JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 67, NO. 2, JUNE 2021
DOI - 10.29974/JTAE.202106_67(2).0005

應用卷積神經網路於用印文件辨識之研究

VERIFICATION OF APPROVED PDF USING CONVOLUTIONAL NEURAL NETWORK

亞洲大學
經營管理學系
副教授
蔡 存 孝*
Tswn-Syau Tsay

亞洲大學
資訊工程學系
學生助理
林 映 辰
Ying-Chen Lin

摘　要

全台各農田水利署管理處(改制前農田水利會)每年度需在「農田水利類公務統計報表與資料輯整合系統」上，填報有關灌溉管理、主計、財務、人事等相關統計資料，最後將經過呈報單位主管核准的「用印」報表掃描為 PDF 檔後上載到系統，但當填報人員誤載「未用印」報表時，系統並無法辨識及時退回，通常需要等到主管單位發現後再退回，造成延誤填報的問題，本研究的初步構想為，應用人工智慧的方法，讓機器可以自動辨識被上載的報表是否「用印」。近年來應用卷積神經網路(CNN)可自動萃取影像特徵的特性，在許多競賽中均證實，CNN 可大幅降低影像辨識的錯誤率，本研究的目的為應用 CNN 建立報表是否用印的辨識模型，因為辨識模型的架構經常決定模型的優劣，本文共測試 4 種 CNN 包含：6 層 CNN 架構、11 層 CNN 架構、VGGNet 及 Pre-trained VGGNet。模型的優劣評估以 Categorical Cross-Entropy Loss 函數作為評估標準。訓練的資料共蒐集 2,536 張「未用印」報表及 978 張「用印」報表，以機器自動標註後合併作為資料集，再處理為全部正置報表的資料集 A，及包含正置/反置/右旋 90 度/左旋 90 度報表的資料集 B。4 種卷積神經網路分別以前述 2 類資料集進行模型訓練，訓練的方法應用 Early Stopping，並假設如果連續 5 次 Loss 沒有降低，則停止訓練以防止過度擬合的問題。結果顯示 4 種 CNN 均可訓練出 99 %以上準確率的辨識模型，其中以 Pre-trained VGGNet 模型最佳，可以訓練 99.9 %準確率的辨識模型。

關鍵詞：卷積神經網路、用印文件辨識、人工智慧。

* 通訊作者，亞洲大學經營管理學系副教授
41354 台中市霧峰區柳豐路 500 號，tswnsyau@gmail.com

VERIFICATION OF APPROVED PDF USING CONVOLUTIONAL NEURAL NETWORK

Tswn-Syau Tsay*
Asia University
Department of Business Administration

Ying-Chen Lin
Asia University
Computer Science & Information Engineering

ABSTRACT

The Management Offices of Irrigation Agency in Taiwan has to report statistical data on irrigation management, accounting, finance, personnel, etc. to "Statistical Report System for Irrigation Associations, SRSIA" each year. The approved reports with "Stamps" have to be scanned as PDF files and upload to SRSIA in the end. However, SRSIA is not able to reject mis-uploaded file without "Stamps", which brought thoughts of applying artificial intelligence method to generate a model to identify whether the uploaded PDF file is with "Stamps" or not. In recent years, the application of convolutional neural network (CNN) can automatically extract the characteristics of image features. CNN has been proved in many competitions that it can greatly reduce the errors of image recognition. The purpose of this research is to use CNN to establish an identification model of whether the report is with "Stamps" or not. Because the architecture of the identification model often determines the pros and cons of the model, this paper tests a total of 4 CNN models including: 6-layer CNN architecture, 11-layer CNN architecture, VGGNet and Pre-trained VGGNet. Categorical Cross-Entropy Loss function is used as the evaluation method to evaluate the model. The training and testing data included a total of 2,536 PDF files without "Stamps" and 978 files with "Stamps", which were automatically labeled by the machine and merged together to be a dataset. PDF files in the dataset then modified to be all upright ones as Dataset-A, and to be including upright/reverse/right rotation 90 degrees/left rotation 90 degrees ones as Dataset-B. 4 CNN models were trained and tested on Dataset-A and Dataset-B. The training method applied Early Stopping with a patience of 5 to avoid overfitting. The results indicated that all 4 models can train identification models with an accuracy rate of more than 99 %. Among them, the Pre-trained VGGNet model is the best, the trained identification model can reach an accuracy of 99.9 %.

Keywords: CNN, Approved PDF, Artificial intelligence.

Tsay, T.S.* & Lin, Y.C. (2021). "Verification of Approved PDF Using Convolutional Neural Network." *Journal of Taiwan Agricultural Engineering,* 67(2), 44-55. https://doi.org/10.29974/JTAE.202106_67(2).0005

JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 67, NO. 2, JUNE 2021
DOI - 10.29974/JTAE.202106_67(2).0005

一、前言

「農田水利類公務統計報表與資料輯整合系統」(本系統)整合農田水利會聯合會「網路資料輯系統」及行政院農業委員會「農田水利類公務統計報表系統」，其功能為彙整年度灌溉管理、主計、財務、人事等相關統計資料，提供農田水利主管單位作為相關決策依據。2019 年後，各管理處資料填報人員在系統填報資料後，不再需要寄送簽核後用印報表至主管單位，改以電子檔案上載至本系統，亦即將用印後紙本報表掃描為 PDF 檔案，再將該 PDF 檔案上載到系統中備查，主管單位承辦人員即可直接從系統中下載用印後報表，核對資料庫輸出結果，以確保呈報資料的一致性。由於「農田水利類公務統計報表與資料輯整合系統」目前並無法辨識填報人員所上載的 PDF 檔案是否為「用印」報表或「未用印」報表，當發生填報人員誤載「未用印」報表時，而需重新辦理呈報作業，以致延誤報表呈報。本研究之目的為建立用印報表辨識模型，未來可以將此辨識模型部署在本系統中，當資料填報人員誤載「未用印」報表時，能由機器判讀及時退回，以降低報表誤載機率。為使機器達到自動判讀辨識「用印」報表或「未用印」報表，本研究以卷積神經網路作為建立辨識模型的工具。

讓機器學習達到與人類有相同的圖像判讀能力，一直是人類努力的目標之一，在傳統的機器學習影像辨識方法中，通常以物件辨識(Objection Recognition)的方法進行辨識模型的建立，首先必須蒐集辨識資料，再先擷取物件的特徵，最後進行模型訓練。著名的物件特徵擷取方法如星座模型(Constellation Model)[1]，認為物件由不同的特徵區塊及其幾何位置組合而成，例如腳踏車必須有 2 個輪胎、一個控制方向的龍頭及一個坐墊組成；另外，定向梯度直方圖法(Histograms of oriented gradients, HoG)則擷取影像輪廓作為辨識的特徵[2]；而 SIFT 以某一點為準，紀錄其周遭像素變化梯度及方向作為影像特徵[3]；DPM 模型則以機器學習，從星座模型、影像輪廓或 SIFT 選擇最佳的物件辨識的特徵[4]，從上述文獻回顧可以歸納，傳統的影像辨識模型需要以人工(handcrafted)的方法決定影像特徵，再以機器學習建立影像分類器，因此，人工所決定的影像特徵為物件辨識模型成敗關鍵，而近年來發展的深度學習(神經網路)則是讓機器自動萃取特徵，再建立影像分類器[5]，成為深度學習與傳統影像辨識(機器學習)最大的區別。

雖然深度學習可以自動萃取影響特徵，但以一張 224 × 224 像素的灰階圖片為例，如果要以神經網路(Neural Network, NN)建立辨識模型，需要將 224 × 224 的矩陣攤平後，再連結到數層的隱藏層，若每一個隱藏層包含 256 個神經元，則所需計算的權重參數將會暴增，即使應用平行計算與 GPU，仍須耗費許多計算時間才能完成權種參數的計算，因此，產生了專為影像辨識的卷積神經網路(Convolutional Neural Network, CNN)[6, 7]，CNN 的功能即在神經網路中自動擷取影像特徵，作為影像辨識的依據。本文列舉近年來數個具代表性的 CNN 模型為例作為 CNN 發展的說明[8]，包含 AlexNet, VGGNet, GoogleNet, ResNet 及 DenseNet；AlexNet[9]提出 8 層 CNN 結構，以兩張 GPU 卡訓練大型的模型以減少計算時間，並提出 ReLU 作為激勵函數(Activation Function)以獲得較快的收斂速度，AlexNet 同時提出 Data Augmentation 及 Dropout 技術解決模型過度擬合(overfitting)的問題，AlexNet 所提出的技術在後來應用 CNN 作為影像辨識都經常被使用。VGGNet[10] 由牛津大學 Visual Geometry Group 提出，該研究發現越多層數的 CNN 可獲得較低的錯誤率，因此提出了 16 層及 19 層 CNN 的架構，VGGNet 在 ILSVRC2012[11]2014 年的競賽中，相較於 AlexNet 的 top-5 error (= 16.4 %)，VGGNet 獲得了大幅的改善(top-5 error = 7.32 %)，但同年競賽冠軍為 GoogleNet[12] (top-5 error = 6.67 %)。GoogleNet 所提出的 22 層結構中，包含 Inception Module 及防止梯度消失的輔助分類器(Auxiliary Classifiers)，因此獲得最佳的成績。ResNet[13]更提出 152 層 CNN 結構，在 ILSVRC2015[14]競賽中，以 top-5 error = 3.57 %，超越 VGGNet 及 GoogleNet。而 DenseNet[15]不僅減緩梯度消失問題並強化各層間的特徵傳遞，因此較 ResNet 更有效益，且可以獲得較低的錯誤率。除了前述的經典 CNN 模型外，還有 ZF Net[16]、R-CNN[17]、Fast R-CNN[18]、Faster R-CNN[19]及其他有關影像辨識發展的文獻[20-23]。由前述文獻回顧可以了解，CNN 為影像中辨識的重要技術。

本研究在辨識模型訓練上，屬於二元分類問題，從系統中可以取得「用印」報表及「未用印」報表，所以屬於監督式學習，「用印」報表為填報人員上載的 PDF 檔案，報表可能為灰階或彩色(紅色用印)、向左旋轉 90 度或向右旋轉 90 度或旋轉 180 度(反置)等 6 種情形的組合。本研究先將 PDF 檔案轉換為圖像檔案，再以本研究建立的卷積神經網路模型及 VGGNet 建立辨識模型，結果發現，無論上載的 PDF 檔案為灰階或彩色及是否經過旋轉，以 Pre-trained VGGNet 建立之辨識模型，可達到 99.9 %的辨識正確率，本文其

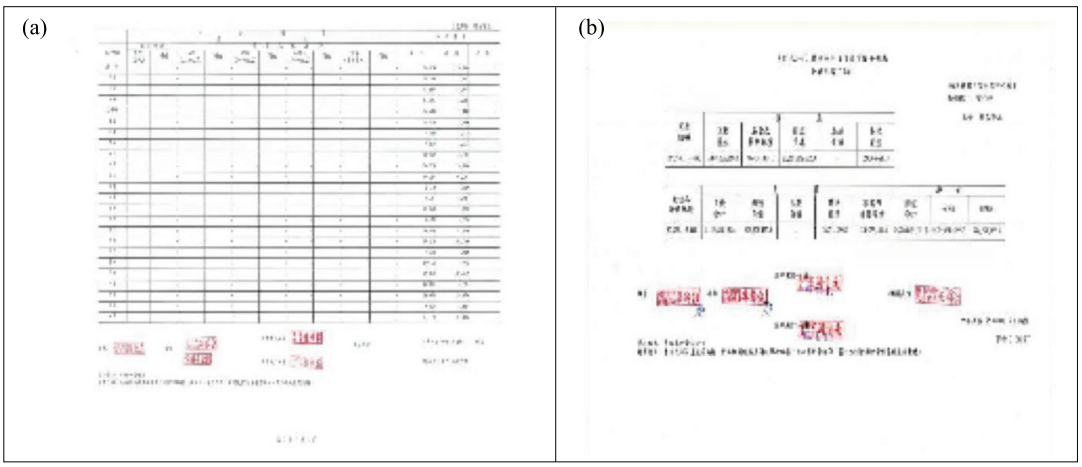

圖 1 (a)理想的上載報表-正置彩色影像範例 1；(b)理想的上載報表-正置彩色影像範例 2

他 3 個 CNN 模型也可以達到 99 %以上的辨識正確率。

二、研究材料

本文需辨識是否用印的報表共兩類，一類為農田水利會資料輯報表，另一類為農田水利類公務統計報表，該兩類報表均在「農田水利類公務統計報表與資料輯整合系統」進行填報作業。當填報人員產製 PDF 報表時，整合系統即同步備份相同的 PDF 報表儲存在系統中，填報人員呈閱報表用印後，即進行上載用印後 PDF 報表到系統中，系統將 PDF 報表轉換成二進位檔案儲存在資料庫，由於填報人員在填報過程中，可能因為修改資料經過多次列印，最後列印的報表才呈閱用印，因此，未用印的報表較用印的報表多，截至 2020 年 4 月，本研究進行訓練、驗證及測試的報表共 3,514 張(如表 1)，雖然兩類的報表格式略有不同，但在訓練辨識模型時將兩類報表合併後一起訓練，訓練後的模型可以同時辨識兩類的報表為「用印」報表或「未用印」報表。

進行辨識模型訓練時，為提高模型的辨識精確率，期望填報人員上載的報表為正置彩色影像報表如圖 1 (a-b)。而所有的「未用印」報表均為灰階影像，辨識模型進行前處理時，將報表轉換為 JPG 檔案格式後，

表 1 本研究驗證模型訓練資料數量表

是否用印數量 / 報表類別	未用印數量	用印數量	合計
農田水利會資料輯報表	1,635	550	2,185
農田水利類公務統計報表	901	428	1,329
合計	2,536	978	3,514

以 RGB 三個圖層作為模型訓練的訓練、驗證及測試資料，若填報人員上載的「用印」報表為彩色影像報表(用印為紅色)，則灰階為「未用印」報表，彩色影像報表則為「用印」報表。但實際上，填報人員可能上載的報表為「反置且/或灰階」報表，或「左旋/右旋」90 度的報表，如圖 2 (a-h)，表 1 中 978 張「用印」報表，共包含 212 左旋 90 度報表、155 張右旋 90 度報表及 5 張反置報表，填報人員上載未正置的用印報表約佔 38 %。

三、研究方法

卷積神經網路主要使用以下幾個重要的技巧：Convolution(卷 積) 及 Pooling(池 化) 因 應 Local Connectivity 及降低影像的解析度等特徵。由於影像具有 Local Connectivity 的特徵[7]，判斷一張報表是否用印的特徵，只要觀察報表的某部分而不必觀察整張報表即可判斷是否用印，且用印的特徵可能出現在報表任一個位置(不會只在最下方)，Weight Sharing[24]則假設卷積的濾波器(filter)的參數都一樣，以減少計算的參數量。另一個特徵是，降低影像的解析度不會影響到眼睛判斷報表是否用印，在神經網路的計算中，辨識低解析度的影像所需的權重參數也較高解析度的影像少。

3.1 卷積(Convolution)

卷積在影像處理中相當於 OpenCV[25]的 blur()函數，定義為以遮照(mask，在這裡稱為濾波器: filter)，在特定影像周圍進行低率波的動作，數學上稱之為卷

JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 67, NO. 2, JUNE 2021
DOI - 10.29974/JTAE.202106_67(2).0005

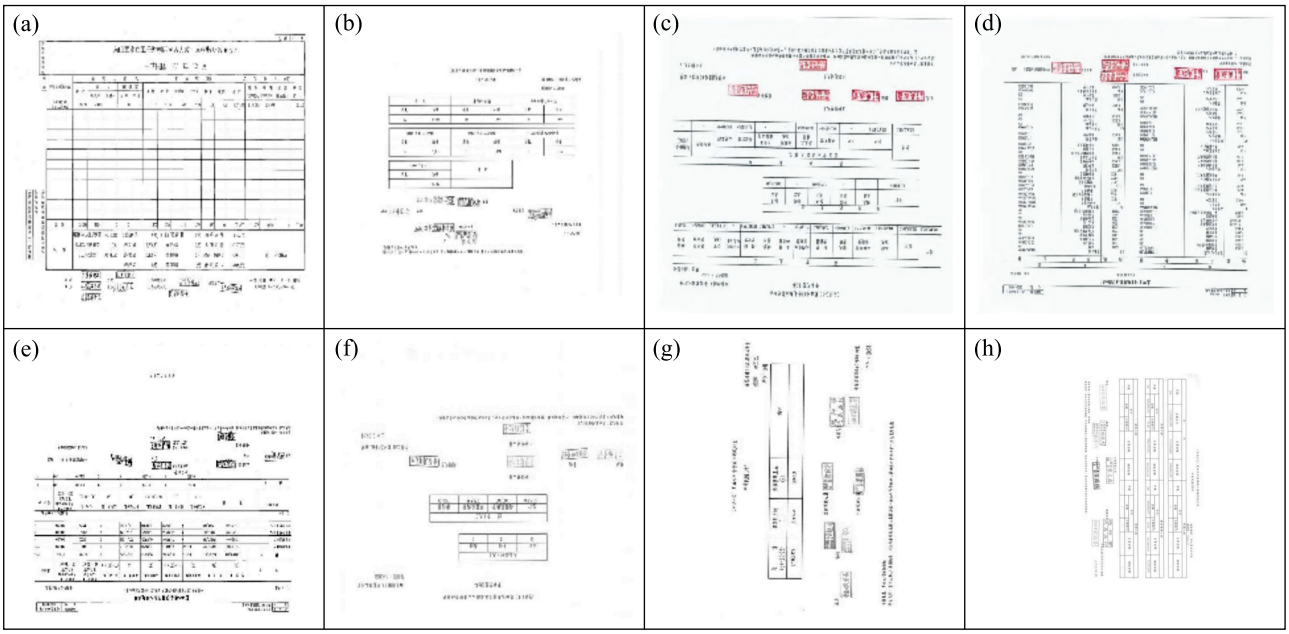

圖 2 (a)實際的上載報表-正置灰階影像範例 1；(b)實際的上載報表-正置灰階影像範例 2；(c)實際的上載報表-反置彩色影像範例 3；(d)實際的上載報表-反置彩色影像範例 4；(e)實際的上載報表-反置灰階影像範例 5；(f)實際的上載報表-反置灰階影像範例 6；(g)實際的上載報表-左旋 90 度灰階影像範例 7；(h)實際的上載報表-右旋 90 度灰階影像範例 8

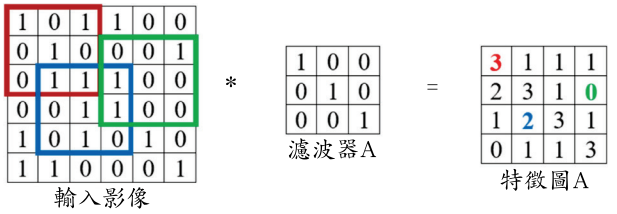

圖 3 6 × 6 輸入影像經過 3 × 3 濾波器 A 後，產生 4 × 4 特徵圖 A

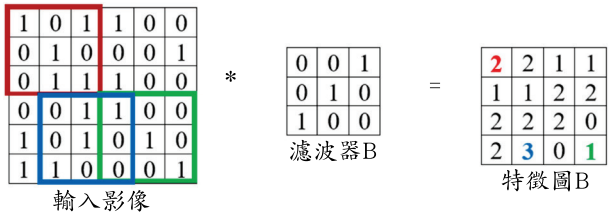

圖 4 6 × 6 輸入影像經過 3 × 3 濾波器 B 後，產生 4 × 4 特徵圖 B

積(convolution)[26, 27]，卷積的內容是計算濾波器與輸入影像的內積，並生成該濾波器的特徵圖。典型的卷積可以下列範例說明，假設某一個影像可以表示為 6 × 6 的矩陣，其中 0 代表白色，1 代表黑色，使用 3 × 3 的濾波器 A，經過卷積後產生特徵圖 A(圖 3)。若再經過另一個濾波器 B，經過卷積後則產生特徵圖 B(圖 4)。

從圖 3 及圖 4 的說明，卷積的計算可以表達為

$$Z_k[m,n]=\sum_{i=0}^{2}\sum_{j=0}^{2} f_k[i,j]I[m+i,n+j]$$

其中，Z_k 為特徵圖 k，f_k 為濾波器 k，I 為輸入影像，m 為 I 由上而下的指標= 0,1,2,..,5，n 為 I 由左至右的指標= 0,1,2,..,5，因此輸入影像的最左上方的點為 $I[0,0]=1$，同理，濾波器 $f_k[i,j]$的表示也是一樣。

圖 3 及 4 中濾波器的值是為了解釋卷積作用因此設定為固定的值，實際上 CNN 卷積層濾波器的值類似神經網路的權重係數，卷積的運算呼應了影像 Local Connectivity 的特徵，加上共用權重參數(Shared Weights)大大的減少了所需要計算的參數。增加卷積層中的濾波器數量可以萃取出更多的影像特徵，而濾波器的大小也可以調整，例如 AlexNet 的第 1 層卷積層使用了 96 個 11 × 11 的濾波器，第 2 層卷積層則設計為 256 個 5 × 5 的濾波器[9]。

3.2 池化(Pooling)

經過卷積後的結果，即進行池化。池化可以擷取影像的重要特徵，降低影像的解析度，藉以減少神經網路所需計算的權重參數。常用的池化分為最大池化(Max Pooling)與平均池化(Average Pooling)，以圖 3 所產生的特徵圖 A 為例，經過最大池化與平均池化的結

圖 5　特徵圖 A 分別經過 Max Pooling 及 Average Pooling 結果

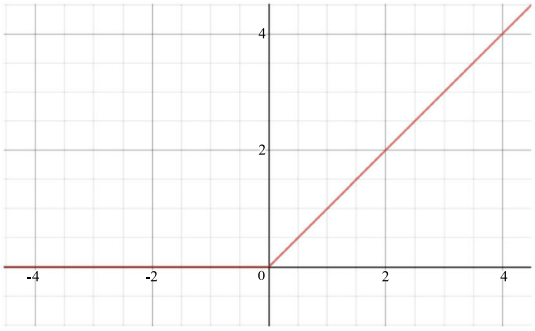

圖 6　ReLU 函數圖[29]

果如圖 5，本文的計算以最大池化作為降低影像解析度的方法。

3.3　ReLU 函數(Rectified Linear Unit Function)

在神經網路的非線性轉換函數中，常用的激勵函數為 sigmoid 函數或 hyperbolic tangent 函數，而 ReLU 函數對於影像辨識的 CNN 可以獲得較佳的收斂效果[9, 28]，當進行 ReLU 函數轉換時，若輸入值小於或等於 0，則輸出 0，若輸入值大於 0，則輸出值等於輸入值，如圖 6。

3.4　歸一化指數函數(softmax function)

softmax 函數[30]是最常被用於分類神經網路最後輸出的函數，softmax 函數可表示為：

$$\text{softmax}(x_i)=\frac{\exp(x_i)}{\sum_{i=1}^{n}\exp(x_i)}$$

x_i 為最後輸出的神經元的值，在 CNN 的模型訓練、驗證及測試中，以 softmax 函數結果作為判釋該輸入影像所屬的類別。

3.5　典型 CNN 結構

典型的 CNN 結構包含數個卷積層(convolution layer)及池化層(pooling layer)的組合，在卷積後則以 ReLU 函數進行非線性轉換，再將影像攤平後以全連接層輸出，最後再以歸一化指數函數(softmax function)進行分類，以本文為例，最後分類為「用印」或「非用印」2 類，如圖 7。

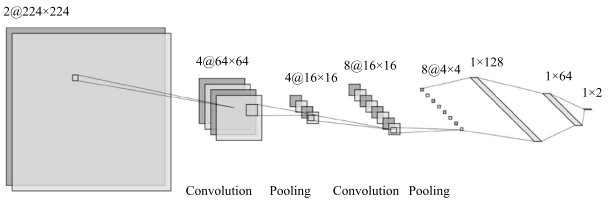

圖 7　典型 CNN 結構圖

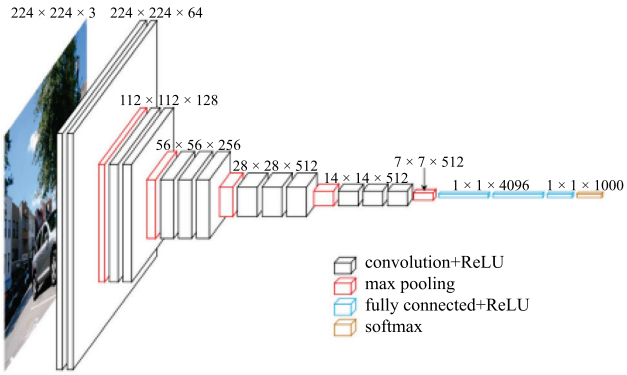

圖 8　VGGNet 網路結構圖[31]

3.6　VGGNet

VGGNet[10]的每一層卷積層均使用 3 × 3 的濾波器，雖然需要較多層但經過較 AlexNet 多次的非線性轉換，可以從影像中萃取較多的特徵，獲得較佳的預測模型。VGGNet 證實了較深的 CNN 網路結構可以獲得較佳的效能。本文使用類似圖 8 為 VGGNet 結構圖[31]，分別包含 13 層卷積層及 3 層卷積層，之後為 3 層全連接層(fully connected layer)，其中各卷積層及全連接層之後的激勵函數為 ReLU 函數，最後以 softmax 函數輸出。

3.7　模型優劣評估

CNN 模型的優劣評估以 Categorical Cross-Entropy Loss (簡稱為 Loss) 函數作為評估標準，定義為：

$$\text{Loss}=-\sum_{i=1}^{\text{類別數}} y_i\cdot log\hat{y}_i$$

其中，$\hat{y}_i$ 為 softmax 輸出的結果(預測值)，y_i 為實際類

JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 67, NO. 2, JUNE 2021
DOI - 10.29974/JTAE.202106_67(2).0005

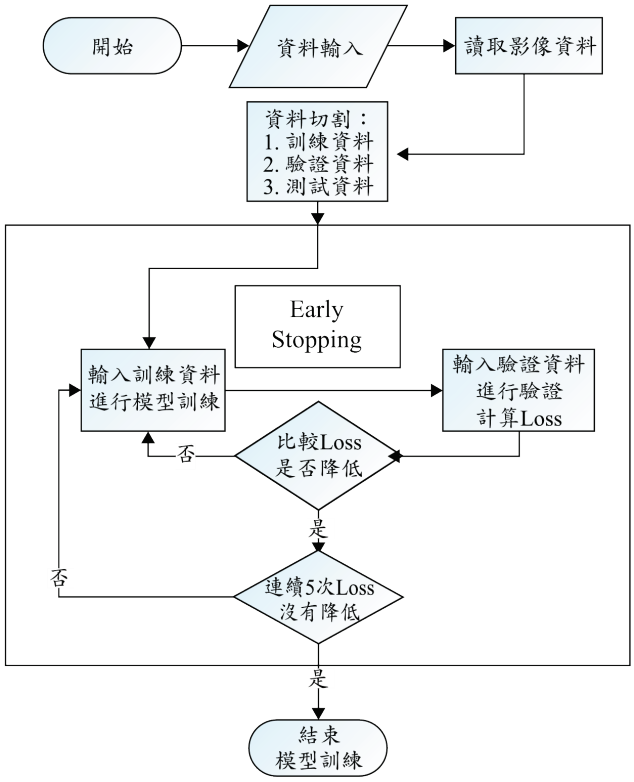

圖 9　Early Stopping 訓練流程

別標籤。以本文而言，因為只有「用印」及「非用印」2 類，若模型對於某一類報表而言，softmax 的結果為 $\hat{y}_1 = 0.8$，$\hat{y}_2 = 0.2$，而實際標籤為 $y_1 = 1$，$y_2 = 0$，則

$$\text{Loss} = -\left(1 \cdot \log 0.8 + 0 \cdot \log 0.2\right) \cong 0.097$$

當 $\hat{y}_1$ 越接近實際標籤 $y_1 = 1$ 時，Loss 就會越接近 0，表示該預測的結果為「用印」報表。

3.8　模練流程

本文應用 Early Stopping[32] 作為模型訓練方法，避免產生過度擬合的問題，如圖 9，訓練流程中，本文假設如果連續 5 次 Loss 沒有降低，則停止訓練。神經網路計算中，將所有的訓練資料看過 1 次，稱為 1 次 Epoch，當模型結束訓練後，繪製 Epoch vs 訓練及驗證 Loss 圖形，以了解模型訓練過程。

四、結果與討論

4.1　資料前處理

作為訓練的原始報表中，「未用印」報表可以從「農田水利類公務統計報表與資料輯整合系統」取得，

表 2　Dataset-A 與 Dataset-B 差異說明

內容 \ 資料集	Dataset-A	Dataset-B
包含「用印」報表	是	是
包含「未用印」報表	是	是
包含灰階報表	是	是
包含彩色報表	是	是
包含正置報表	是	是
包含反置報表	否	是
包含右旋 90 度報表	否	是
包含左旋 90 度報表	否	是

而且所有「未用印」報表皆為正置報表，所有「未用印」報表可由機器直接標示為「未用印」報表，共 2,536 張，而「用印」報表共 978 張(表 1)。由於填報人員上載的報表不一定為用印後的報表，且上載的報表包含「反置/右旋 90 度/左旋 90」報表或「彩色用印/灰階用印」報表，為使訓練後模型能辨識各種狀況下為「用印」報表或「未用印」報表，需以人工進行報表標示「反置/右旋 90 度/左旋 90」報表，再用機器將所有「反置/右旋 90 度/左旋 90」修改為正置報表，然後標示為「用印」報表，再合併「未用印」報表及「用印」報表作為 Dataset-A。為使辨識模型能夠辨識「反置/右旋 90 度/左旋 90」報表是否用印，將 Dataset-A 中所有報表進行隨機「反置/右旋 90 度/左旋 90」作為 Dataset-B。因此，Dataset-A 中的報表均為正置報表，Dataset-B 的報表包含正置、反置、右旋 90 度或左旋 90 度報表，如圖 2 (a-h)。Dataset-A 與 Dataset-B 差異說明如表 2。

4.2　模型結構

因為辨識模型的架構經常決定模型的優劣，本文共比較 3 個 CNN 結構，如圖 10 (a-c)，分別訓練 dataset-A 及 dataset-B 建立辨識模型，圖 10 (a)為 6 層 CNN 架構包含 3 層卷積層及 3 層全連接層，圖 10 (b)為 11 層 CNN 架構包含 8 層卷積層及 3 層全連接層，圖 10 (c)為 VGGNet 包含 13 層卷積層及 3 層全連接層，本文同時比較 VGGNet 模型及 Pre-trained VGGNet[33, 34]，以了解 Pre-train 參數在模訓練上的效益，因此共比較 4 個模型在辨識模型的比較：6 層 CNN 架構、11 層 CNN 架構、VGGNet 及 Pre-trained VGGNet。

4.3　模型訓練與測試

進行模型訓練時，以所有資料(3,514 筆)的 70%作為訓練及驗證，30 % (1,054 筆)作為測試，所有計算均

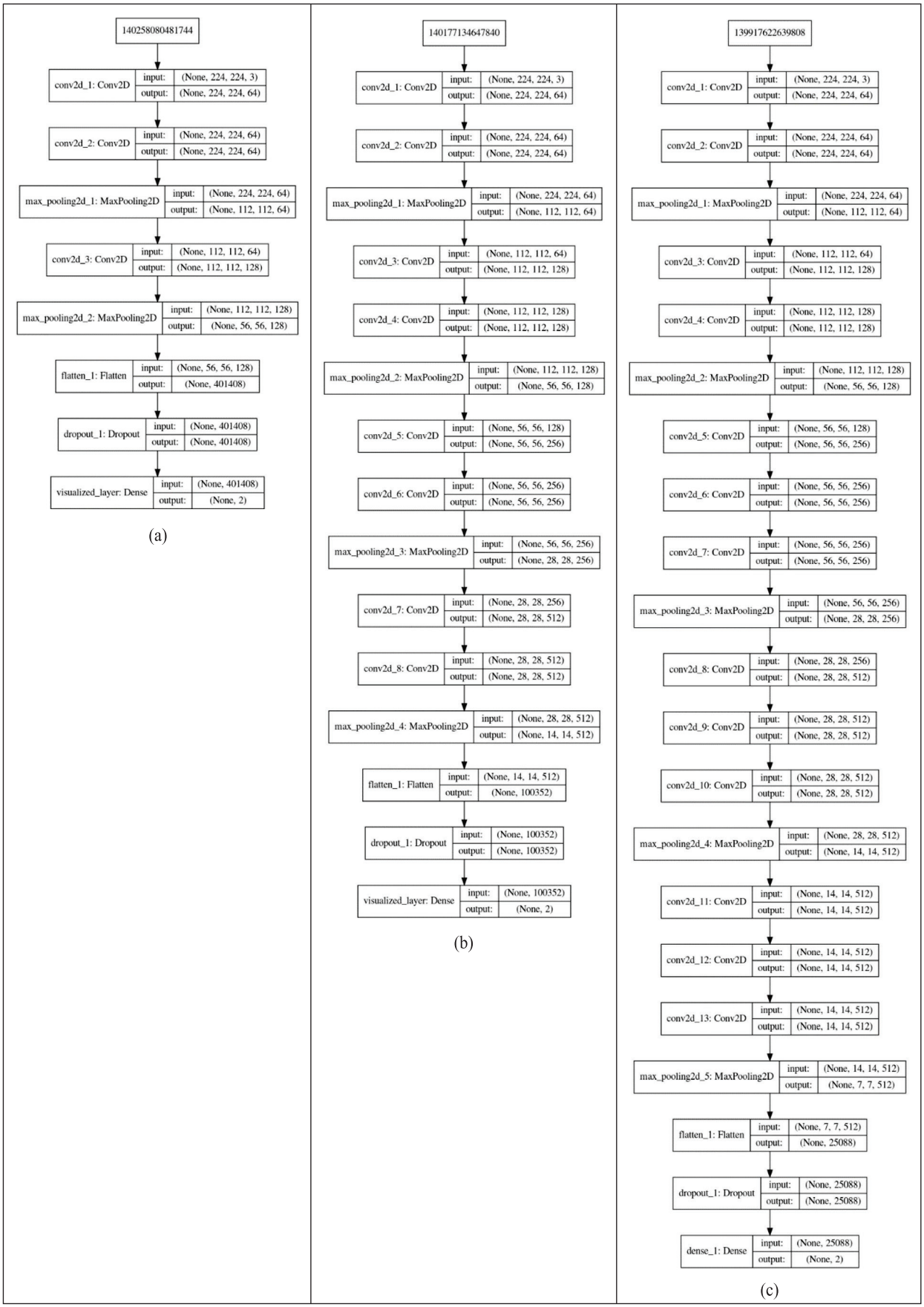

圖 10　(a) 6 層 CNN 架構；(b) 11 層 CNN 架構；(c) VGGNet 架構

JOURNAL OF TAIWAN AGRICULTURAL ENGINEERING
VOL. 67, NO. 2, JUNE 2021
DOI - 10.29974/JTAE.202106_67(2).0005

表 3　不同 CNN 架構以不同資料集訓練精計算權重數及計算時間比較表

CNN 架構 / 資料集	6 層 CNN 架構	11 層 CNN 架構	VGGNet	Pre-trained VGGNet
Dataset-A	915,394/254 秒	4,886,082/489 秒	14,764,866/557 秒	14,764,866/343 秒
Dataset-B	915,394/316 秒	4,886,082/494 秒	14,764,866/471 秒	14,764,866/442 秒

表 4　不同 CNN 架構以不同資料集訓練準確率比較表

CNN 架構 / 資料集	6 層 CNN 架構	11 層 CNN 架構	VGGNet	Pre-trained VGGNet
Dataset-A	99.9% ($\frac{1}{1055}$ 誤判)	99.9% ($\frac{0}{1055}$ 誤判)	99.4% ($\frac{6}{1055}$ 誤判)	100% ($\frac{0}{1055}$ 誤判)
Dataset-B	99.0% ($\frac{11}{1055}$ 誤判)	99.2% ($\frac{8}{1055}$ 誤判)	99.0% ($\frac{10}{1055}$ 誤判)	99.9% ($\frac{1}{1055}$ 誤判)

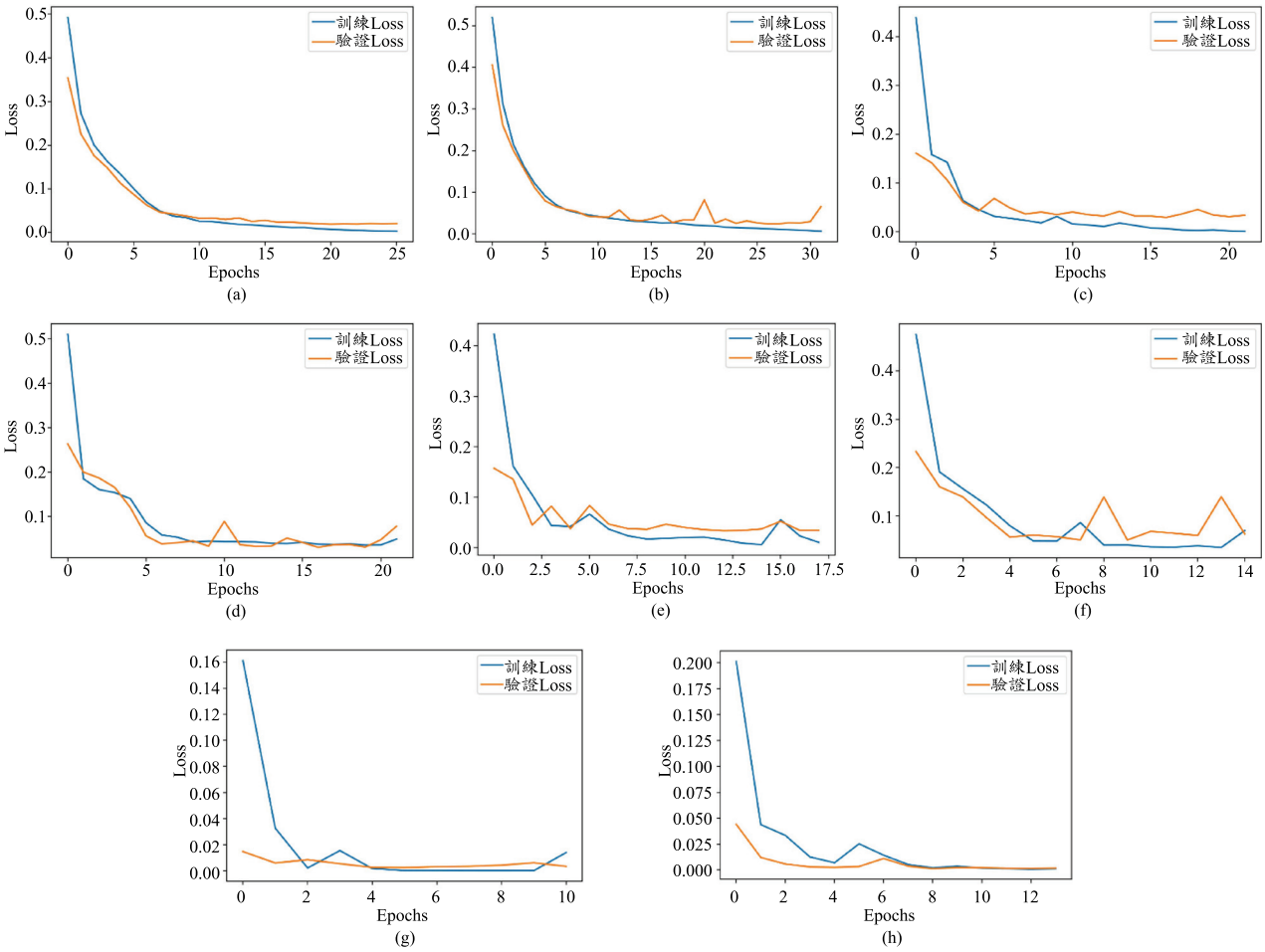

圖 11　(a) 6 層 CNN 架構(Dataset-A)學習曲線；(b) 6 層 CNN 架構(Dataset-B)學習曲線；(c) 11 層 CNN 架構(Dataset-A)學習曲線；(d) 11 層 CNN 架構(Dataset-B)學習曲線；(e) VGGNet 架構(Dataset-A)學習曲線；(f) VGGNet 架構(Dataset-B)學習曲線；(g) Pre-trained VGGNet 架構(Dataset-A)學習曲線；(h) Pre-trained VGGNet 架構(Dataset-B)學習曲線

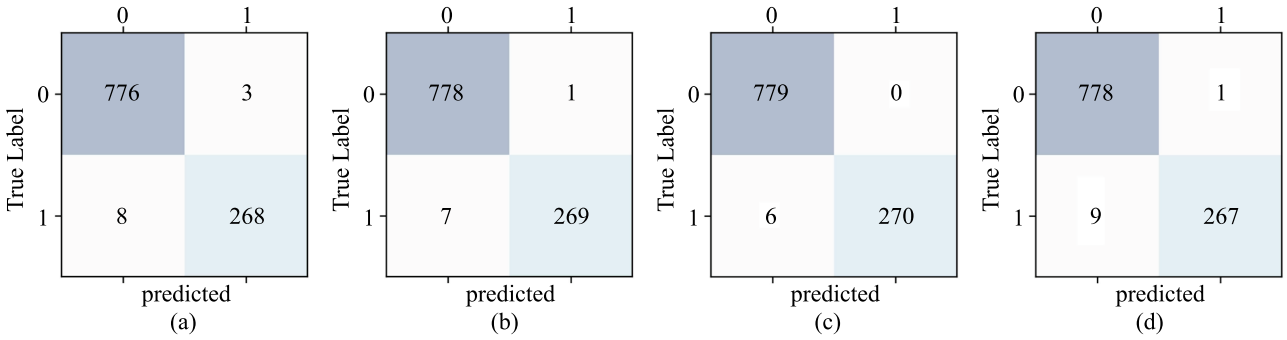

圖 12　(a) 6 層 CNN 架構(Dataset-B)混淆矩陣(0:未用印，1:用印)；(b) 11 層 CNN 架構(Dataset-B)混淆矩陣(0:未用印，1:用印)；(c) VGGNet(Dataset-A)混淆矩陣(0:未用印，1:用印)；(d) VGGNet (Dataset-B)混淆矩陣(0:未用印，1:用印)

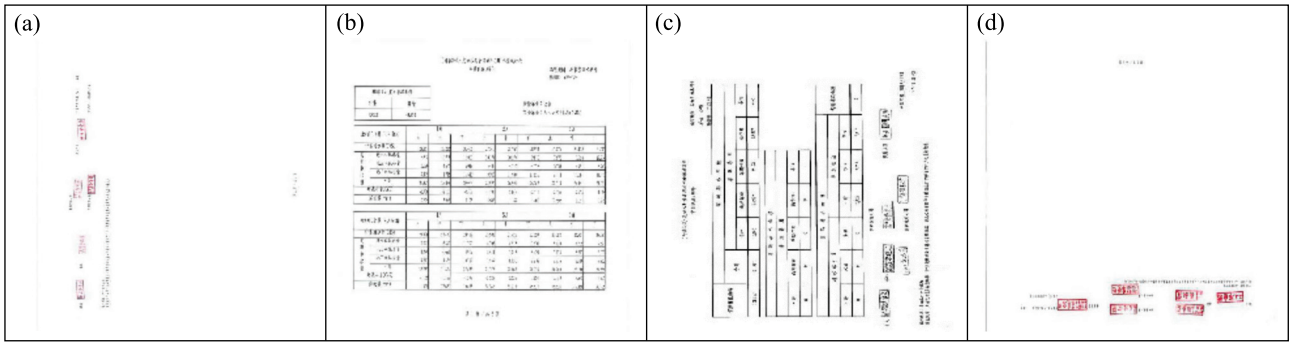

圖 13　(a) 6 層 CNN 架構(Dataset-B)誤判報表例 1(1 誤判為 0)；(b) 11 層 CNN 架構(Dataset-B)誤判報表例 1(0 誤判為 1)；(c) 6 層 CNN 架構(Dataset-B)誤判報表例 2(1 誤判為 0)；(d) 11 層 CNN 架構(Dataset-B)誤判報表例 2(1 誤判為 0)

以 GeForce RTX 2080 完成。各模型對於不同的資料集訓練所需計算的權重參數量及計算時間如表 3。從計算結果可以發現，層數越少的 CNN 結構，所需計算的權重參數量也越少，同一個 CNN 架構對於複雜的資料集(Dataset-B)所需花費的計算時間也較 Dataset-A 多，雖然 VGGNet 為 16 層的網路結構，所需計算的參數量高達 14,764,866，但因為 Pre-train VGGNet 使用預先訓練的權重參數作為起始值[33, 34]，因此計算時間並沒有因為需要計算參數量的增加而增加，比較特別的是，以 Dataset-B 作為訓練資料時，VGGNet 因可以較快獲得收斂效果，所需的計算時間較 11 層 CNN 架構少，各模型的學習曲線如圖 11 (a-h)，其中 Pre-trained VGGNet 所需的 Epoch 數量最少。以訓練後模型在測試資料集(1,054 筆)的表現而言，Pre-trained VGGNet 具有最佳的模型表現，在 Dataset-A 及 Dataset-B 的預測表現上分別可以達到 100 %及 99.9 %的準確率，如表 4。表現較差的模型為 6 層 CNN 架構及 11 層 CNN 架構，以 Dataset-B 作為訓練資料所產生的模型訓練測試結果，以混淆矩陣表示如圖 12 (a)及 12 (b)，其中 0 代表「未用印」報表、1 代表「用印」報表，部分誤判的報表如圖 13 (a-d)，從誤判的報表發現，不論是 0 誤判為 1 或 1 誤判為 0，均顯示模型沒有訓練到 100 %可以合理判斷「未用印」報表或「用印」報表，未來在增加訓練資料的蒐集後，期望可以訓練出可以 100 %辨識「未用印」報表或「用印」報表模型。

五、結論與建議

由於近幾年在人工網路計算方法的演進及應用 GPU 進行平行計算的普及，使卷積神經網路(CNN)在自動萃取特徵及建立影像分類器，大幅超越傳統影像辨識(傳統機器學習)的鑑別能力。本文共比較 4 種卷積網路模型於是否用印報表的辨識模型訓練，研究發現 Pre-trained VGGNet 對於包含「正置/反置/右旋 90 度/左旋 90」影像的資料集(Dataset-B)可達到 99.9 %辨識準確率，其他 3 個 CNN 結構所建立的模型也可以達到 99 %以上的辨識準確率。歸納結論及建議如下：

1. 從文獻回顧中可以歸納，自 AlexNet 及 VGGNet 開發後，CNN 已經成為影像辨識模型的主流，在

ILSVRC2012[11]2014 年的競賽中顯示其錯誤率較傳統機器學習低了許多，且在之後的競賽，也只有使用 CNN 的變形結構，才能獲得較前的名次。

2. 本文所訓練的 4 種卷積神經網路辨識模型皆可達到 99%以上辨識準確率。
3. 使用全部報表影像皆為正置的影像資料集(Dataset-A)較容易訓練高準確率的驗證模型，但包含「正置/反置/右旋 90 度/左旋 90」影像的資料集(Dataset-B)需要更強大的影像辨識模型才能提高影像辨識能力。
4. Pre-trained VGGNet 雖然包含 16 層網路大於 11 層 CNN 結構，但可以獲得較佳的計算收斂效果，因此計算時間也較 11 層 CNN 結構少，但因為 16 層網路結構可以擷取較多的影像特徵，因此準確率也較 11 層 CNN 結構高。
5. Pre-trained VGGNet 使用預先訓練的權重參數作為模型訓練的起始權重參數值，大幅減少訓練的 Epoch 次數，對於包含「正置/反置/右旋 90 度/左旋 90」影像的資料集(Dataset-B)其辨識準確率可以達到 99.9%。未來將隨著系統填報作業的進行，取得更多「用印」報表及「未用印」報表後再進行訓練，期望可以提高訓練精確率到 100%。
6. 依據目前所蒐集的訓練用 978 張「用印」報表均為「全部用印」報表，本研究現階段成果僅局限於「全部用印」報表及「全部未用印」報表的辨識，尚無法辨識「部分用印」報表模型，未來將虛擬建立「部分用印」報表，進一步開發「部分用印」報表辨識模型，以提醒使用者誤傳未全部用印報表到系統的問題。
7. 本研究現階段發現，只要是「用印」報表，其主管用印的位置都非常接近報表的規定用印位置，且僅少數報表用印後，主管會加註文字(如日期)，目前的「加註文字用印」報表數量仍太少，不足作為模型之用，未來可以虛擬建立「加註文字用印」報表及「用印位置異常」報表，以開發多元辨識模型，未來的辨識內容應包含：是否全部用印、用印位置是否正確及是否加註用印日期，以滿足主管機關對於用印完整性的要求。

參考文獻

1. Fergus, R., P. Perona, and A. Zisserman. *Object class recognition by unsupervised scale-invariant learning*. in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*. 2003. IEEE.
2. Dalal, N. and B. Triggs. *Histograms of oriented gradients for human detection*. in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. 2005. IEEE.
3. Lowe, D.G.J.I.j.o.c.v., *Distinctive image features from scale-invariant keypoints*. 2004. 60(2): p. 91-110.
4. Felzenszwalb, P.F., *et al*., *Object detection with discriminatively trained part-based models. 2009*. 32(9): p. 1627-1645.
5. Zhao, Z.-Q., *et al*., *Object detection with deep learning: A review.* 2019. 30(11): p. 3212-3232.
6. LeCun, Y., *et al*. *Handwritten digit recognition with a back-propagation network.* in *Advances in neural information processing systems*. 1990.
7. LeCun, Y., *et al*., *Gradient-based learning applied to document recognition*. 1998. 86(11): p. 2278-2324.
8. Gu, J., *et al*., *Recent advances in convolutional neural networks*. 2018. **77**: p. 354-377.
9. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. in *Advances in neural information processing systems*. 2012.
10. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*. ICLR, 2015.
11. Lab, S.V. *Large Scale Visual Recognition Challenge 2012*. 2012; Available from: http://image-net.org/challenges/LSVRC/2012/index.
12. Szegedy, C., *et al*. *Going deeper with convolutions.* in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
13. He, K., *et al*. *Deep residual learning for image recognition.* in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
14. Lab, U.C.H.V. *Large Scale Visual Recognition Challenge 2015*. 2015; Available from: http://image-net.org/challenges/LSVRC/2015/.
15. Huang, G., *et al*. *Densely connected convolutional networks.* in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
16. Zeiler, M.D. and R. Fergus. *Visualizing and understanding convolutional networks.* in *European conference on computer vision*. 2014. Springer.
17. Girshick, R., *et al*. *Rich feature hierarchies for accurate*

object detection and semantic segmentation. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.

18. Girshick, R. *Fast r-cnn*. in *Proceedings of the IEEE international conference on computer vision*. 2015.
19. Ren, S., *et al*. *Faster r-cnn: Towards real-time object detection with region proposal networks.* in *Advances in neural information processing systems*. 2015.
20. Frid-Adar, M., *et al*., *GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification*. 2018. 321: p. 321-331.
21. Zheng, Z., L. Zheng, and Y. Yang. *Unlabeled samples generated by gan improve the person re-identification baseline in vitro.* in *Proceedings of the IEEE International Conference on Computer Vision*. 2017.
22. Luo, Z., S. Cheng, and Q. Zheng. *GAN-Based Augmentation for Improving CNN Performance of Classification of Defective Photovoltaic Module Cells in Electroluminescence Images.* in *IOP Conference Series: Earth and Environmental Science*. 2019. IOP Publishing.
23. Karpathy, A. and L. Fei-Fei. *Deep visual-semantic alignments for generating image descriptions.* in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
24. Han, S., *et al*., *EIE: efficient inference engine on compressed deep neural network*. 2016. 44(3): p. 243-254.
25. OpenCV-team. *Open Source Computer Vision Library*. 2020; Available from: https://opencv.org/.
26. Culjak, I., *et al*. *A brief introduction to OpenCV*. in *2012 proceedings of the 35th international convention MIPRO*. 2012. IEEE.
27. Benedetti, A., A. Prati, and N. Scarabottolo. *Image convolution on FPGAs: the implementation of a multi-FPGA FIFO structure.* in *Proceedings. 24th EUROMICRO Conference (Cat. No. 98EX204)*. 1998. IEEE.
28. Nair, V. and G.E. Hinton. *Rectified linear units improve restricted boltzmann machines*. in *ICML*. 2010.
29. Agarap, A.F., *Deep learning using rectified linear units (relu)*. 2018.
30. Bishop, C.M., *Pattern recognition and machine learning*. 2006: springer.
31. Theme, s.Y. *VGGNet*. 2018; Available from: http://simtalk.cn/2016/09/25/VGGNet/.
32. Caruana, R., S. Lawrence, and C.L. Giles. *Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping.* in *Advances in neural information processing systems*. 2001.
33. Russakovsky, O., *et al*., *Imagenet large scale visual recognition challenge*. 2015. 115(3): p. 211-252.
34. Ashqar, B.A. and S.S.J.I.J.o.A.E.R. Abu-Naser, *Identifying Images of Invasive Hydrangea Using Pre-Trained Deep Convolutional Neural Networks*. 2019. 3(3): p. 28-36.

收稿日期：民國 109 年 07 月 28 日
修正日期：民國 109 年 12 月 20 日
接受日期：民國 110 年 04 月 01 日