

國立臺灣大學電機資訊學院資訊網路與多媒體研究所

碩士論文

Graduate Institute of Networking and Multimedia
College of Electrical Engineering & Computer Science

National Taiwan University

Master Thesis



利用遷移式學習與植物器官獨立模型進行植物影像辨識
Plant Image Recognition by Transfer Learning and Plant Organ
Separated Model

蕭至恆

Chih-Heng Hsiao

指導教授：張智星博士

Advisor : Jyh-Shing Roger Jang, Ph.D.

中華民國 107 年 6 月

June, 2018

國立臺灣大學碩士學位論文
口試委員會審定書

利用遷移式學習與植物器官獨立模型進行植物影像辨識
Plant Image Recognition by Transfer Learning and Organ
Separated Models

本論文係蕭至恆君（學號 R05944033）在國立臺灣大學資訊網路與多媒體研究所完成之碩士學位論文，於民國一百零七年六月廿七日承下列考試委員審查通過及口試及格，特此證明

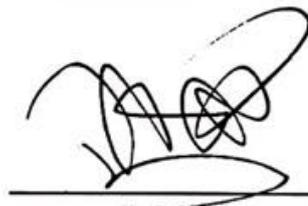
口試委員：

張智星

(簽名)

(指導教授)

王鈺恆



楊佳玲

所長：



致謝

在台灣大學的這兩年，使我轉變了許多。大學時期對於未來充滿了願景卻又迷茫，不懂得自身的程度在哪，因為害怕失敗而常常沒有自信，感謝指導教授張智星老師，採取非常開放的態度照顧每一位學生，如果一件事情雖盡了全力，卻仍然失敗，老師並不會因此責怪，而是會想要知道下一步可以怎麼走，想要知道我們有沒有想到其他的解決辦法，如果前方困難重重，「那就先做做看阿！」老師總是如此鼓勵，離開學校的未來中一定會出現難關，「那就先做做看吧」記得這句話並且勇敢向前。

感謝我的父母親對我在學業、生活上的支持與鼓勵，作我的堅強後備，使我能專注在學業上，厥功甚偉；另外要感謝實驗室的學長與同學們，感謝各位在這兩年對我的幫助，不論是在研究所的考試、計畫或最後的論文研究，有 MIR 實驗室的各位陪伴，這兩年我過的很順利也很開心，感謝工業技術研究院的丁川偉博士在這篇論文上提供了非常多的幫助，使辨識率能持續提升，也感謝范哲誠學長會一起開會討論研究方向。



摘要

本論文為植物影像的辨識與分類，利用深度神經網路 (deep neural network, DNN) 中的卷積神經網路去訓練出能對植物影像進行分類的模型，論文裡第一階段將會使用起源網路 (InceptionV3)、起源殘差網路 (Inception-Resnet) 與極端起源網路 (Xception) 作為基礎模型進行訓練，訓練過程將會使用資料增強 (data augmentation) 與遷移式學習 (transfer learning) 進行模型訓練，找出有最佳辨識率的模型進行下一階段器官獨立模型的訓練。第二階段為這篇論文提出的器官獨立模型的植物辨識方法，首先把資料集依照器官標籤分成多個子資料集，利用上一階段訓練辨識正確率最高的卷積神經網路模型，作為各自器官的分類模型，依照對應的子資料集訓練出專精於各自植物器官的子分類模型，並再訓練一個器官分類器，與多個子分類模型組合成器官獨立模型，嘗試使整體的辨識率能夠再次上升，並對分類錯誤的資料進行分析。

關鍵字：植物影像辨識、卷積神經網路、遷移式學習、特徵向量抽取、起源神經網路、分散式模型



Abstract

This paper focus on plant image recognition and classification. Briefly, we use extended deep neural network - convolution neural network(CNN) to train models for classifying plant images. For CNN model selection, InceptionV3, Inception-Resnet and Xception are very powerful condidates. These models can achieve state-of-the-art accuracy. First stage, we train these models via data augmentation and transfer learning. After experiment, model with the highest accuracy will be selected to the next stage. In the next stage, trying to reduce error rate, we propose a new method called "Organ Attribute Separated Model". First of all, we divide the original dataset to organ separated datasets by organ labels. After plant subdataset generated, we will train multiple CNN models for every subdatasets and an organ classifier. Combining all these models to complete organ separated model. Last but not least, we will do error analysis by extracting features from CNN models.

Key word: Plant image recognition, Convolution neural network, Transfer learning, Feature extraction, Inceptin network, Separated model



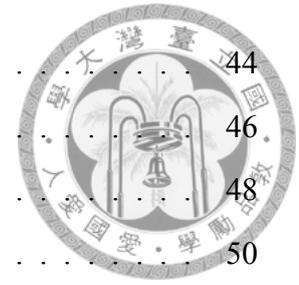
目錄

口試委員會審定書	i
致謝	ii
摘要	iii
Abstract	iv
目錄	v
圖表目錄	viii
表格目錄	xi
1 緒論	1
1.1 主題簡介	1
1.2 方法簡介	2
1.3 章節概述	2
2 研究方法	4
2.1 卷積神經網路	4
2.1.1 卷積層	4
2.1.2 池化層	5
2.1.3 激活函數	6
2.1.4 全連接層	7
2.2 基於 AlexNet 的葉子辨識	8



2.2.1	AlexNet 模型架構	9
2.3	起源神經網路	10
2.3.1	1x1 卷積核	11
2.3.2	輔助分類器	12
2.3.3	卷積核空間分解	12
2.3.4	標籤平滑化	13
2.4	起源殘差神經網路	14
2.4.1	區段正規化	14
2.4.2	起源殘差網路架構	15
2.5	極端起源神經網路	17
2.5.1	極端起源網路架構	17
2.5.2	分離式卷積	18
2.5.3	極端神經網路的其他嘗試	20
2.6	損失函數	21
2.6.1	均方誤差	21
2.6.2	交叉熵	23
3	實驗設置與結果	24
3.1	資料集	24
3.1.1	Top-500 資料集	25
3.1.2	Top-500 器官資料集	26
3.2	實驗環境	28
3.3	影像前處理	29
3.3.1	重縮放	29
3.3.2	資料增強	30
3.4	遷移式學習與模型選取	32
3.4.1	監督式影像分類的遷移式學習	33
3.4.2	卷積神經網路模型選取	35
3.4.3	遷移式學習特徵圖群分析	41
3.5	器官獨立模型	43

3.5.1	器官子模型訓練	44
3.5.2	器官分類器訓練	46
3.5.3	整合器官獨立模型	48
3.6	錯誤分析	50
4	結論與未來展望	53
4.1	結論	53
4.1.1	辨識率	53
4.1.2	實際應用	54
4.2	未來展望	54
	參考文獻	56





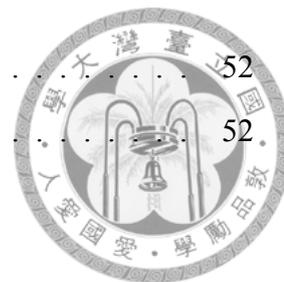
圖表目錄

2.1	卷積神經網路模型訓練架構圖	5
2.2	卷積層運算示意圖	5
2.3	對特徵圖做最大池化的例子	6
2.4	邏輯函式示意圖	7
2.5	線性修正單元示意圖	8
2.6	全連接層示意圖	8
2.7	Malaya 樹葉影像	9
2.8	AlexNet 模型架構示意圖	10
2.9	起源網路卷積層初版示意圖	11
2.10	起源網路卷積層改良示意圖	11
2.11	輔助分類器	12
2.12	普通卷積核與分解卷積核	13
2.13	殘差網路模塊	15
2.14	改良殘差網路模塊	16
2.15	起源殘差網路模塊	17
2.16	起源網路模塊簡化示意圖	18
2.17	極端起源網路模塊	19
2.18	極端起源卷積層的不同激活函數之訓練結果	20
2.19	極端起源網路模型架構	21
2.20	極端起源結合殘差路徑的訓練紀錄	22
3.1	工業技術研究院所提供的樹葉資料集	24
3.2	資料集分佈	25

3.3	較為特殊的測試資料	26
3.4	數量最多的前 500 類分佈	26
3.5	資料集器官標籤	27
3.6	包含各項器官影像之數量分佈圖	28
3.7	切割資料集示意圖	28
3.8	縮小輸入值與否對訓練網路比較圖	30
3.9	資料增強	31
3.10	有無資料增強對於測試資料準確度之影響	32
3.11	資料增強個別的影響	33
3.12	水平性遷移式學習	34
3.13	垂直性遷移式學習	35
3.14	起源神經網路訓練過程中的辨識率	37
3.15	起源神經網路訓練過程中的損失函數變化	37
3.16	起源殘差神經網路訓練過程中的辨識率	38
3.17	起源殘差神經網路訓練過程中的損失函數變化	38
3.18	極端起源神經網路訓練過程中的辨識率	39
3.19	極端起源神經網路訓練過程中的損失函數變化	39
3.20	抽取特徵圖群示意圖	42
3.21	無遷移式學習的特徵圖群	42
3.22	遷移式學習的特徵圖群	43
3.23	器官獨立模型的流程圖	44
3.24	各器官子模型不使用垂直性遷移式學習的 Top-1 正確率	45
3.25	各器官子模型垂直性遷移式學習的 Top-1 正確率	46
3.26	各式器官分類模型訓練示意圖	47
3.27	各式器官分類模型訓練過程中的辨識率	48
3.28	整合模型	49
3.29	抽取特徵向量示意圖	50
3.30	類別間相似所造成的分類錯誤	51
3.31	影像相似的分類錯誤	51



3.32 影像不相似的分類錯誤	52
3.33 影像重複所造成的分類錯誤	52





表格目錄

2.1 AlexNet 模型細節	10
3.1 ImageNet2012 資料集的辨識率	36
3.2 實驗參數設定	36
3.3 有無使用遷移式學習的各類模型辨識率	41
3.4 原模型對各器官的辨識率	45
3.5 無垂直性遷移式學習辨識率	45
3.6 器官子模型綜合比較	46
3.7 器官分類器辨識率	47
3.8 器官獨立模型辨識率	49



Chapter 1

緒論

1.1 主題簡介

影像的辨識與分類是計算機電腦科學研究中重要的一環。長久以來，如何使電腦有效解讀圖片中的資訊，自動化分類影像，進而抽取並提供人類所需的資訊是眾多資訊人才努力耕耘的目標。自從類神經網路提出後，巨量資料解析的準確度有了大幅度的提升，而卷積神經網路 (convolution neural network) 進一步特化了一般深層網路 (deep neural network) 後，更是讓圖片影像的辨識與分類之準確度一樣有了大幅度的提升，是電腦視覺中重要的突破，以此為基礎，隨著硬體資源的進步，電腦運算量上升，越來越多的類神經網路架構被提出，更多元多變的網路模型架構更是讓準確率再度上升，進步到在幾項電腦視覺的領域能超越人類的判斷，使之能夠在日常生活中所使用，如自動駕駛的周遭物體偵測、商品自動結帳等技術，進而改變人類的科技使用模式。

植物影像的自動分類有助於一般民眾對於常見的花卉盆栽理解與辨識，同時也可以應用於野生生態區中的植物辨識，推動相關的科普教育，使社會大眾能更簡單且方便的了解自身周圍的植栽，快速地得到相關的知識。在卷積神經網路應用在植物分類之前，植物影像需要先經過人為定義的特徵提取工程 (feature extraction)，如攤平並去除背景的葉子圖片，計算以葉子中心點為基準對 360 度延伸至葉邊緣的距離為特徵向量去做葉子影像分類，而此種特徵抽取法使用的限制繁多，只有特定經過攤平去除背景過後的圖片可以進行分類，無法應用於日常植物照片的分類，然而卷積神經網路可以針對訓練資料自動學習、偵測出特定特徵，

使得特徵提取工程對於一般影像的適用性大幅提升，受惠於此，植物影像辨識與分類開始實用於一般可得的照片，不再受限於特殊處理過的影像，更加貼近人類的需要。



本研究會分成兩個部分，第一部分為利用現今辨識率較高的卷積神經網路模型進行遷移式學習，訓練資料為臺灣工業技術研究院之臺灣植物辨識計畫所提供的資料集，內容為臺灣常見或特有植物的影像，利用這些圖片對卷積神經網路模型進行訓練並測試其模型準確度。第二部分為器官獨立模型，利用多個各自特化為器官加以訓練之卷積神經網路模型，集成一大型分散式神經網路模型，藉以嘗試提高單一卷積神經網路模型的準確率。

1.2 方法簡介

第一部分的遷移式學習模型利用起源網路模型第三版 (InceptionNet ver.3) 與其網路模型之改良版的起源網路模型第四版 (Inception Residual Net) 和極端起源網路模型進行資料訓練，從隨機生成網路中的權重 (weight) 加以訓練，去與以經過良好訓練的模型為基礎再加以訓練的遷移式學習模型進行實驗比較，不同的網路權重起源點對於最後網路模型的辨識率影響，經過一些參數調變之後，選出對於此一資料集辨識率最高的卷積神經模型，利用此模型去進行第二部分的分散式模型架構。

分散式模型架構需要訓練四個卷積神經模型，其中三個卷積神經模型為器官特化模型，其模型為利用植物影像較具辨識度的花、葉、果實各自訓練出專門分類所屬器官的卷積神經模型，模型架構為第一部分對於此一資料集所能得到最高辨識率的卷積神經模型，另外一個模型為植物器官分類模型，利用已標示為花、葉、果實的影像作為訓練資料，訓練出模型去辨識影像應分類為花、葉或是果實，進而當作該器官特化模型的輸入，該模型的準確度大幅影響分散式模型架構的總體準確度。

1.3 章節概述

本論文分成四個章節，各章節分成：

第一章 -緒論：介紹本論文主題的背景，以及方法的概述。

第二章 -相關研究介紹：說明卷積神經網路的相關演算法與其衍伸模型架構的特殊與歷程，其他使用到的演算法也會在這章加以概述。

第三章 -研究方法介紹: 說明這篇論文提出的方法之詳細介紹，如何預處理影像、模型架構與流程圖等，本次論文所提出的實驗結果將會記錄在這個章節，並分析模型準確率。

第四章 -結論與未來展望：分析論文所提出的方法如何應用於現實情形，有什麼優點與困難，並提出未來可能改進的方向。





Chapter 2

研究方法

2.1 卷積神經網路

卷積神經網路，簡稱為 CNN，是現今深度學習應用在影像辨識或分類上的重要演算法，許多相關的論文在探討卷積神經網路在進行深度學習時，每層之間的激活函數 (activation function)、區段正規化 (batch normalization) 等等的細節演算法對於網路架構的影響，其辨識率的增強或是運算效能的節省。與一般深度神經網路不同，卷積神經網路模仿人類視覺，若訓練得宜，模型會先注意到影像中較明顯的點、線、面，從點的運算變成影像局部的比對，針對卷積層抽取出來的特徵組合成特徵圖群 (feature maps)，透過一層層的特徵圖群去判斷，逐步堆疊比對得到最後的各項機率值，決定經由這個網路模型的輸入影像應該為哪一類別。架構上大致可分為卷積層 (convolution layer)、全連接層 (fully connected layer) 與池化層 (pooling layer)，如 Figure 2.1 所示。

2.1.1 卷積層

在電腦視覺領域中經常使用卷積 (convolution) 來進行影像處理，譬如去除雜點、影像銳化、以及邊緣偵測等。卷積層為輸入影像與卷積核 (convolution kernel) 進行卷積運算故而得名，卷積核為影像特徵抽取的核心，採用滑動視窗 (sliding window)，對影像做掃描式的卷積運算，影像中每一個像素點會被卷積核運算一次，得到一個該點對卷積核的運算結果，對應到特徵圖的相對位置，持續掃描直

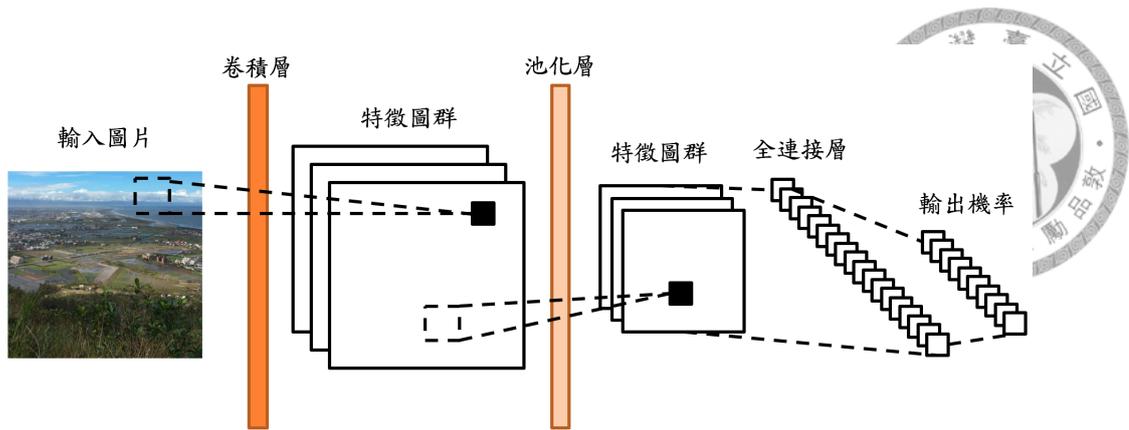


Figure 2.1: 卷積神經網路模型訓練架構圖

到輸入影像的像素點都被運算過後，輸出的結果成為下一層的卷積層或全連接層的輸入。在訓練的過程中，卷積核中的各項權重數值皆會隨著訓練資料變動，經由訓練學習出訓練資料的特徵，使卷積核的權重能夠偵測出正確的特徵，如 Figure 2.2 所示，卷積核學習出「X」的圖形為輸入矩陣的特徵，輸入矩陣出現特徵的點與卷積核運算後，會產生比其他點更高的數值，越高的數值，代表該點對應於此卷積核的特徵越明顯，而出現越多較高的數值，在深度網路的向下傳遞時，被留住的可能性也會越高，以利於較深層的網路去做更細微的空間關聯性判斷，得出正確的答案。

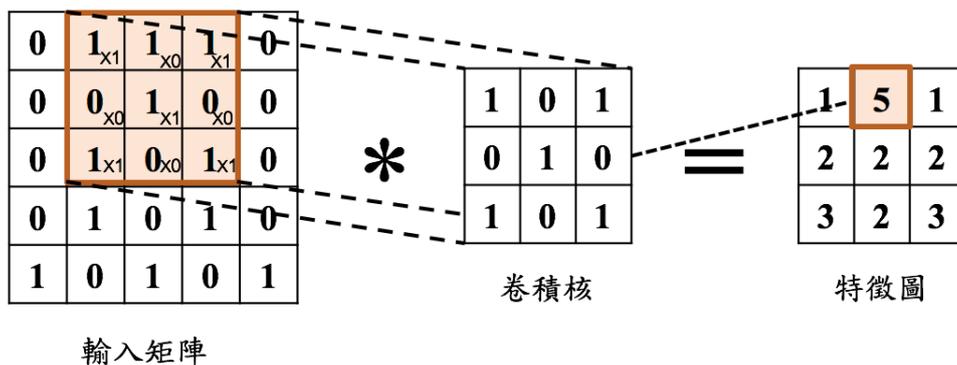


Figure 2.2: 卷積層運算示意圖

2.1.2 池化層

池化是卷積神經網路中重要的一環，池化是壓縮特徵圖並保存重要資訊的方式，與卷積核一樣，池化採用滑動視窗，當滑動視窗長寬皆為二，且滑動步數長 (stride) 也為二時，池化會使特徵圖縮小為原本的四分之一，最常使用的池化

是取滑動視窗中的最大值 (max pooling)，如 Figure 2.3，也有取滑動視窗的平均 (average pooling)，目的為降低運算量與雜訊的同時，能夠保留局部範圍的最大可能性，理解上為池化更專注於特徵有無出現在特徵圖或影像中，而不在意特徵出現的位置，只要有出現並被卷積核偵測到，就會有較大的數值，而池化依然會保留較大的數值，因此特徵在影像或特徵圖中的位置較不重要，偏移的特徵依然可以被偵測並傳遞到深層網路。



Figure 2.3: 對特徵圖做最大池化的例子

2.1.3 激活函數

激活函數為增加深度網路中的非線性因素，使得深度網路能夠更趨近非線性模型，對於深度神經網路理解並學習複雜且非線性的模型有著重要的作用，如果不使用激活函數，最終層的輸出訊號就會是一個線性的多項式，複雜程度有限，從數據中學習出複雜函數的趨近程度較差，只能視為一個線性迴歸 (linear regression) 模型，因此需要添加激活函數，增加模型的非線性因素，藉此增加模型的複雜程度，使任何的輸入值能夠映射到對應且接近希冀的非線性輸出。

邏輯函數

數值經過邏輯函數 (sigmoid) 如公式 (2.1) 使輸出值區間為 $[0, 1]$ 之間，為一 S 型曲線如 Figure 2.4，單一數值會有對應的輸出，通常應用在深度網路的最後一層輸出層，使被邏輯函數激活的每一點神經元擁有一個 $[0, 1]$ 之間的機率值，神經元互相之間並不影響。

$$P(t) = \frac{1}{1 + e^{-t}} \quad (2.1)$$

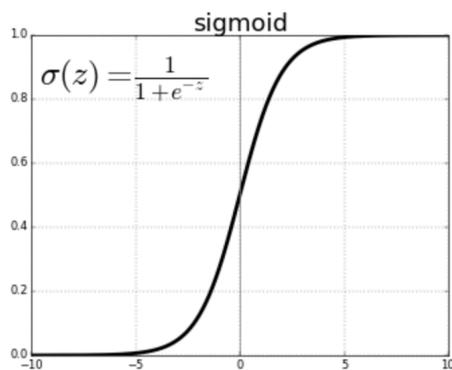


Figure 2.4: 邏輯函式示意圖

歸一化指數函數

與邏輯函數相似，為邏輯函數的推廣函數，通常也使用於深度網路的最後一層，不同於邏輯函數，歸一化指數函數 (softmax) 將含任意實數 K 維向量中的值經由公式 (2.2) 映射到另一個 K 維向量中，使 K 維向量中各值在 $[0, 1]$ 之間，且各項值加起來為 1，歸一化指數函數運算過後會使 K 維向量中最大值更加明顯，進而壓抑其他較小的分量。

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}, j = 1, \dots, K \quad (2.2)$$

線性修正單元

線性修正單元 (ReLU) 看似比起上述的兩種激活函數簡單，但對於深度網路的隱藏層 (hidden layer) 來說，線性修正單元實用很多，對於網路的收斂效果也較高，且運算效能需求低。線性修正單元為判斷數值正負，正的數值直接向下傳遞，而負的數值則直接歸 0，如公式 (2.3)。

$$R(z) = \max(0, z) \quad (2.3)$$

2.1.4 全連接層

在眾多模型中，全連接層通常會接在卷積神經網路較後端，接收網路前端的卷積層或全連接層所擷取出來的特徵值向量，把這些特徵值資訊向量每點乘上一加

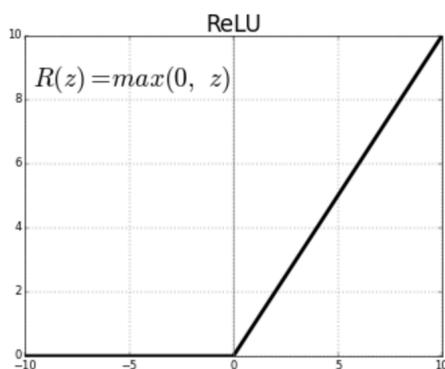


Figure 2.5: 線性修正單元示意圖

權值並相加，再選擇性加上偏差值 (bias) 後，通過激活函數並傳遞到下一層全連接層的一個神經元中，成為下一層特徵值向量的其中一個數值，持續運算直到最後的輸出層。中間的全連接層通過的激活函數通常為線性修正單元，以加速整體網路的收斂，與增加模型非線性因素，使模型複雜化，得到正確率更高的輸出結果。最後一層的激活函數視目標而定，若為二元分類，則多使用邏輯函數，並設定閾值 (threshold) 分類；若為多項分類，則多使用歸一化指數函數，使目標類別的輸出值能相較其他輸出值最大化。

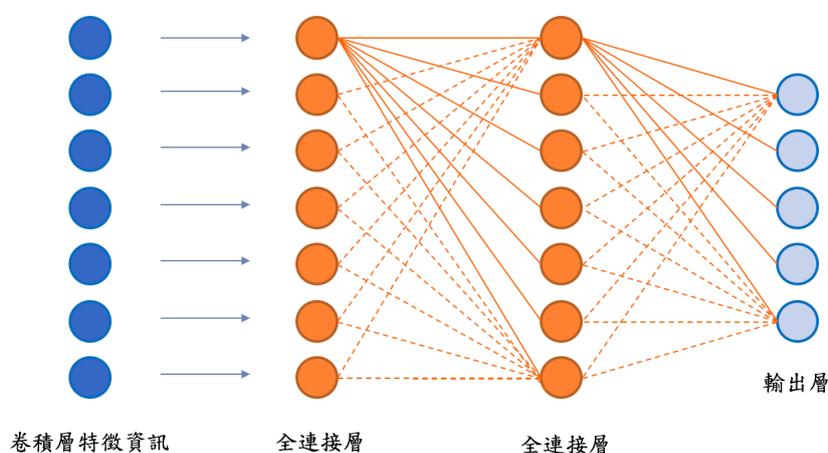


Figure 2.6: 全連接層示意圖

2.2 基於 AlexNet 的葉子辨識

植物影像辨識行之有年，在卷積神經網路應用於影像分類後，從原本需要人去定義的特徵抽取轉變為自動特徵抽取，中間不再需要人為的介入，而得到的效果

比之前的結果好上很多，在 2012 年的圖像辨識分類比賽 ILSVRC (ImageNet large scale visual recognition competition) [1] 中，一個對於當時是非常震撼的卷積神經模型-AlexNet [2] 出現在影像辨識的世界，辨識結果比起 2011 年的最低錯誤率 25% 以上，大幅的進步了 10% 左右，達到只剩將近 15% 的錯誤率，是跨幅度的大成長，同時也宣告了以卷積神經網路為基礎架構的相關模型，將會在未來的影像辨識這個研究領域大放異彩。

在這個模型提出後，許多領域也以此模型為基礎開始了在其他影像方面的應用，包含一些影像偵測、影像分類等等，而其中植物的影像辨識也引用了 AlexNet 的模型進行分類實驗 [3]，實驗中的資料集是由英國馬來亞大學 (Malaya) 所提供的，資料包含 44 種樹葉類別，訓練資料共 2288 張，而測試資料為 528 張，影像為攤平去背的樹葉圖片如 Figure 2.7。

使用馬來亞大學進行訓練及測試，在 AlexNet 的模型架構下，模型準確度達到 97.7%，效果十分卓越，以此為發想，本篇論文所使用的方法亦是使用深層卷積網路模型進行分類，往後的模型架構將會越來越複雜，之後本篇章節將會有詳細的描述。

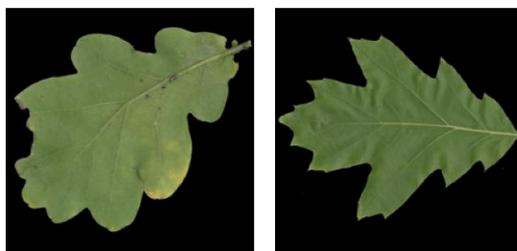


Figure 2.7: Malaya 樹葉影像

2.2.1 AlexNet 模型架構

AlexNet 的模型架構由卷積層、池化層與全連接層組合而成，細節如 Table 2.1 所述，Figure 2.8 為 AlexNet 視覺模型架構，雖然為基礎的卷積神經網架構，但在應用上使得卷積神經網路有一個基本的架構概念，由卷積層去找出影像或特徵圖群中的關聯性並加以學習，用最大池化層去透析出最重要的資訊，捨棄較為不重要的雜訊，最後由全連接層去整理卷積層與池化層輸出的特徵向量，以達到最佳的結果。



Table 2.1: AlexNet 模型細節

輸入層	影像大小為 [227, 227, 3]
卷積層	96 個長寬為 [11, 11] 的卷積核，滑動步數為 4
池化層	最大池化層，長寬為 [3,3]，滑動步數為 2
卷積層	256 個長寬為 [5, 5] 的卷積核，滑動步數為 1
池化層	最大池化層，長寬為 [3,3]，滑動步數為 2
卷積層	384 個長寬為 [3, 3] 的卷積核，滑動步數為 1
卷積層	384 個長寬為 [3, 3] 的卷積核，滑動步數為 1
卷積層	384 個長寬為 [3, 3] 的卷積核，滑動步數為 1
池化層	最大池化層，長寬為 [3,3]，滑動步數為 2
全連接層	由 4096 個神經元組成
全連接層	由 4096 個神經元組成
輸出層	輸出預測值

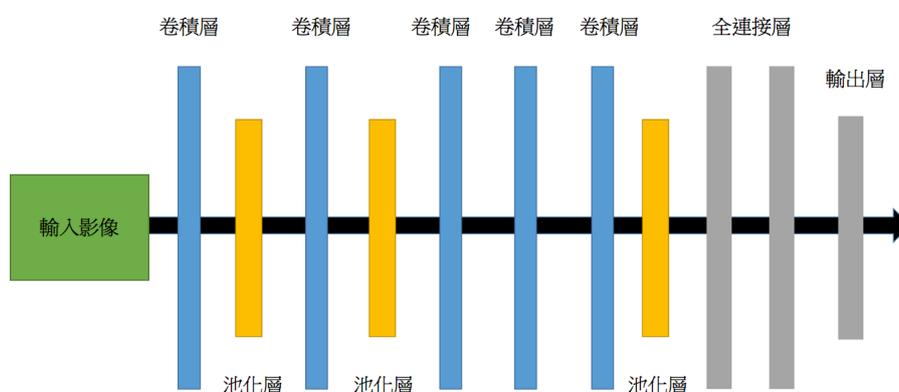


Figure 2.8: AlexNet 模型架構示意圖

2.3 起源神經網路

起源神經網路 [4] 首次出現於 2014 年 ILSVRC 的比賽，以 Top-5 錯誤率 6.67% 取得頭籌。不同於之前的卷積神經網路，起源神經網路每層卷積層不再是單一大小的卷積核，而是使用不同大小的卷積核對特徵圖群分別進行卷積，得出個別的特徵圖群後疊加在一起，增加網路對不同尺寸的適應性，並傳遞到下一層卷積層，如 Figure 2.9 所示，不同大小的卷積核對於特徵圖群的學習有不一樣的作用，較大的卷積核對於特徵圖空間上的學習較為全面，預期上能學習出範圍較大的特徵，使較遠的點與點之間的關聯性能被連結、學習；而較小的卷積核則能特化學習特徵圖群中的通道特徵，提高網路資訊的表達能力，使特徵圖群相對應的點之間的關聯性能被當成特徵特化學習，亦可以對輸入圖群做資訊的降維或升維，使原本的圖群能映射到更多或更少的特徵圖群。

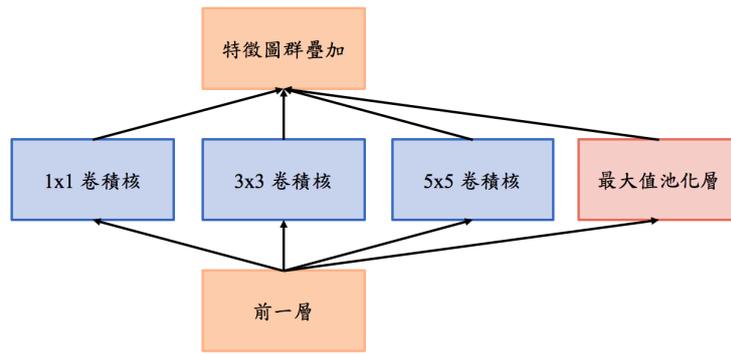


Figure 2.9: 起源網路卷積層初版示意圖

2.3.1 1x1 卷積核

受到赫布理論 (Hebbian theory) 的啟發，在架構深度類神經網路時，傾向使相關性較高的神經元相連在一起運算，如同一般的卷積神經網路的卷積核運算原理，在空間上相近的地方理論上相連程度較高，而後起源網路的卷積層在每個主要卷積核分支之前，新加入了 1×1 的極小卷積核， 1×1 的卷積核主要為學習跨通道資訊，對於特徵圖群中的相同位置的點，先作一次資訊組織，不同特徵圖的相同位置，彼此間的相關性也很高，經由 1×1 卷積層，先學習出特徵圖群通道的相連性，再通過主要的卷積核，去學習更全面的資訊，而一個分支等於經過兩次卷積的特徵變換，用計算代價很小的 1×1 卷積層就能有效地增加網路的非線性因素。架構如 Figure 2.10 所示，

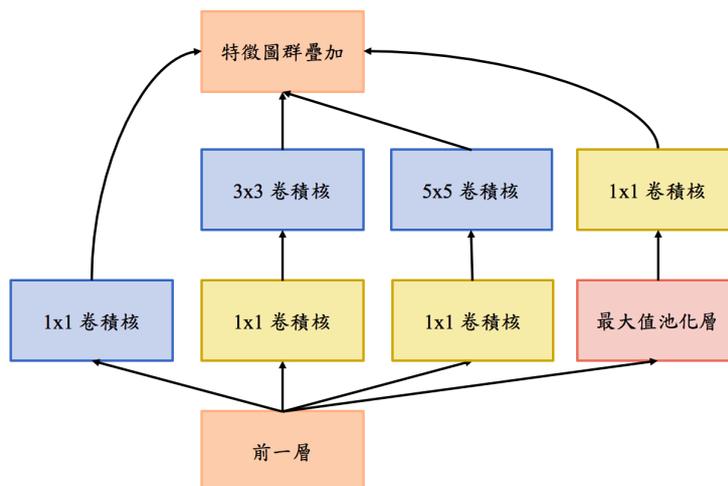


Figure 2.10: 起源網路卷積層改良示意圖



2.3.2 輔助分類器

輔助分類器 (Auxiliary classifier) 是起源網路重要的特色之一，一般深度網路的分類器接在網路的最後一層，也就是激活函數為歸一化指數函數、邏輯函數等的全連接層，而輔助分類器則是在深度網路的前、中段先分出一支分類器，如 Figure 2.11 中匡出的區塊，使整體網路有多個反向且不同深度的運算路線，在起源網路初版中提到輔助分類器有助於深度網路的收斂，原本的效用是希望能把有用的梯度資訊更有效且直接的推向更深層的網路，當網路深度化的時候，常常會出現較前面的網路層在訓練迭代次數高的情況下，權重數值無法再更新，而較後面的網路層則因為激活函數如線性修正單元等原因，深度網路的梯度數值非常的微小，不容易訓練且出現梯度消失的問題，輔助分類器有助於解決此項深度網路問題，利用較短的反向運算路線，調整網路前段的運算，使之在迭代次數高的情況下，依然能調整前段網路，進而使資訊能更有效的傳遞到深層的網路，提升整體網路的準確率。

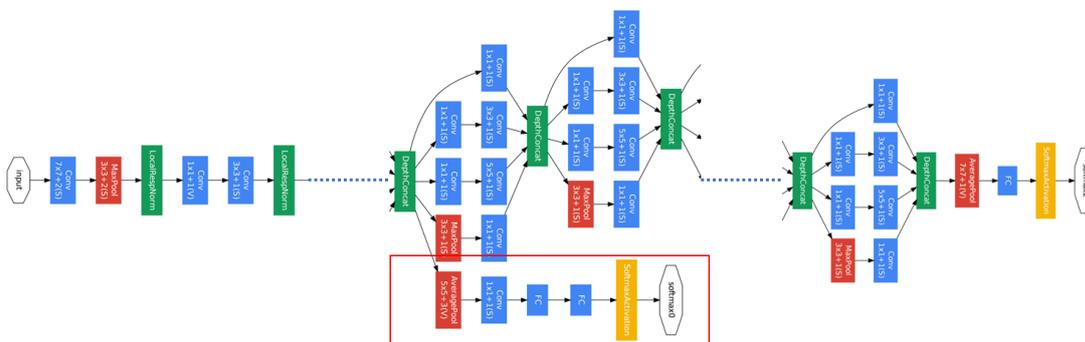


Figure 2.11: 輔助分類器¹

2.3.3 卷積核空間分解

起源網路把卷積核作空間上的分解，把 $[n, n]$ 的卷積核分解為 $[n, 1]$ 的卷積核後接著 $[1, n]$ ，卷積核的分解有助於降低整體網路運算量的同時，不至於失去太多空間關聯性上的學習，以 $[3, 1]$ 接 $[1, 3]$ 的卷積核為例，如 Figure 2.12 右圖所示，第一層會先學習橫列的空間相關性，並輸出 3 個數值供下一層學習直列的關聯性，兩層過後的維度與原本 $[3, 3]$ 的卷積核一樣，但是運算量降低了 33%，但與一般卷

¹部分圖片來自於：<https://arxiv.org/pdf/1409.4842.pdf>

積核不一樣，經由實驗，卷積核空間分解並不適用於較淺層的網路，而是適用在中後段的深層網路，當卷積核的目標特徵圖群長寬為 12 至 20 時，有較佳的表現。

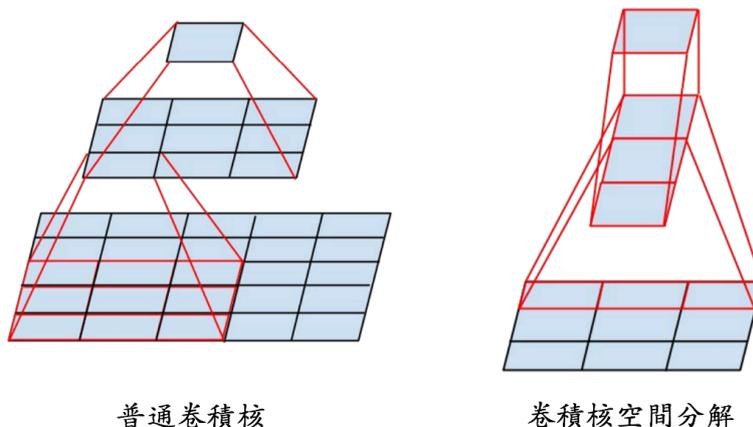


Figure 2.12: 普通卷積核與分解卷積核²

2.3.4 標籤平滑化

不同於其他的改進方式，標籤平滑化 (Label smoothing) 不著手於整體網路架構，而是在訓練資料的標籤上進行修改，作者認為目標標籤如果為 1，那麼將會造成兩種使網路降低辨識能力的可能性。第一種可能為過度訓練 (over-fitting)，因為所有訓練資料的標籤都把全部的機率值分配到一個目標項目，可能會造成網路過度相信訓練資料，在訓練的過程中可能發現不出太大的異狀，但是可能會造成在測試網路辨識率的時候，出現過度依賴訓練資料，而無法辨識出沒有出現在網路過的其他測試資料，造成過度訓練，反而降低整體辨識率。第二種可能，作者認為訓練資料的標籤為全機率的 1，會擴大訓練資料的預測目標值與其他值的差距，在有限的梯度訓練下，反而下降了模型的適應性，因此作者使用了以下的公式 (2.4) 去平滑化訓練資料的標籤，使標籤不再是 1。

$$q'(k) = (1 - \varepsilon)\delta_{k,y} + \frac{\varepsilon}{K} \quad (2.4)$$

其中 k 為第 k 項標籤， $\delta_{k,y}$ 為標籤值 1， K 為總共的標籤類別數目， ε 為自訂的參數，預設為 0.1。

²圖片來源：<https://arxiv.org/pdf/1512.00567.pdf>



2.4 起源殘差神經網路

起源殘差神經網路為 [5] 起源神經網路的再次改版，參考並引用了微軟所發表的殘差網路 (Residual net) [6]，微軟所發表的這篇殘差網路是為了使網路加深所提出，深度神經網路基於硬體層面的進步而逐漸深度化，網路越深，整體網路的複雜度將會逐漸提高，使網路複雜化以提高網路的可適性與辨識率，但深層網路在網路後端反而出現了梯度消失 (gradient vanishing)、梯度膨脹 (gradient exploding) 等問題。

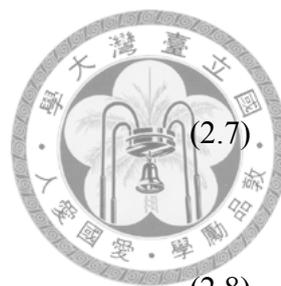
梯度消失為當網路過深時，網路深層輸出的結果是由許多小於 1 的數值運算而來，其中多為相乘，造成的結果為後段深層的網路所持有的數值極小或接近於 0，使前向傳導完過後的反向傳導所進行的偏微分無法更新網路數值，進而癱瘓了後段深層網路的學習，後段網路幾乎停滯無法繼續更新，整體網路也就無法收斂了；而梯度爆炸則是相反，當初始權重數值過大時，前段的網路因為因應學習改變較為快速，而數值隨著網路加深而指數性的膨脹，整體網路也如同梯度消失，無法收斂。

2.4.1 區段正規化

為了解決這個問題，許多學者提出了相關的解決方案，其中聲名遠播且效能卓越的論文為區段正規化 (batch normalization) [7]，是起源神經網路的第二次改版，利用以下公式使輸出訊號的平均值為 0，平方差為 1，藉此使網路保持穩定，在正向與反向的訓練傳播時，網路中的權重值對梯度消失與梯度膨脹影響甚巨，區段正規化對每次的輸出權重經過比例放大縮小、平移，把輸出值的平均與平方差固定在 0 與 1，使每次的訓練偏差值大量的降低，從而穩定網路的輸出。

$$\mu_{\beta} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad (2.5)$$

$$\sigma_{\beta}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\beta})^2 \quad (2.6)$$



$$\widehat{x}_i \leftarrow \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}} \quad (2.7)$$

$$y_i \leftarrow \gamma \widehat{x}_i + \beta \equiv BN_{\gamma, \beta}(x_i) \quad (2.8)$$

其中 μ_β 為每區段輸出值的平均值， σ_β^2 為每區段輸出值的平均差， m 為每段輸出之總數， γ 與 β 則為訓練的參數，隨著訓練找出最適合的值使每段區段能輸出在理想的範圍內。

受助於這些方法，網路的學習可以越來越深度化，網路越深，能過學習並表達出的特徵應該會一起上升，但實驗結果反而是準確率下降了，為了解決這個問題，殘差式的網路被提出，利用跳躍式的方式，傳遞層之間的資訊如 Figure 2.13，梯度值可以快速回遞到前幾層的網路，使幾個反向傳遞的路線比原先短，可以視為淺層的網路，保有淺層網路時所學到的，但也有反向傳遞極深層的路線，可以去學習訓練資料的精細特徵，比起淺層網路的可適應性又較高。

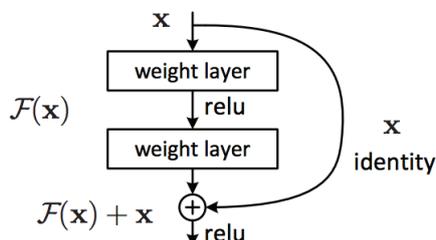


Figure 2.13: 殘差網路模塊³

$$F(x) \Rightarrow F(x) + x \quad (2.9)$$

一般網路層的傳遞公式為 $F(x)$ ，則殘差網路的跳躍連接層公式則為 $F(x) + x$ ，其中後面的 x 為前面層向前捷徑式傳遞的值。

2.4.2 起源殘差網路架構

受到殘差網路的啟發，結合了起源網路與殘差網路的特性，嘗試融合兩者間的優點去突破各自的缺陷，於是有了起源殘差網路。在殘差路徑 (Residual path) 的

³<https://arxiv.org/pdf/1512.03385.pdf>

幫助下，比起原本的起源網路，起源殘差網路嘗試了使用更多的起源模塊，以組成更深的起源網路，在模塊上，起源殘差神經網路也做了大量的細部修改，原本的殘差路徑為 Figure 2.13所示，向前傳遞了兩個一般的卷積層，而起源殘差網路結合了原本的起源模塊，這邊也使用了 1×1 的卷積層，先經過 1×1 的卷積層，去學習通道間的關係如前述，再通過一般的卷積核去連結空間之間的關係，如 Figure 2.14所示。

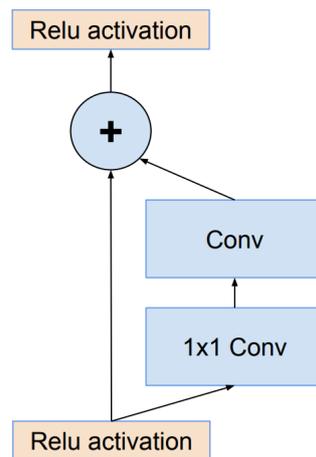


Figure 2.14: 改良殘差網路模塊⁴

之後結合原本起源網路的特色，由多個卷積層路徑去結合出一個起源模組，在加上一條直接從輸入傳遞到輸出的殘差路徑，作為縮短部分網路的反向傳遞的路徑，使網路可以更加深層化，如 Figure 2.15所示，起源殘差網路的模塊有許多種組合，這邊僅列出其中一種，起源殘差網路由多種類別的模塊組合而成，各有不同的卷積層路徑與自己的架構，反覆堆疊組合成起源殘差網路，作者亦嘗試了多種耗費不同運算資源的組合，去比較與起源神經網路相似的運算量下，所得到的準確率，發現有殘差路徑的組合，在網路訓練前期的收斂速度較快，特殊的組合下，準確率也比起源神經網路略高。詳細的起源殘差網路模型可以參照 [5] 的論文。

⁴圖片來源：<https://arxiv.org/pdf/1602.07261.pdf>

⁵圖片來源：<https://arxiv.org/pdf/1602.07261.pdf>

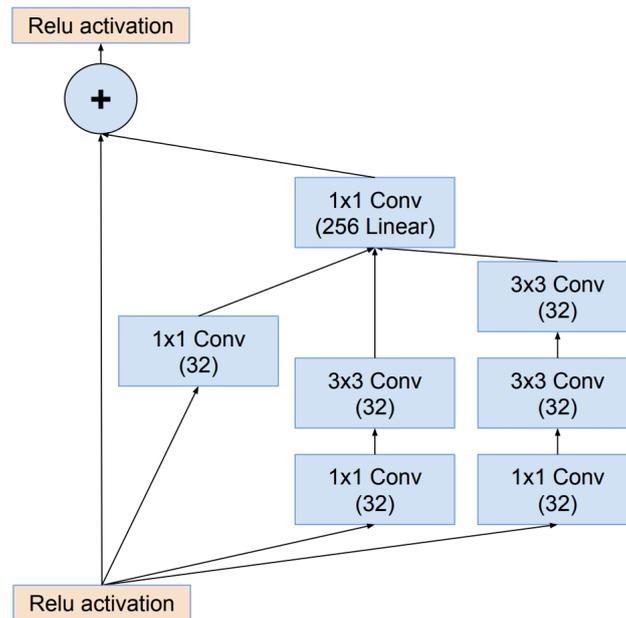


Figure 2.15: 起源殘差網路模塊⁵

2.5 極端起源神經網路

極端起源神經網路 (Xception net) [8] 從起源網路的概想延伸而來，起源神經網路的一大特點為 1×1 的卷積核，用來學習跨通道間的關聯性，一般的卷積核學習的為三維空間的特徵關聯性，其中二維為空間向量，為特徵圖群的長與寬，而剩下的一維則為通道向量，雖然起源網路有使用 1×1 的卷積核特別特化學習通道向量，但接在 1×1 卷積核後的卷積層依然為一般的卷積核，學習著三維向量空間的關聯性，通道向量的關聯性在兩次卷積核都被當作重點學習，而極端起源網路的「極端」就是特別把通道向量與空間向量分開學習，讓兩者各自有自己的卷積核去連結彼此間的關聯性，使空間向量與通道向量完全的被分開來鏈結學習，基於赫布理論，使更相近的相同向量空間的權重值能束在一起學習，藉以提高相近權重向量間的學習能力，進而提升整體網路的辨識率。

2.5.1 極端起源網路架構

在模型的架構方面，與起源網路相似，極端起源網路也是使用模組化堆疊而成。極端起源網路先是簡化了起源網路的模塊，如 Figure 2.16 的左圖，對比先前起源神經網路的模組圖 Figure 2.10，極端起源網路先是把所有卷積核層改為 3×3

與 1×1 ，使架構上看起來更為清楚簡單，與起源網路的不同大小卷積核不同，極端起源網路並不著重在同一向量空間的不同大小卷積核所帶來的效益，而著重在分開空間向量空間與通道向量空間後，分開運算所帶來的效益與影響。

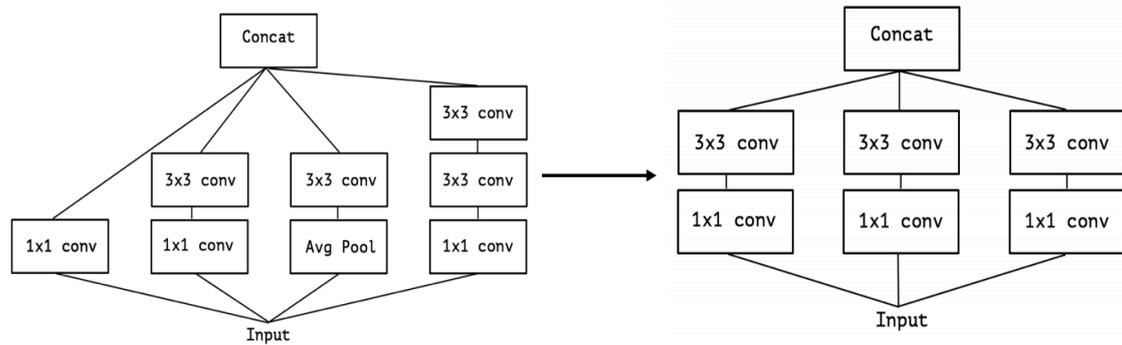


Figure 2.16: 起源網路模塊簡化示意圖⁶

基於上述的原因，所以極端起源網路的模組又再次簡化，成為 Figure 2.16 的右圖所示，每個通道都成為一個 1×1 的卷積層連接一個 3×3 的卷積層，但如果只是這樣做，將會變成幾乎一樣的卷積層分成三個路徑去學習，當卷積核多了以後，也就是下一層的特徵圖形變多時，特意分出去的分支將會顯得多餘且不重要，所以這邊特別提出了非常不同於一般卷積核，而與分離式卷積（separable convolution）[9] 相似的作法。

2.5.2 分離式卷積

分離式卷積不同於一般卷積，一般卷積所使用的卷積核為 $[x, y, z]$ 共 3 個向度的核心權重（kernel weights），其中 x, y 分別代表長與寬的空間向量，而 z 則是代表通道向量，以 60×60 的特徵圖群 28 張為例，一般 3×3 的卷積核代表的是 $[3, 3, 28]$ 的核心權重，每個卷積核將會直接學習 28 張特徵圖的相同空間向量位置的資訊，而分離式卷積則是會先通過一個空間向量的卷積核。

以上述例子為例，每個分離式卷積核的空間向量卷積核的 z 只為 1，也就是 $[3, 3, 1]$ 的核心權重，此種空間向量卷積特別學習並侷限於特定一張特徵圖，每個卷積核針對一張特徵圖作長與寬方面的學習，所以稱為空間向量卷積核，之後再通過 $[1, 1, 28]$ 的通道向量卷積核，與空間向量卷積核相反，卷積核的核心權重向

⁶圖片來源：<https://arxiv.org/pdf/1610.02357.pdf>



量 $[x, y]$ 均為 1，通道向量卷積核只針對特徵圖群間同一點位置去做學習，也就是在相同特徵圖群裡的相同長寬座標的特徵圖點，針對這些圖群間的相同位置點去做連結與學習。

經過兩種不同向量空間的卷積核學習後，新的特徵圖一樣具備完整的三種核心權重所連結出來的關係，而且更加理想化了赫布理論，使較有關聯性的特徵權重能結合在一起，經由不同的卷積核掃出來的特徵圖，疊加在一起成為新的特徵圖群，成為下一層的輸入，與一般卷積核一樣容易使用，卻又有些微但重要的不同。

極端起源網路所使用的分離式卷積核再次修改了上述的分離式卷積核，第一是順序的調換，極端起源網路的分離式卷積核為先通過 1×1 的通道向量卷積核，再對滑動視窗掃出來的特徵圖群，作 3×3 的空間向量卷積，每一個卷積核只對 1×1 的通道向量卷積核得出來的特徵圖群中的其中一張圖作學習與連結，如 Fig 2.17 所示。

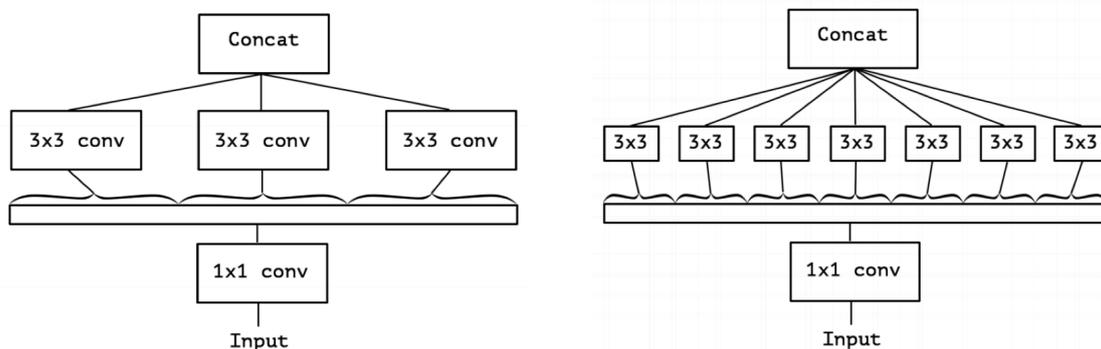


Figure 2.17: 極端起源網路模塊⁷

左圖為特徵圖群經過 1×1 的通道向量卷積核後，產生 3 個特徵圖，而針對這 3 個特徵圖，將產生 3 個 3×3 的空間向量卷積核，去各自學習各自的特徵圖，經過滑動視窗卷積後，產生的特徵圖再堆疊成新的特徵圖群，右圖則是經過 1×1 的通道向量卷積核後，產生了 7 個特徵圖，針對這 7 個特徵圖，產生 7 個 3×3 的空間向量卷積核，一樣去學習各自的特徵圖，之後疊加成為新的特徵圖群。

第二個不一樣的地方則是激活函數使用時機的選擇，極端起源網路的分離式卷積並沒有使用激活函數去增加非線性因素，而是讓兩層的卷積核所構成的分離式

⁷圖片來源：<https://arxiv.org/pdf/1610.02357.pdf>



卷積做完運算後，再經過區段正規化，才使用線性修正單元。兩層卷積層中間有無使用激活函數對辨識率的影響如 Figure 2.18所示，可以發現，不使用激活函數得到最佳的結果。

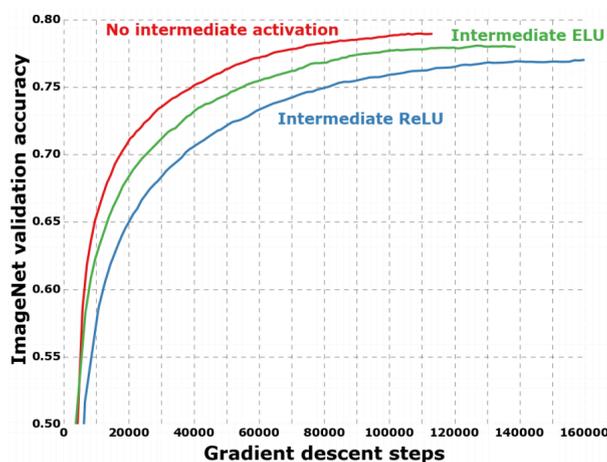


Figure 2.18: 極端起源卷積層的不同激活函數之訓練結果⁸

整體的網路架構比起源網路簡化很多，利用極端起源網路模塊組合而成，可以很容易地利用 Keras [10] 及 Tensorflow [11] 實踐，如 Figure 2.19所示，以 Keras 而言，運用許多的 *SeparableConv* 模型函式可以很輕鬆的堆疊出模型架構。

2.5.3 極端神經網路的其他嘗試

通過極端網路的卷積層後，可以選擇是否加入全連接層，在現今眾多的網路模型中，除了最後一層的激活函數為歸一化指數函數的全連接層外，幾乎沒有在卷積層後再使用全連接層的網路，因為全連接層的運算代價極高，且依照架構的不同，全連接層未必會帶來正面的效益，所以在使用極端起源網路時，可以自己決定是否添加全連接層。

極端起源網路也有嘗試使用殘差路徑，但在實作上，並沒有找到合適的使用方式使辨識率能提高，然而，殘差路徑依然提供了加速收斂的優點，除此之外，作者認為殘差路徑並不為加深網路的必要要素，使用類似 VGG [12] 的網路堆疊方式，依然能疊出非常深層的網路，如極端起源網路是類似於 VGG 網路的堆疊方式，只是把一般卷積層變換成為分離式卷積層，詳細的訓練記錄如 Figure 2.20。

⁸圖片來源：<https://arxiv.org/pdf/1610.02357.pdf>

⁹圖片來源：<https://arxiv.org/pdf/1610.02357.pdf>

¹⁰圖片來源：<https://arxiv.org/pdf/1610.02357.pdf>

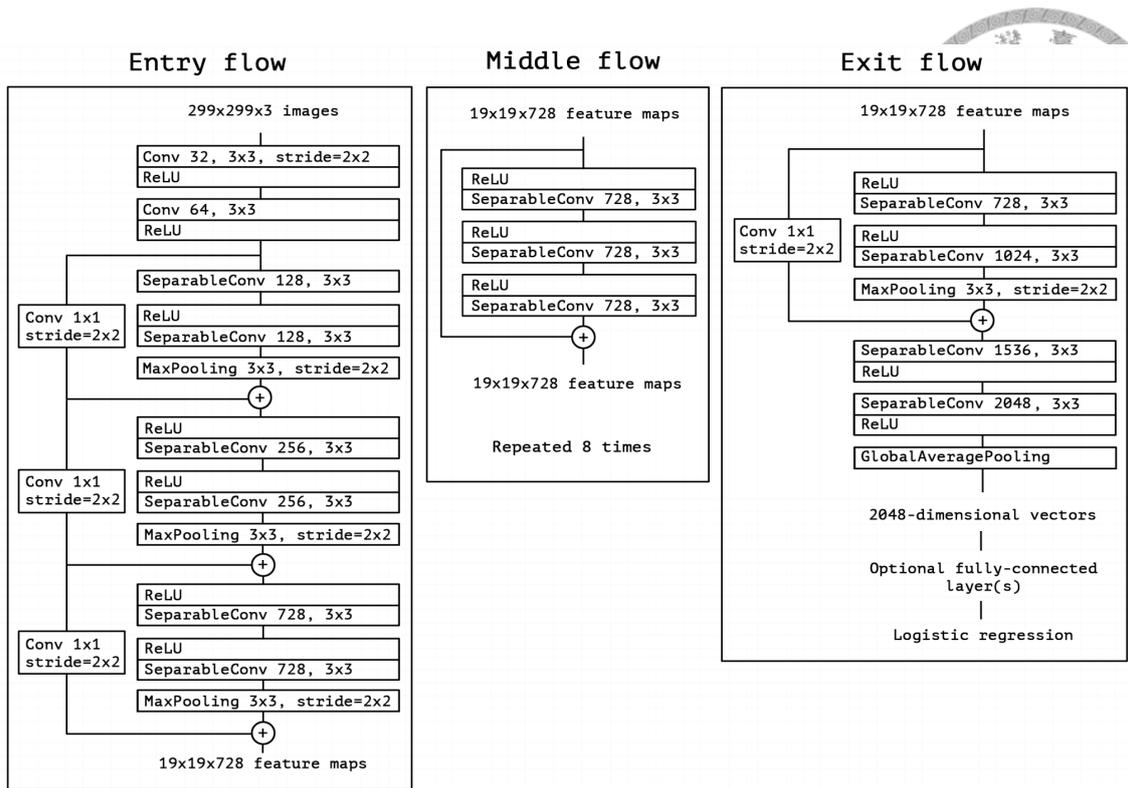


Figure 2.19: 極端起源網路模型架構⁹

2.6 損失函數

損失函數 (loss function) 為數學上迴歸常見的問題，不管在訓練深度神經網路，或單純的線性迴歸，都需要使用到損失函數，當線性乘積或神經網路輸出一個輸出值後，如何使這個輸出值與目標值 (ground truth) 越來越接近就是下一個難題，需要一個量化的指標去衡量輸出值與目標值之間的差距，針對該量化指標去做偏微分更新權重值，常用的損失函數有均方誤差 (mean-square error, MSE) 與交叉熵 (cross-entropy, CE)，不同的損失函數對於整體網路的訓練也會有不同的效果。

2.6.1 均方誤差

較常用於線性迴歸問題，為目標與預測值之間差的量度，其值為預測的值與目標值樣本間的標準差 (standard deviation) 偏移量，其公式如下：

$$MSE = \frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n} \quad (2.10)$$

其中 n 為線性迴歸或深度網路輸出值的總數， \hat{y} 為最後一層網路的實際輸出的

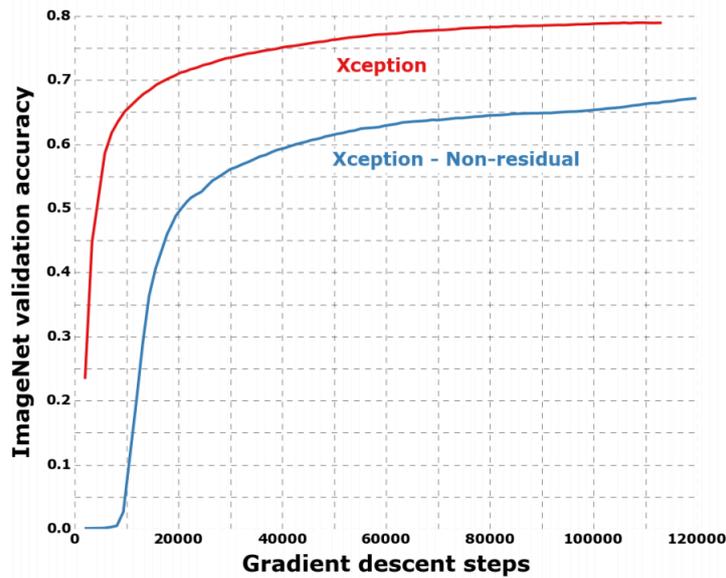


Figure 2.20: 極端起源結合殘差路徑的訓練紀錄¹⁰

各項值， y 為目標的各項值。由公式可以得知，當兩者間的差異性越小時，MSE 的值將會越來越小，反之當 MSE 較大時，則代表輸出值與目標值間的差異較大。

在對均方誤差做偏微分時， \hat{y} 為神經元的最後輸出，也就是 $\sigma(wx + b)$ ， w 為權重， x 為上層輸出， b 為偏差值 (bias)， σ 為邏輯函數， L 為 MSE 函數，訓練時需要針對 w 與 b 做偏微分以更新，公式如下：

$$\frac{\partial L}{\partial w} = (\hat{y} - y)\sigma'(wx + b)x = \hat{y} + \sigma'(wx + b) \quad (2.11)$$

$$\frac{\partial L}{\partial b} = (\hat{y} - y)\sigma'(wx + b) = \hat{y} + \sigma'(wx + b) \quad (2.12)$$

而更新時如下公式，其中 η 為學習速率 (learning rate)：

$$w \leftarrow w - \eta \frac{\partial L}{\partial w} = w - \eta * \hat{y} * \sigma'(wx + b) \quad (2.13)$$

$$b \leftarrow b - \eta \frac{\partial L}{\partial b} = b - \eta * \hat{y} * \sigma'(wx + b) \quad (2.14)$$

又因為對邏輯函式作微分的 $\sigma'(wx + b)$ 在取有一大部分值時會出現極小值，因為如前面 Figure 2.4 所示，邏輯函式的兩極端偏平，微分後出現在極大極小值的數

值將會非常難以更新，因為 $\sigma'(wx + b)$ 將會趨近於 0，導致 $\eta * \hat{y} * \sigma'(wx + b)$ 的結果亦是趨近於 0。



2.6.2 交叉熵

交叉熵在現今網路中常用於分類問題，也能解決上述均方誤差所帶來的缺點，與均方誤差一樣的地方在於兩者運算的結果必然不是負數，且輸出值與目標值越相近時，函數結果越趨近 0，其公式如下：

$$CE = -\frac{1}{n} \sum_x (y \log \hat{y} + (1 - y) \log(1 - \hat{y})) \quad (2.15)$$

在對交叉熵做偏微分時， \hat{y} 為神經元的最後輸出，也就是 $\sigma(\sum_{j=1}^n w_j * x_j + b)$ ， w 為權重， x 為上層輸出， b 為偏差值 (bias)， σ 為邏輯函數， L 為 CE 函數，訓練時需要針對 w 與 b 做偏微分以更新，公式如下：

$$\frac{\partial L}{\partial w_j} = \frac{1}{n} \sum_x x_j (\sigma(\sum_{j=1}^n w_j * x_j + b) - y) \quad (2.16)$$

$$\frac{\partial L}{\partial b_j} = \frac{1}{n} \sum_x (\sigma(\sum_{j=1}^n w_j * x_j + b) - y) \quad (2.17)$$

仔細比較均方誤差與交叉熵可以發現，在做導數運算時，均方誤差會出現 σ' 的運算，也就是對邏輯函數作微分，導致在極大以及極小值時難以更新權重，而交叉熵並沒有對邏輯函數作微分，權重的更新主要只受到輸出值與目標值之間的差所影響，因此在數值差距大時，調整權重的幅度會較大，而數值差距小時，調整權重的幅度會縮小，有利於神經網路的訓練，以加快找到最小值並逐漸趨近之。

實驗方面，[13] 的結論也指出交叉熵在大部分的情況能找到比均方誤差更為理想的區域最小值，因為權重的變更在極大值時，交叉熵能更有效率的調整權重。



Chapter 3

實驗設置與結果

3.1 資料集

本次研究所使用的資料集為工業技術研究院（Industrial Technology Research Institute, ITRI）資通所所提供的植物影像資料集，包含約 2400 種植物的種類，植物影像圖片數量約為 240,000 張，來源多為網路搜尋圖片，一個種類中可能包含植物的各項器官，也可能是局部的植物圖片或植物的整體影像，如 Figure 3.1 所示，有 3 張圖片，圖中植物皆為同一種類，左邊是花的影像，中間是葉子的影像，最右邊則為整體的植物影像，可以見得資料集影像間彼此的差異程度大，有些為植物器官的特寫，如左邊與中間的影像，而有些則是拍攝距離較遠的植物的整體影像，不同的圖片中包含的特徵有很大的差異。



Figure 3.1: 工業技術研究院所提供的樹葉資料集

總數為 2,400 類的資料，共有超過 240,000 張圖片，但每種類別間的圖片數目差距也很大，訓練資料中單一類影像資料最多為 450 張，但單一類影像資料最少的數量可以少至個位數的資料量，如 Figure 3.2 所示，其中橫軸為單一類植



物所擁有的資料的區間，縱軸為類別的數量，而長條圖則是表示擁有的資料影像數量落入該區間的類別有多少，大部分的植物類別所擁有的數量落在 $[0, 45]$ 之間，第二名多的區間則是在 $(45, 90]$ 之間，而超過 315 張影像的類別僅僅只有最後的極短的長條圖，共 33 類而已。

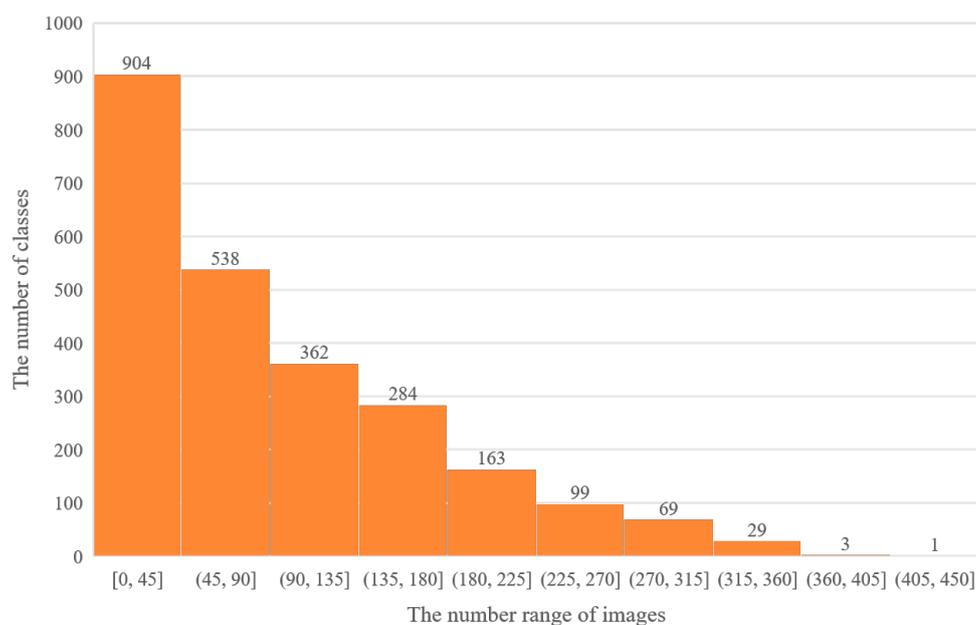


Figure 3.2: 資料集分佈

3.1.1 Top-500 資料集

可見在超過 2400 類圖片中，大部分類別的訓練資料是非常不足的，而對於深度神經網路而言，資料量的多寡嚴重地影響了模型對於該類別的辨識率，過於懸殊的資料量，影響了模型的學習能力。

又測試時，我們會從工研院所提供的資料集中，針對每個類別依照 9 : 1 的比例切出訓練資料與測試資料，如此在資料集中，數量區間落在較稀少的類別，測試資料也會極端的少，測出來的準確率可能只有全對或全錯，影響總體準確率的同時，也無法針對該類別去做適當的錯誤分析，因為測試資料是隨機選擇的，有一些機率會選到該類別中較特別的資料如 Figure 3.3，圖中資料相對於 Figure 3.1 獨特許多，可能圖片中不重要的雜訊太多或與該類別中其他資料相差太遠。

為了避免資料量的不足與差距過大導致模型的辨識率下降與研究上的困難，我們依照資料量的多寡選出了最多資料量的 500 類，之後以 Top-500 簡稱之，Figure



Figure 3.3: 較為特殊的測試資料

3.4為 Top-500 資料集的圖片數量分佈圖，其中圖片量最少的含有數量為 158 張，最多的為 450 張，其中圖片數量最多的類別，其圖片數量介於 [158, 187] 之間，Top-500 資料集包含了 110,934 張圖片，每個類別都依照 9:1 的比例切出訓練資料與測試資料，訓練資料的總數為 99,673 張，測試資料為 11,261 張。

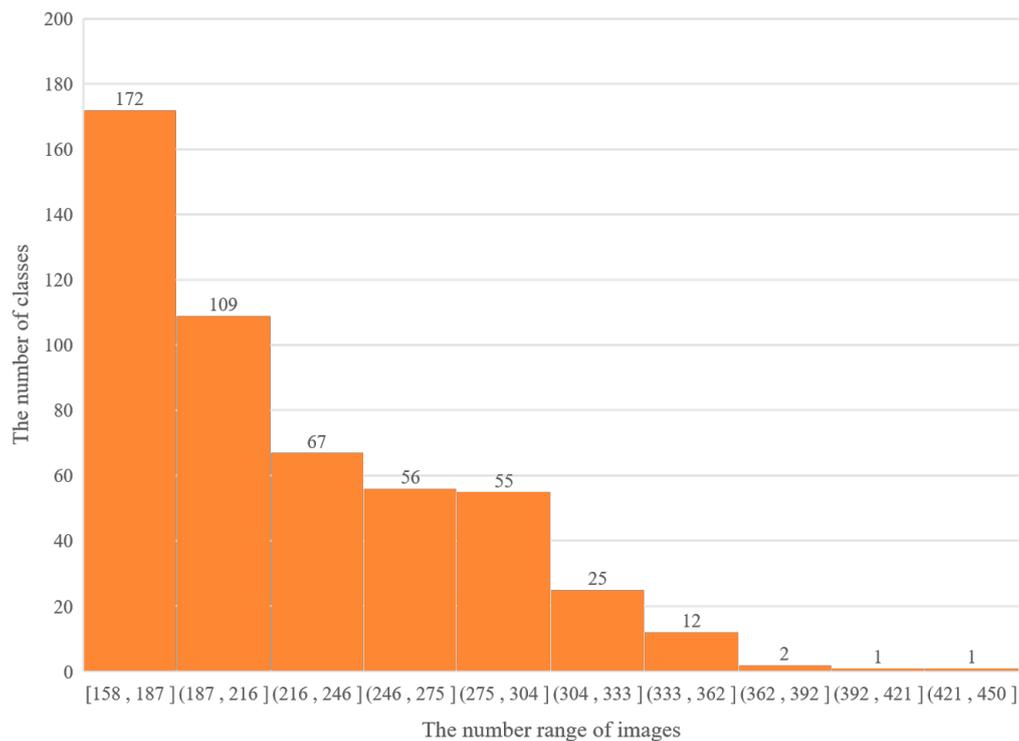


Figure 3.4: 數量最多的前 500 類分佈

3.1.2 Top-500 器官資料集

本篇論文嘗試以著重於植物器官的分離式獨立模型進行植物影像辨識，受惠於工研院所提供的資料整理，此資料集有針對每張圖片去標示該張圖片是否出現植

物的根、莖、葉、花、果實或其他器官，如 Figure 3.5所示，若圖片中出現花，則該圖片花的欄位為 1，若無則為 0；若圖片中出現葉子，則該圖片葉的欄位為 1，若無則為 0；若圖片中出現根，則該圖片根的欄位為 1，若無則為 0；若圖片中出現莖，則該圖片莖的欄位為 1，若無則為 0；若圖片中出現果實，則該圖片果實的欄位為 1，若無則為 0；而如果沒有出現這些人為定義裡較重要的器官，則可能出現了其他器官，則歸類為其他，其他為 1。

plant	植物	檔案	花	葉	莖	根	果	其它
Otanthera sca	糙葉耳葯花	C3025_267.jp	1	0	0	0	0	0
Otanthera sca	糙葉耳葯花	C3025_368.jp	0	0	0	0	0	1
Otanthera sca	糙葉耳葯花	C3025_15.jpg	0	0	0	0	0	1
Otanthera sca	糙葉耳葯花	C3025_244.jp	1	1	0	0	0	0
Otanthera sca	糙葉耳葯花	C3025_205.jp	1	1	0	0	0	0
Otanthera sca	糙葉耳葯花	C3025_203.jp	1	0	0	0	0	0
Otanthera sca	糙葉耳葯花	C3025_8.jpg	0	0	0	0	0	1
Otanthera sca	糙葉耳葯花	C3025_10.jpg	0	0	0	0	0	1
Otanthera sca	糙葉耳葯花	C3025_1.jpg	1	0	0	0	0	0
Otanthera sca	糙葉耳葯花	C3025_269.jp	1	0	0	0	0	0
Otanthera sca	糙葉耳葯花	C3025_7.jpg	0	0	0	0	0	1
Otanthera sca	糙葉耳葯花	C3025_5.jpg	1	0	0	0	0	0

Figure 3.5: 資料集器官標籤

對於所有的資料集統計，發現在 240,000 以上的圖片中，出現最多的器官為葉子，數量為 134,838 張，而第二多的為包含花的圖片，數量為 121,099 張，第三多的為 36,340 張，其次為莖與根，分別為 4780 張與 518 張，如 Figure ??所示，如圖可知，根與莖的數量比起最多的葉子，已然非常非常稀少，分布 2400 以上的種類中，每種種類平均甚至不到 1 張，因此莖與根無法提供足夠的訓練資料去訓練出獨立的器官模型，因此本篇論文的資料集將會從擁有最多圖片的 500 類中，去根據資料集器官的標籤，把擁有花、葉、果實之任一種標籤的圖片選擇出來，去嘗試建立花、葉、果實的獨立器官模型，選擇出的資料之訓練資料有 97,054 張，而測試資料有 10,787 張。

而之後的實驗會提到花的辨識率在三中器官中為最高的，果實次之，葉子的辨識率最低，所以若影像中同時出現花、果實、葉子，則會把該影像分類在花的類別中，而若沒有花的標籤出現，但是出現葉子與果實，則把該圖片分類在果實，最後只有葉子的標籤的圖片才會歸類在葉子的圖片，以此為分類，如 Figure 3.7所示，其中標籤有花的訓練資料有 61,402 張，測試資料有 6,816 張；葉子的訓練資料有 22,210 張，測試資料有 2,482 張；果實的訓練資料有 13,442 張，測試資料有

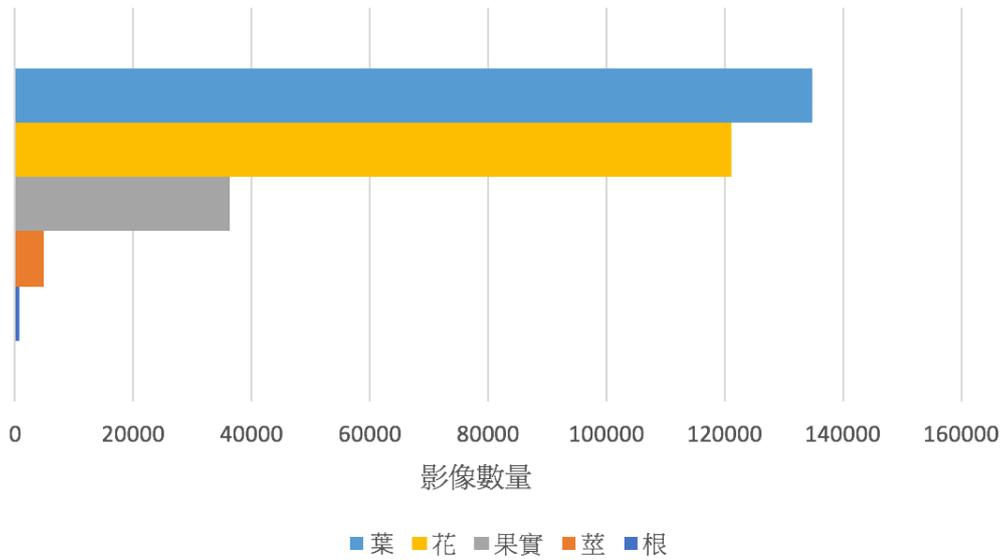


Figure 3.6: 包含各項器官影像之數量分佈圖

1,489 張，之後將會以 Top-500 器官資料集代稱之。

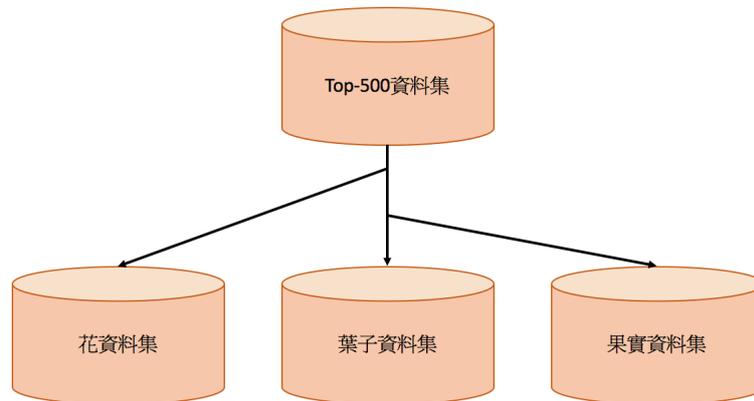


Figure 3.7: 切割資料集示意圖

3.2 實驗環境

OS: Ubuntu 16.04

CPU: Intel Xeon CPU E5-2630 v4

GPU: Nvidia GeForce GTX 1080 Ti

Host memory: 128GB

Device memory: 12GB

Programming language: Python 3.6.3 & TensorFlow 1.3.0 & Keras



3.3 影像前處理

為了使實驗能順利進行或加速以及穩定模型，在訓練資料與測試資料分別進行訓練與測試之前，會先經過資料影像的前處理，本實驗在影像前處理中分為兩部分，第一部分為影像像素值的重縮放 (rescale)，與第二部分為資料增強 (data augmentation) 去增加訓練資料的豐富度，使網路能學習到更多的正確資訊。

3.3.1 重縮放

一般而言，輸入的彩色圖片會有 RGB 三個通道 (channel)，其中 R 代表紅色，G 代表綠色而 B 代表藍色，每個通道的每個像素值 (pixel) 為 0-255 之間的正整數，0 通常為不顯示該顏色，而 255 代表該顏色的色度最深，一張彩色圖片的一個像素值是由 3 個通道對應的不同的數值所組成，而深度卷積神經網路的對應則為卷積核去掃描輸入的影像，為 $[x, y, 3]$ 大小的卷積核，其中 x 與 y 為長與寬的大小，不論大小為何，最後都會需要掃描的通道為 3，而卷積神經網路的輸出值通常為某個類別的機率值，在經過歸一化指數函數或邏輯函數後，輸出的機率值介於 0-1 之間，0 為可能性最小，而 1 為可能性最大。

輸入層為 0-255 之間的數值，而輸出值為 0-1 之間，這樣深度神經網路在更新每層的權重值時，同時也需要使每層的數值逐漸下降使輸出值能介於 0-1 之間，在沒有區段正規化或其他輸出正規化之前，每層輸出數值逐漸縮小只能全由每層的權重值去改變，當學習速率 (learning rate) 較大時，網路前端層的輸入乘上一開始的權重值，經過更新後，可能產生梯度爆炸的問題，而較小的學習速率，對於更新神經網路又十分緩慢，且可能會陷入區域最小值 (local minimum)。受惠於區段正規化，有使用的深度卷積神經網路較不會出現梯度爆炸或消失的問題，但輸入卷積神經網路的值介於 0-1 之間仍然會在訓練初期時對整體網路的辨識率產生一些影響。

把 0-255 重縮放至 0-1 之間對於訓練資料的準確率幾乎沒有影響，若把訓練的

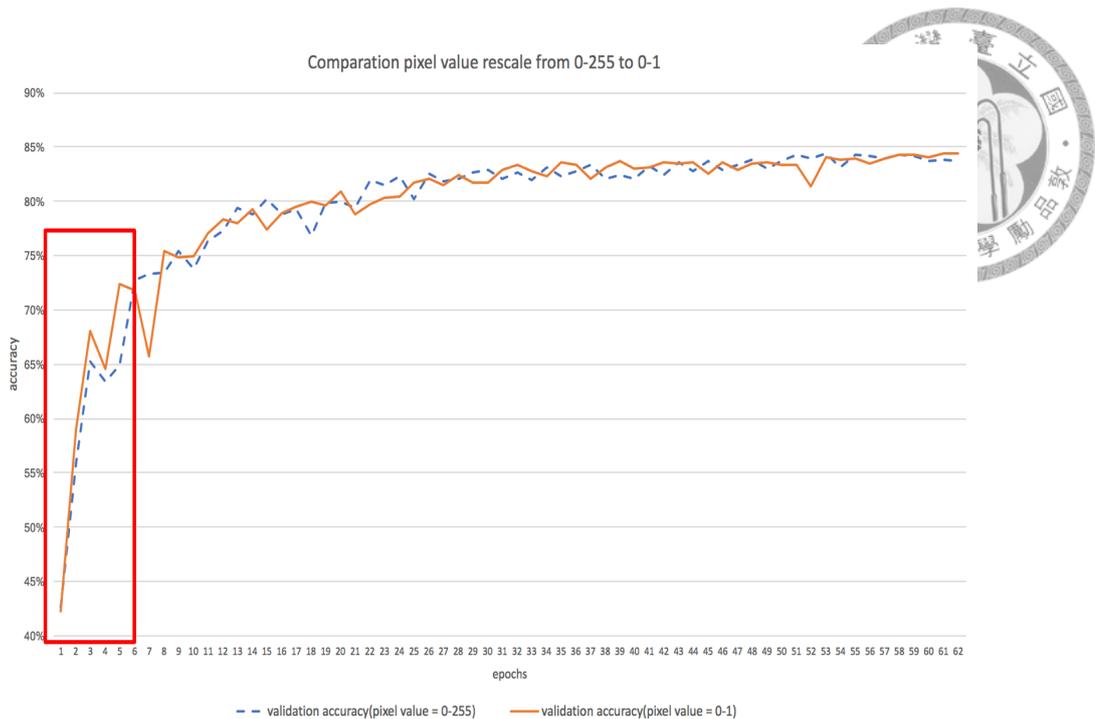


Figure 3.8: 縮小輸入值與否對訓練網路比較圖

準確率作折線圖來看，兩條線幾乎是重疊在一起，然而，重縮放對於測試資料的影響卻比較顯而易見，如 Figure 3.8所示，實線為輸入影像的每個像素單位值在 0-1 之間，而虛線則為輸入影像的每個像素質在 0-255 之間，在其他實驗參數及環境都相同的情況下進行實驗，在框出來的地方中，可以發現訓練初期時，實線部分的準確率相比虛線部分要來得更高，實線部分準確率上升較虛線更為快速，然而網路收斂後，兩者的辨識率卻相差不多，可見縮小像素單位值至 0-1 之間有助於網路前期的訓練，在更新權重值能對測試資料有更佳的辨識率。

3.3.2 資料增強

在現今深度卷積網路的訓練中，需要非常大量的資料去支持整個網路去做學習，太少的訓練資料容易發生過度訓練的問題，使模型在預測訓練資料時十分準確，但在使用訓練中從未出現過的測試資料時預測正確率非常的低，但大量資料的收集十分耗費人力成本，整理大量資料又十分耗時，在資源有限的情況下，資料增強能些為稍微彌補訓練資料不足所帶來的窘境。

資料增強為針對原有訓練資料作水平翻轉、垂直翻轉、旋轉、平移、放大縮小等等的處理之後，再把經過處理過後的影像放入深度神經網路之中，使神經網路

接受與原本訓練資料有些微不同的影像當作新的訓練資料，雖然這些訓練資料不是真實存在的資料，但運用了這些新的資料，使神經網路能嘗試學習不同的特徵，降低過度學習的可能性，藉以提高辨識率。



Figure 3.9: 資料增強

水平翻轉：隨機

放大縮小：倍率在 10% 之內隨機

旋轉：旋轉角度在 +20 度到-20 度之間隨機

本次實驗中所使用的資料增強，包括水平翻轉、放大縮小與旋轉。只使用水平翻轉，而不使用垂直翻轉是因為人在對植物進行拍照時，不會倒立拍照，植物多為正立的照片，然而水平的翻轉依然是正立的植物照片，且對於影像而言，卻是很大的不同，旋轉角度的取捨也與此相關，旋轉角度範圍限制在-20 度到 +20 度共 40 度以內，屬於正常的拍照角度範圍，放大縮小的倍率限制在 10% 以內，有鑒於影像為解析度 227*227 的圖片，過於放大將會重度影響影像的清晰程度，也會造成網路訓練不良的原因。

在資料進入卷積神經網路之前，會先經過亂數決定這張影像是否水平翻轉、在 10% 以內的放大縮小的倍率，以及在正負 20 度之內的旋轉角度，因此對於卷積神經網路而言，每次的影像輸入皆是不一樣的，如 Figure 3.9 中，也許在人類看來，那幾張圖片幾乎一模一樣，不仔細看甚至看不出太大的差別，但是就是這些細小差別，使卷積神經網路能提高對於測試資料的辨識準確率，對於神經網路來說，上述的資料增強方式，對於卷積核去掃描輸入圖片而言，都是相差甚大的。

實驗使用起源神經網路，結合下個小章節會提到的遷移式學習，在其他條件相同下，去比較有無資料增強對於測試資料的準確率會帶來什麼樣的影響，從

Figure 3.10所示的實驗中，其中藍線為使用資料增強，而橘線為不使用資料增強，可以很明顯的發現有使用資料增強的測試準確率幾乎都高於無資料增強的測試準確率，由實驗可知，資料增強對於訓練卷積神經網路有非常理想的助益。

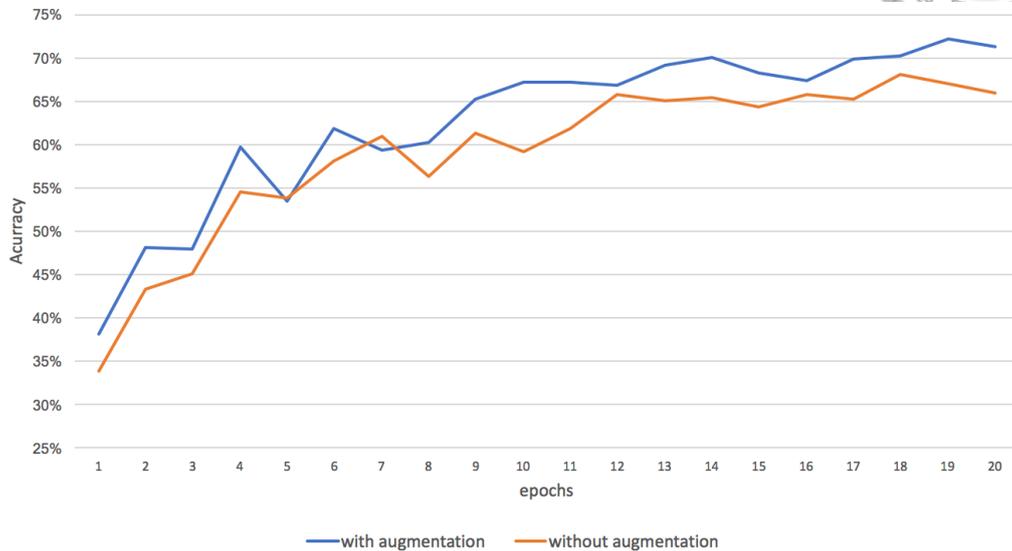


Figure 3.10: 有無資料增強對於測試資料準確度之影響

而資料增強個別的影響如 Figure 3.11，在這次的個別實驗中，一樣在其他條件不變下，去測試每種不同的資料增強方式對於卷積神經網路辨識率的影響，藍線為不使用任何的資料增強，橘線為單獨使用隨機水平翻轉的資料增強方式，黃線為單獨使用放大縮小的資料增強方式，最後灰線為單獨使用隨機旋轉角度的增強方式，從實驗中可以發現，在訓練後期，有使用任何資料增強的實驗組決大部分會比不使用任何資料增強的實驗對照組辨識率高上一截，其中資料增強中的水平翻轉為影響幅度較高者，而對比於 Figure 3.10綜合使用資料增強的實驗提高的辨識率又比個別使用更高，其中個別的資料增強的參數設定則是由多次實驗設計而出，並無特殊限制在這組資料增強的參數，加入其他方式或針對不同的資料集需要有不同的設定。

3.4 遷移式學習與模型選取

遷移式學習在人類學習上是非常重要的環，當某種知識或技能的學習對於其他方面的學習產生幫助的時候，可以視為兩種學習間的刺激或反應相似，就是一種正向的遷移式學習，如先學習英文再學習美語，兩者間有些為不同，但大部分

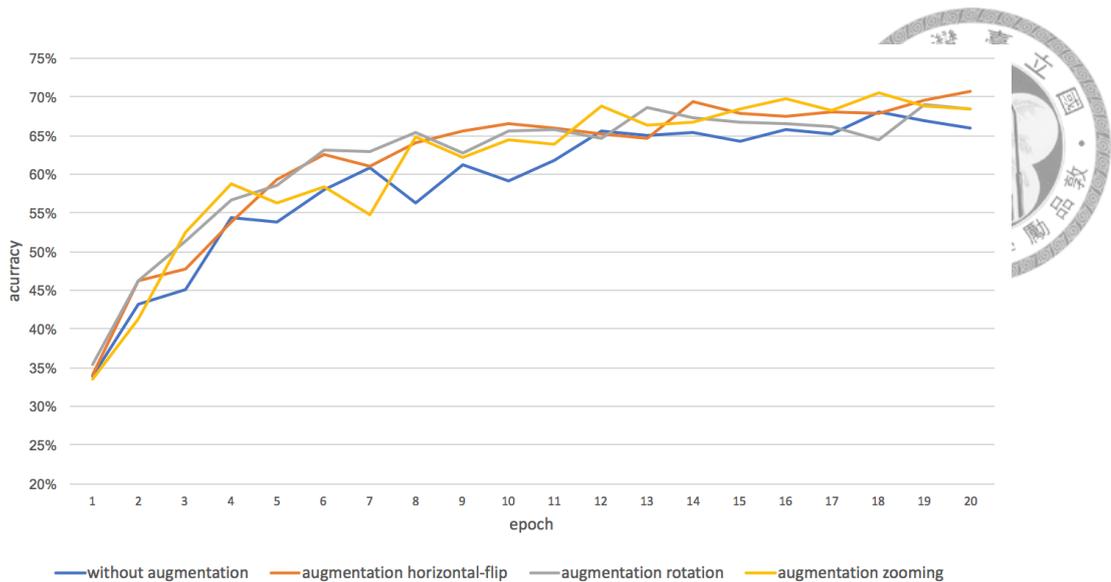


Figure 3.11: 資料增強個別的影響

的片語相似，把英文學好，再去學美語就會變得十分快速且精確，然而，若是該種知識或是技能的學習與目標學習的刺激或反應極度不相似，則可能會造成阻礙目標學習，這種稱為負面的遷移式學習，如先學習英文再去學習日語，兩者間極度不相似，不論是單字或文法，都有極度的不同，可能會在學習時造成混亂，先學習英文反而可能成為學習日文的絆腳石。

接下來將會介紹何為遷移式學習，遷移式學習如何應用於這次的實驗，與深度卷積網路模型的比較與選取，針對有使用遷移式學習與不使用遷移式學習去做實際的實驗比較兩者差異，針對這些差異去看看網路內部卷積層的輸出值有何不同，造成辨識率的高低影響。

3.4.1 監督式影像分類的遷移式學習

遷移式學習應用於深度網路模型已經行之有年，相當多的研究分佈於幾乎所有的深度網路領域 [14] [15]，舉凡監督式學習 (supervised learning)、非監督式學習 (unsupervised learning)、生成對抗網路 (generative adversarial network, GAN) 等等，本篇論文應用於監督式影像的分類，監督式影像分類的遷移式學習可以在大致上分為以下兩類：



水平性遷移式學習

水平性的遷移式學習為定義域間的轉換，從較不相關的資料集所訓練出來的模型，將其已訓練好的參數作為新的模型的參數起始點，用不同於原本資料集的新資料集去訓練這個網路，使新的網路能基於原本的所學習出的特徵轉換成適合新的資料集適合的特徵，以原本資料集所訓練出來的定義域為 $D_S = \{X_S, f_S(X)\}$ ，該資料集訓練的目標為 T_S ，而新資料集的定義域則為 $D_T = \{X_T, f_T(X)\}$ ，新的資料集訓練目標為 T_T ，其中 $D_S \neq D_T$ 且 $T_S \neq T_T$ ，水平性的遷移式學習，就是希望藉由 D_S 和 T_S 去改善 $f_T(*)$ 的定義域使 T_T 能有更好的表現。

Figure 3.12為本次實驗所使用的水平性的遷移式學習示意圖，Model A 為 Keras 利用前一章所介紹的網路模型所訓練，模型開始時使用 2012 年的 ImageNet 所提供的資料集，該年資料集提供 1,000 類別的影像，共計 100,000 張以上的影像資料，經過多次迭代收斂後，訓練出針對 ImageNet2012 的深度卷積網路，之後把這個訓練好的網路權重值移植到 Model B 作為起始的權重值，訓練資料更新為 Top-500 器官資料集，重新經過多次迭代收斂，使基於 Model A 權重的 Model B 之辨識率能高於直接使用 Top-500 器官資料集去做訓練的 Model C。

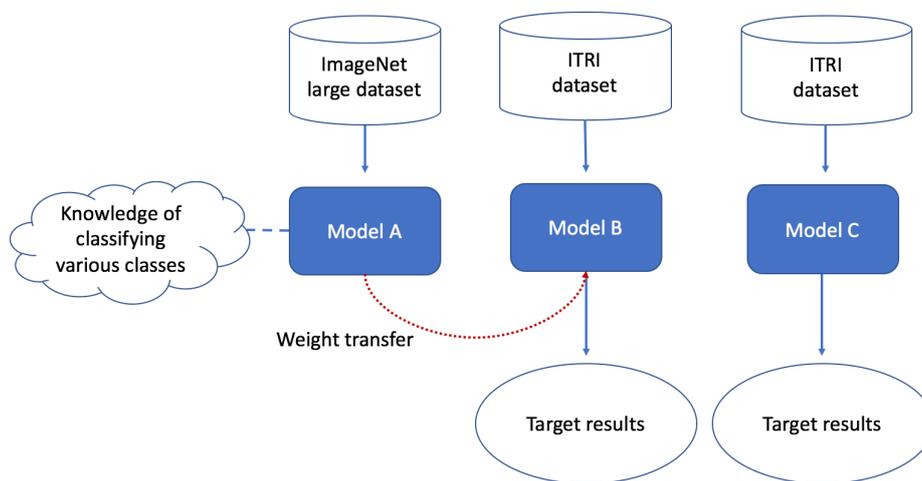


Figure 3.12: 水平性遷移式學習

垂直性遷移式學習

垂直性的遷移式學習不同於水平性的遷移式學習，原本資料集與目標資料集的相關性較高，從原本較大的資料集去學習訓練出模型，將已經訓練好的權重參

數值作為新的模型的參數起始點，而新的資料集為原本資料集的子集合或相關性極高的其他資料集，在這次的實驗中 Figure 3.13，先使用 Top-500 器官資料集訓練出 Model A，使用訓練完成的 Model A 的權重值為起始點，針對 Top-500 器官資料集的子資料集-花資料集、果實資料集、葉資料集分別重新訓練出 Model B、Model C 與 Model D，分別使原本的 Model A 特化成器官的分類模型，試圖以此把整體的辨識率再度提高。

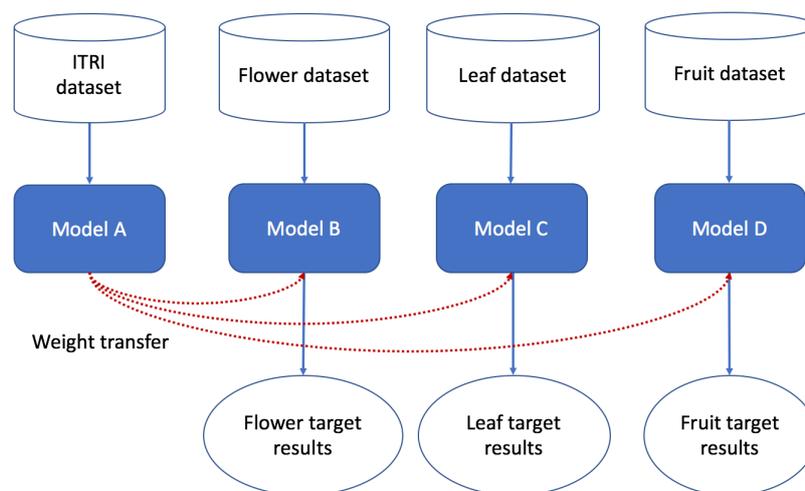


Figure 3.13: 垂直性遷移式學習

3.4.2 卷積神經網路模型選取

訓練複雜的深度卷積神經模型是相當困難的，需要耗費大量的時間以及運算資源，在一般小公司或實驗室中，很少有硬體能夠支援訓練需要非常多資料的網路模型，遷移式學習的應用稍微解決了這項難題，Google 旗下的 Keras 以及許多其他的套件提供了預訓練 (pre-trained) 的深度卷積網路模型的架構與訓練完成的權重值供一般大眾使用，受惠於這些公司利用大量的機器與時間訓練出來的模型，遷移式學習較容易被實際應用，否則從頭開始訓練 ImageNet 資料集將會是非常大的挑戰，一般機器要訓練到相近 Keras 釋出的模型權重值，可能就要花費一個月以上。

在上一章節，我們提到了由 Google 公司研究團隊所提出的起源網路、起源殘差網路與極端起源網路，是現今網路模型中，辨識率相當高的其中 3 種網路，在 ImageNet2012 年的資料集辨識率如 Table 3.1 所示。Top-1 正確率表示模型預測出

最高機率的類別與目標標籤相同才算正確，而 Top-5 正確率則表示當模型預測出的前 5 高的類別中，有與目標標籤相同的，即可算是正確。最低辨識率的起源神經網路也有 0.788% 的正確率，而 Top-5 正確率更是高達 0.944%，在總類別數為 1000 類的分類中，這些神經網路所達到的辨識成功率已經十分逼近人類的辨識成功率。

Table 3.1: ImageNet2012 資料集的辨識率

模型名稱	Top-1 正確率	Top-5 正確率	訓練參數數量
起源神經網路	0.788	0.944	23,851,784
起源殘差神經網路	0.804	0.953	55,873,736
極端起源神經網路	0.790	0.945	22,910,480

接下來將會以這 3 種模型架構為基礎，替換掉最後一層的 1,000 個神經元的全連接層，加入新的 500 個神經元的全連接層作為網路的輸出層，個別使用 Top-500 器官資料集進行遷移式學習，去比較有使用遷移式學習與不使用遷移式學習對於模型辨識率的影響，不同的模型將會有自己的實驗對照組。

3 種模型在 ImageNet 的資料集中，辨識率最高的為起源殘差網路，但考慮 ImageNet 資料集中類別間的影像差異程度較 Top-500 器官資料集類別間的影像差異程度還要大上許多，Top-500 器官資料集的類別全為植物，需要學習出的特徵可能需要更為精細，在此資料集中，哪種模型能達到最高的辨識率需要實驗去個別查證。

Table 3.2: 實驗參數設定

Batch size	32
Epoch	100
Learning rate	0.045
Decay	1e-6
Momentum	0.9
Nesterov	True
Loss	Cross entropy

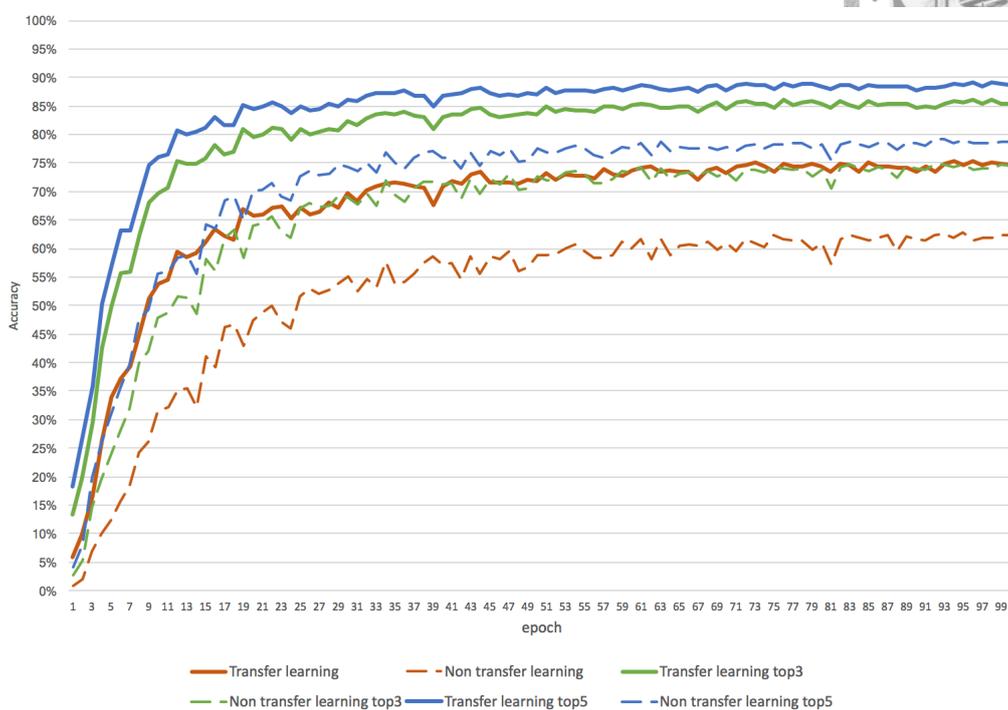
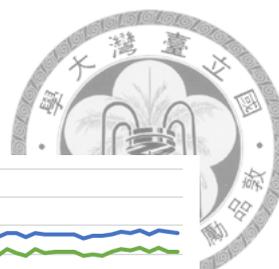


Figure 3.14: 起源神經網路訓練過程中的辨識率

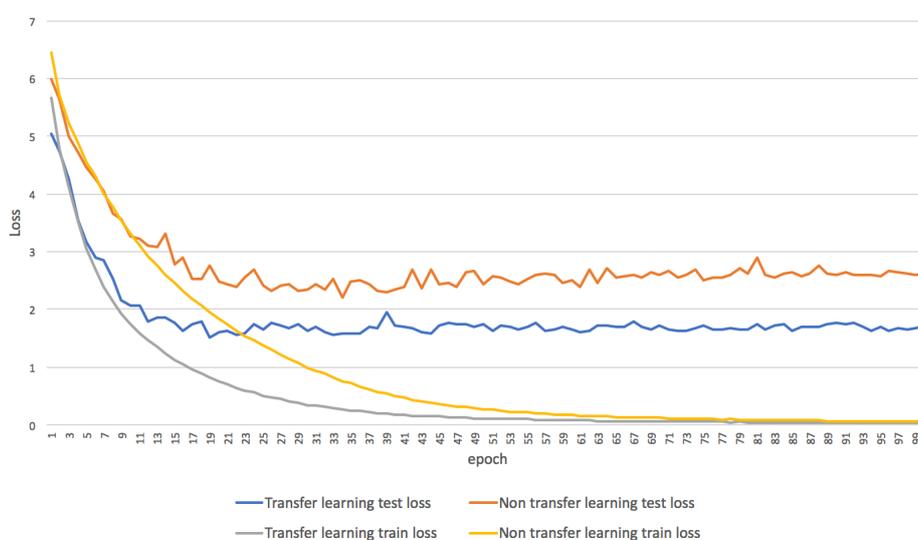


Figure 3.15: 起源神經網路訓練過程中的損失函數變化

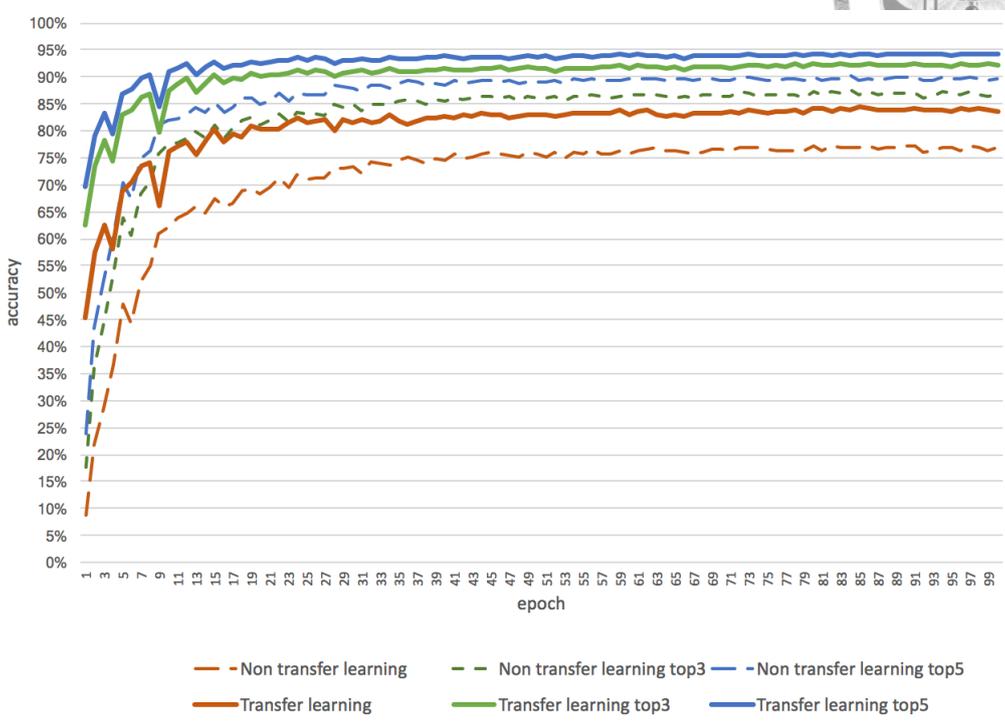
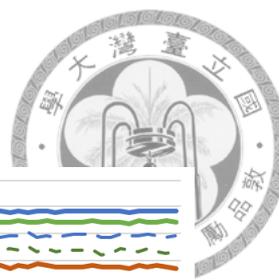


Figure 3.16: 起源殘差神經網路訓練過程中的辨識率

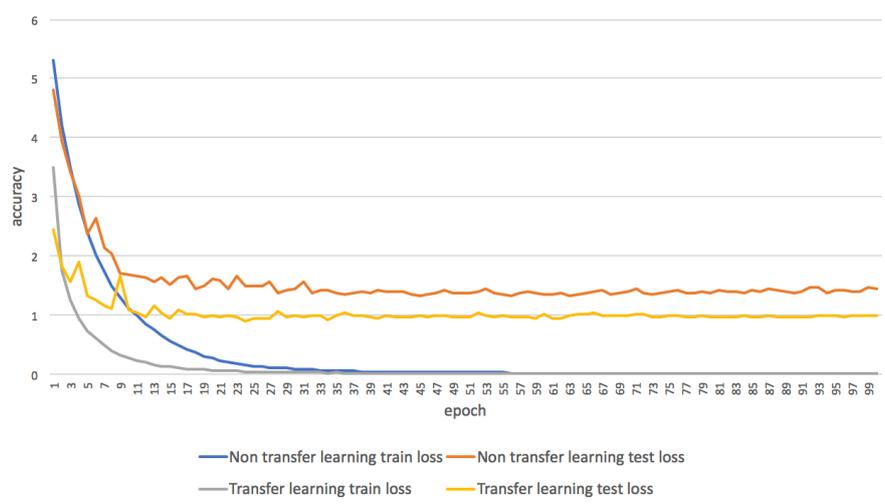


Figure 3.17: 起源殘差神經網路訓練過程中的損失函數變化

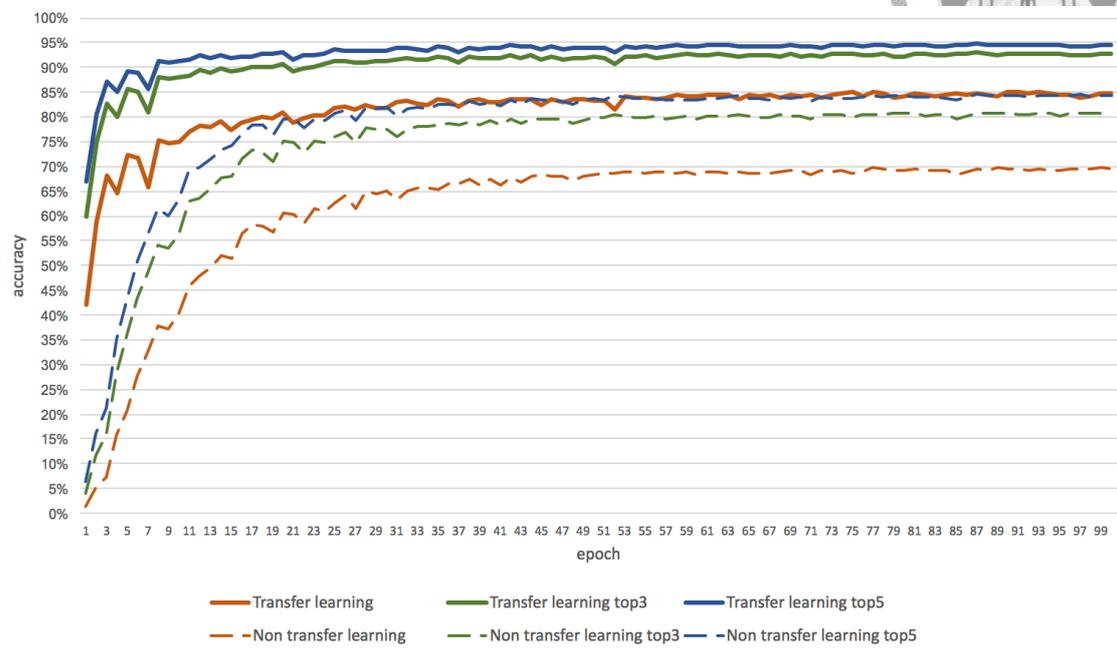
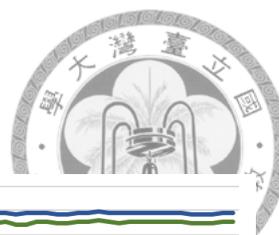


Figure 3.18: 極端起源神經網路訓練過程中的辨識率

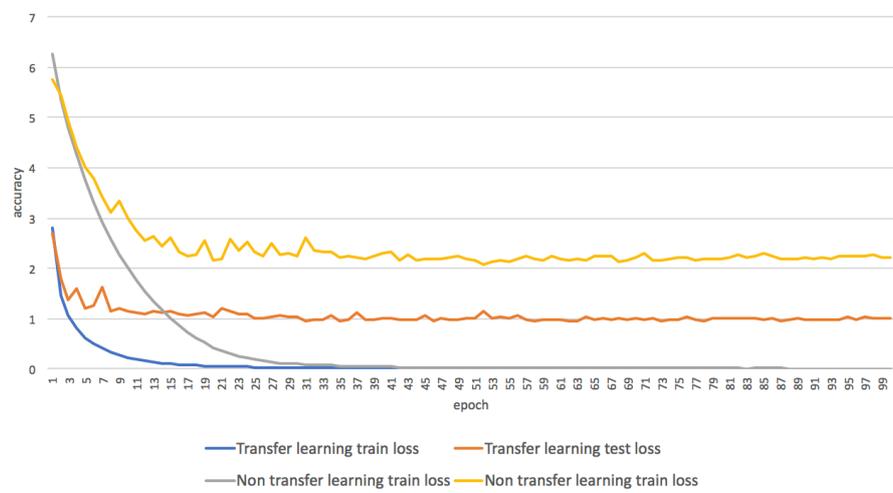


Figure 3.19: 極端起源神經網路訓練過程中的損失函數變化

上面三頁為三種網路的訓練過程辨識率的紀錄與損失函數的變化，每頁的上圖 Figure 3.14 3.16 3.18 為 3 種網路模型各自訓練過程中辨識率的記錄，其中實線為有使用 ImageNet2012 年的資料集作遷移式學習的訓練過程，虛線為沒有使用遷移式學習的訓練過程，紅色的線為 Top-1 正確率，綠色線為 Top-3 正確率，藍色線為 Top-5 正確率。每頁的下圖 Figure 3.15 3.17 3.19 則為 3 種網路模型訓練中的損失函數紀錄，包含有使用遷移式學習的訓練損失函數與測試損失函數，與沒有使用遷移式學習的訓練損失函數與測試損失函數。

使用 ImageNet 資料集的模型訓練 Top-500 器官資料集可以視為水平性的遷移式學習，資料集間彼此的相關性非常低，先從每頁的上圖開始解讀，Figure 3.14 3.16 3.18 的網路模型訓練紀錄中，可以很明顯的發現有使用遷移式學習的起源神經網路不論在 Top-1、Top-3、Top-5 的正確率都比沒有使用遷移式學習的對照組高很多，且辨識率的上升幅度十分快速，其中起源殘差網路的辨識率上升速度最快，因為殘差網路的結構，能夠加速網路的收斂，不論有無使用遷移式學習，收斂速度皆為這三種網路最快的。值得注意的是極端起源網路的訓練過程，有使用遷移式學習的極端起源網路，其效果比沒有使用遷移式學習的好非常多，其 Top-1 正確率幾乎與對照組中的 Top-5 辨識率相差不遠，可見遷移式學習對於 Top-500 器官資料集的訓練為正向的影響。

每頁的下圖為訓練過程中損失函數的變化，從 Figure 3.15 3.17 3.19 的損失函數下降速度，可以很明顯的發現，有使用遷移式學習的模型訓練速度較沒有使用遷移式學習的快很多，當訓練趨近收斂時，不論有無使用遷移式學習，訓練資料的損失函數數值都非常趨近於 0，針對訓練資料而言，兩種訓練方式的能都達到近乎 100% 辨識率，然而有無使用遷移式學習對於測試資料的損失函數影響很大，從圖中可知，有使用遷移式學習的模型的測試資料損失函數皆低於沒有使用的，可能為遷移式學習應用於此資料集時，因為已經是訓練良好的模型，所以再訓練時能夠避免陷入較不理想的損失函數區域最小值，比沒有使用遷移式學習更能找到相對的最小值，在另一方面也可以解讀為訓練良好的模型已經學會非常多且重要的細節，因此在學習其他資料集時，較不會出現過度擬合的狀況。

Table 3.3 整理了有無使用遷移式學習的起源神經網路、起源殘差網路與極端起源網路的各項最佳表現值，表中紅字為該列表現最佳者。在不使用遷移式學習的



Table 3.3: 有無使用遷移式學習的各類模型辨識率

模型名稱	Top-1 正確率	Top-3 正確率	Top-5 正確率	損失函數
起源網路	0.6251	0.7474	0.7880	2.562
遷移式起源網路	0.7533	0.8609	0.8908	1.624
起源殘差網路	0.7726	0.8728	0.8996	1.3866
遷移式起源殘差網路	0.8401	0.9227	0.9423	0.9663
極端起源網路	0.6975	0.8081	0.8433	2.2028
遷移式極端起源網路	0.8514	0.9273	0.9464	0.9739

狀況下，三種網路中表現最好的為起源殘差網路，Top-1 辨識率比第二名的極端起源網路都要高出將近 8%，而在使用遷移式學習後，極端起源網路的辨識率提高了非常多，不論是 Top-1、3、5 的辨識率皆超過起源殘差網路，0.8514 的 Top-1 辨識率比第二名的 0.8401 高了超過 1% 的正確率，且參考 Table 3.1 的訓練參數數量，可以發現極端起源網路所需的訓練參數值最少，而加入遷移式學習後表現反而還比其他兩種網路要更好。

在資料集類別之間較為相似的情況下，不使用遷移式學習，起源殘差網路能訓練出最佳參數模型，而使用遷移式學習，起源殘差網路的測試損失函數還是有最佳的表現，然而其訓練參數值最多，訓練時間與運算負荷要求最高，所以接下來的器官獨立模型，我們將會使用極端起源網路作為四種器官獨立模型所需要的模型的架構，以使用遷移式學習的極端起源網路作為標準值，嘗試使辨識率能再次提高。

3.4.3 遷移式學習特徵圖群分析

遷移式學習在數據上贏過沒有使用遷移式學習的模型非常多，而遷移式學習如何影響了辨識率，這次分析嘗試從類神經網路內部去了解遷移式學習使用與否的差別，在使用遷移式學習訓練好極端起源網路後，抽出極端起源網路架構中第二層的卷積層輸出的特徵圖群，如 Figure 3.20 所示，藉由抽取出的特徵圖群去了解兩者間的差異。

Figure 3.21 為沒有使用遷移式學習的特徵圖群，Figure 3.22 則為使用了遷移式學習的特徵圖群，每一張圖片為一個卷積核的輸出，從 36 張特徵圖中挑出了 28 張作為分析的依據，沒有使用遷移式學習的特徵圖群在邊角上的描述較遷移式學

¹部分圖片來自：<https://arxiv.org/pdf/1409.4842.pdf>

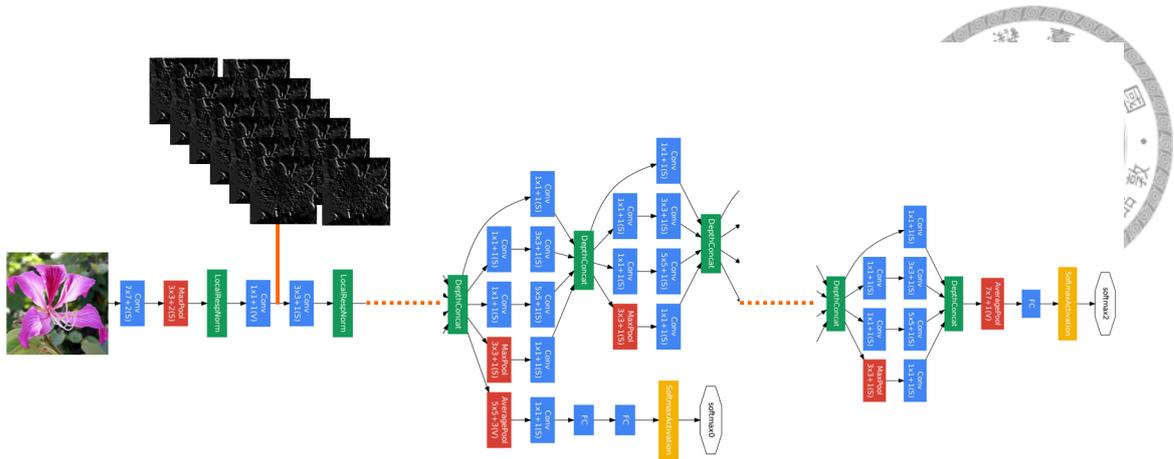


Figure 3.20: 抽取特徵圖群示意圖¹

習的特徵圖群薄弱，而其中又有絕大部分的特徵圖所掃描出來的特徵不是植物的特徵，如圖群中的下半部圖群，幾乎沒有學到植物應該有的特色，因為此特徵圖群傳到接下來的網路中，其他卷積層能得到的資訊也不是植物本身的特徵，以此為特徵資訊繼續學習下去，反而更難以幫助判斷植物的種類。

有使用遷移式學習的特徵圖群在邊角上的描述就非常細膩，如花蕊的部分表現就十分良好，每張特徵圖幾乎都包含了目標植物的特徵，因此在向下傳遞時，至少能比沒有使用遷移式學習的特徵圖群表現更好，因為其他卷積層也能得到正確的該學習的植物特徵，進而往下傳遞，使網路的正確率能有效提升，分析網路的前面幾層的輸出，就可以很容易的分辨出有無使用遷移式學習所帶來的優劣。

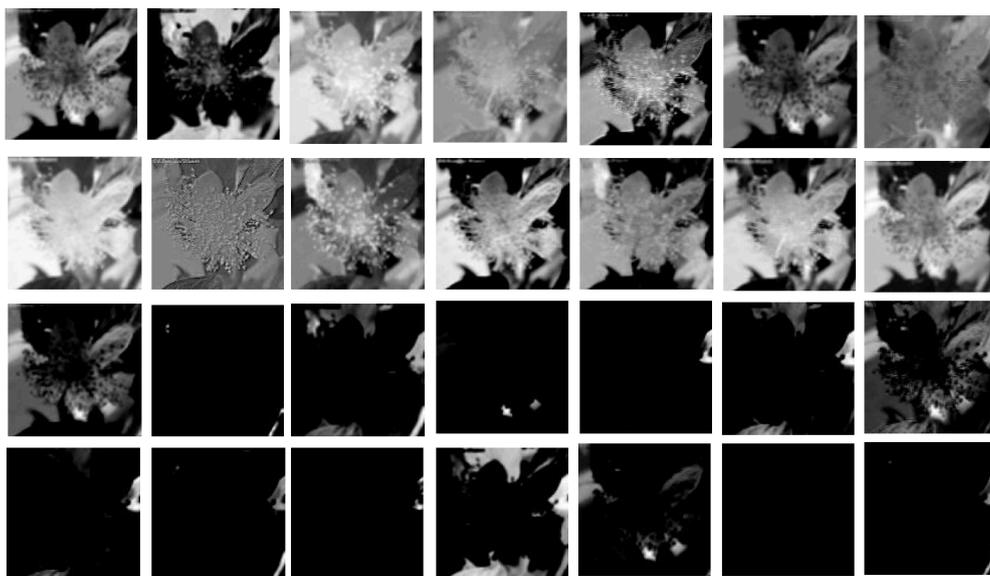


Figure 3.21: 無遷移式學習的特徵圖群

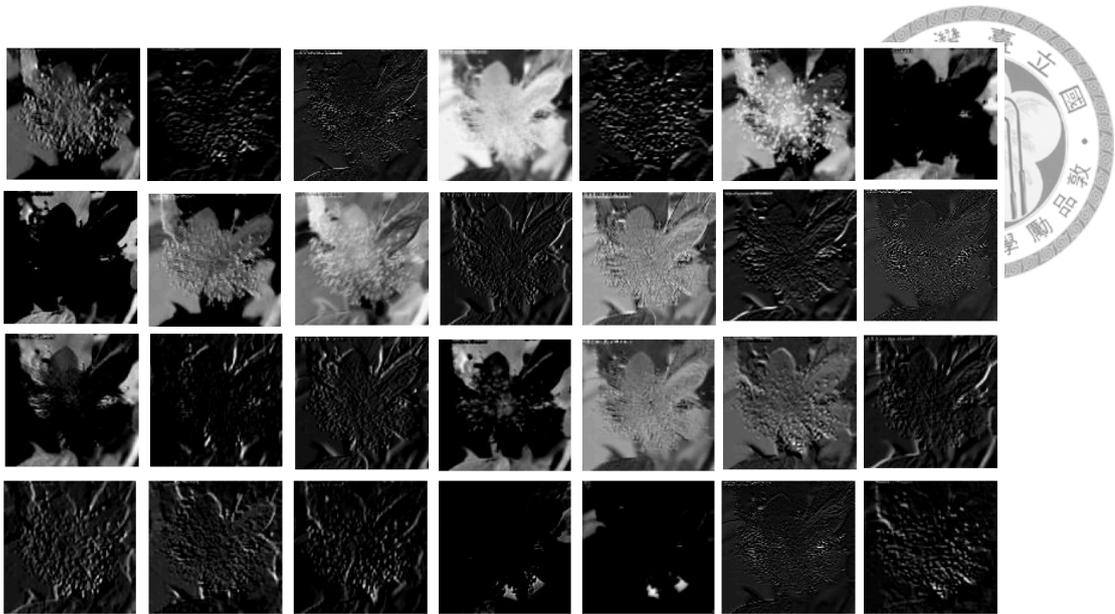


Figure 3.22: 遷移式學習的特徵圖群

3.5 器官獨立模型

工業技術研究院所提供的植物資料集如前所述，有特別針對每張圖片進行有出現的器官去做標籤，既然在資料集中多出了這項資訊，我們嘗試使用多出的這項資訊去作方法上的改進與嘗試，因為在介紹資料集時提過的原因，實驗的資料主要為花、葉、果實的影像，希望能基於多出的資訊再次提高辨識率，於是提出了器官獨立模型，希望能訓練出專門針對各自器官的深度卷積模型，我們期望這些獨立的模型能在各自子資料集的領域中訓練測試的結果能比全資料集中的結果更好。

除此之外，我們仍然需要一個器官分類模型，在圖片影像輸入後，能分析出該影像應該偏重哪個模型，在前面的實驗中，我們得知在目標類別數目為 500 甚至 1,000 時，現今的深度卷積模型已經可以達到超過 80% 的辨識正確率，那分類器官的模型辨識率也應該會非常高，因為目標類別數目比之前要低上非常多，在本次實驗中，目標類別數目為 3 類，器官分類器的準確率很大的影響了整體的辨識率。

實驗流程

整體實驗流程如 Figure 3.23 所示，利用 Keras 訓練好權重的極端起源網路 (ImageNet 模型) 去做水平性的遷移式學習，得到水平遷移的極端網路模型，利用

此模型去分別訓練出器官分類模型、花分類模型、葉分類模型與果實分類模型，器官分類模型的輸出值為 $P_{flower}, P_{leaf}, P_{fruit}$ 分別代表此影像屬於花、葉、果實分別的機率，而花分類模型輸出值為 $P_{1-1}, P_{1-2}, \dots, P_{1-500}$ ，葉分類模型的輸出值為 $P_{2-1}, P_{2-2}, \dots, P_{2-500}$ ，果實輸出值為 $P_{3-1}, P_{3-2}, \dots, P_{3-500}$ ，其中 P_{x-y} 代表輸入影像屬於 y 類別的機率。

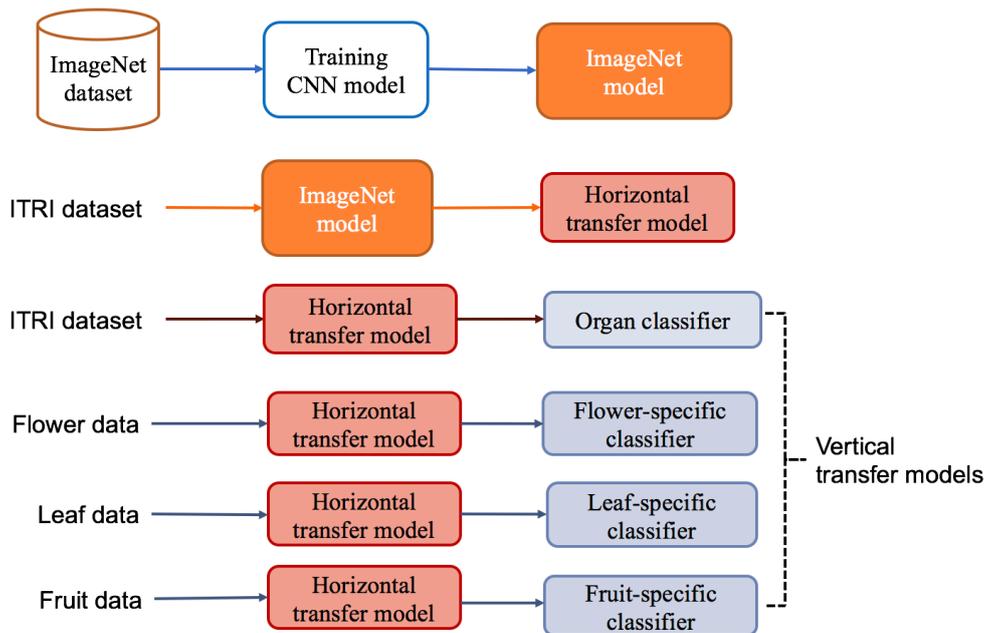


Figure 3.23: 器官獨立模型的流程圖

得到這些模型的預測值後，依照器官分類模型的輸出作為比重值，將各自器官的分類器 $P_{flower}, P_{leaf}, P_{fruit}$ 乘上器官分類器的權重之後相加如公式 (3.1)，得到的 Prediction 為一向度為 500 的數值陣列，取陣列中最高值為預測的植物種類。

$$\begin{aligned}
 Prediction &= P_{flower} * (P_{1-1}, P_{1-2}, \dots, P_{1-500}) \\
 &+ P_{leaf} * (P_{2-1}, P_{2-2}, \dots, P_{2-500}) \\
 &+ P_{fruit} * (P_{3-1}, P_{3-2}, \dots, P_{3-500})
 \end{aligned} \tag{3.1}$$

3.5.1 器官子模型訓練

子模型的訓練包括花分類模型、葉分類模型與果實分類模型，使用水平性遷移式學習訓練好的極端起源網路 (稱為水平模型) 進行垂直性遷移式學習，分別對花、葉、果實的子資料集重新訓練。首先，器官子模型對各自器官影像的辨識正



確率要高於原本的水平模型，如果各自器官的辨識能力沒有高於原本的水平模型，那麼器官獨立模型的方式將不可行。

Table 3.4: 原模型對各器官的辨識率

	花	葉	果實
辨識率	0.9015	0.7433	0.8025

各自的器官模型如果不使用垂直性遷移式學習，而是直接使用 ImageNet 模型進行水平性遷移式學習，每個器官模型的辨識率非常不理想，過度擬合的情形很嚴重，Figure 3.24為訓練的紀錄，Table 3.5為各自模型針對對應器官測試資料的辨識率，比對 Table 3.4，每一項的辨識率皆下降很多，只使用 ImageNet 資料集進行水平性遷移學習得到的效果非常差，無法作為器官獨立模型的子模型訓練。

Table 3.5: 無垂直性遷移式學習辨識率

	花	葉	果實
辨識率	0.7816	0.6829	0.7380

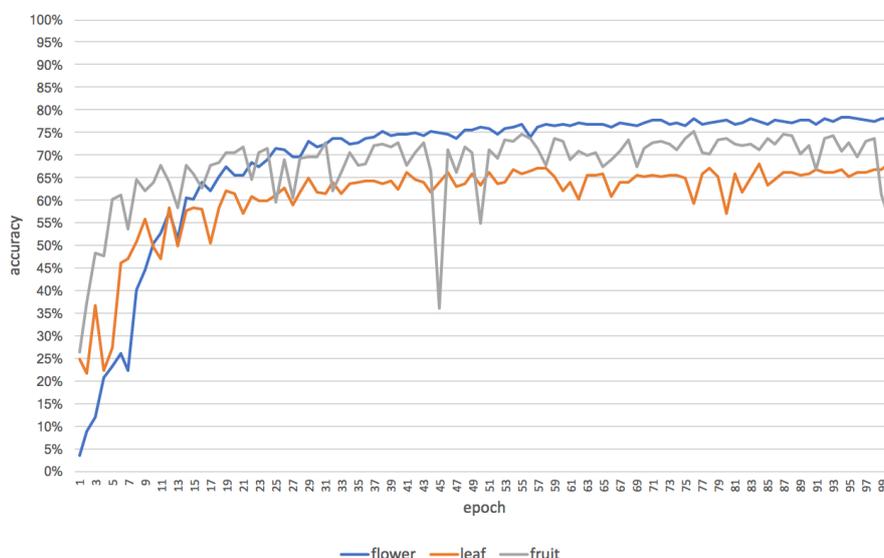


Figure 3.24: 各器官子模型不使用垂直性遷移式學習的 Top-1 正確率

嘗試使用垂直性遷移式學習，各自的器官子模型利用自己的子資料集對已經訓練完成的極端起源網路水平模型進行再訓練，因為子資料集原本就為水平模型的訓練資料，因此訓練模型的辨識率起始點就相當的高，是利用子資料集對水平模型進行微調訓練，訓練出各自的器官子模型，等於是使能辨識所有器官影像的水平模型專注於某一器官的辨識，使微調的過程中能提升辨識率，經過 100 次迭代

後，花分類模型最高辨識率為 90.65%，葉分類模型最高辨識率為 73.40%，而果實最高辨識率達到 80.92%，微調的訓練紀錄如 Figure 3.25。

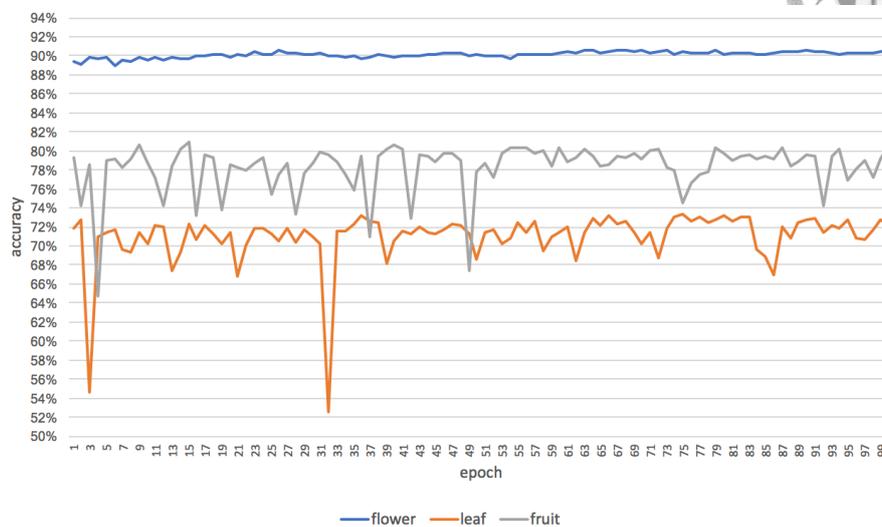


Figure 3.25: 各器官子模型垂直性遷移式學習的 Top-1 正確率

Table 3.6為器官子模型的綜合比較，使用垂直性遷移式學習的器官子模型在花與果實的子資料集中，辨識率分別些微上升了 0.5% 與 0.67%，上升幅度不小，然而葉的子資料集，經過垂直性遷移訓練後，最高辨識率反而下降了，可能為葉子的圖片類別間較其他子資料集相似許多，導致在微調訓練時，反而發生了過度擬合的情況。在器官子模型的選擇方面，花與果實的分類模型選擇垂直性遷移式模型，而葉分類模型經過垂直性遷移學習後辨識率反而下降，因此選擇原本的水平模型作為葉的子分類模型。

Table 3.6: 器官子模型綜合比較

	花辨識率	葉辨識率	果實辨識率
原水平模型	0.9015	0.7433	0.8025
無垂直性遷移	0.7816	0.6829	0.7380
垂直性遷移	0.9065	0.7340	0.8092

3.5.2 器官分類器訓練

器官分類器為器官獨立模型中非常重要的一環，但對於模型而言，也是相對較為簡單的一環，因為在這次的實驗中目標輸出的分類為 3 類，在辨識率方面是較為容易提升的，實驗部分為三種器官分類器的訓練方式，皆以 3 個神經元替換掉

極端起源網路最後一層的全連接層。第一種為 Type A：以不使用水平性遷移式學習訓練完成的極端起源網路權重值作為訓練起始點，嘗試訓練出器官分類器；第二種為 Type B：以經過水平遷移式學習的水平模型進行類垂直性遷移學習，嘗試訓練出器官分類器；第三中為 Type C：以 ImageNet 模型為基礎直接對 Top-500 器官資料集作水平性的遷移式學習，如 Figure 3.26所示。

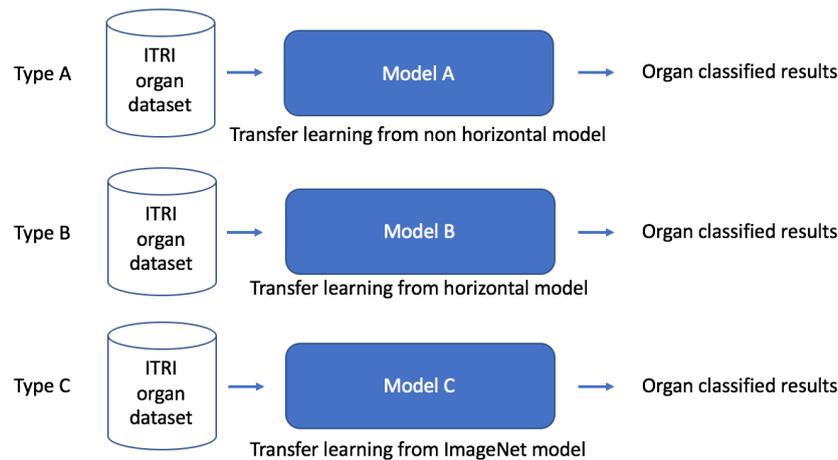


Figure 3.26: 各式器官分類模型訓練示意圖

Figure 3.27為三種訓練方式的訓練過程紀錄，可以發現 Type A 為辨識率最低的，其起始的模型沒有經過第一次的水平性遷移式學習，是 Table 3.3中的極端起源網路，辨識率在一開始就較低，即使經過了第二步驟的垂直性遷移式學習，其先前學到較差的卷積核依然限制了模型的發展，辨識率最高的為 Type B，經過第一步驟的水平性遷移式學習時，網路中的卷積核就學習到了良好的特徵權重，為 Table 3.3中的遷移式極端起源網路，起始辨識率就極高，有了這些特徵權重再經過垂直性遷移式學習，使目標分類為 500 類下降到 3 類，辨識率為最高，Type C 的辨識率為中間，比起從未經過水平性遷移學習的 Type A 要好上很多，但比起 Type B 經過兩步驟水平與垂直性的遷移式學習，依然要差了一截，Table 3.7為器官分類器的辨識率，最高為 92.89%，在這次的實驗中，水平性結合垂直性遷移式學習得到了最佳的結果。

Table 3.7: 器官分類器辨識率

	Type A	Type B	Type C
Accuracy	0.8903	0.9289	0.9163

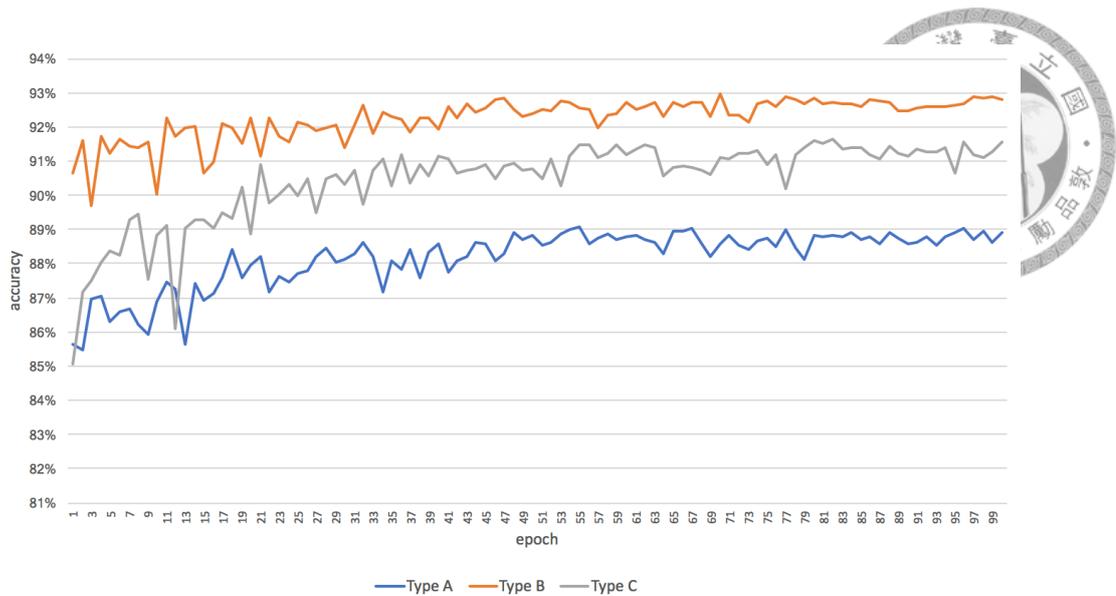


Figure 3.27: 各式器官分類模型訓練過程中的辨識率

3.5.3 整合器官獨立模型

經過上面的實驗得到了 4 個模型，分別是器官分類模型、花分類模型、葉分類模型與果實分類模型，在整合時，如 Figure 3.28 與公式 3.1 所示，輸入影像先各類器官子模型，得到子模型對於該影像的預測值，但我們仍不知道應該倚重哪一個子模型，所以影像再經過器官分類模型，得到 P_{flower} , P_{leaf} , P_{fruit} 三個機率值，分別為影像屬於花、葉、果實的機率，使用這些機率作為權重值乘上各自對應的子模型預測值，器官分類模型的輸出值作為權重可以使對應的子模型的輸出值升高，而其他較不相關的模型輸出值降低，藉此得出新的預測值，這種整合方式稱為方法 A (Method A)。

另外實驗將會嘗試不同的模型整合方法 B (Method B)，在方法 A 中，三個子模型對於預測的結果都會有或大或小的影響，主要只是在權重方面的差異；而方法 B 將會只使用一個子模型去預測，若經過器官分類模型得到的機率值最高為花，則影像只通過花的子模型去做預測，其他兩個模型將不被使用。兩種方法間的差異為絕對相信或相對相信器官分類模型的結果。

Table 3.8 為這次實驗之結果，因為比起原本只使用水平性遷移式學習的模型，器官獨立模型的方法又多訓練了 100 次迭代，為了實驗的公平性，針對水平模型進行了再訓練 100 次迭代，為表中的第二列模型。從第三列開始，列表中的數據為上述的 2 種整合方式 Method A、Method B 對應器官分類器的三種訓練方式

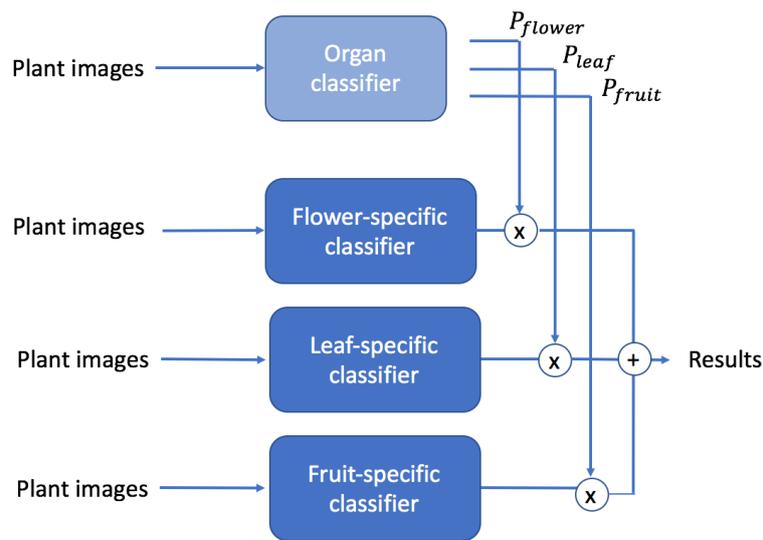


Figure 3.28: 整合模型

Type A、Type B、Type C 所作出的模型分類器的預測準確度。

其中，可以看出只使用水平遷移式學習再進行 100 次迭代的訓練，其結果與原本的模型辨識率相差很小，模型已趨近收斂，進步空間已經非常小，幾乎無法進步，而使用器官獨立模型的成果皆比原模型效果更好，方法 A 的整合方式結合辨識率最高的 Type B 器官分類器得到的辨識率最佳，比起原本的模型 Top-1 辨識率進步了 0.52%，Top-3 與 Top-5 的辨識率也都有些微上升，又所有對照組中，整合方法 A 比起整合方法 B 效果又更卓越，使用器官分類器輸出的機率值作為權重值分配給子模型作整合，由多個模型預測的辨識率較佳。綜合以上，證明器官獨立模型的方法在這個資料集是可以再度提升辨識率。

Table 3.8: 器官獨立模型辨識率

	Top-1 accuracy	Top-3 accuracy	Top-5 accuracy
Original model	0.8514	0.9273	0.9464
Re-train original model	0.8520	0.9277	0.9457
Method A - Type A	0.8545	0.9272	0.9475
Method B - Type A	0.8530	0.9260	0.9466
Method A - Type B	0.8564	0.9283	0.9477
Method B - Type B	0.8539	0.9265	0.9470
Method A - Type C	0.8548	0.9268	0.9473
Method B - Type C	0.8538	0.9263	0.9474



3.6 錯誤分析

植物影像辨識利用遷移式學習與器官獨立模型在 Top-1 的最佳辨識率為 85.64%，接下來我們將會進行分析輸入的植物影像為何會被分在錯誤的類別，在該錯誤的類別中是什麼樣的訓練資料使網路對植物影像判斷錯誤，接下來的實驗如 Figure 3.29 將會抽取卷積神經模型倒數第二層的輸出值作為特徵向量，使用歐幾里得距離，公式 3.2 去計算圖片的特徵向量在空間中的距離，取圖片與錯誤類別中最相近的兩張圖片作為相似圖片進行分析。

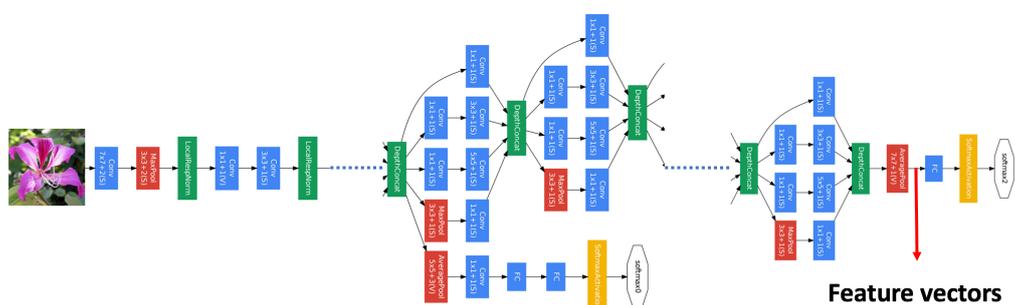


Figure 3.29: 抽取特徵向量示意圖²

$$Euclidean\ distance = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.2)$$

Figure 3.30 共有三種不同的植物類別，上排圖片為分類錯誤的植物影像，下排圖片為錯誤類別訓練資料中最相似的植物影像，這三類植物間訓練資料集中的花非常相似，即便是人類也很難分辨其中的差異，這幾個類別的測試資料在 Top-1 正確率時十分容易互相預測錯誤，這樣的分類錯誤在 Top-500 器官資料集中佔了不小的比例，因此雖然在 Top-1 時正確率為 85.64%，但 Top-3 正確率可以達到 92.83%，Top-5 正確率更是達到了 94.77%，在 Top-1 正確率時非常相似的幾類植物被錯誤地交互預測到彼此的類別，因為訓練資料真的非常相似，這類的預測錯誤是可以被接受的。

Figure 3.31 為錯誤預測的資料與對應的訓練資料十分相似的分類錯誤，這些圖片即使是人類在乍看之下也可能會分辨錯誤，但仔細看後可以發現兩張圖片還是有很大的不同，這類型的錯誤是網路沒有學出正確的特徵去分辨這些植物影像，

²部分圖片來自：<https://arxiv.org/pdf/1409.4842.pdf>

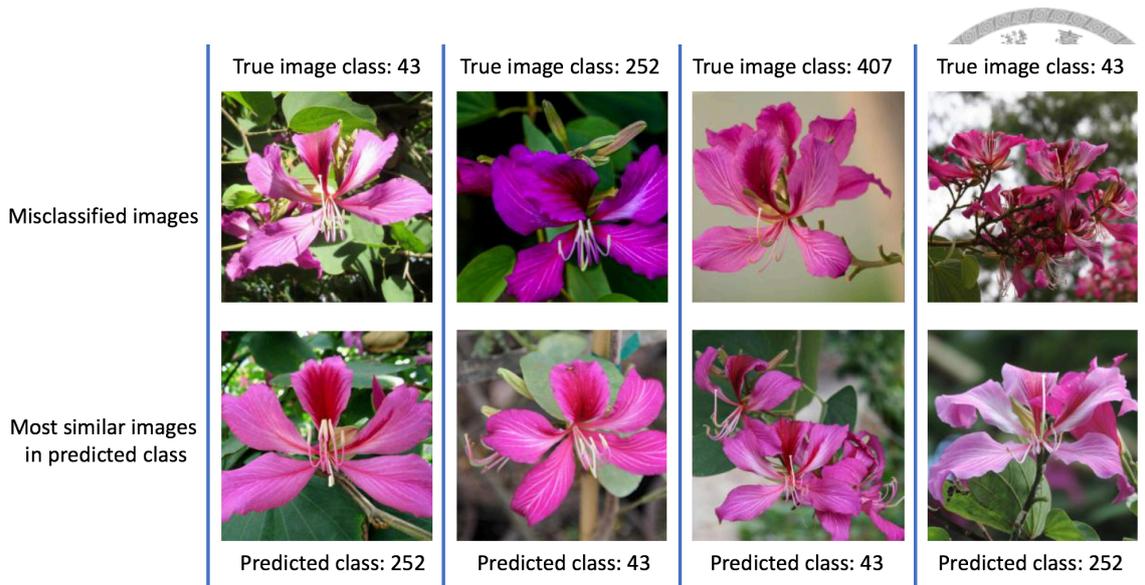


Figure 3.30: 類別間相似所造成的分類錯誤

每一組圖片如果用局部微觀地來看，其實小地方的特徵非常地相似，因此卷積神經模型很難分辨出彼此間的植物特徵差異，從而認為兩者為相似的影像，這類型的預測錯誤是希望可以避免的。



Figure 3.31: 影像相似的分類錯誤

Figure 3.32為測試資料本身就較為特殊的分類錯誤，有些植物影像甚至不為正常的植物照片，而是由多個植物圖片剪貼而成，不論是在訓練資料或測試資料中，這類型的資料都非常稀少，因此卷積神經網路對於此類型的影像圖片非常難以分類成功，或許這些測試圖片在某些色塊組成或小細節上被網路模型認為與對應的訓練資料相對相似，但對人類來說，組圖間並不屬於相似的範圍，這類型的預測

錯誤很難以被避免，可能要直接對資料集再次整理，過濾一些過於特殊的影像。

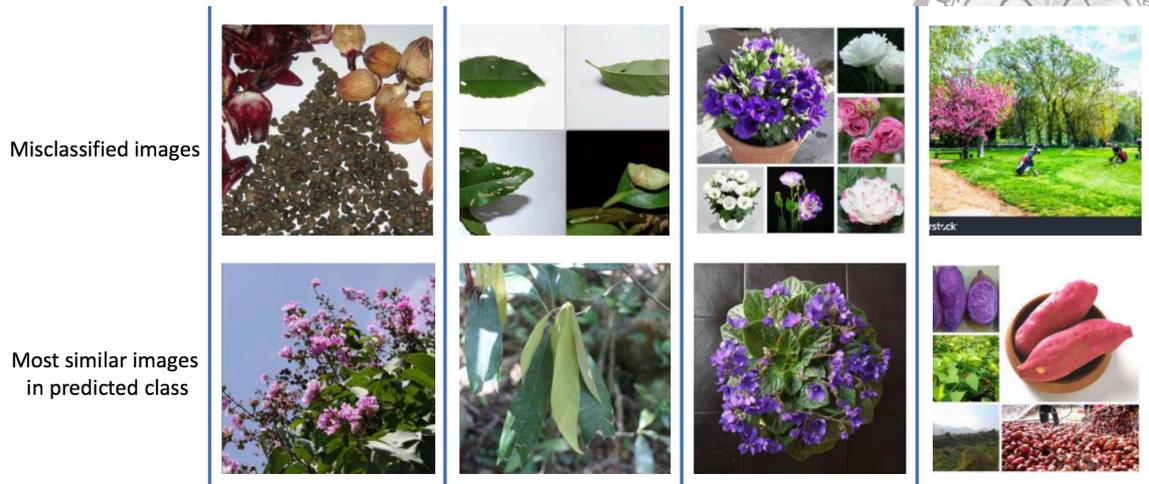


Figure 3.32: 影像不相似的分類錯誤

收集與整理如此龐大的資料集非常困難，非常少部分的影像資料可能在標籤的時候會被重複標籤到多個類別中，如 Figure 3.33 中，進行測試時，發現測試資料分配到錯誤的類別，而在對應的訓練資料中也出現了一樣的植物影像，因此分類錯誤無可厚非，這樣的分類錯誤較影響 Top-1 正確率，但對 Top-3、Top-5 正確率而言卻是不太影響，因為重複標籤的類別彼此間也十分相似，模型輸出的前幾高的機率值幾乎包括了所有的相似類別。為了提高 Top-1 正確率，這類的錯誤必須由植物專家解決，使植物影像能確實的只有一個標籤，否則很難避免分類錯誤。

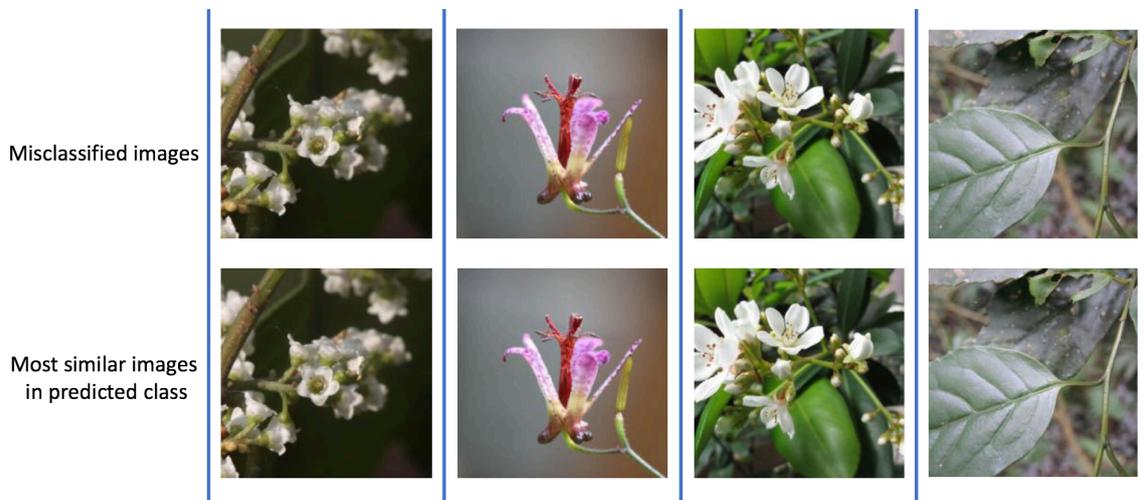


Figure 3.33: 影像重複所造成的分類錯誤



Chapter 4

結論與未來展望

4.1 結論

植物影像辨識在現實世界能幫助許多人用更快速且便捷的方式瞭解植物，幫助對植物有興趣的孩童或大眾快速的查詢植物類別，在野外能依照相片快速地查詢植物的特性。要做到這些事情，辨識率至關重要，在本次的實驗中，我們著重在辨識率的提升，越正確的分類出植物種類，才能更正確的提供植物的特性供科普或興趣使用。

4.1.1 辨識率

Google 公司研發的起源網路、起源殘差網路、極端起源網路是現今眾多深度卷積網路模型中辨識率首屈一指的，單純地使用這些網路去做訓練，加上一些資料增強與參數調整後，起源殘差網路得到的 Top-1 辨識率有 77.26%，Top-5 辨識率有接近 90%，在實用上已然相當不錯，接下來實驗了使用水平性遷移式學習對這些網路做訓練，表現最好的網路為極端起源網路，Top-1 辨識率達到了 85.14%，Top-5 辨識率有 94.64%，這段實驗說明了使用良好訓練過的模型進行水平性遷移式學習對於辨識率的影響有正向的幫助，論文中使用的三種模型的辨識率皆有大幅度的上升。

最後論文中提出了器官獨立模型，這樣的模型雖然對於資料要求較高，在現實中應用可能較為侷限，但在辨識率方面又有些微的提升，使用極端起源網



路所組成的器官獨立模型組合，再經過垂直性遷移式學習的再訓練後，得到的 Top-1 辨識率為 85.64%，Top-5 辨識率達到了 94.77%，使用器官獨立模型應用在這 Top-500 器官資料集時，得到的辨識率為最佳，若純粹以辨識率評斷方法的好壞，那麼利用遷移式學習與器官獨立模型進行植物影像辨識能夠得到非常好的效果。

4.1.2 實際應用

在學術上而言，辨識率的高低為最重要的評斷因素之一，然而辨識率的提升往往牽涉到了許多其他層面，在實際應用上，時間與硬體設備的需求反而可能是主要的評斷因素，考慮到 Table 3.1 中的參數數量，起源網路與極端起源網路為參數數量最低的兩種高辨識率的網路模型，而起源殘差網路的參數數量則為兩者的兩倍，是訓練所需的硬體設備需求最高的網路模型，然而假設在沒有良好訓練過的模型去進行遷移式學習的情況下，單純訓練模型時起源殘差網路能達到最高的辨識率。

在有良好訓練的網路模型進行遷移式學習時，極端起源網路成為了最佳的網路模型應用，在三種模型中，其辨識率最高且佔用的記憶體容量最小，在實際應用上是最佳的選擇，而器官獨立模型在實際應用上需要高強度的硬體支持，因為應用時需要四個模型同時運作，雖然實驗時辨識率能達到最高，但需要消耗將近四倍於單一極端起源網路的運算量，在硬體能夠同時運算四個器官網路輸出時，才能在實際速度上與單一極端網路相近。

4.2 未來展望

這篇論文實驗中使用的資料集為 Top-500 器官資料集，使用到的資料為資料量前 500 多種類的集合，而工業技術研究院所提供的資料超過 2,400 類，未來在硬體許可的情況下，要嘗試訓練 2,400 類的資料訓練，且植物器官在現實中也不只是花、葉、果實，還有根、莖或其他器官，在資料量不足的情況下，只能暫時以花、葉、果實三種器官進行器官獨立模型訓練，若未來其他器官的資料也有一定幅度的增加，那麼器官獨立模型在實際應用上將也會有更好的結果。



起源網路、起源殘差網路、極端起源網路為現今最好的幾種網路模型，受惠於這些模型的高辨識率，器官獨立模型才能有這麼高的辨識率，然而卷積神經網路的相關研究仍然持續不斷的更新，2017年NASnet [16] 帶來了震撼的研究結果，利用深度神經網路自行學習出最適合該資料集的卷積網路架構是個很新的領域，在Keras釋出模型程式碼後，有稍微的測試了一下此種方法應用於此次資料集的結果，然而其硬體的要求非常高，在Nvidia GeForce GTX 1080Ti上進行一次迭代的訓練時間就需要將近2個小時，漫長的等待結果得到的辨識率也沒有很理想，但畢竟這是一個很新的領域，未來如果在演算法或其他方面又有所改進，很有可能再次衝擊電腦視覺的這塊研究領域，對於植物影像的辨識也會是一個新的希望。



參考文獻

- [1] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
- [2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [3] Sue Han Lee, Chee Seng Chan, Paul Wilkin, and Paolo Remagnino. Deep-plant: Plant identification with convolutional neural networks. *CoRR*, abs/1506.08425, 2015.
- [4] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015.
- [5] Christian Szegedy, Sergey Ioffe, and Vincent Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. *CoRR*, abs/1602.07261, 2016.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.
- [7] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.

- 
- [8] François Chollet. Xception: Deep learning with depthwise separable convolutions. *CoRR*, abs/1610.02357, 2016.
- [9] Laurent Sifre and Stéphane Mallat. Rigid-motion scattering for texture classification. *CoRR*, abs/1403.1687, 2014.
- [10] François Chollet et al. Keras. <https://keras.io>, 2015.
- [11] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [12] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
- [13] Pavel Golik, Patrick Doetsch, and Hermann Ney. Cross-entropy vs. squared error training: a theoretical and experimental comparison.
- [14] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *CoRR*, abs/1802.03601, 2018.
- [15] Jindong Wang et al. Everything about transfer learning and domain adaptation. <http://transferlearning.xyz>.
- [16] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. Learning transferable architectures for scalable image recognition. *CoRR*, abs/1707.07012, 2017.