國立臺北科技大學

資訊與財金管理系碩士班

碩士學位論文

利用深度學習進行人臉情緒辨識之研究

A Study on Facial Emotion Recognition by Using Deep Learning

研究生：蔡雯惠

指導教授：翁頌舜 博士

中華民國一百零九年六月

摘要

論文名稱：利用深度學習進行人臉情緒辨識之研究

頁　　數：四十九頁

校 所 別：國立臺北科技大學 資訊與財金管理系

畢業時間：一百零八學年度 第二學期

學　　位：碩士

研 究 生：蔡雯惠

指導教授：翁頌舜 博士

關鍵字：人臉情緒辨識、深度學習、卷積神經網路

在生活水平、醫療技術及公共衛生觀念提升等因素影響下，台灣國人平均壽命呈上升趨勢，老年照護也是一個越來越重視的議題，現今有很多機器人輔助人們的日常生活，由於機器人需要與人互動，為了要更人性化，情緒的感測是非常重要的，若能偵測到人們的情緒，就能給予友善的回應與安慰。

本研究透過預訓練卷積神經網路模型，訓練其辨識情緒，訓練完後，再輸入真實圖片，進行人臉偵測，偵測出照片中人臉的位置後，再去分類其情緒，並將該圖片預測的情緒結果回傳顯示在圖片上。本研究使用資料集之作者 Goodfellow et al. (2015) 指出，人類在該資料集的辨識準確率大約是65%，而本研究之模型準確率是69.3%，高於此準確率，且模型在快樂、驚訝的表情準確率達80%，足見本研究之成果良好。

ABSTRACT

Title: A study on Facial Emotion Recognition by Using Deep Learning

Pages: 49

School: National Taipei University of Technology

Department: Information and Finance Management

Time: June, 2020

Degree: Master

Researcher: Wen-Huei Tsai

Advisor: Sung-Shun Weng, Ph.D.

Keywords: Facial Emotion Recognition, Deep Learning, Convolutional Neural Network

Under the influence of living standards, medical technology and the improvement of public health concepts, the Taiwanese people's average life is rising. Elderly care is also an important issue. Nowadays, there are many robots to assist people's daily life. In order to interact with people more humanized and friendly, emotional recognition is very important. If robot or machine can detect human emotions, it can give human a friendly and comfortable feeling during interaction.

Therefore, in this study, we train convolutional neural network model to recognize emotion of human face. After training, we input the image, face detection is performed first, then crop the face and put it into model to analyze its emotions. At the end, we return the emotion result on the original picture. Goodfellow et al. (2015), the authors

of the Fer2013 dataset, described that the accuracy rate of human's classification is about 65%, and the accuracy of our model is higher than this accuracy rate. Moreover, the model's accuracy rate on recognizing happy and surprised emotion is nearly 80%. It shows that the results of our study are good.

誌謝

能完成碩士學位，首先要感謝的是我的家人，謝謝爸爸媽媽讓我能衣食無憂，專心於課業上的學習，謝謝哥哥、嫂嫂、姐姐與姊夫總是帶著我亂吃亂喝亂走跳，謝謝可愛的兩位姪子，在我回家時一直黏著我，讓我能保持著純真的心。

很榮幸能有翁頌舜教授當我的指導老師，因為有老師細心的指導與溫暖的帶領，論文才能順利完成，謝謝老師時常提醒我的論文進度，讓有拖延症候群的我能如期完成論文；感謝口試委員楊欣哲老師與鄭麗珍老師，因為有你們的建議，給了我許多不同的觀點，讓我的論文更詳盡全面；謝謝所有教導過我的老師們，給予我許多靈感與知識基礎，啟發我做這個論文主題。

謝謝我的碩班朋友袁嘉妮、陳儀真與黃彥翔，跟我一起度過這兩年的碩班時光，陪我一起哭一起笑，有你們陪伴，真的很幸運；也謝謝碩班的同學與學弟妹們，讓我亂串門子、餵食我各種食物與叮嚀我記得各種被我遺忘的事情與物品；謝謝跟我一起住過的眾多室友們，在半夜聽著我敲打的鍵盤聲入眠、總是帶家鄉名產給我吃跟我分享自己國家的人文風情。

謝謝我的大學朋友游念柔與陳雅倚，就算工作繁忙還是維持著每個月都會相聚一次，講著從前從前，每次聊天就像市場大媽一樣吵到不行；也謝謝前男友讓我更成熟更加懂得替別人著想，謝謝你體諒我的選擇；最後，謝謝所有在我成長過程中幫助過我的貴人與緣分們，願自己能夠一直保持善良並做一個懂得感恩與祝福的人。

蔡雯惠 謹誌

2020年7月

目 錄

圖目錄

表目錄

第一章　緒論

人們會透過臉部表情、聲音、語言、肢體動作等有意或無意的表露出自己的內心情感、想法，進而去建立彼此間良好友善的關係，在這麼多表達方式中，臉部表情是我們在表達情感時最直觀、自然的信號(Tian et al., 2001)，我們能夠透過表情表達自己與識別他人的情緒。

1.1 研究背景與動機

在生活水平、醫療技術及公共衛生觀念提升等因素影響下，台灣國人平均壽命呈上升趨勢，在人口結構變化上，在前年邁入高齡化社會（高齡人口比例超過14%），預計 6 年後就會邁入超高齡社會(高齡人口超過 20%)，如圖 1.1 所示；其實不光是台灣，由於戰後嬰兒潮世代（1946 到 1964 年出生者）已經在 2011 年陸續達到 65 歲，因此未來二、三十年間，他們都將邁入高齡或超高齡，會有行動不便、慢性疾病與長者照護等問題浮現而出，因此全球長期照護需求將會巨額增長。

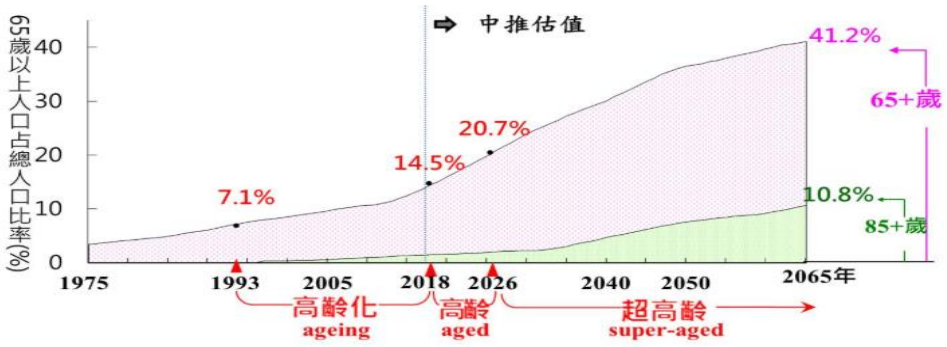


圖 1.1 高齡化時程[1]

資料來源：國家發展委員會「中華民國人口推估（2018 至 2065 年）」

[1] 高齡化時程，https://www.ndc.gov.tw/Content_List.aspx?n=695E69E28C6AC7F3

與不斷增加的醫療照護需求相反，我們的醫護人員人力不足，世界衛生組織提出，到 2035 年，全球將會缺少 1,290 萬的醫療工作者，年長的醫療工作者陸續退休後，醫療照護人力將面臨嚴重缺乏，而 2015 年台灣約每 6 位青壯年可以共同扶養一位老人，但在 2025 年將降到每 3.4 位（國家發展委員會人口推估[2]，2018），因此青壯年的扶老負擔也會越發沉重，為了因應這些變化，目前有許多類型的機器人投入來支援人力缺口，協助實際的工業流程、醫療行為、日常生活照應監測、關懷陪伴、居家照護等。

其實除了長者照護議題，現在許多家庭都是雙薪家庭，父母常常在外工作、加班，能陪伴小孩的時間也比較少，因此許多孩子在成長過程中，常常是孤身一人的玩耍、讀書，因應這種現象，多家廠牌的陪伴型機器人在這幾年陸續出現，若在陪伴小孩時，也能偵測到小孩的情緒，那麼在與小孩互動的過程中，想必能更加有趣與可愛。

由於機器人需要與人們互動，為了要更人性化，情緒的感測是非常重要的，若能偵測到當下對方的情緒，就能給與友善的回應與安慰，使其不再只是擁有冰冷外殼，而是能夠更加貼近人心、給予陪伴溫暖的科技。

人臉偵測、人臉辨識與人臉情緒辨識(Facial Expression Recognition)等相關技術是計算機視覺處理領域中很活躍的研究話題，其也有些應用，像是駕駛疲勞監測、影像監控，可以提升駕駛員的安全性，監測可疑人物。而在商業方面，可以幫助公司進行市場反應分析和影片廣告監測，透過瞭解人們的情緒資訊，使機器在與人類互動的過程中，能更加細膩的協助人們的日常生活，帶給人們許多便利性。

在情感分類上，在 20 世紀，Ekman et al. (1987)在跨文化研究的基礎上定義了六種基本情感，分別是憤怒(Anger)，厭惡(Disgust)，恐懼(Fear)，快樂(Happiness)，悲傷(Sadness)和驚訝(Surprise)，此研究表明無論文化差異如何，人類都以相同的

[2] 扶老比，https://www.ndc.gov.tw/Content_List.aspx?n=84223C65B6F94D72

方式感表達某些基本情感，代表這些情感都有普遍性與跨文化性，除了這六類表情，實際生活中也有心情比較平靜、起伏不大的時候，因此本研究也將中性(Neutral)表情納入所要辨識的情緒之一，故總共七類表情。

1.2 研究目的

情緒辨識是一個很重要的技術，若是能達到一定的辨識率，相信對於機器人應用層面也有一定的貢獻，故本研究期望能實作出能檢測人臉情緒的介面，透過影像輸入，先進行人臉偵測，偵測出照片中人臉的位置後，再進一步用卷積神經網路(Convolutional Neural Networks, CNN)去分析其情緒，其情緒分類分成七類：憤怒、厭惡、恐懼、快樂、悲傷、驚訝和中性，並將結果顯示在圖片上，研究目的說明如下：

1.使用深度學習(Deep Learning, DL)提取特徵與分類，提升模型性能。

過去在人臉情緒辨識上常用手工提取特徵再搭配 SVM 等分類器進行分類，但自從深度學習被應用在其他領域且辨識率多有提升後，也可在情緒辨識上運用深度學習，來提升模型性能。

2.探討情緒辨識技術，並將其技術應用在真實照片上。

本研究探討多種 CNN 模型，並將訓練後的模型應用在實際圖片上，並分析其效果。

1.3 研究架構

第一章緒論，包含研究背景與動機、研究目的、研究架構等三個節次，用以介紹本文的研究方向。第二章文獻探討，包含辨識應用、人臉偵測、預處理、情緒辨識等四個節次，用以介紹本文所套用的技術其背後理論內容與相關的文獻研究。第三章研究方法，包含研究架構、研究方法等兩個節次。第四章實驗結果與

分析。第五章結論與建議，包含研究結論、研究建議，用以整合本文分析結果來回應研究目的並提出研究建議，研究流程如圖 1.2。

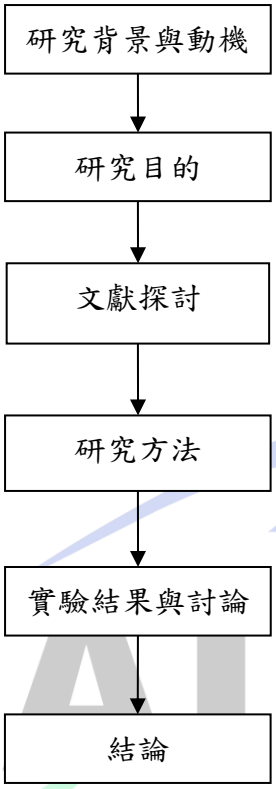


圖 1.2 研究流程

第二章　文獻探討

以下先介紹本研究情緒辨識方法的相關文獻，再去探討有關於人臉偵測與預處理相關文獻，最後再連結辨識的相關應用。

2.1 情緒辨識運用方法

在辨識臉部情緒時，演算法有分為統計相關方法跟深度學習，以下將介紹過去文獻中是如何利用這兩種算法去辨識情緒。

2.1.1　統計相關方法

常見情緒辨識流程是人臉偵測、特徵提取最後分類，特徵提取方面，有些研究會因為特徵提取後有維度過高的問題，必須先降維後，再分類（劉曉、譚華春、章毓晉，2006）；在特徵提取方面大多使用手工提取，主要可分為兩類：基於外觀特徵和基於幾何特徵；基於幾何的特徵描述了面部的形狀及其組成部分（例如嘴或眉毛），而基於外觀的特徵則描述了由表情引起的面部紋理變化；最後情緒分類，則常使用 SVM, AdaBoost 等進行分類。

幾何特徵上，Ghimire & Lee (2013)根據 52 個面部界標點的位置和角度使用了兩種類型的幾何特徵。首先，計算一幀內每界標之間的角度和歐幾里得距離(Euclidean Distance)，其次，從視頻序列的第一幀中的相應距離和角度中減去該距離和角度。在分類器上，使用具有動態時間扭曲的多類 AdaBoost，或者對增強特徵向量使用 SVM。

外貌特徵方面，Happy et al. (2012)將面部圖像分成幾小塊並以局部二進制模式(Local Binary Patterns, LBP)直方圖作為特徵向量，將這些特徵連結起來得到該

圖像的特徵，最後使用主成分分析(Principal components analysis, PCA)對各種面部表情進行分類。

就如同劉曉等人(2006)所說，在傳統方法中有很多模型需要手工介入來標記區域，雖然外貌特徵不用手工標記，但其提取的信息往往不夠可靠且易受干擾，因此也許可以考慮用機器學習的方式去提取特徵與分類表情。

2.1.2 深度學習

隨著晶片處理能力大幅提升(Graphics Processing Unit, GPU)以及各種神經網路結構不斷地湧現，許多不同領域的研究、應用開始使用深度學習方法，而深度學習也確實大幅提升了辨識的準確率，以圖像辨識來看，CNN 擁有良好的辨識率，其模型架構組成主要由卷積層(Convolutional Layer)、池化層(Pooling Layer)與全連接層(Fully Connected Layer)等核心層組合而成。

卷積層主要是對輸入的圖像提取特徵，將一個 N*N 大小的卷積核(Kernel, Filter)參數，透過滑動視窗的方式，在輸入圖像上，由上而下、由左到右掃描過去，此動作就像捲毛巾一樣，因此稱作卷積，最後會生成特徵圖(Feature Map)，操作如圖 2.1 所示。

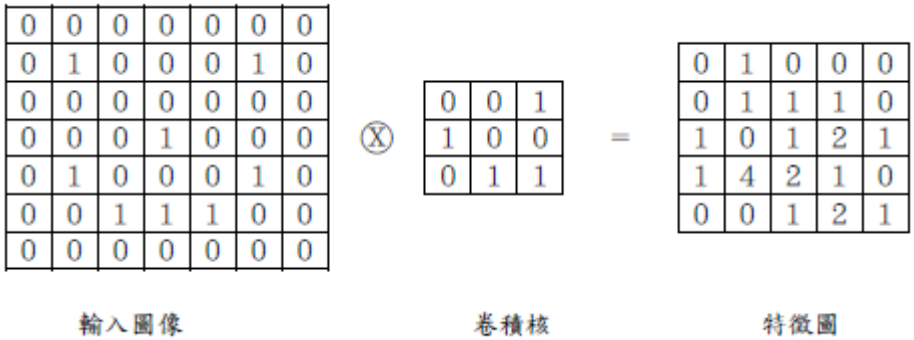


圖 2.1 卷積層示意圖

資料來源：本研究整理

池化層通常接在卷積層後使用，主要功用是壓縮圖片、保留重要資訊且能減少參數數量，常見的算法有最大池化(Max Pooling)與平均池化(Average Pooling)，池化層也像卷積層一樣有一個滑動視窗，但它裡面沒有參數，在視窗內依照算法不同，將卷積層出來的原特徵圖中取出值最大或將值平均，最後生成的特徵圖則保有最顯著的特徵，操作如圖 2.2 所示。

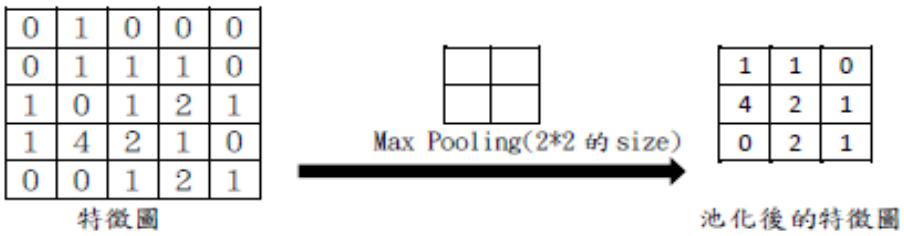


圖 2.2 池化層示意圖

資料來源：本研究整理

全連接層會把從卷積層、池化層中產生的特徵組合們，從二維的空間中轉換成一維的清單，並且透過那些特徵組合來去做分類，輸出各分類的可能性值作最後結果，其操作如圖 2.3 所示。

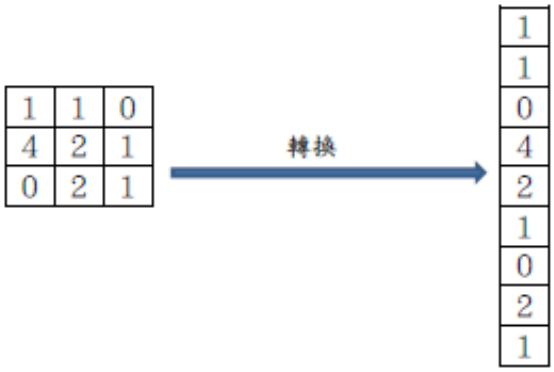


圖 2.3 全連接層示意圖

資料來源：本研究整理

在臉部情緒辨識上，有用自己從頭開始訓練的模型去辨識、有拿知名的預訓練模型、也有把經過手工特徵後的圖像當作輸入輸進模型訓練等各種各樣的方法，來嘗試讓辨識率能夠提高。

Ramdhani et al. (2018) 用自行創建的兩種 CNN 模型（層數分別為 9 與 10 層）與不同的批次大小(Batch Size)來進行識別高興，失望，生氣和自然，最後依照不同的方法得到的準確率在 FER2013 數據集中大概落在 60%~70%。

Akash et al. (2019) 提出提出的模型由六個卷積層，兩個最大池化層和兩個全連接層組成，在調整各種超參數後，該模型在 FER2013 數據集中的最終準確度為 60%。

Ng et al. (2015) 透過 ImageNet 資料集上進行預訓練的網路當作初始化，並提出兩階段微調，第一階段用 FER2013 數據集微調，第二階段再用真正要測的目標數據集來做調整，使訓練模型更加貼近目標數據集。

Ding et al. (2017) 認為雖然在大型資料集上進行預訓練的網路再微調的策略表現良好，但是仍有一些問題，像是調整過後的情緒辨識模型可能仍然包含對人臉識別的資訊，進而削弱了網路表示不同表情的能力，此外為人臉識別領域設計的網路通常對於情緒辨識任務而言太大，過擬合問題仍然很嚴重，因此他們提出在預訓練階段，只訓練卷積層，並對其進行正規化；在提煉階段，將全連接層附加到預訓練的卷積層上，並共同訓練整個網路。

Levi & Hassner (2015) 提出與將 RGB 圖像當作模型輸入的大部分研究不同的新穎想法，就是把圖像轉換為 LBP 並將這些值映射到 3D 度量空間來做為模型的輸入。

2.2 預處理(Pre-processing)

由於一張圖片中會含有人臉與非人臉的背景部分，再加上光線、頭部姿勢等因素，容易影響後續模型的特徵提取與分類效果，因此常在進行模型訓練前，會

先進行預處理，以下介紹本研究有使用到的預處理步驟，主要有資料擴充(Data Augmentation)和特徵正規化(Feature Normalization)等處理方法。

（一）資料擴充：Shorten & Khoshgoftaar (2019)整理了有關於在深度學習中資料擴充的背景、相關技術、設計時需考慮的地方等內容，他們提到因為有限的資料量、模型泛化能力不佳導致的過擬合(Overfitting)、圖片因角度、光照、背景等原因導致模型誤判為不同類別與不平衡的數據等現象，資料擴充可以是其中一個解決方案，透過增加樣本的多樣性，降低上面對模型的影響，進而提升模型性能，降低過擬合。

資料擴充方法有分為幾類基於圖像的處理與基於深度學習的處理，基於圖像的有隨機剪裁(Random Cropping)、旋轉(Rotation)、翻轉(Flipping)、隨機擦除（Random Erasing）、色彩空間改變(Color Space Transformation)及加入噪聲(Noise Injection)等，基於深度學習的有生成式對抗網路(Generative Adversarial Network, GAN)、神經風格轉換(Neural Style Tranfer)等，藉由上述的方法來產生各式各樣的資料來擴充。

Porusniuc et al. (2019) 在把 Fer2013 資料集的資料丟到自行創建及 ResNet50 兩模型前，除了做直方圖均衡，增強圖片對比度外，還有有將圖像做隨機翻轉、移動、旋轉等資料擴充步驟，增加資料多樣性。

（二）特徵正規化：Singh & Singh (2019)分析了在 21 種資料集下，各種特徵正規化方法、特徵選擇(Feature Selection)跟特徵權重(Feature Weighting)對其分類效能的影響，他們提到經特徵正規化後可以降低資料集中數值較大的特徵點對於模型權重的影響，分類出來的結果就比較不會偏向數值較大的特徵的結果，且特徵正規化對於建構機器學習模型，像人工神經網路(ANN)、支援向量機等的準確性有很大的重要性；他們實驗結果發現 z 分數(z-score)跟 Pareto Scaling 在特徵集與特徵選擇上表現最好，表現最差的是均值集中(Mean Centered)、中位數(Median)、中位數絕對偏差(Median Absolute Deviation)跟未標準化(Un-Normalized)數，他們提到每個資料集可能具有的不同特性，所以特徵正規化具有主觀而非客

觀的性質，但可以參考他們實驗得出的結果從各式各樣的正規化方法中找到適合該模型的方法。

以下介紹本研究所使用的正規化方法最大正規化(Max Normalization)，其公式如 2.1 所示。

$$X'_{i,n} = \frac{X_{i,n}}{\max(|X_i|)} \tag{2.1}$$

$X_{i,n}$：尚未正規化的第n筆資料，其中i表示為該資料特徵點。

$X'_{i,n}$：正規化後的第n筆資料，其中i表示為該資料特徵點。

$\max(|X_i|)$： 取i個特徵點中最大的那個特徵點值。

由於本研究是分析圖片，所以資料特徵點就是圖片的每個像素值，而最大正規化就是該圖片的每個像素值除以該圖片中最大像素值，該類方法可以保存原始輸入資料之間的關係，但易受原始資料的異常值或極值的影響，而圖片的最大像素值就是 255，因此受異常值或極值的影響非常小。

2.3 人臉偵測(Face Detection)

不管是人臉情緒辨識還是人臉辨識，在一張圖像中，最基本的還是要找到那張臉，如果找不到那張臉或把物品、景物辨識成人臉，那後面的辨識就會錯的一塌糊塗，因此人臉偵測是在辨識前一個很基本卻也很重要的步驟；在人臉偵測上有幾個常用的用法，如下所述：

（一）Viola & Jones (2004)提出實時的人臉檢測，其作法首先使用 Haar 特徵(Haar-like Features)來去提取特徵，也就是在視窗的某個位置取一個矩形框，並將該框分為白色與黑色兩個部分，如圖 2.4，用白色部分像素點灰度值的和減去黑色部分像素點灰度值的和，得到一個特徵值，再用積分圖(Integral Image)去加快其像素計算部分，分類器則採用自適應增強(Adaptive Boosting, AdaBoost)並將多

個分類器級聯(Cascade)在一起，由排在最前面也最簡單的分類器對其進行分類，如果這個視窗被分為非人臉視窗，那麼就將此視窗排除，不會送到後面的分類器分類，經過層層的篩檢後，需要判別的視窗就少很多，因此只要付出很少的計算代價就能夠排除大部分非人臉窗口，該方法 OpenCV 套件已經寫好程式碼，因此可以直接套用。

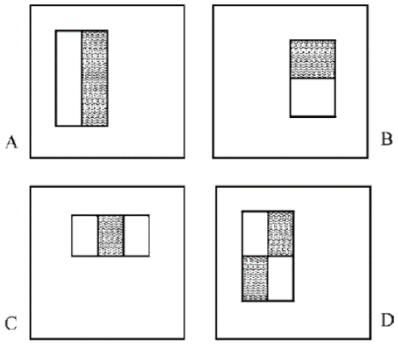


圖 2.4 Haar 特徵

資料來源：Viola & Jones (2004)

（二）Dalal & Triggs (2005)提出使用方向梯度直方圖(Histogram of Oriented Gradients，HOG)加上支援向量機(Support Vector Machine, SVM)來做行人偵測，後來很多研究就以他們的這個方法來做人臉偵測、目標偵測等應用，HOG 是一種局部特徵描述子(Descriptor)，能夠描述人體的邊緣且對偏移和光照變化不敏感，其作法是先計算圖像每個像素的梯度（包括大小和方向）值，主要是為了捕捉目標的輪廓資訊，再將圖像分成一個個小的連通區域，叫細胞單元(Cell)，然後計算每個 Cell 的梯度直方圖，最後把這些直方圖組合起來就可以構成該 Cell 的特徵描述子，把多個 Cell 的特徵描述子組合起來變成一個大區域(Block)的特徵，最後多個 Block 組合起來，就變成該圖像的特徵描述子，並當作 SVM 分類的特徵向量，該方法在 Dlib 上也有寫好的程式碼可以進行套用。

（三）Zhang et al. (2016)提出一個基於 CNN 級聯框架，他將三個 CNN 聯合在一起去進行人臉偵測，首先第一個是架構比較簡單的淺層 CNN，主要從圖像

中快速產生候選人臉視窗並過濾掉高度重疊的候選人臉視窗，第二個是較複雜的 CNN，主要是排除掉從上一個 CNN 篩選出的候選窗中大量非人臉視窗，最後一個是最複雜的 CNN 模型，根據第二個輸出結果再進一步作改善，並輸出人臉關鍵點位置(Facial Landmark Localization)。

2.4 辨識相關應用

過去在辨識技術上，因視角變異及辨識網路過於單調導致偵測效能低落、準確率無法提升，再加上數據集也是以實驗室環境為主，光線、陰影等外在因素也會影響辨識的效能，現如今因神經網路的發展，再加上大量標記好的各種不同姿態、表情、光線等資料集，能訓練出一個用於真實場景的偵測模型，因此也產生出很多應用，如下所述。

在學生學習情緒上，吳政德(2017)在學生觀看學習影片時利用網路攝影機錄下觀看影片時的臉部表情，並將影片切成三秒、一秒、半秒三種形式去辨識學生的學習情緒，此外法國 ESG 商學院[3]的兩個線上課程也有用攝影機追蹤眼球移動和臉部表情，對其進行分析，來衡量學生的注意力集中程度和課堂參與度。

在門禁系統上，傳統的基於 RFID 的訪問控制系統只能通過 RFID 卡識別員工，所以不管你是不是該公司的員工，只要持有該公司員工已註冊 RFID 卡都能通過身份驗證，直接進入公司，因此 Chang & Liao (2015)利用人臉辨識結合 RFID 來檢查誰有權進入工廠，且所有操作都隨時間記錄，使人力資源經理可以檢查資料庫中的記錄。

在影像監控、出勤管理上，Harikrishnan et al. (2019)於實驗室和教室中進行實時監視，會辨識有修該堂課的學生人臉，當你因翹課而沒被辨識到時，則不會有出勤紀錄，最後會將出勤記錄傳送到後端的 Excel 保存結果。

[3] 上課分心逃不掉！臉部追蹤技術揪出學生「恍神時刻」，幫助老師改進課程，https://www.inside.com.tw/article/9451-new-facial-recognition-to-point-out-when-do-students-not-paying-attention

在機器人照護上，工研院李國徵等人(2018)開發出一套機器人可以去監測長者的睡眠狀況、平常的活動力、日常三餐菜色及語音情緒等，當子女日常工作繁忙，無法回家照顧家中長者時，可以透過該系統去關心長者居家生活狀況，且如果長者有意外發生時，也會立即通知子女。

廣告推播部分，其背後所需的年齡與性別技術，Levi et al. (2015)利用 CNN 來去進行年齡與性別的分類，而現實中英特爾的 AIM Suite[4]則利用攝影機與軟體偵測人臉，判斷消費者的年齡與性別，並給出該消費者可能有興趣的廣告。

在觀看電影時的觀眾情緒上，迪士尼研究中心的 Deng et al. (2017)利用深度學習，可以在影廳內辨別觀眾的臉部表情，並學習分辨大笑、微笑等不同程度的情緒。

最後，在病人監控上，工研院的高志忠等人(2018)與醫療單位合作，開發一套用於失智症患者的情緒辨識模型，以協助病患情緒狀態評估，上述相關應用、背後技術與其內容整理於下表 2.1。

表 2.1 辨識應用

應用	背後技術	內容
門禁系統 Chang & Liao (2015)	人臉辨識	識別員工誰有權進入工廠、公司工作，並透過系統警示通知相關人員進行處理。
廣告推播 Levi et al. (2015)	年齡、性別辨識	根據廣告受眾的年齡與性別來顯示動態的廣告、促銷，進而促進銷量。
機器人照護 李國徵等人(2018)	人臉辨識 姿勢辨識 語音辨識	透過影像去觀察父母的進食狀況、活動力、異常狀態(跌倒)、睡眠品質、語音情緒等

[4] 臉部偵測軟體 能判年齡可讀心，https://news.ltn.com.tw/news/world/paper/556579

表 2.1 辨識應用(續)

應用	背後技術	內容
學生學習情緒 吳政德(2017)	眼球追蹤 人臉辨識 情緒辨識	用臉部表情來評估學生的學習情緒，藉由補捉學習情緒給教師可以幫助教師了解學習者的學習狀況。
病患情緒監控 高志忠等人(2018)	情緒辨識	有些病症的情緒表現較為明顯、不會隱藏(如：失智症)，若能偵測病患或家人的情緒狀況，將提醒醫護人員關懷與處理病患，有助於醫院分配人力，減輕照護上的負擔。
影像監控 出勤管理 Harikrishnan et al. (2019)	人臉辨識	實時識別上課學生進行出缺勤管理，並將結果傳回後端紀錄。
觀眾情緒 Deng et al. (2017)	人臉辨識 情緒辨識	識別影廳內觀眾臉上的表情，調查你對該劇情的喜愛程度並預測觀眾是否會喜歡後面的劇情，其能依據學習結果，在開演的前幾分鐘就能預測觀眾的情緒。

資料來源：本研究整理

第三章　研究方法

本研究主要是透過臉部情緒去進行情緒分類並將結果顯示在圖像上，因此本章節根據本文研究目的與文獻探討進行架構設計，並說明本研究所採用的研究方法。

3.1 研究架構

配合本研究的主題以及相關理論內容，其架構設計如圖 3.1 所示

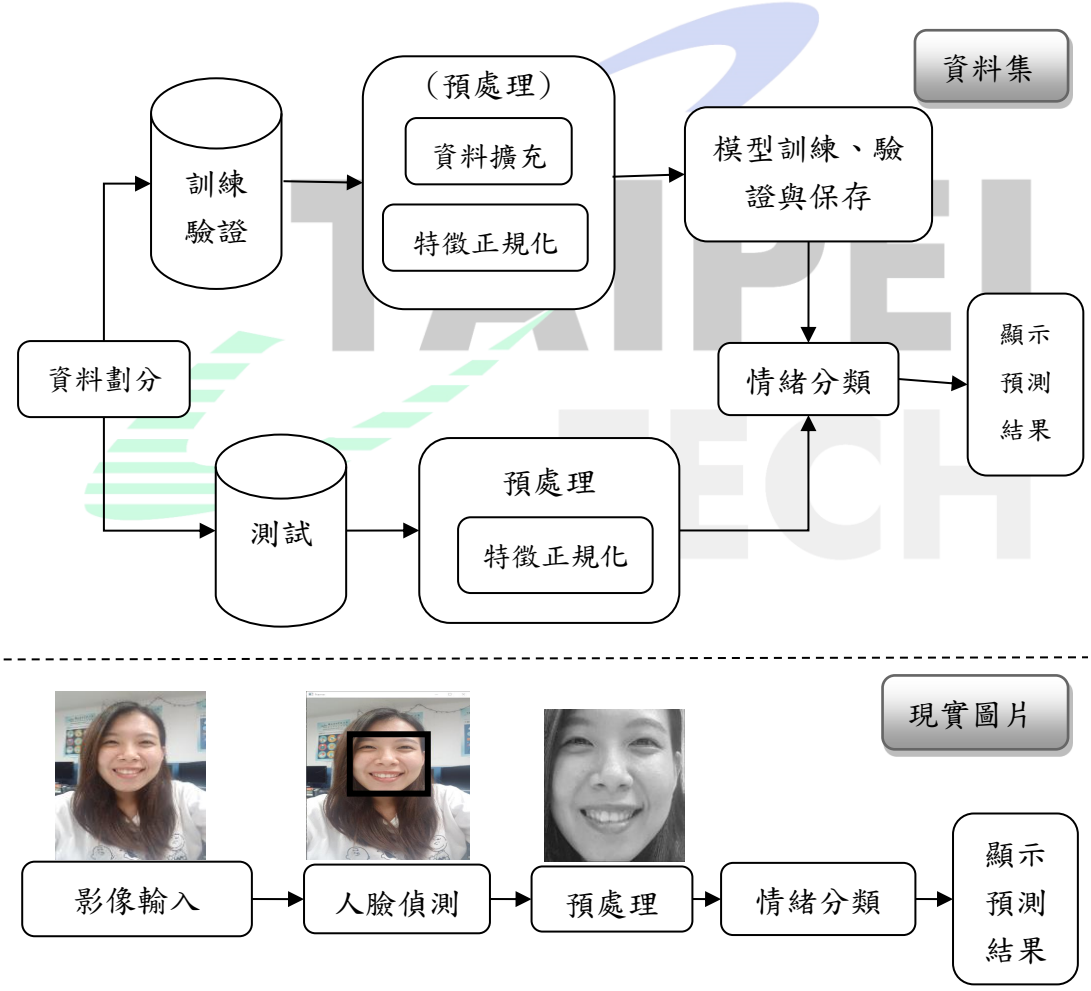


圖 3.1 研究架構

資料來源：本研究整理

本研究架構分為資料集(訓練驗證與測試)與真實圖片，首先資料集部分，在訓練驗證階段時，會將訓練與驗證集資料先進行預處理，處理完後丟入 CNN 模型進行訓練、參數調整，最後將訓練好的模型保存，供測試階段載入使用；在測試階段，將測試集資料也進行預處理，並將處理完的資料放入已經訓練好的神經網路模型，預測出該圖片的情緒類別，最後將測試集結果以混淆矩陣顯示出來；另外，真實圖片部分，會輸入現實中的照片，擷取照片中的人臉，以訓練驗證階段訓練好的模型來預測其情緒，最後將情緒結果輸出在圖片上。

3.2 研究方法

此節會說明資料劃分、人臉偵測、調整大小及資料擴充與特徵正規化等預處理，也會介紹預處理完後輸入的神經網路模型。

3.2.1 資料劃分與影像輸入

在資料劃分上，該資料集已經分好訓練集、驗證集與測試集，因此本研究就直接採用做訓練，另外，由於資料集不是圖檔而是 csv 檔，其檔案內容如圖 3.2 所示，emotion 代表七類表情，由 0 到 6 數字組成，依序為憤怒、厭惡、恐懼、快樂、難過、驚訝與中性，pixels 代表圖片的像素值，Usage 中有 Training、Public 與 Private 代表訓練集、驗證集與測試集，由於訓練、驗證及測試集等所有的資料都放在同一個檔案中，因此需要將這些資料分開，再將裡面一串串由數字與空白組合而成的像素值(Pixel)轉換成 48*48 的矩陣；在真實圖片部分，則是需要將圖片上傳，進行後續人臉偵測、預處理與情緒預測。

	A	B	C
1	emotion	pixels	Usage
2	0	70 80 82 72 58 58 60 63 54 58	Training
3	0	151 150 147 155 148 133 111	Training
4	2	231 212 156 164 174 138 161	Training
5	4	24 32 36 30 32 23 19 20 30 41	Training
6	6	4 0 0 0 0 0 0 0 0 0 0 0 3 15 23	Training
7	2	55 55 55 55 55 54 60 68 54 85	Training
8	4	20 17 19 21 25 38 42 42 46 54	Training
9	3	77 78 79 79 78 75 60 55 47 48	Training

圖 3.2 訓練資料檔案內容

資料來源：Challenges in Representation Learning: Facial Expression Recognition[5]

3.2.2 人臉偵測(Face Detection)

本研究使用 OpenCV 與 Dlib 來進行人臉偵測，並以多張圖片去比較兩者的偵測效果，最後選擇效果最好的當作人臉偵測的偵測器。

OpenCV 全名是 Open Source Computer Vision Library，是一套電腦視覺和機器學習等功能的函式庫，它是由 Intel 參與開發並發起，以其開源、免費，可跨平台使用(Linux, Mac OS, Windows)，提供了機器學習的基礎演算法呼和 cuda, python, Java 等的使用介面，使影像處理與分析更容易上手。

Dlib 是一套包含了圖像處理、機器學習算法、計算機視覺等的函式庫，使用 C++開發而成，它不但開源、免費，而且還可跨平台使用(Linux, Mac OS, Windows)，在工業及學術界廣泛使用，應用在手機、機器人、嵌入式系統與大型的運算架構中，另外提供 Python API。

OpenCV 在人臉偵測上所使用的演算法是 Haar 特徵加上自適應增強並將多個分類器級聯(Cascade)在一起，用 Haar 特徵進行特徵提取，再透過一個個分類器將非人臉視窗排除，最後偵測出人臉。

Dlib 在人臉偵測上所使用的演算法是 HOG 加上 SVM 進行偵測，先將圖片轉為灰階降低所需的參數量，接著利用 HOG 進行特徵提取，最後再使用 SVM 進

[5] Kaggle Challenges in Representation Learning: Facial Expression Recognition Challenge, https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data

行分類，偵測出的人臉會以一個方框框起來，如圖 3.3。

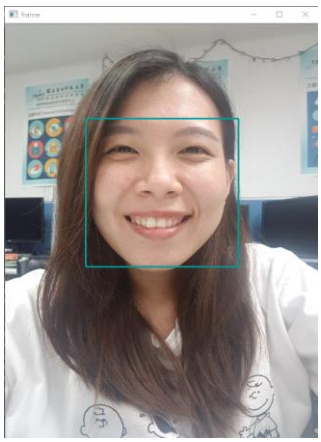


圖 3.3 人臉偵測 Dlib

資料來源：本研究整理

3.2.3 預處理

為了要降低圖片中因為背景、雜訊(Noise)等因素造成辨識準確率下降的影響，需要在放入 CNN 模型辨識前，先將這些因素處理完畢，相關處理內容依照真實圖片與資料集分別敘述如下。

（一）真實圖片：在真實圖片中，由於只要偵測表情的部分，臉部以外的背景資訊易對分類有雜訊的影響，因此須將臉部以外的背景進行裁剪，降低雜訊影響，裁剪完後、調整圖像大小成 48*48，轉成灰階，並將其正規化後再丟入模型辨識，圖 3.4 為裁剪、調整大小並轉成灰階後的處理結果。

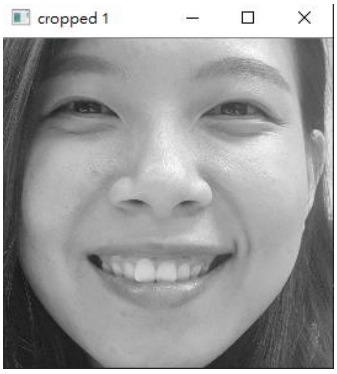


圖 3.4 裁剪結果

資料來源：本研究整理

（二）資料集：在資料集上，只對訓練集做資料擴充，考慮到現實中的圖片不會有臉倒著拍的畫面，因此在角度旋轉上只做±40 度，再加上左右翻轉、水平與垂直平移的處理；由於神經網路模型為對權重敏感的模型，因此需要進行特徵正規化，將圖片中的每個像素處以該張圖片的最大像素值，將其縮放為 0 到 1 之間，來去平衡各像素間值較大對權重的影響，此外經過特徵正規化後，也會增加模型收斂的速度，在此本研究選用最大正規化法來作特徵縮放。

3.2.4 神經網路模型

由於 CNN 需要調整大量權重，為此通常需要大量的訓練數據。如果在數據較小的資料集上訓練神經網路容易造成過擬合的現象，因此為了減輕過擬合的現象與加快模型收斂的速度，本研究選擇多種 CNN 模型來當預訓練模型，分析這些模型的分類效能，最後選出準確率最好的作為真實圖片預測的模型；在訓練時，會調整圖像丟入模型的尺寸大小，重新訓練全連接層，而模型的卷積層則會去比較完全凍結、微調及完全解凍此三種，來找出最好的模型；模型介紹如下：

（一）VGG16、VGG19 與 VGG-Face

VGG16 與 VGG19 是用 ImageNet 資料集—其包含各式各樣的物品、人物、景物等圖片綜合而成的資料集，而 VGG-Face 是用人臉資料集去訓練的網路，想比較綜合資料集訓練的特徵與人臉資料集訓練的特徵對於表情特徵的影響，因此選用這三種當作預訓練模型。

VGG16、VGG19 與 VGG-Face 皆由牛津視覺幾何學小組(Oxford Visual Geometry Group)的人所建立的，VGG16 與 VGG19 是 Simonyan & Andrew (2015) 建立的，VGG16 由 2 塊 2 層卷積層與 3 塊 3 層卷積層共 13 層，加上 3 層全連接層與 Softmax 輸出總共 16 層，層與層之間使用最大池化分開，激活函數都採用 ReLU 函數，卷積層的卷積核大小為 3*3，輸入的圖像大小為 224*224，VGG19 則是在 VGG16 的配置上，在 3 塊 3 層卷積層上各多加一個卷積層，變成 3 塊 4

層卷積層；而 VGG-Face 是 Parkhi et al. (2015)建立的，是以 VGG16 體系架構下調整而成的，模型是使用自行收集的人臉數據集進行訓練，兩者的差別就是背後資料集與全連接層的不同，詳細的模型配置與架構可見圖 3.5 與 3.6。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

圖 3.5 VGG16 與 VGG19 配置（D 模型與 E 模型）

資料來源：Simonyan & Andrew (2015)

layer	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
type	input	conv	relu	conv	relu	mpool	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv
name	-	conv1_1	relu1_1	conv1_2	relu1_2	pool1	conv2_1	relu2_1	conv2_2	relu2_2	pool2	conv3_1	relu3_1	conv3_2	relu3_2	conv3_3	relu3_3	pool3	conv4_1
support	-	3	1	3	1	2	3	1	3	1	2	3	1	3	1	3	1	2	3
filt dim	-	3	-	64	-	-	64	-	128	-	-	128	-	256	-	256	-	-	256
num filts	-	64	-	64	-	-	128	-	128	-	-	256	-	256	-	256	-	-	512
stride	-	1	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	2	1
pad	-	1	0	1	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1

layer	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
type	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	softmx
name	relu4_1	conv4_2	relu4_2	conv4_3	relu4_3	pool4	conv5_1	relu5_1	conv5_2	relu5_2	conv5_3	relu5_3	pool5	fc6	relu6	fc7	relu7	fc8	prob
support	1	3	1	3	1	2	3	1	3	1	3	1	2	7	1	1	1	1	1
filt dim	-	512	-	512	-	-	512	-	512	-	512	-	-	512	-	4096	-	4096	-
num filts	-	512	-	512	-	-	512	-	512	-	512	-	-	4096	-	4096	-	2622	-
stride	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1
pad	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0

圖 3.6 VGG-Face 架構

資料來源：Deep Face Recognition[6]

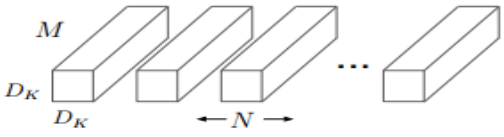
[6] Deep Face Recognition, http://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/poster.pdf

VGG16 與 VGG19 要學會分 1000 類物品種類，而 VGG-Face 要學會從 2622 張臉中辨識出正確的那張臉，因此兩者的全連接層會把前面學到的特徵組合在一起，然後傳到輸出層輸出可能性最高的結果。

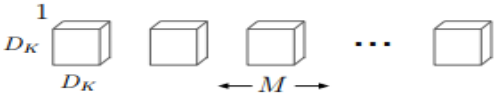
（二）MobileNet

由於 VGG 系列的模型參數都非常的龐大，而 Google 旗下的研究團隊 Howard et al. (2017)所提出的 MobileNet 則可以降低大量的參數且執行速度也快速許多，使其在行動裝置上能夠執行神經網路的相關應用。

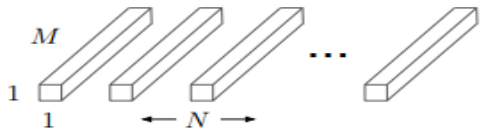
MobileNet 之所以能夠降低大量的參數計算量是因為他在卷積層的部分採用深度可分離卷積(Depthwise Separable Convolution)，也就是把標準卷積層分解成深度卷積(Depthwise Convolution)和逐點卷積(Pointwise Convolution)，如圖 3.7 所示，傳統的卷積層圖 3.7 (a)計算量為特徵圖的邊長(D_f)*輸入通道數量(M)*卷積核邊長(D_k)*輸出通道(N)，也就是D_f*D_f*M*N*D_k*D_k，而 3.7 (b)深度卷積會將 M@D_f*D_f的特徵圖切成 M 個D_k*D_k*1 的單一通道特徵圖，因此計算量為D_f*D_f*M*D_k*D_k，其輸出的特徵圖通過 3.7 (c)逐點卷積的 M@1*1 的卷積來去組合，變成新的輸出通道 N 個D_f*D_f的特徵圖，其計算量為 M*N*D_f*D_f。



(a) Standard Convolution Filters



(b) Depthwise Convolutional Filters



(c) 1 × 1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

圖 3.7 深度可分離卷積

資料來源：Howard et al. (2017)

將深度卷積與逐點卷積的計算量相加再除以傳統卷積的計算量就可以得到 MobileNet 降低的參數計算量，如圖 3.8 所示，其所使用的參數量比標準卷積少 8 到 9 倍。

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2}$$

圖 3.8 MobileNet 參數計算量

資料來源：Howard et al. (2017)

MobileNet 的網路架構如圖 3.9 所示，dw 代表深度卷積，除了最後一個完全連接層直接進入 softmax 層進行分類外，所有層經過批次正規化(Batch Normalization)與 ReLU 激活函數，扣掉平均池化與 softmax 輸出層，總共有 28 層。

Table 1. MobileNet Body Architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	3 × 3 × 3 × 32	224 × 224 × 3
Conv dw / s1	3 × 3 × 32 dw	112 × 112 × 32
Conv / s1	1 × 1 × 32 × 64	112 × 112 × 32
Conv dw / s2	3 × 3 × 64 dw	112 × 112 × 64
Conv / s1	1 × 1 × 64 × 128	56 × 56 × 64
Conv dw / s1	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 128	56 × 56 × 128
Conv dw / s2	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 256	28 × 28 × 128
Conv dw / s1	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 256	28 × 28 × 256
Conv dw / s2	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 512	14 × 14 × 256
5× Conv dw / s1	3 × 3 × 512 dw	14 × 14 × 512
5× Conv / s1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw / s2	3 × 3 × 512 dw	14 × 14 × 512
Conv / s1	1 × 1 × 512 × 1024	7 × 7 × 512
Conv dw / s2	3 × 3 × 1024 dw	7 × 7 × 1024
Conv / s1	1 × 1 × 1024 × 1024	7 × 7 × 1024
Avg Pool / s1	Pool 7 × 7	7 × 7 × 1024
FC / s1	1024 × 1000	1 × 1 × 1024
Softmax / s1	Classifier	1 × 1 × 1000

圖 3.9 MobileNet 架構

資料來源：Howard et al. (2017)

本研究會使用 Keras 深度學習套件來訓練模型，其是一個用 Python 編寫的神經網路 API，專門用在模型的建立、訓練與預測等功能，模型的網路層、損失函數等都是一個一個模塊可以任意結合、添加，且模型背後的矩陣運算由 Keras 提供支援的後端引擎，如：Theano、Tensorflow 等進行運算，因此擁有對使用者友善、模組化、易擴展等優點。

3.3 小結

根據以上對人臉偵測、預處理、神經網路的說明，我們得出以下結論:

(1)人臉偵測比較 OpenCV 與 Dlib 從中選出最適合的偵測器來找出圖像中的人臉；(2)將人臉與非人臉部分裁剪開來並調整裁剪後的人臉大小，經過灰階與正規化後，當作圖像的輸入；(3)情緒辨識部分參考各種 CNN 模型採用其預訓練的權重當作初始化再以表情數據集去訓練網路全連接層、輸出層與調整權重，神經網路以 Keras 來建構、訓練；關於後續的實驗設計、結果與結論，將在第四章與第五章進行說明。

第四章　實驗結果與分析

本章內容依序為實驗環境、實驗數據、實驗設計與實驗結果等四項內容，描述資料集、模型架構、評估方法與分析實驗結果。

4.1 實驗環境

本研究實驗環境為:Intel(R) Core(TM) i7-6770 CPU @ 3.40GHz, RAM 32.00GB, GPU Nvidia GeForce GTX1060 6GB，作業系統 Windows7 專業版 64 位元，開發語言 Python 3.6.10，開發工具 Visual Studio Code。

虛擬環境與套件管理皆使用 Anaconda 來做創建與管理，Anaconda 是一個開源跨平台的環境與套件管理軟體，可與像是 VSCode、Jupyter、Spyer 與 Rstudio 等多個開發工具一起使用，其內含許多有關資料科學相關套件。

本研究實驗所用的套件與版本如表 4.1 所示：

表 4.1 實驗套件與版本

套件名稱	版本	說明
Python	3.6.10	開發語言
Numpy	1.18.2	維度陣列與矩陣運算
Keras	2.2.5	模型建構用
Tensoflow	1.14.0	Tensorflow-gpu，模型底層的引擎
Cuda	10.0	Gpu 加速
Cudnn	7.4.1	Gpu 加速
Sklearn	0.22.2.post 1	全名為 Scikit-learn，計算混淆矩陣
Cv2	3.4.2	全名為 Opencv，讀取圖片、繪畫
Matplotlib	3.2.1	準確率與損失線圖的視覺化工具
Pandas	1.0.3	讀取 csv 檔案
Dlib	19.8.1	人臉偵測
PyQt	5.13.0	GUI 介面

資料來源：本研究整理

與模型建構、矩陣運算相關的套件有 Keras、Tensorflow、Cuda、Cudnn 與 Numpy，模型評估指標像是準確率、混淆矩陣等則是使用 Matplotlib 跟 Sklearn，讀取資料在 csv 是使用 Pandas，而圖片則是 Cv2，人臉偵測則是使用 Dlib 與 Cv2，人臉偵測的矩形框與情緒預測的文字皆是使用 Cv2 顯示，介面則是 PyQt。

4.2 實驗數據

本研究使用 FER2013 資料集，其圖片是用 Google 圖片搜索 API 從網路上收集來的，以情緒加上年齡、性別與種族有關的關鍵字一起在圖像搜索 API 上查詢，保留每個查詢結果的前 1000 張圖片，用 OpenCV 偵測出人臉邊界框，再以人工貼標的方式檢查貼錯的標籤與重複的圖像等錯誤，資料集包含 7 種情緒分別是憤怒(Anger)，厭惡(Disgust)，恐懼(Fear)，快樂(Happy)，悲傷(Sad)，驚訝(Surprise)和中性(Neutral)，總共 35887 筆資料，圖片為 48*48 像素灰階圖。

該資料集的優點是圖片中的人物年齡分布廣，從嬰兒到老人的表情都有涵蓋，人臉的角度方向涵蓋範圍也很廣，有在人臉稍微微揚、側臉等角度方向的表情且資料量相較於其他表情資料集多，但將其轉為圖片後，發現有些表情類別含有非表情與卡通手繪的圖片，如圖 4.1(a)所示，因此該資料集有雜訊的現象，此外該資料集作者 Goodfellow et al. (2015)描述人類在該資料集的辨識準確率大概是 65%。

該資料集已切分訓練集、驗證集、測試集分別是 28709 筆、3589 筆、3589 筆，不管是訓練、驗證或測試集情緒資料量落差很大，厭惡約占資料量 1.5%，而開心卻占 25%，其他情緒則有 13±3%，因此有資料量分布不均的現象，資料集圖片與資料分布如下圖 4.1(b)與圖 4.2。

圖 4.1(a) 非表情與卡通圖片

圖 4.1(b) 資料集圖片

資料來源：Challenges in Representation Learning: Facial Expression Recognition[7]

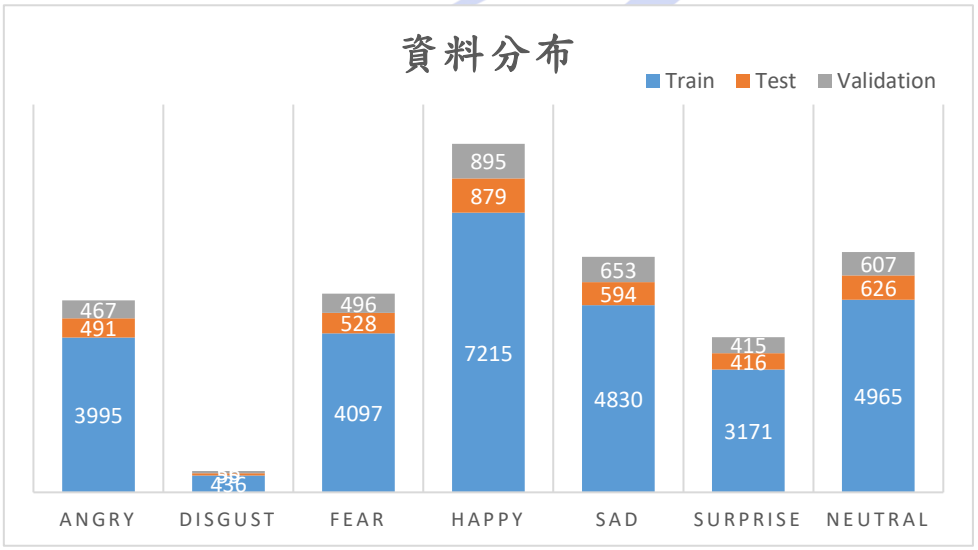


圖 4.2 資料分布

資料來源：本研究整理

4.3 實驗設計

根據第三章的研究架構與模型介紹，本節分作三段來詳細說明對其模型的修改之處、預處理及評估模型效能的指標，其描述如下所示。

[7] Kaggle Challenges in Representation Learning: Facial Expression Recognition Challenge, https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data

4.3.1 預處理

在丟入模型訓練、辨識前須先將圖片做一些處理，降低過擬合現象，第一個是資料擴充，在擴充資料時須考慮到該圖片的角度是否在現實中會出現，以人臉為例，拍照時我們並不會將臉倒過來（180 度）拍照，通常都只會將臉的角度傾斜一點拍而已，因此我們只將圖片做正負傾斜 40 度、左右翻轉、沿水平、垂直方向平移等處理，處理後的圖片如圖 4.3。

原圖　　資料擴充

圖 4.3 資料擴充圖片

資料來源：本研究整理

第二個是特徵正規化，使用最大正規化，原圖片像素值落在 0 到 255 之間，將圖片中每個像素值除以該張圖片中的最大像素值，使其落在 0 到 1 之間，平衡因像素值大小造成對模型權重的影響。

4.3.2 神經網路模型

本研究依據第三章的對 VGG16、VGG19、VGG-Face 的 VGG 系列與 MobileNet 的介紹，會在此四種模型結構下會做幾點調整，其說明如下:

（一）調整圖像大小:模型的圖像輸入大小為 224*224，本研究的資料集大小為 48*48，因此圖像輸入會以 48*48 當作輸入值。

（二）調整全連接層與更換輸出層:VGG 系列全連接層有三層，分別要分 1000 類與 2622 類，而本研究則是要將情緒分成七類，因此會調整全連接層的層數與神經元數量，調整成兩層，也不會採用兩模型的 Softmax 輸出層，會重新訓

練輸出層，修改後的結構如下表 4.2。

（三）調整卷積層權重:本研究想比較完全凍結、微調與完全解凍這三種方法,哪個能得到較好的準確率,完全凍結就是卷積層的權重是預訓練的初始權重,因此在訓練全連接層進行反向傳播更新權重時，並不會更新卷積層的權重，只會更新全連接層的權重；而微調就是只固定卷積層的第一、二層，因為前面幾層是模型在學基礎特徵,如邊緣、線條等,而後面幾層則是學比較複雜、高階的特徵,因此後面幾層卷積層的權重則會更新;完全解凍則是以預訓練模型的權重做初始化，而不會以隨機權重做初始化，所有卷積層權重與全連接層都會進行更新。

表 4.2 圖片大小、全連接層與輸出層比較

Layer	VGG-Face		VGG16、VGG19		MobileNet		本研究	
輸入層	(224,224,3)						(48,48,3)	
完全連接層	Fc1	4096			Fc1	1024	Fc1	512
	Fc2	4096						
	Fc3	2622	Fc3	1000	Fc2	1000	Fc2	7

資料來源：本研究整理

4.3.3 評估方法

為了得知模型分類效能，使用準確率(Accuracy)與混淆矩陣(Confusion Matrix)來去分析整體與各自情緒分類結果,混淆矩陣以可視化的方式來傳遞模型預測的結果與真實結果的情況，其相關的定義及公式說明如下：

（一）混淆矩陣

根據 Tharwat (2018) 對三類別 3*3 的混淆矩陣圖的說明，為了更貼近表情分類，因此本研究將其擴充為七類別 7*7 的混淆矩陣，混淆矩陣如表 4.3 所示，以 angry 類當作範例，TP_a代表模型預測為 angry 而實際也是 angry 樣本數量，而E_{ad}

則代表實際的 angry 樣本被模型錯誤分類為 disgust，即分類錯誤的樣本，因此 angry 類中的假陰性(FN_{angry})就是$E_{ad} + E_{af} + E_{ah} + E_{as} + E_{asu} + E_{an}$，也就是實際 angry 被錯誤分類的樣本總合，而假陽性(FP_{angry})就是$E_{da} + E_{fa} + E_{ha} + E_{sa} + E_{sua} + E_{na}$，也就是預測是 angry 但實際上是別的情緒的樣本總合。

表 4.3 混淆矩陣

		True						
		C_{angry}	$C_{disgust}$	C_{fear}	C_{happy}	C_{sad}	$C_{surprise}$	$C_{neutral}$
Predict	C_{angry}	TP_a	E_{da}	E_{fa}	E_{ha}	E_{sa}	E_{sua}	E_{na}
	$C_{disgust}$	E_{ad}	TP_d	E_{fd}	E_{hd}	E_{sd}	E_{sud}	E_{nd}
	C_{fear}	E_{af}	E_{df}	TP_f	E_{hf}	E_{sf}	E_{suf}	E_{nf}
	C_{happy}	E_{ah}	E_{dh}	E_{fh}	TP_h	E_{sh}	E_{suh}	E_{nh}
	C_{sad}	E_{as}	E_{ds}	E_{fs}	E_{hs}	TP_s	E_{sus}	E_{sn}
	$C_{surprise}$	E_{asu}	E_{dsu}	E_{fsu}	E_{hsu}	E_{ssu}	TP_{su}	E_{nsu}
	$C_{neutral}$	E_{an}	E_{dn}	E_{fn}	E_{hn}	E_{sn}	E_{sun}	TP_n

資料來源：本研究整理

透過混淆矩陣，我們可以得到每個情緒類別真實與預測的數量或機率，更加了解模型在每個情緒的預測能力。

（二）準確率

根據 Tharwat (2018) 對準確率的說明，準確率就是被正確分類的樣本與樣本總數之間的比率，其公式如下：

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{4.1}$$

其中 P 和 N 分別表示正樣本和負樣本的數量。

4.4 實驗結果

本節將分為卷積層、預處理與批次大小對 VGG-Face、VGG16、VGG19 與 MobileNet 四模型的影響進行比較分析，並以準確率、混淆矩陣比較各模型的分類效能，最後再以人臉實際圖片去探討分類結果是否與模型在測試集上的結果一致。

4.4.1 參數設定

過往在訓練模型時常常是設定一個數字比較大的 epoch，等跑完所有 epoch 後再根據準確率等資料去調整模型與參數，但是可能在某一回 epoch 時就已經過擬合了，後面再繼續訓練只是擬合更嚴重，浪費訓練時間，而 Keras 有提供 callback 回調函數，可以用來解決這個問題，透過監控訓練時的損失或準確率等這些評量指標，來中斷訓練、調整學習率、保存模型與指標等。

本研究使用 Keras 的 CSVLogger、EarlyStopping、ModelCheckpoint 與 ReduceLROnPlateau 這四個函數來調整模型；CSVLogger 是將每一回 epoch 的訓練及驗證損失與準確率記錄在 CSV 檔裡面直到訓練結束，EarlyStopping 是監控訓練時的評量指標，當評量指標停止進步時，也就是損失沒有繼續下降或準確率沒有繼續上升的情況持續幾回 epoch 後，則停止訓練；ModelCheckpoint 則是在指定時刻，將模型權重存起來，最後 ReduceLROnPlateau 則是當評量指標停止進步時，降低一定比例的學習率，並以調整後的學習率繼續訓練。

本研究對於此四種回調函數設定的情況分別是，CSVLogger 每次訓練會產生一個 CSV 檔紀錄模型每回的訓練與驗證損失與準確率，EarlyStopping 當驗證損失停止下降持續 15 回 epoch 時則中斷訓練，ModelCheckpoint 在訓練時，當驗證準確率創新高時，則保存其模型權重，最後 ReduceLROnPlateau 當驗證損失停止下降持續 4 回 epoch 時則下降 40%的學習率，而初始學習率設定為 0.0001。

4.4.2 各模型卷積層比較

本研究在有特徵正規化與批次大小為 16 的參數設定下，比較四種模型的卷積層在完全凍結、微調與完全解凍三種方法的測試準確率與損失，結果如表 4.4 所示。

表 4.4 模型卷積層比較

Ways	VGG-Face		VGG-16		VGG-19		MobileNet	
	Acc (%)	Loss	Acc (%)	Loss	Acc (%)	Loss	Acc (%)	Loss
凍結	35.85	1.62	45.91	1.45	44.97	1.49	25.38	1.86
微調	60.23	4.15	64.86	3.53	65.70	3.13	49.59	2.02
解凍	61.60	4.21	66.78	3.36	64.25	3.42	61.79	1.476

資料來源：本研究整理

表 4.4 中凍結代表卷積層權重完全凍結，解凍代表卷積層權重完全解凍；完全凍結預訓練模型的卷積層權重，只訓練全連接層，每個模型的準確率都非常低，都不到 50%，代表預訓練模型原始權重所萃取出來的特徵並不適合表情分類任務；微調則是固定前兩層卷積層的權重，後面卷積層權重都會更新，準確率比完全凍結好很多，代表其所萃取的特徵有比較適合表情分類；最後完全解凍則是以預訓練模型的權重當作初始權重，不會以隨機權重當作初始權重，全部的卷積層權重都會更新，其準確率表現較好，因此後續模型的訓練都採用完全解凍的方式訓練。

4.4.3 預處理對模型的影響

本研究在批次大小為 16 的參數設定下，將未經過預處理、特徵正規化與特徵正規化加上資料擴充(FN+DA)對模型的影響，以準確率來評估，如表 4.5 與圖 4.4 所示。

表 4.5 模型預處理準確率比較

Model	VGG-Face			VGG16		
Ways	Raw	FN	FN+DA	Raw	FN	FN+DA
Test Acc (%)	61.32	61.60	65.95	66.02	66.78	69.23
Test Loss	4.31	4.21	1.18	3.54	3.36	1.10
Model	VGG19			MobileNet		
Ways	Raw	FN	FN+DA	Raw	FN	FN+DA
Test Acc (%)	64.18	64.25	67.51	61.66	61.79	64.78
Test Loss	3.45	3.42	1.03	1.45	1.476	0.95

資料來源：本研究整理

表中 Model 代表模型、Ways 代表處理方法、Test Acc 與 Test Loss 分別代表測試集準確率與損失，Raw 代表未經過任何預處理，DA 代表資料擴充，FN 代表特徵正規化，在準確率與損失部分，各模型經過正規化後的準確率都略微提升一點，損失除了 MobileNet 以外，其他都略微下降，而經過資料擴充後，不但準確率都有所提升而且損失都大幅下降，因此可知正規化與資料擴充對於模型的準確率提升與損失下降是有幫助的。

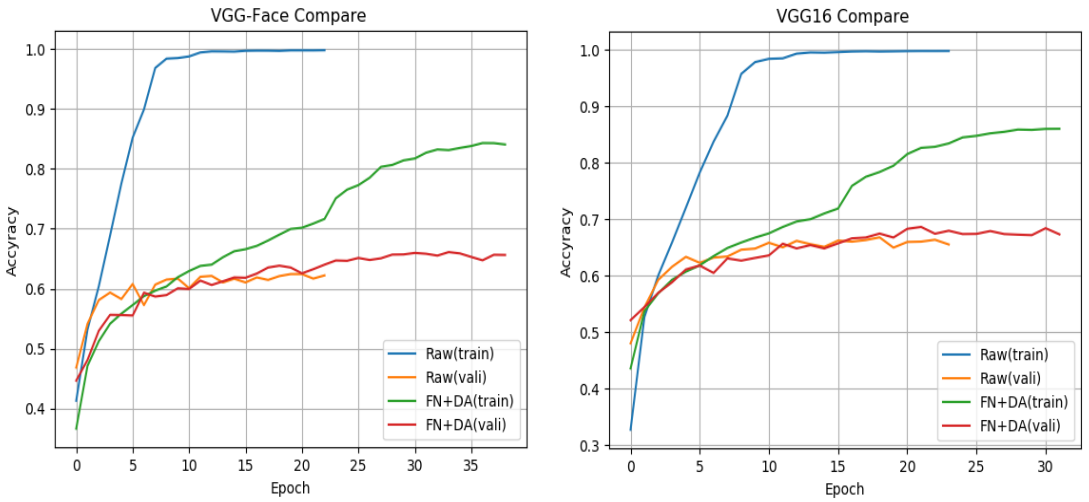


圖 4.4 模型預處理曲線比較

資料來源：本研究整理

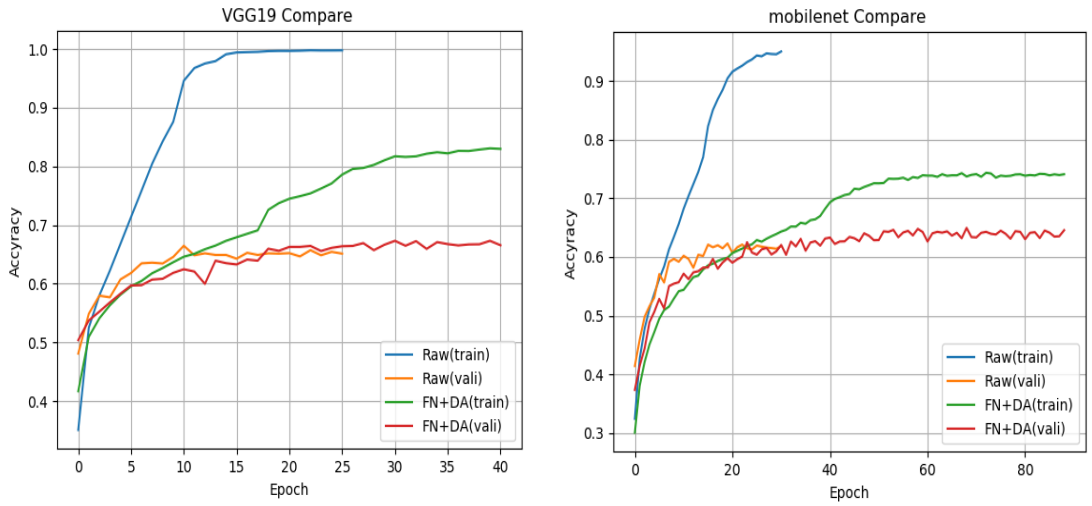


圖 4.4 模型預處理曲線比較（續）

資料來源：本研究整理

圖 4.4 是在比較特徵正規化加上資料擴充與未經預處理在訓練集及驗證集上對模型過擬合的影響，每個模型在未經處理的情況下，訓練集準確率在前 1/4 Epoch 時就已達 90%以上了，訓練集與驗證集準確率相差近 40%，過擬合情況最為嚴重，但經過特徵正規化加上資料擴充的步驟後，訓練集的準確率緩慢增加，過擬合情況減緩許多，因此可知預處理能有效減緩過擬合情況。

由於預處理對模型準確率、損失與過擬合都有幫助，因此每個模型都會經過特徵正規化與資料擴充這兩種預處理方法。

4.4.4 模型在不同批次大小下比較

本節的模型都已經經過特徵正規化與資料擴充，卷積層則是完全解凍，比較在批次大小為 8、16、32、64 與 128 下各模型的表現，如表 4.6 與圖 4.5 所示。

表 4.6 模型在不同批次大小比較

Batch	VGG-Face		VGG-16		VGG-19		MobileNet	
	Acc (%)	Loss	Acc (%)	Loss	Acc (%)	Loss	Acc (%)	Loss
8	66.31	1.10	67.81	1.08	66.95	1.04	60.04	1.24
16	65.95	1.18	69.23	1.10	67.51	1.03	64.78	0.95
32	65.33	1.20	68.40	1.12	67.79	0.95	64.39	0.98
64	65.08	1.17	68.73	1.20	67.14	1.03	62.58	1.06
128	64.25	1.15	67.34	1.31	68.06	1.32	59.12	1.12

資料來源：本研究整理

表 4.6 中 Batch 代表批次大小，Acc 代表測試集準確率，Loss 代表測試集損失，VGG-Face 的準確率約 64%~66%，損失與準確率在批次大小為 8 的準確率最好，VGG16 的準確率約 67%~69%，損失隨著批次大小增加而增加，準確率在批次大小為 16 的準確率表現最好，VGG19 的準確率約 66%~68%，損失在批次大小為 32 時最低，準確率則是 128 時最好，MobileNet 的準確率約 59%~64%，損失與準確率在批次大小為 16 的準確率最好；整體來看，在準確率方面，VGG16 的表現最好，MobileNet 的表現最差。

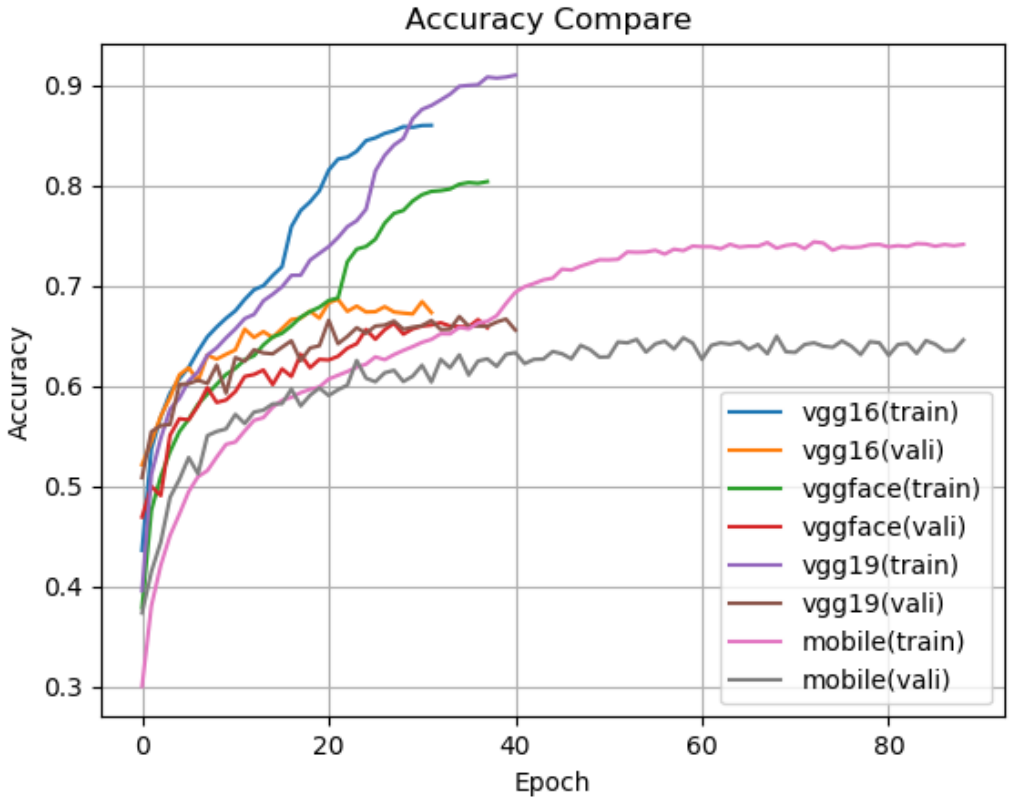


圖 4.5 模型準確率比較

資料來源：本研究整理

圖 4.5 的準確率曲線是以表 4.7 中每個模型準確率最高的批次大小去畫的曲線，右下角的圖標是各模型依訓練集與驗證集而有不同顏色的準確率曲線，train 為訓練集，vali 為驗證集；在訓練集上每個模型的準確率隨著 Epoch 增加而提高，在驗證集上，VGG16 的準確率曲線高於其他模型，VGG-Face 與 VGG19 相近，最後是 MobileNet；在 Epoch 上，MobileNet 需花費較長的時間，VGG16 則花費最少的時間，但是在訓練集與驗證集過擬合程度上，MobileNet 是擬合程度最小的。

綜合了四個模型的表現，雖然 MobileNet 所需的參數量最少，但準確率是最低的，而 VGG16 準確率最高，考量到是在電腦去進行真實圖片的情緒辨識，在速度上並不需要實時，因此後續的混淆矩陣與真實圖片情緒辨識的模型都會以 VGG16 為主。

4.4.5 情緒辨識文獻比較

表 4.7 是 FER2013 相關的期刊論文的準確率及使用方法與本研究所使用方法的比較，Tang (2013)與 Ionescu et al. (2013)是 FER2013 競賽的第一名與第四名，Tang 是使用簡單的 CNN 去提取特徵，而分類器則是用 L2SVM 去分類，Ionescu et al.在特徵提取是使用詞袋表示法提取 SIFT 的特徵子，分類器則是跟 Tang 一樣，使用 SVM 做分類；Saeed et al. (2018)在特徵提取方面是使用梯度直方圖，再以 SVM 去分類，上述三種文獻特徵提取的方式都不太一樣，但皆是以 SVM 作為分類器。

Mollahosseini et al. (2016)使用自己創建的 CNN 模型，以兩層傳統的卷積層再加上 Inception 模塊，也就是 1*1、3*3 和 5*5 並行的卷積層，最後加上兩層完全連接層做分類；Devries et al. (2014)使用多分類任務網路，此多分類任務是分類情緒與人臉特徵點(Facial Landmarks)，他們使用三層卷積層，在輸出層部分採取並行的兩個輸出層，一層輸出情緒預測，一層輸出特定人臉特徵點（眉毛、嘴巴等）在原始圖像中的機率，此研究表明透過特徵點的位置與形狀可以提高情緒辨識；Zhang et al. (2015)使用多分類任務去預測圖片中的人物之間的社會關係（領導、競爭、信任、友好等八類關係），其採用四層卷積層與一層完全連接層，此外為了得知人物間的關係，其聯合多種不同屬性（性別、情緒、姿勢與年齡）的資料集與自己創建的社會關係資料集，但資料集間的標記屬性與統計分布都不同，因此不能直接聯合訓練，所以此研究提出一個橋接層(Bridging layer)，利用資料集間的共通特徵，也就是鼻子、嘴巴等面部形狀，先用人臉特徵點找出其位置，利用 K-means 分群，找出每群的 HOG 特徵與距離當作特徵描述子，最後將這些特徵描述子與卷積層所萃取的特徵一起丟入完全連接層中去分類該人物的性別、情緒、姿勢與年齡等權重特徵，至於人物之間的社會關係，會聯合兩個人物的神經網路，將兩人的性別、情緒等特徵加上空間訊息，得出兩人的社會關係；上述是跟 CNN 相關的文獻，有自己創建的模型，也有多分類任務的模型。

表 4.7 其他情緒辨識文獻比較

Paper Author	Method	Test Accuracy(%)
Saeed et al. (2018)	HOG + Cubic SVM	57.17
Mollahosseini et al. (2016)	CNN (DeeperCNN)	61.10
Devries et al. (2014)	Multitask Network	67.21
Tang (2013)	SVM	71.162
Zhang et al. (2015)	Multitask Network	75.1
Ionescu et al. (2013)	SIFT+SVM	67.48
本研究	**Modified CNN**	**69.23**

資料來源:本研究整理

根據 Devries et al. (2014)在分類層使用 softmax 還是 L2SVM 的描述，在時間花費上 L2SVM 所需的時間是 softmax 的兩倍，而我們分類層是使用 softmax，因此所付出的時間成本較少。

雖然我們的方法在準確率上沒有達到 State-of-the-art，但比大部分的方法好，我們不需要像 Zhang et al. (2015)整合多個不同類別的資料集，只需使用一個資料集，也不需要訓練多個 CNN 分類模型或去調整卷積層架構，更不需要手工萃取特徵。

另外，我們的模型由於預處理只需要做基於圖像的資料擴充與特徵正規化，因此在預處理上所需的時間非常少，模型上的訓練週期也只要不到 40 個 Epoch 即可達到 69.23%的準確率，與 Mollahosseini et al. (2016)在訓練週期要 150000 Epoch 相比，本研究的預處理與模型訓練在時間花費上有很大的優勢。

4.4.6 混淆矩陣

準確率是整體的結果，但並不是每個情緒的個別準確率都跟整體準確率一樣，

可能有些情緒預測比較好，有些情緒預測比較差，所以如果要看模型對於每個情緒的預測狀況，就可以用混淆矩陣來得知，如圖 4.6。

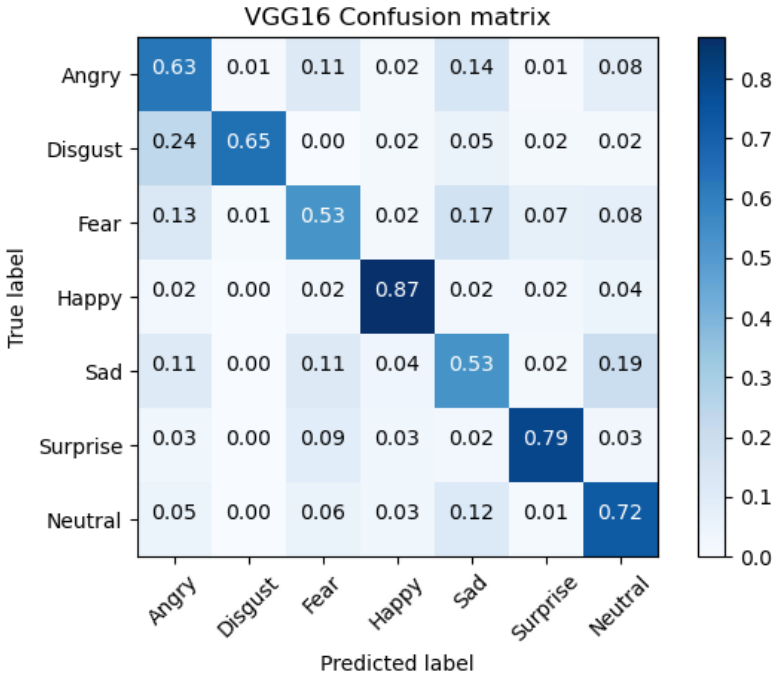


圖 4.6 模型混淆矩陣比較

資料來源:本研究整理

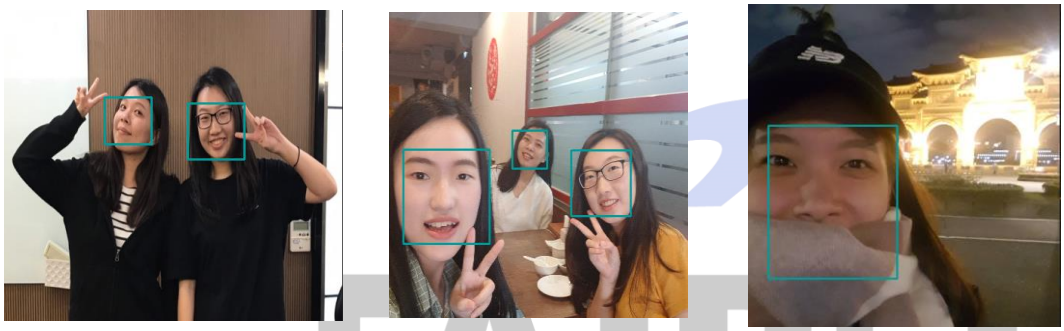
圖 4.6 是 VGG16 的混淆矩陣，其中表現最好的是快樂與驚訝，其次是中性，再來是憤怒與厭惡，最後是難過跟恐懼表情；快樂與驚訝不易預測為其他表情，中性有 1.2 成會分類到難過，憤怒約有 1.3 成會分類為難過與恐懼，難過有約 2 成的機率分類為中性表情，恐懼有 1.7 成會分類為難過，厭惡約有 2.5 成會分類為憤怒。

4.4.7 人臉偵測比較

在進行情緒辨識前，要先偵測出人臉，下圖 4.7 是 OpenCV 與 Dlib 在不同環境、人臉角度與遮蔽物下的偵測結果，上排是 OpenCV 的結果，下排是 Dlib 的

結果。

OpenCV 偵測結果



Dlib 偵測結果

圖 4.7 OpenCV 與 Dlib 偵測結果比較

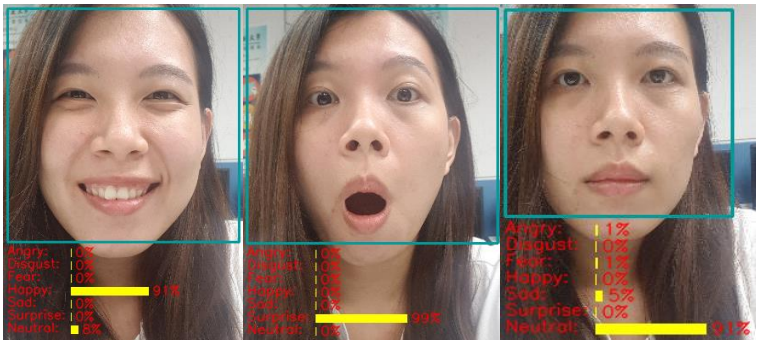
資料來源:本研究整理

如上圖所示，在最左側的圖中，左邊的人臉稍微往上微揚，OpenCV 沒有偵測到而 Dlib 有偵測到；在中間的圖中，環境有點曝光，OpenCV 只有偵測到中間的人臉，而 Dlib 三個都有偵測到；最右邊的圖中，該張人臉有戴帽子與圍巾，在有遮擋住額頭與嘴巴的情況下，OpenCV 沒有偵測到而 Dlib 有偵測到，綜合上述，在一些人臉角度、環境與遮擋的情況下，Dlib 的表現比 OpenCV 好，因此本研究使用 Dlib 作為人臉偵測的偵測器。

4.4.8 人臉圖片結果分析

模型訓練完後，本研究想得知其預測能力是否能在真實圖片中預測準確，因

此以手機拍照後的七類情緒圖片丟入模型分類預測，預測模型為 VGG16，其預測的結果如圖 4.8 到圖 4.10。

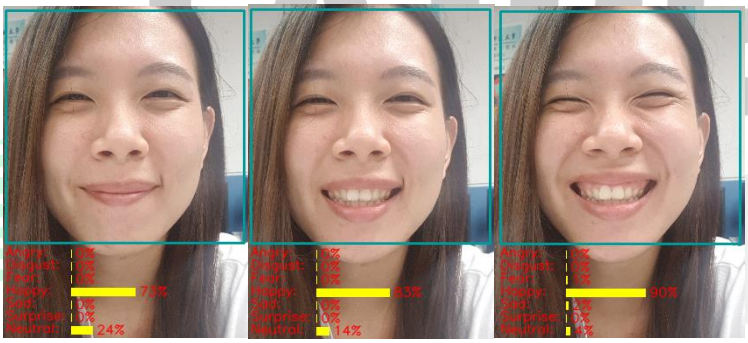


快樂　　驚訝　　中性

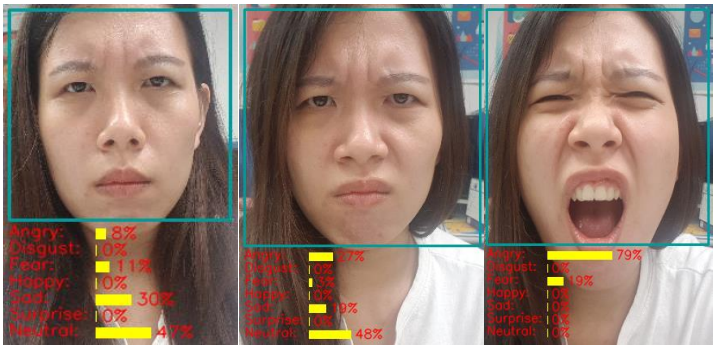
圖 4.8 快樂、驚訝與中性表情

資料來源:本研究整理

如圖 4.8 所示，圖片依序為快樂、驚訝與中性三張表情，模型在預測快樂、驚訝與中性時準確率約在 90%，很小的機率會預測到其他情緒，就如同混淆矩陣所描述，在此三種情緒的預測是最好的。



快樂（弱）　　快樂（中）　　快樂（強）



憤怒（弱）　　憤怒（中）　　憤怒（強）

圖 4.9 憤怒與快樂在不同強度下比較

資料來源:本研究整理

如圖 4.9 所示，上排是快樂情緒依強度從左到右為弱到強，下排則是憤怒情緒，從圖中可以得出兩個結果，第一個是如模型混淆矩陣所示，容易把憤怒情緒歸類為難過或中性表情，第二個是隨著表情的強度增加，情緒起伏變大，模型在預測該情緒的準確率就越高，就像快樂從左到右的準確率為 73%、83%與 90%，而憤怒則是 8%、27%與 79%，由此可知，模型在情緒起伏比較大時預測能力最好。

難過　　厭惡　　恐懼

圖 4.10 難過、厭惡與恐懼表情

資料來源:本研究整理

如圖 4.10 所示，圖片依序為難過、厭惡與恐懼三張表情，難過準確率約 60%，有一定機率會預測為恐懼，厭惡則有一定機率分類成憤怒，恐懼雖然不像混淆矩陣所述容易預測為難過，但其有一定機率預測為憤怒。

第五章　結論

本章依第四章的實驗結果與分析作後續的研究結論，並說明實驗的研究貢獻、限制與未來展望。

5.1研究結論及貢獻

本研究使用 FER2013 資料集作為模型訓練的資料，經過預處理後，放入預訓練模型訓練，並去比較 VGG-Face、VGG16、VGG19 與 MobileNet 四種模型的分類效能，最後將訓練好的模型進行情緒分類預測，並以真實人臉照片去測試其分類結果，結論如下：

（一）比較卷積層在完全凍結、微調與完全解凍下的準確率，發現完全解凍下權重更新所萃取出的特徵比較適合該表情資料集的分類任務。

（二）資料集像素值經特徵正規化將其縮放至 0 到 1 之間，可以略為提升準確率與略為降低損失，此外再經過資料擴充後，準確率有提升 3%~4%，損失則大幅下降，過擬合的情況也從 40%的訓練集與驗證集差距縮小至 20%左右。

（三）經過特徵正規化與資料縮放後，比較各模型在不同批次大小下的結果，發現 VGG16 的表現最好，準確率高於其他模型。

（四）模型混淆矩陣在快樂的準確率達 9 成，在驚訝的準確率達 8 成，在中性達 7 成，憤怒與厭惡則為 6 成，恐懼與難過則為 5 成，因此模型在快樂、驚訝與中性表情的預測能力最好。

（五）在人臉偵測方面，比較 OpenCV 與 Dlib 的偵測結果，在一些人臉角度、環境與遮擋的情況下，Dlib 的表現比 OpenCV 好，因此本研究使用 Dlib 作為人臉偵測的偵測器。

（六）情緒的強度也會影響模型的分類準確率，在情緒張力幅度不大時，容易有預測為其他情緒的狀況發生, 但隨著幅度的加強，對於表情分類能力也會與之提升，究其原因是資料集的表情圖片是由網路抓取，圖片中的情緒幅度都較為

強力，因此模型在情緒幅度較強時，預測能力最好。

貢獻方面：

（一）模型最高準確率為 69.3%，該準確率比人類實際在辨識 FER2013 資料集的準確率還高，因此在資料集上該模型性能比人眼辨識還好。

（二）與其他文獻相比，雖然我們在準確率上沒有達到 State-of-the-art，但比大部分的方法好，此外不需要整合多個不同類別的資料集、手工萃取特徵，模型上也不需要去調整卷積層的架構，或是訓練多個 CNN 分類模型，此符合研究目的的第一點，利用深度學習提取特徵與分類，提升模型性能。

（三）本研究詳細的比較卷積層、批次大小的與參數設定等 CNN 模型的超參數，並在最後用真實圖片去測試結果是否與測試集一致，此符合研究目的第二點，探討情緒辨識技術並將其用在真實照片中。

（四）本研究的預處理只需要做基於圖像的資料擴充與特徵正規化，因此在預處理上所需的時間非常少，模型上的訓練週期也只要不到 40 個 Epoch 即可達到 69.23%的準確率，因此在時間成本上較其他使用 CNN 的文獻還具有優勢。

5.2研究限制與未來展望

首先，研究限制方面，描述如下：

（一）受限於資料集的大小，CNN 模型的輸入大小有規定須大於一定尺寸，但將資料集的尺寸放大，其圖片含有很多雜訊且非常模糊，並不能很好的萃取出特徵，此外資料集的品質與數量有點參差不齊，此會影響模型的效能。

（二）由於資料集的資料大多都是情緒幅度較強的圖片，模型在情緒幅度較弱的圖片上並沒有很好的辨識能力，因此在預測情緒幅度較弱的表情時，會容易預測出其他情緒。

（三）人臉偵測上並沒有百分之百能完全偵測出圖片的人臉，且有時也會有偵測到非人臉情況發生，因此對於未偵測到的人臉沒有辦法辨識其情緒，對於偵

測錯誤的非人臉部分，也會有錯誤的情緒預測情況發生。

接著是未來展望的部分，描述如下：

（一）本研究之情緒辨識，僅以 FER2013 資料集作為模型背後的資料，然情緒的資料集依蒐集的方式不同，也會有更多不同強度的情緒、環境、角度等多元豐富的資料，所以可以嘗試結合多個情緒資料集來去當模型訓練材料。

（二）本研究只有研究靜態圖片的表情，對於動態表情的預測並沒有涉獵，因此可以加上時間的因素，結合 LSTM、RNN 等模型來去預測在時間維度上的情緒。

（三）本研究只有預測七種情緒，但是人類的情緒並不只有七種，因此可以嘗試預測更多情緒種類，使其更加貼近人類情感。

（四）本研究在預處理方面只有用特徵正規化與資料擴充，若還有其他預處理方法能提升情緒辨識的準確率，也可以嘗試使用。

參考文獻

中文部分

吳政德（2017）。利用卷積神經網路預測學習情緒之研究。國立中興大學資訊管理學系所碩士論文，台中市。

劉曉，譚華春，章毓晉（2006）。人臉表情識別研究的新進展。中國圖像圖形學報，第 11 卷，第 10 期，第 1360-1366 頁。

李國徵，陳瑞文，徐雅蕙，劉翠萍，張瑞玲，黃瑞星，吳俊賢，張傑智，邱碧貞（2018）。智慧照護-樂齡陪伴機器人之智能感知技術探索。電腦與通訊,智慧城市智能感知與物聯網專輯，第 175 期。

高志忠，宋美盈，蘇奕宇，吳佳樺，彭子芸，鄭名宏，蕭裕憲（2018）。智慧監控再升級-導入人臉情緒識別。電腦與通訊,智慧城市智能感知與物聯網專輯，第 175 期。

英文部分

Akash S., Gurudutt P. & Dr. K. S. Gayathri (2019). Facial Emotion Recognition using Convolutional Neural Networks, International Symposium on Artificial Intelligence and Computer Vision. College of Engineering, Guindy. Chennai, India.

Chang, C.-Y., & Liao, J.-J. (2015). Combination of RFID and face recognition for access control system. 2015 IEEE International Conference on Consumer Electronics - Taiwan.

Dalal, N., & Triggs, B. (n.d.). Histograms of Oriented Gradients for Human Detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05).

Deng, Z., Navarathna, R., Carr, P., Mandt, S., Yue, Y., Matthews, I., & Mori, G. (2017).

Factorized Variational Autoencoders for Modeling Audience Reactions to Movies.2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6014–6023.

Devries, T., Biswaranjan, K., & Taylor, G. W. (2014). Multi-task Learning of Facial Landmarks and Expression. 2014 Canadian Conference on Computer and Robot Vision.

Ding, H., Zhou, S. K., & Chellappa, R. (2017). FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 118–126.

Ekman, P., Friesen, W. V., Osullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K. & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. Journal of Personality and Social Psychology, 53(4), 712–717.

Ghimire, D., & Lee, J. (2013). Geometric Feature-Based Facial Expression Recognition in Image Sequences Using Multi-Class AdaBoost and Support Vector Machines. Sensors, 13(6), 7714–7734.

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., … Bengio, Y. (2015). Challenges in representation learning: A report on three machine learning contests. Neural Networks, 64, 59–63.

Happy, S. L., George, A., & Routray, A. (2012). A real time facial expression classification system using Local Binary Patterns. 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI), 1–5.

Harikrishnan, J., Sudarsan, A., Sadashiv, A., & Ajai, R. A. (2019). Vision-Face Recognition Attendance Monitoring System for Surveillance using Deep Learning Technology and Computer Vision. 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN).

Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H.(2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861

Ionescu, R. T., Popescu, M., & Grozea, C. (2013). Local learning to improve bag of visual words model for facial expression recognition. Workshop on Challenges in Representation Learning. (ICML).

Mollahosseini, A., Chan, D., & Mahoor, M. H. (2016). Going deeper in facial expression recognition using deep neural networks. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV).

Levi, G., & Hassncer, T. (2015). Age and gender classification using convolutional neural networks. 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

Levi, G., & Hassner, T. (2015). Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI 15, 503–510.

Ng, H.-W., Nguyen, V. D., Vonikakis, V., & Winkler, S. (2015). Deep Learning for Emotion Recognition on Small Datasets using Transfer Learning. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI 15, 443–449.

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. Procedings of the British Machine Vision Conference 2015.

Porusniuc, G. C., Leon, F., Timofte, R., & Miron, C. 2019. “Convolutional Neural Networks Architectures for Facial Expression Recognition.” 2019 E-Health and Bioengineering Conference (EHB).

Ramdhani, B., Djamal, E. C., & Ilyas, R. (2018). Convolutional Neural Networks Models for Facial Expression Recognition. 2018 International Symposium on

Advanced Intelligent Informatics (SAIN), 96–101.

Saeed, S., Baber, J., Bakhtyar, M., Ullah, I., Sheikh, N., Dad, I., & Ali, A. (2018). Empirical Evaluation of SVM for Facial Expression Recognition. International Journal of Advanced Computer Science and Applications, 9(11).

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1).

Simonyan, K. & Andrew, Z. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations.

Singh, D., & Singh, B. (2019). Investigating the impact of data normalization on classification performance. Applied Soft Computing, 105524.

Tian, Y.-I., Kanade, T., & Cohn, J. (2001). Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2), 97–115.

Tharwat, A. (2018). Classification assessment methods. Applied Computing and Informatics.

Viola, P., & Jones, M. J. (2004). Robust Real-Time Face Detection. International Journal of Computer Vision, 57(2), 137–154.

Tang, Y. (2013). Deep learning using linear support vector machines. Workshop on Challenges in Representation Learning (ICML).

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. IEEE Signal Processing Letters, 23(10), 1499–1503.

Zhang, Z., Luo, P., Loy, C.-C., & Tang, X. (2015). Learning Social Relation Traits from Face Images. 2015 IEEE International Conference on Computer Vision (ICCV).

網站部分

高齡化時程，https://www.ndc.gov.tw/Content_List.aspx?n=695E69E28C6AC7F3，（2019）。

扶老比，https://www.ndc.gov.tw/Content_List.aspx?n=84223C65B6F94D72，（2019）。

上課分心逃不掉！臉部追蹤技術揪出學生「恍神時刻」，幫助老師改進課程，https://www.inside.com.tw/article/9451-new-facial-recognition-to-point-out-when-do-students-not-paying-attention，（2019）。

臉部偵測軟體 能判年齡可讀心，https://news.ltn.com.tw/news/world/paper/556579，（2019）。

Deep Face Recognition, http://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/poster.pdf,(2019).

Kaggle Challenges in Representation Learning: Facial Expression Recognition Challenge,https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data,(2019).