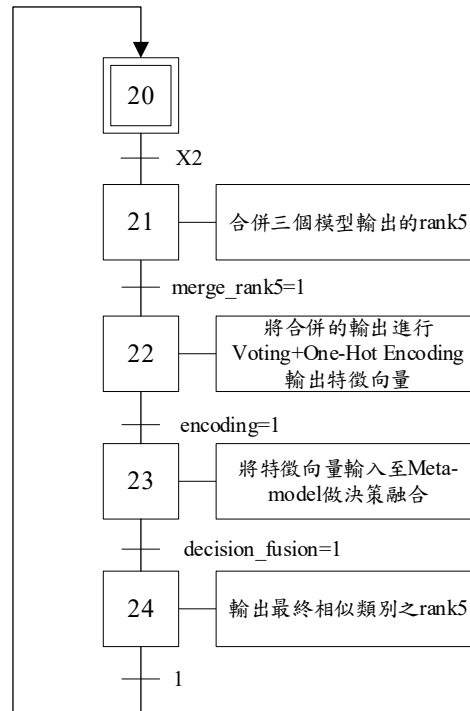


圖 3.9 決策融合離散事件建模

第四章、系統整合與驗證

本節將使用三個不同的資料集來進行上一章節所提出的使用特徵增強策略在 MLP-Mixer 影像分類器之實驗，並逐步地與不同的實驗進行比較，探討其性能與結果。本章節將分成四個小節，第一小節先介紹開發環境，第二小節為資料集介紹，第三小節將分別與其他實驗進行評估比較。

4.1 實驗開發環境介紹

本研究所使用之軟體訓練平台如表 4.1 所示。為了避免資料集過少導致過擬合 (overfitting) 與提升開發效率，採用了 MLP-Mixer_L/16_224 於訓練時進行遷移學習，其模型參數如表 4.2。而軟體測試平台如表 4.3。

表 4.1 訓練環境

項目	規格
電腦主機	處理器：Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz 記憶體：32GB DDR4 硬碟：Intel SSDPEKKW256G8 256GB Seagate ST1000DM010-2EP102 1TB/7200 轉 作業系統：Ubuntu 20.04.1 圖形處理器：NVIDIA GeForce GTX 1080 Ti
程式開發環境	編譯器：Visual Studio Code version1.67 程式語言：Python 3.7
運行環境	Pytorch 1.10.1 Opencv-python 4.5.3.56 Scikit-learn 1.0.2
訓練模型	MLP-Mixer_L/16_224 pretrained by ImageNet-21k

表 4.2 MLP-Mixer_L/16_224 規格表

參數規格	MLP-Mixer_L/16_224
Image input size	224×224
Number of layers	24
Patch resolution $P \times P$	16×16
Hidden size C	1024

Sequence length S	196
MLP dimension D_C	4096
MLP dimension D_S	512
Parameters (M)	229
Pretrained dataset	ImageNet-21K

表 4.3 圖形化介面應用環境

項目	規格
電腦主機	處理器：Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz 記憶體：16GB DDR3 硬碟：KINGSTON SUV400S37240G WDC WD10EZEX-75ZF5A0 作業系統：Windows10 專業版 64-bits
程式開發環境	編譯器：Visual Studio Code version1.67 程式語言：Python3.7
運行環境	Pytorch 1.10.1 Opencv-python 4.5.3.56 Scikit-learn 1.0.2

4.2 實驗資料集介紹

本研究將採用三個不同的影像資料集進行實驗驗證，分別為魚類影像資料集、種子影像資料集與中歐森林影像資料集。其中魚類與種子資料集皆為 MIAT 實驗室過去與現今使用於影像辨識上之資料集影像，而中歐森林影像數據則是來自於開源資料[37]。

4.2.1 魚類資料集

魚類資料集由行政院農業委員水產試驗所提供，共計 40 類，每類皆包含 30 張影像，其中訓練集共有 800 張，驗證集與測試集皆為 200 張影像。圖 4.1 為魚類資料集影像之範例。



圖 4.1 魚類資料集範例

4.2.2 種子資料集

種子資料集由行政院農業委員會種苗改良繁殖場所提供，其種類為非延伸種，共計 560 類種子，其中訓練集共有 16800 張影像，驗證集 1680 張，測試集 2800 張，

圖 4.2 為種子資料集影像之範例。

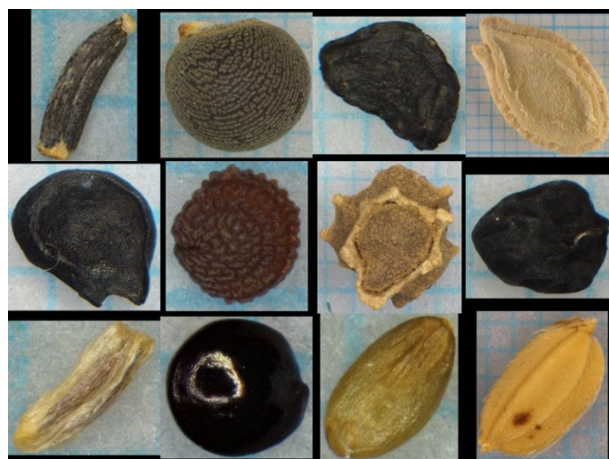


圖 4.2 種子資料集範例

4.2.3 中歐森林資料集

中歐森林資料集(Middle European Woods , MEW)[37]包含了 153 種中歐地區常見的木本植物，其資料集內容為木本植物的樹葉，由樹葉影像進行辨識，每種類別至少 50 張影像不等，共有 9745 個樣本，其中訓練集共有 6120 張影像，驗證集與測試集分別有 1530 張和 3825 張影像，圖 4.3 為中歐森林資料集影像之範例。



圖 4.3 中歐森林資料集影像範例

4.3 特徵增強策略在 MLP-Mixer 影像分類驗證

此小節將進行系統實驗驗證，並分別採用上個小節所介紹的資料集進行驗證，首先將先進行在 MLP-Mixer 影像分類器上有無使用特徵增強之比較其辨識結果，再比較本系統與混合式神經網路之辨識結果，最後進行本系統與單個神經網路之辨識結果比較。

4.3.1 有無特徵增強於 MLP-Mixer 影像分類之比較

此小節將進行在 MLP-Mixer 影像分類器上有無先行影像增強之比較其辨識結果，使用 MIAT 的 40 類魚類資料集進行驗證。表 4.4 為魚類資料集於本系統的訓練參數與結果。

表 4.4 魚類資料集於本系統訓練參數與結果



Epochs	30
Batch_size	32
Learning Rate	0.0003
Momentum	0.9
Accuracy	99%
Precision	99.17%
Recall	99%
F1-score	98.99%

表 4.5 為 MIAT 魚類資料集在 MLP-Mixer 影像分類器上是否有先進行影像增強之辨識率比較，由表 4.5 實驗結果可得知採用影像增強策略在 MLP-Mixer 分類器上優於未採用特徵增強的影像分類器約 3%，表 4.6 為與土魷魚相似的魚種和土魷魚在經過三種特徵增強後的分類結果，魚類資料集通常於 Top-2 就可獲得正確結果，故在此表格呈現上僅檢視 Top-2 的分類結果，可觀察出土魷魚和花腹鯖在形狀上皆為扁細長魚身，在顏色上兩者皆為灰銀色，且在魚背上顏色較深，雖在顏色上分辨錯誤成花腹鯖，但在最終決策融合的輸出結果仍為土魷魚。

表 4.5 MIAT 魚類資料集在有無採用特徵增強於 MLP-Mixer 比較

神經網路架構	使用特徵增強策略在 MLP-Mixer 影像分類器	未採用特徵增強之 MLP-Mixer 影像分類器
Top-1 accuracy	99%	96%

表 4.6 土魷魚特徵增強分類結果

種類名稱	土魷魚	花腹鯖
影像		
Top-2 分類結果	1	2
形狀特徵增強	土魷魚	花腹鯖
紋理特徵增強	土魷魚	黃鰭鯖
顏色特徵增強	花腹鯖	土魷魚
決策融合結果	土魷魚	

4.3.2 本系統與混合式神經網路之比較

此小節將進行使用特徵增強策略在 MLP-Mixer 影像分類器與混合式神經網路之 Top-1 辨識率比較，採用 MIAT 的種子資料集進行驗證。混合式神經網路結合了 ResNet 中的 ResNet-50 與孿生神經網路(Siamese Network, Siamese)，先由 ResNet-

50 進行初步分類，輸出最相似之前五類(Top-5)種子，再由 Siamese 將測試資料與最 Top-5 的種子進行兩兩比較，並計算其歐式距離(Euclidean Distance)，針對 Top-5 的種子進行重新排序其相似順序。表 4.7 為種子資料集於本系統訓練參數與結果。

表 4.7 種子資料集於本系統訓練參數與結果

Epochs	50
Batch_size	32
Learning Rate	0.0001
Momentum	0.9
Accuracy	90.65%
Precision	91.6%
Recall	90.13%
F1-score	89.67%

表 4.8 為 MIAT 種子資料集在本系統與混合式神經網路(ResNet-50+Siamese)之 Top-1 準確率比較，由表 4.8 可得知本系統在種子分類上優於混合式神經網路約 20%，表 4.9 為與韭蔥相似的種子和韭蔥在經過各別特徵增強後分類之結果，種子資料集通常於 Top-2 就可獲得正確結果，故在此表格呈現上僅檢視 Top-2 的分類結果，可觀察出韭蔥與洋蔥在形狀上皆為橢圓且上方會有個凹陷，在紋理上洋蔥有較為明顯的皺褶，而在顏色上韭蔥與洋蔥皆為黑灰色，雖然在形狀上分辨錯誤成洋蔥，但在最終決策融合的輸出結果仍為韭蔥。

表 4.8MIAT 種子資料集在本系統與混合式神經網路之比較

神經網路 架構	使用特徵增強策略在 MLP-Mixer 影像分類器	未採用特徵增強之 MLP-Mixer 影像分類器	混合視神經網路 (ResNet-50 +Siamese)
Top-1 accuracy	90.65%	89.43%	70.23%

表 4.9 韭蔥特徵增強分類結果



種類名稱	韭蔥	洋蔥
影像		
Top-2 分類結果	1	2
形狀特徵增強	洋蔥	韭蔥
紋理特徵增強	韭蔥	洋蔥
顏色特徵增強	韭蔥	洋蔥
決策融合結果	韭蔥	

表 4.10 為與越瓜相似的種子和越瓜在經過各別特徵增強後分類之結果，在形狀上可觀察出兩者皆為橢圓形，但越瓜的底又比胡瓜來的圓弧，而顏色上兩者皆為黃色，雖在顏色上分辨錯誤成胡瓜，但在最終決策融合的輸出結果仍為越瓜。

表 4.10 越瓜特徵增強分類結果



種類名稱	越瓜	胡瓜
影像		
Top-2 分類結果	1	2
形狀特徵增強	越瓜	胡瓜
紋理特徵增強	越瓜	胡瓜
顏色特徵增強	胡瓜	越瓜
決策融合結果	越瓜	

表 4.11 為與油菜相似的種子和油菜在經過各別特徵增強後分類之結果，可觀察出油菜與芥菜在形狀上皆為圓形，而油菜在紋理上有較為明顯的格狀，芥菜則是多了一小塊黑蒂，顏色上油菜為黑灰色而芥菜為紅棕色，雖在形狀上分辨錯誤成芥菜，但在最終決策融合的輸出結果仍為油菜。

表 4.11 油菜特徵增強分類結果


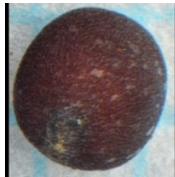


種類名稱	油菜	芥菜
影像		
Top-2 分類結果	1	2
形狀特徵增強	芥菜	油菜
紋理特徵增強	油菜	芥菜
顏色特徵增強	油菜	芥菜
決策融合結果	油菜	

表 4.12 為與甜椒相似的種子和甜椒在經過各別特徵增強後分類之結果，可觀察出甜椒與辣椒在形狀上皆近似圓形，而辣椒在紋理上有較為明顯的橫條，顏色上甜椒為橘黃色而辣椒則為黃色，雖在形狀上分辨錯誤成辣椒，但在最終決策融合的輸出結果仍為甜椒。

表 4.12 甜椒特徵增強分類結果

種類名稱	甜椒	辣椒
影像		
Top-2 分類結果	1	2
形狀特徵增強	辣椒	甜椒
紋理特徵增強	甜椒	辣椒
顏色特徵增強	甜椒	辣椒
決策融合結果	甜椒	

4.3.3 本系統與單個神經網路之比較

此小節將進行使用特徵增強策略在 MLP-Mixer 影像分類器與單個神經網路的 Top-1 辨識率比較，採用開源的中歐森林資料集進行驗證，與 J. Gu 等人[38]所採用

的單個神經網路 VGG16 實驗進行比較。表 4.13 為中歐森林資料集於本系統訓練參數與結果。

表 4.13 中歐森林資料集於本系統訓練參數與結果

Epochs	50
Batch_size	32
Learning Rate	0.0001
Momentum	0.9
Accuracy	97.91%
Precision	98.21%
Recall	97.81%
F1-score	97.87%

表 4.14 為開源的中歐森林資料集在本系統和單個神經網路(VGG16)之 Top-1 準確率比較，由表 4.14 可得知本系統在中歐森林分類上優於單個神經網路約 4.5%。

表 4.14 開源中歐森林資料集在本系統與單個神經網路之比較

神經網路 架構	使用特徵增強策略在 MLP-Mixer 影像分類器	未採用特徵增強之 MLP-Mixer 影像分類器	單個神經網路 (VGG16)
Top-1 accuracy	97.91%	96.86%	93.4%

表 4.15 為與 Hedera helixSTERILE 相似的樹葉和 Hedera helixSTERILE 在經過各別特徵增強後分類之結果，中歐森林資料集通常於 Top-2 就可獲得正確結果，故在此表格呈現上僅檢視 Top-2 的分類結果，可以觀察到 Hedera helixSTERILE 在紋理與顏色上皆較容易與 Liquidambar styraciflua 分辨錯誤，而在形狀上則可以明確辨別出差異，雖在紋理與顏色上辨識錯誤成 Liquidambar styraciflua，但在最終決策融合的輸出結果仍為 Hedera helixSTERILE。

表 4.15 Hedera helix STERILE 特徵增強分類結果






種類名稱	Hedera helix STERILE	Liquidambar styraciflua
影像		
Top-2 分類結果	1	2
形狀特徵增強	Hedera helix STERILE	Liquidambar styraciflua
紋理特徵增強	Liquidambar styraciflua	Hedera helix STERILE
顏色特徵增強	Liquidambar styraciflua	Hedera helix STERILE
決策融合結果	Hedera helix STERILE	

表 4.16 為與 *Sophora japonica* 相似的樹葉和 *Sophora japonica* 在經過各別特徵增強後分類之結果，可以觀察出 *Sophora japonica* 在形狀、顏色上與 *Lycium barbarum* 較為相似，兩者形狀皆是前方較尖的橢圓形，而在紋理上則可以觀察出 *Lycium barbarum* 和 *Robinia pseudacacia* 的葉脈較 *Sophora japonica* 明顯且間距較大，雖在形狀與顏色上辨識錯誤成 *Lycium barbarum*，但在最終決策融合的輸出結果仍為 *Sophora japonica*。

表 4.16 *Sophora japonica* 特徵增強分類結果

種類名稱	<i>Sophora japonica</i>	<i>Lycium barbarum</i>	<i>Robinia pseudacacia</i>
影像			
Top-2 分類結果	1		2
形狀特徵增強	Lycium barbarum		<i>Sophora japonica</i>
紋理特徵增強	<i>Sophora japonica</i>		<i>Robinia pseudacacia</i>
顏色特徵增強	Lycium barbarum		<i>Sophora japonica</i>
決策融合結果	<i>Sophora japonica</i>		

第五章、 結論與未來展望

5.1 結論

由於 CNN 模型擁有多層卷積層，固在自動擷取特徵通常十分耗費計算資源且後續較難以針對其分類進行解釋，固本研究提出透過增強影像特徵的方式並結合 MLP-Mixer 分類器，增加神經網路的可解釋性與提升準確度。

本研究針對形狀、紋理與顏色三個較具廣泛且直觀的特徵進行特徵增強，所採用的特徵增強方式分別為方向梯度直方圖、局部二值模式與單尺度視網膜增強算法，將增強後的影像輸入至 MLP-Mixer 分類器進行分類，分別輸出 Top-5 的類別後，再將三個特徵增強方式的 Top-5 作為決策融合的輸入，經由多類別羅吉斯回歸分類並輸出最終的決策結果。

本研究所提出的架構在魚類資料集、種子資料集和中歐森林資料集上皆有良好的表現。在 40 類 MIAT 魚類資料集上本系統能達到 99% 的辨識率優於未進行影像增強的 MLP-Mixer 分類器約 3%；在 560 類 MIAT 種子資料集上能達到 90.65% 的辨識率優於混合式神經網路約 20%；在中歐森林資料集 153 類上可達到 97.91% 的辨識率優於單個神經網路約 4.5%。

本研究設計一個可以廣泛應用於不同物種資料集的影像辨識系統，透過多種的特徵增強方式提取影像形狀、紋理、顏色的特徵，再經由 MLP-Mixer 分類器進行分類且各別輸出 Top-5，再將 Top-5 結果經由決策融合輸出最終決策結果，在不同資料集上皆有良好的辨識效能；由於融合多種且直觀的特徵增強方法之輸出，固能夠對分類器的最終結果進行解釋與分析。

5.2 未來展望

本研究目前僅廣泛應用於魚類、種子與中歐森林這種生物辨識上，希望未來能

嘗試更多不同的資料集，如非生物辨識或醫療影像、瑕疵影像檢測等資料集上，並且可針對不同的資料集選擇所需要的特徵增強模組，來提升其應用表現。

参考文献

- [1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
- [2] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541-551, 1989.
- [3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.
- [5] J. Amara, B. Bouaziz, and A. Algergawy, "A deep learning-based approach for banana leaf diseases classification," *Datenbanksysteme für Business, Technologie und Web (BTW 2017)-Workshopband*, 2017.
- [6] I. Goodfellow, Y. Bengio, and A. Courville, "Convolutional Networks" in *Deep learning*, MIT press, pp. 321-362. 2016.
- [7] M. Elhoushi, Z. Chen, F. Shafiq, Y. H. Tian, and J. Y. Li, "Deepshift: Towards multiplication-less neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2359-2368, 2021.
- [8] R. Maini and H. Aggarwal, "A comprehensive review of image enhancement techniques," *arXiv preprint arXiv:1003.4053*, 2010.
- [9] L. Hong, Y. Wan, and A. Jain, "Fingerprint image enhancement: algorithm and performance evaluation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, no. 8, pp. 777-789, 1998.
- [10] S. Ritter, D. G. Barrett, A. Santoro, and M. M. Botvinick, "Cognitive psychology for deep neural networks: A shape bias case study," in *International conference on machine learning*, pp. 2940-2949, 2017.
- [11] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel,

- "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness," arXiv preprint arXiv:1811.12231, 2018.
- [12] H. Li, X.-j. Wu, and T. S. Durrani, "Infrared and visible image fusion with ResNet and zero-phase component analysis," *Infrared Physics & Technology*, vol. 102, p. 103039, 2019.
 - [13] H. Li, X.-J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *2018 24th international conference on pattern recognition (ICPR)*, pp. 2705-2710, 2018.
 - [14] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, and S. Gelly, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.
 - [15] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, D. Keysers, J. Uszkoreit, and M. Lucic, "Mlp-mixer: An all-mlp architecture for vision," arXiv preprint arXiv:2105.01601, 2021.
 - [16] W. Samek, T. Wiegand, and K.-R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," arXiv preprint arXiv:1708.08296, 2017.
 - [17] R. C. Fong and A. Vedaldi, "Interpretable explanations of black boxes by meaningful perturbation," in *Proceedings of the IEEE international conference on computer vision*, pp. 3429-3437, 2017.
 - [18] D. Smilkov, N. Thorat, B. Kim, F. Viégas, and M. Wattenberg, "Smoothgrad: removing noise by adding noise," arXiv preprint arXiv:1706.03825, 2017.
 - [19] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *International conference on machine learning*, pp. 3319-3328, 2017.
 - [20] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PloS one*, vol. 10, no. 7, p. e0130140, 2015.
 - [21] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2921-2929, 2016.

- [22] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in European conference on computer vision, pp. 818-833, 2014.
- [23] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," arXiv preprint arXiv:1312.6034, 2013.
- [24] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in Proceedings of the IEEE international conference on computer vision, pp. 618-626, 2017.
- [25] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 1, pp. 886-893, 2005.
- [26] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in Proceedings of 12th international conference on pattern recognition, vol. 1, pp. 582-585, 1994.
- [27] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," IEEE transactions on image processing, vol. 6, no. 3, pp. 451-462, 1997.
- [28] C. W. Niblack, R. Barber, W. Equitz, M. D. Flickner, E. H. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin, "QBIC project: querying images by content, using color, texture, and shape," in Storage and retrieval for image and video databases, vol. 1908, pp. 173-187, 1993.
- [29] J. Canny, "A computational approach to edge detection," IEEE Transactions on pattern analysis and machine intelligence, no. 6, pp. 679-698, 1986.
- [30] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 7, pp. 971-987, 2002.
- [31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, pp. 436-444, 2015.
- [32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, vol. 25, 2012.

- [33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9, 2015.
- [34] C.-H. Chen, M.-Y. Lin, and X.-C. Guo, "High-level modeling and synthesis of smart sensor networks for Industrial Internet of Things," Computers & Electrical Engineering, vol. 61, pp. 48-66, 2017.
- [35] M. Mora, O. Adalakun, S. Galvan-Cruz, and F. Wang, "Impacts of IDEF0-Based Models on the Usefulness, Learning, and Value Metrics of Scrum and XP Project Management Guides," Engineering Management Journal, pp. 1-17, 2021.
- [36] R. Julius, T. Trenner, A. Fay, J. Neidig, and X. L. Hoang, "A meta-model based environment for GRAFCET specifications," in 2019 IEEE International Systems Conference (SysCon), pp. 1-7, 2019.
- [37] P. Novotný and T. Suk, "Leaf recognition of woody species in Central Europe," Biosystems Engineering, vol. 115, no. 4, pp. 444-452, 2013.
- [38] J. Gu, P. Yu, X. Lu, and W. Ding, "Leaf species recognition based on VGG16 networks and transfer learning," in 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), vol. 5, pp. 2189-2193, 2021.