

逢甲大學學生報告 ePaper

AI 人臉辨識

Face recognition by Artificial Intelligence

作者：蔡宸銘、黃清風

系級：光電三乙

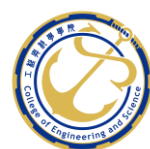
學號：D0789394、D0789381

開課老師：劉榮平、林昱志 老師

課程名稱：數位光學造影技術

開課系所：光電科學與工程學系

開課學年：109 學年度 第 2 學期



中文摘要

隨著人工智慧的發展越趨進步，它也逐漸融入我們的生活當中。其中人臉辨識就是其中一項應用，可以應用在門禁管理，身分核對上等。雖然人工智慧在模型耗時的建立較久，但是有了模型便能快速地對影像進行辨識，藉以達到即時的功能。在本專題我們針對人臉辨識這項主題去進行實作與探討。我們利用鏡頭進行影像擷取，再將擷取到的影像藉由人工智慧建立的模型進行辨識。在辨識神經網路模型我們分別使用特徵學習以及深度學習兩種方法建立，並比較兩者結果的差異。最終我們能夠實現實時的人臉辨識系統。

關鍵字：人工智慧、人臉辨識、深度學習、機器學習



Abstract

With the progress of artificial intelligence (AI), it gradually integrated into our life. Face recognition is one of AI applications. It can be applied in access control management, identification, and so on. Although the construction of an AI model is time-consuming, the image can be recognized quickly with the model to achieve real-time function. In this project, we have implemented face recognition by AI. We used the lens to capture images, and then recognized images with AI. In the artificial neural network, we used feature learning and deep learning, respectively, for face recognition. The difference of the two methods was also discussed. Finally, we have demonstrated a real-time face recognition system.

Keyword : artificial intelligence, face recognition, deep learning, ,machine learning



目次

一、序	p.4
二、影像擷取與人臉追蹤	p.5
三、特徵學習人臉辨識	p.6
四、深度學習人臉辨識	p.8
五、結論	p.14
六、參考文獻	p.14



一、序

人工智慧(artificial intelligence)於 1956 被提出，至今也有許多的發展[1]。其主要的想法為希望電腦或是機器能夠模仿人類的思維，能夠有獨立的思考與判斷能力。最初階的人工智慧像是自動化控制，可以模擬人類的思維行為，但沒有獨立的思考。機器所做的行為僅遵循著開發者的程式指令，無法做指令以外的判斷，因此在程式中須將所有的可能性考量進去，比方說自動做飲料的機器手臂，可以模仿人類做飲料的行為，但是無法自己思考出新的飲料配方。接著 1980 年開始有了機器學習(machine learning)。機器學習為由過往的經驗去尋找規則。以監督式學習為例，我們將要進行訓練的資料庫進行指定的特徵擷取，並在每筆資料加上分類標籤。所有資料的特徵以及對應的分類標籤經由演算法可以找出他們之間的關聯，便能夠使電腦「自主」的學習，建構出一個能夠分類的模型。換個說法我們不需要告訴電腦分類判斷的依據，機器學習能夠由資料特徵與分類標籤之間獨自找到規則。深度學習(deep learning)為機器學習的一種，但建立的模型結構又更為複雜。深度學習模仿了人腦的結構，由許多層神經元所建構而成，且不需要由人類告訴它要去學習的資料特徵，特徵值可由電腦自動去學習得到。雖然機器學習跟深度學習都需要大量的資料庫進行訓練，且模型的建構需要很長的時間，但是有了模型後便可以快速地進行資料分類。

人臉辨識在日常生活中隨處可見，像是大樓的門禁管制、手機上的臉部解鎖。相較於其他身分驗證的方式，臉部辨識擁有非侵入與非接觸式的好處，甚至可以應用在偵測表情、判斷性別或者有沒有戴口罩其他用途。本專題的目標為做出一個能夠即時辨識人臉的系統，人臉辨識有許多種方法[2]，如果我們用傳統的相似度關係計算，便需要大量的時間進行計算，無法達到即時的條件。因此我們使用 AI 的方式去建構一個能夠辨識人臉的模型，能夠快速地且有效的將偵測人臉進行辨識。本次專題主要流程，如圖一所示。第一步為將硬體鏡頭的捕捉到的影像進行分幀擷取，並提取出人臉的部分。第二步為將提取出的人臉進行辨識，我們分別使用機器學習中特徵學習與深度學習的兩種方法去建構模型，並驗證其成效。



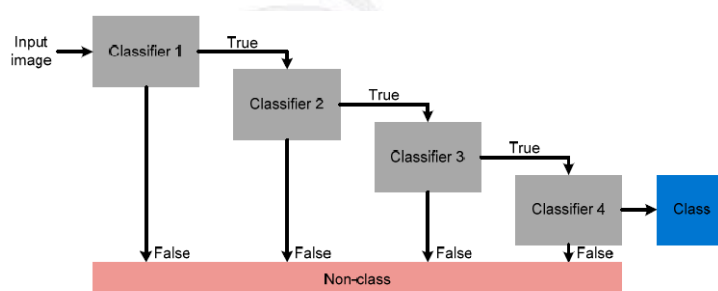
圖一、人臉辨識流程。

二、影像擷取與人臉追蹤

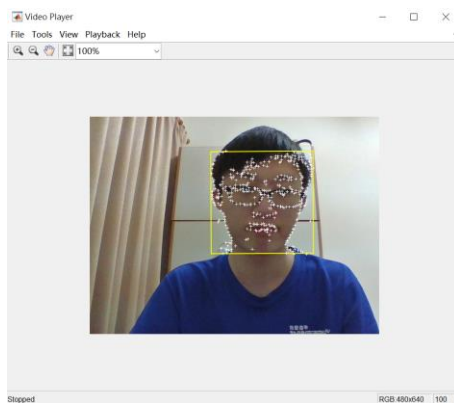
在影像捕捉的部份我們使用的電腦前置鏡頭，規格如表一所示，我們首先將鏡頭捕捉到的畫面分幀擷取成一張張的圖片。分幀出的圖片我們經由 Cascades 瀑布式的辨識方式[3]，判斷該幀畫面裡有沒有人臉。Cascades 的原理如圖二所示。我們將圖片依照不同大小與區域進行切割，在 face Cascades 裡有 6000 個定義是否為人臉分類層，切割後的區塊須滿足每一個分類層條件才能被定義為可能是人臉。由於每一次將區塊與所有 6000 個分類層計算需耗費許多時間，因此每個區塊都是由第一個分類層依序到最後，如果中間有不滿足的就直接淘汰，不須再往下層進行判斷。最後每一個畫面若有符合所有分類層的區塊將被框選，如圖三所示，裁切後進入下一個辨識階段。

表一、電腦鏡頭規格。

parameter	
resolution	640x480
exposure mode	auto
white balance	4000
exposure	-5
contrast	32
saturation	64
gamma	100
brightness	0



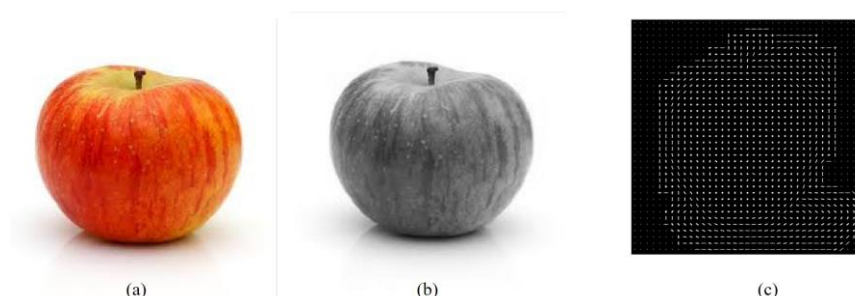
圖二、Cascades 架構概念(擷取自[3])。



圖三、以 Cascades 追蹤偵測人臉。

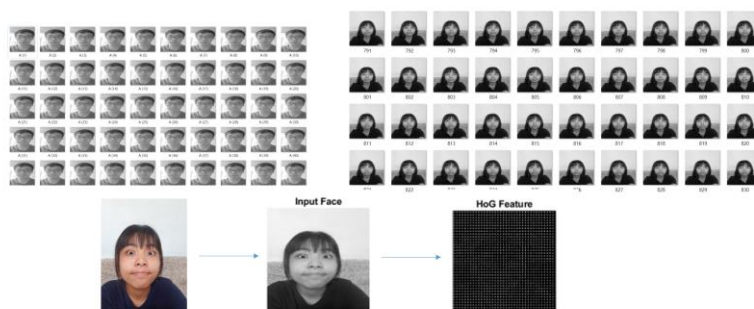
三、特徵學習與人臉辨識

接著我們要建構一個能辨識人臉的 AI 模型，我們先使用機器學習的特徵學習 (Feature learning) 來建構。特徵學習為將原始圖像轉換為能讓機器學習的特徵。常見的特徵演算法有局部二值模式 (Local binary patterns, LBP)、尺度不變特徵轉換 (Scale-invariant feature transform, SIFT) 以及方向梯度直方圖 (Histogram of oriented gradient, HOG) 等。本次我們讓機器所學習的特徵為 HOG [4]，此特徵的擷取方式為將圖像分割成許多小塊，每個區域都會生成一個直方圖，為計算出該區域的梯度與方向值，可用於描述影像的結構與形狀，圖像特徵擷取如圖四所示。



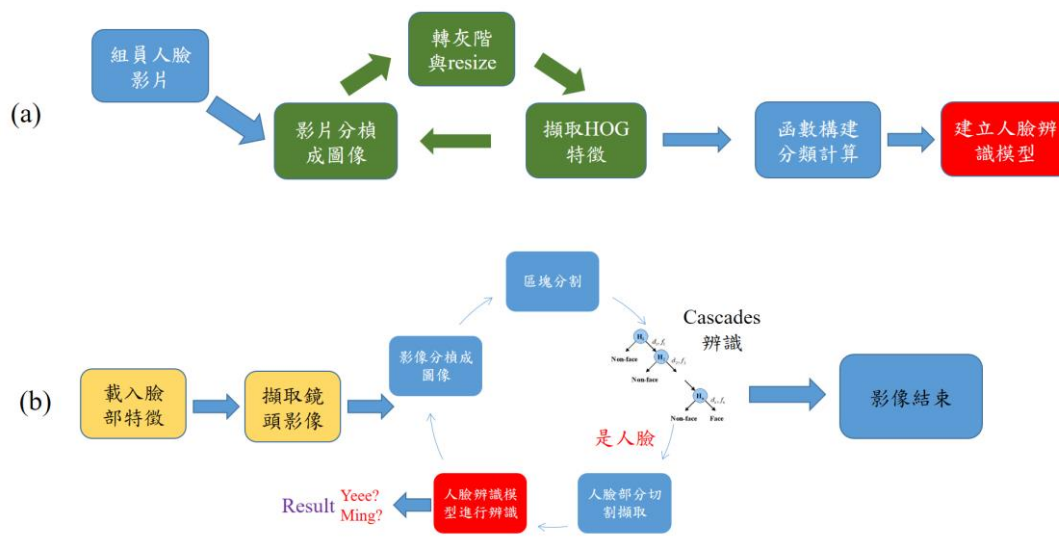
圖四、(a)原始影像，(b)灰階影像，(C)目標 HOG Features。

要建立模型前我們須有訓練的資料庫，資料庫為用於訓練模型的預先準備資料，並以人工的方式將每一筆資料做分類標籤，在這裡每一類標籤即為誰的人臉，如圖五所示。在特徵學習的部分我們一共要辨識兩張臉，為了能更快收集訓練集，我們藉由錄製一段臉部有表情變化的影片，再藉由分幀成一張張圖像，如此一來我們在很短的時間內就可以收集到每類超過 1500 張的訓練集，省下一張張拍照蒐集資料庫的時間。接著將每一張圖片轉成灰階與 resize 成統一大小後，提取 HOG 特徵，然後利用支持向量機分類器演算法建立出人臉辨識模型，建立模型程式流程如圖六(a)所示。



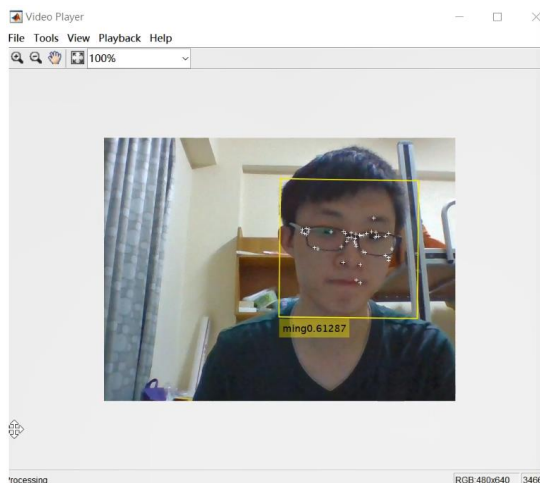
圖五、特徵學習建立之訓練資料庫。

辨識的部分我們接續上一章節的步驟，鏡頭捕捉到的每一幀畫面進入迴圈，經 cascades 判斷出是否有人臉，並將人臉切割出來。切割出的人臉再送入我們建立好的模型進行判斷是誰的人臉，重複步驟直到鏡頭停止捕捉畫面。辨識程式流程如圖六(b)所示。



圖六、特徵學習程式流程，(a)模型建立，(b)影像辨識。

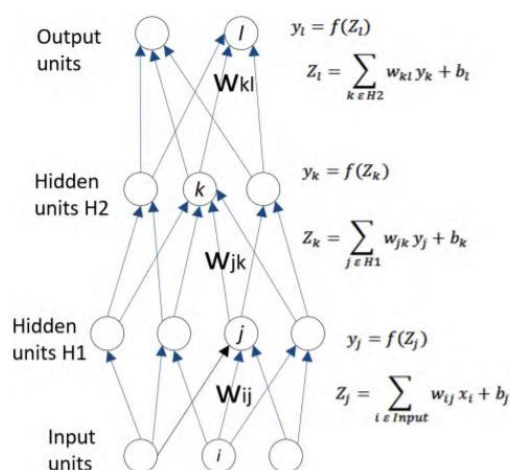
最後辨識的畫面如圖七所示，因為有了模型後辨識的速度可以十分快速，因此能達到實時辨識的條件。接著我們要去定量分析出該模型的準確度，我們將不在訓練集的人臉照片 100 張做為測試集載入模型進行辨識。最終辨識準確率是 100%，雖然辨識都正確，但是平均預測結果的 binary loss 高達 0.3032，這代表模型對它的預測結果沒有自信，而且測試集的圖片都是完整的臉讓它去辨識，背景也較為乾淨。在測試過程中發現若在臉部加入一些其他資訊，像是戴了口罩或戴了墨鏡，辨識準確率就會下降許多。目前辨識的種類只有兩項，我們有嘗試建立辨識更多人臉的模型，但準確率將會變得很低。因此使用 HOG 建立的 AI 辨識模型雖然可行，但還是有它的限制在。



圖七、人臉辨識畫面。

四、深度學習與人臉辨識

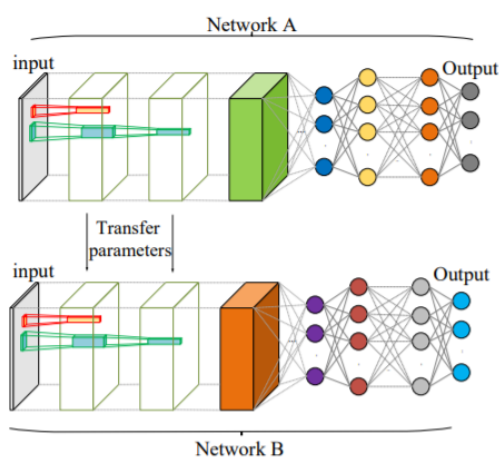
為了建構出更精準的人臉辨識模型，我們接著嘗試使用深度學習的方式去建立[5]。深度學習是一種以人工神經網路架構建立而成的演算法。相較於機器學習由人去定義機器所要學習的特徵，深度學習是在訓練中自己去找出自己所需學習的特徵。一般深度學習網路架構是由輸入、輸出層與中間夾雜的許多隱藏層所建立而成，如圖八所示。建立一個深度學習網路會先去定義它的神經網路架構，接著定義它的學習目標來判斷它學習成果的好壞，像是方均跟誤差(RMSE)或是交叉熵(cross entropy)等，最後在訓練的過程中不斷地去調整架構中每一層的參數與權重，最後優化出最適合的數值組合。



圖八、深度學習網路基本架構(擷取自[5])。

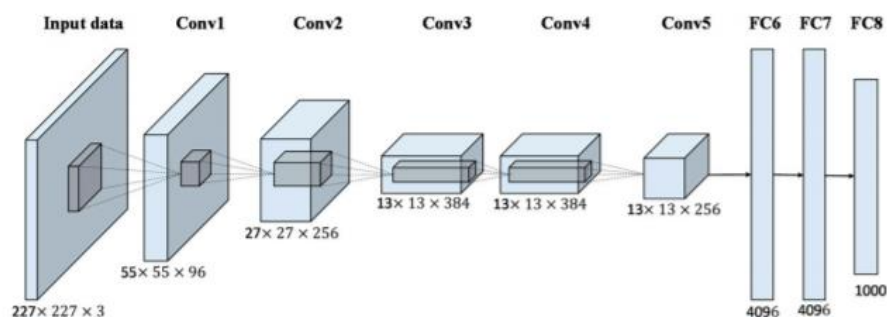
本次我們使用遷移式學習的方式來建構模型[6]。遷移式學習的概念為部分使用別人已經建立好的模型與參數，即使別人分類的種類可能與自己不同，但還是能夠用來訓練自己想要分類的模型。之所以遷移式學習能夠辦到的原因

為，要分類的圖像的特徵在一定程度上都會與別人相似，可以想像為今天我們一開始學的是鋼琴，但是我們可以把學鋼琴得經驗拿去學電子琴或是其他鍵盤樂器。遷移式學習的優點為，引用別人已經架設很好的模型可以省去自己架設時調整架構與參數嘗試的時間，也可以避免自己架設的架構不好，又或者假如自己手上沒有足夠多的訓練集訓練模型，引用別人已經由大量資料訓練的模型參數也是很好的選擇。



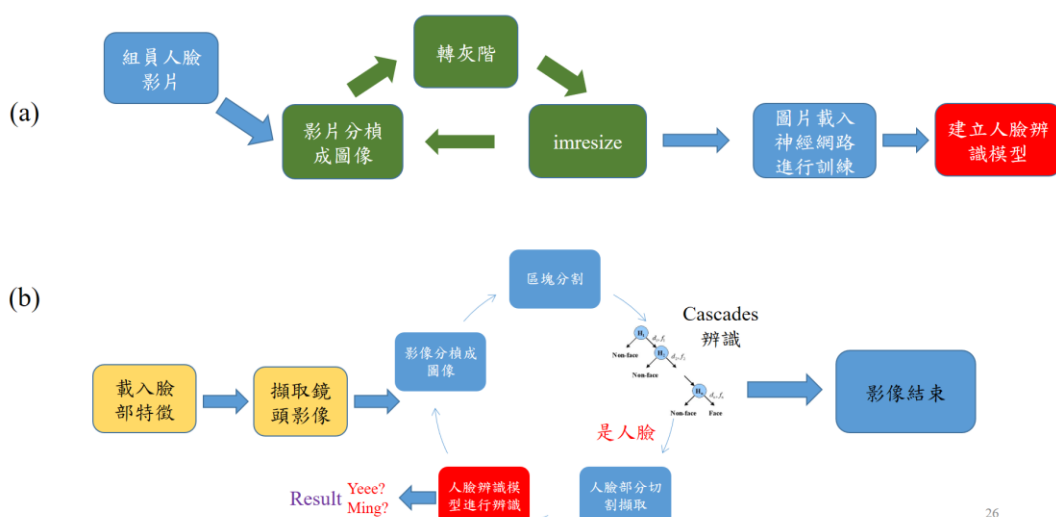
圖九、遷移式學習網路概念(擷取自[6])。

本次我們所使用的遷移式學習模型為 AlexNet [7]。AlexNet 於 2012 年被提出，在當時是一個跨時代的深度學習模型，它是由分成 1000 類的 1400 萬張圖片訓練集所建構而成的卷積神經網路架構，且在分類 1000 類圖片中準確率可達到 85%，因此是一個十分優秀的模型。AlexNet 的架構如圖十所示。它是由輸入、輸出層以及五層卷積層、池化層跟三層全連接層所構成。卷積層為讓特徵遮罩與影像去做卷積計算，藉以得到圖像的特徵圖。池化層則是挑選出特徵圖中最大值，即保留最重要的特徵。最後全連接層將特徵平坦化，將特徵依照權重做最後的分類。



圖十、AlexNet 架構(圖片擷取自[7])。

有了神經網路模型後我們便可以開始訓練。蒐集訓練集的方式也是由影片中分幀擷取。但為了減少訓練時間我們將分幀間距拉長，每一類的訓練集有 300 張以上。訓練集如圖五所示，但這次我們不需要預先提取圖像特徵，並且我們將要辨識的臉增加為三類。辨識模型建立的流程與前一章相似如圖十一所示，訓練的優化函數設定為 adam，minibatchsize 為 10，迭代次數 10 次，initial learning rate 為 0.0001，訓練過程與結果如圖十三所示。

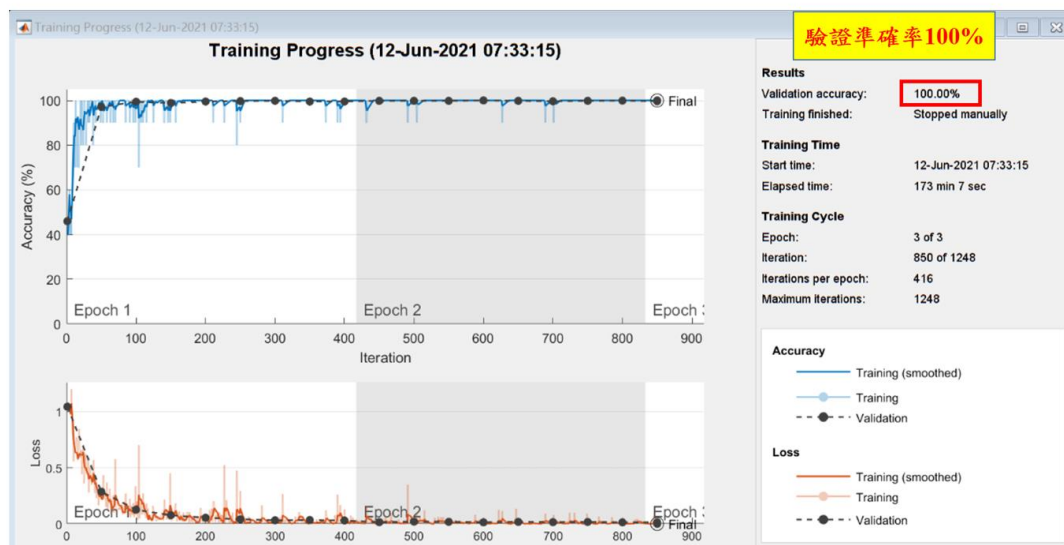


圖十一、深度學習程式流程，(a)模型建立，(b)影像辨識。

Confusion Matrix

	feng	ming	yee	
feng	365 35.0%	0 0.0%	0 0.0%	100% 0.0%
ming	0 0.0%	309 29.7%	0 0.0%	100% 0.0%
yee	0 0.0%	0 0.0%	368 35.3%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
	feng	ming	yee	
	Target Class			

(a)

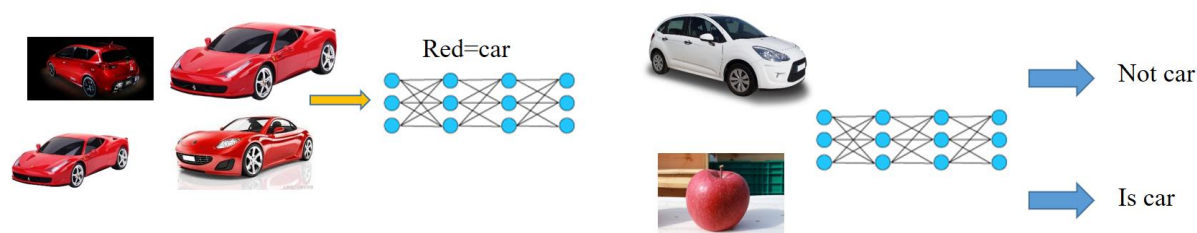


(b)

圖十二、深度學習訓練過程與結果，混淆矩陣與訓練過程。

訓練完的模型我們一樣使用 100 張不在訓練集的圖片當作測試集去做辨識，最後模型測試準確率十分不佳，只有 69% 的準確率。但是在圖十二混淆矩陣與訓練過程中驗證集的準確率卻都有到達 100%，同時我們可以觀察到訓練過程中準確率上升的非常快，因此我們推測模型不準確的原因是因為過擬合導致。

導致過擬合的可能原因為訓練集的圖片相似度太高，概念如圖十三所示。假設今天我們訓練的圖像都是紅色的車，因此我們的模型就很輕易地認為紅色的物體就是車，當我們今天拿紅色的蘋果給模型辨識，模型也會認為是車，反而認為白色的車不是車。我們使用得訓練集(如圖五)是由影片分幀擷取影像，每一張人臉圖片差異度不大且都是在同一背景下拍攝，這使的模型是在「舒適圈」所建立，當今天給模型看一個與訓練集差異較大的圖片，模型便無法準確的辨識。



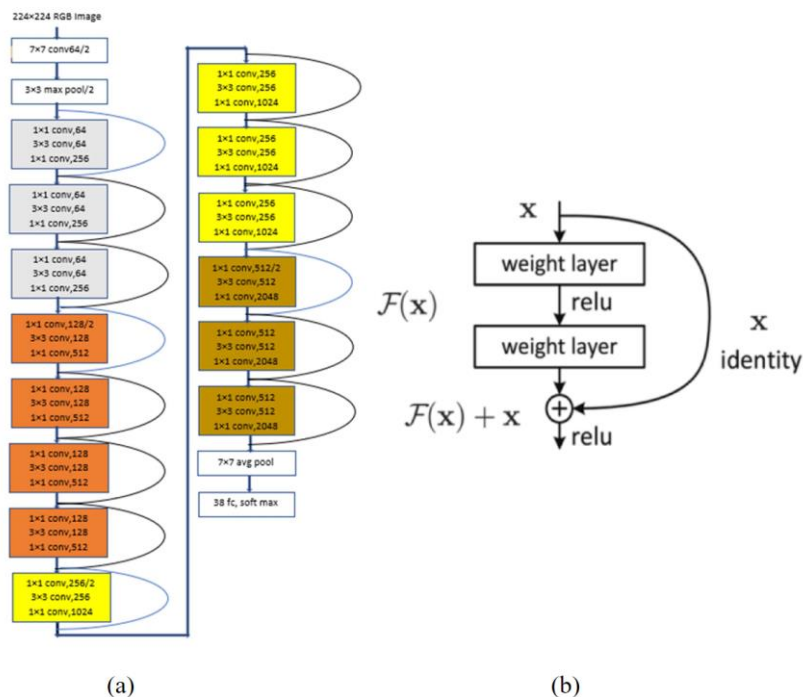
圖十三、過擬合概念。

為解決過擬合的問題我們增加訓練集的多樣性，首先我們將影片擷取改為每 30 幀只截取一張圖像，並在訓練集增加其它除了只有臉部的照片，像是不同背景下的全身照片、不同服裝、臉部局部放大等，藉此增加訓練集的多樣性，如圖十四所示。接著我們將訓練的模型改成 ResNet50[8]，如圖十五所示，藉由架構中的

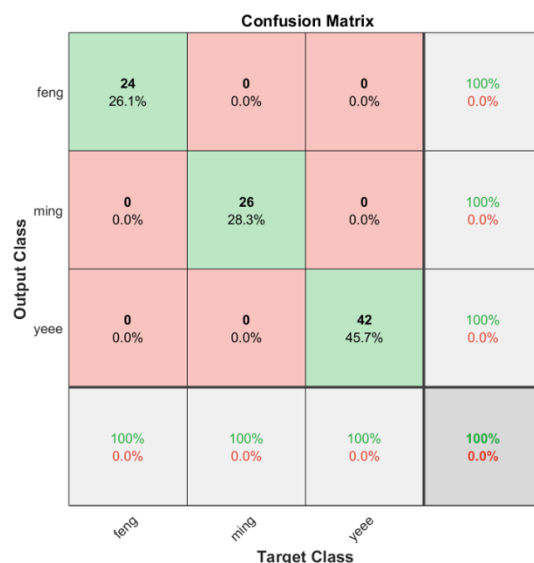
殘差層可以使在增加神經網路深度同時又可以使較深的層數都拿到該有的梯度，避免網路退化。同時我們在訓練過程中加入資料擴增，包括圖像隨機翻轉、平移、旋轉等，因為這對訓練來說都會是新的一張圖片，也可以增加訓練集的數量與多樣性，訓練的過程與結果如圖十六所示。



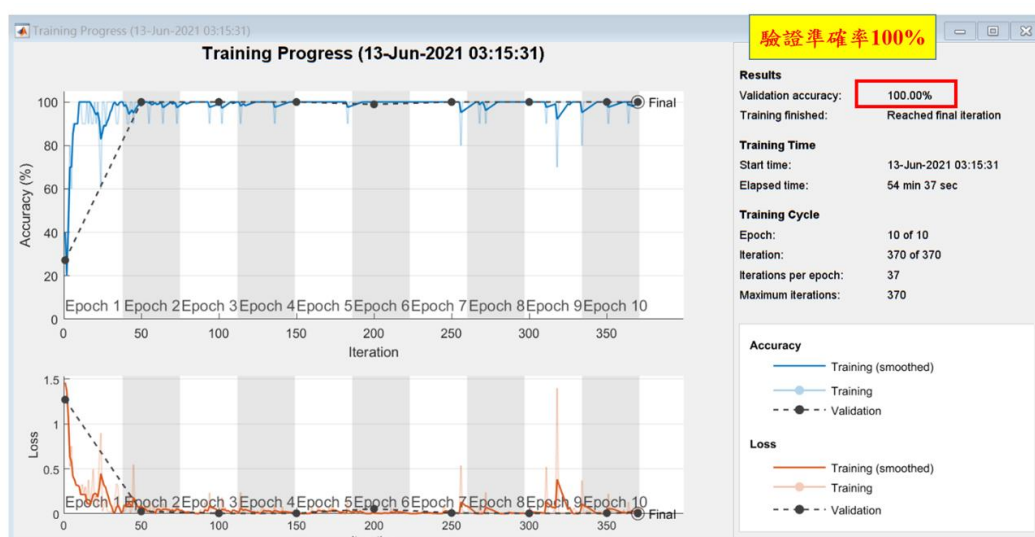
圖十四、增加多樣性後的訓練集。



圖十五、(a)ResNet50 架構(b)殘差層概念示意圖(擷取自[8, 9])。



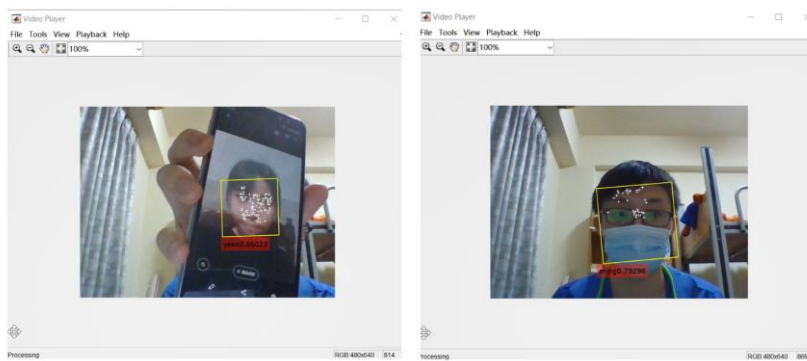
(a)



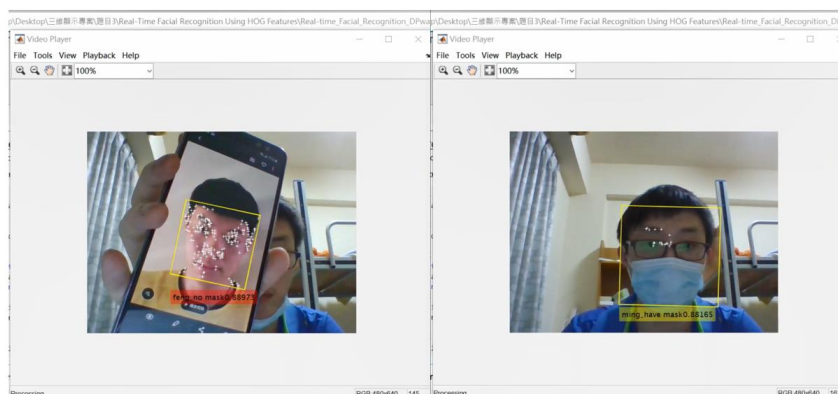
(b)

圖十六、優化後深度學習訓練過程與結果，(a)混淆矩陣，(b)訓練過程。

最終訓練出的辨識模型我們一樣用 100 張測試集去做測試，準確率達到了 100%，且每項辨識結果分類分數都接近 1，代表模型對於此項分類結果十分確定。圖十七為模型實際應用在即時人臉辨識，即使戴上口罩後也能準確的辨識，準確率也十分穩定。圖十八我們在系統中增加新的功能判斷可以判斷這個人是否有戴口罩。雖然深度學習訓練建立模型的時間約要好幾小時，但是有了模型辨識便能快速即時且準確。



圖十七、深度學習人臉即時辨識。



圖十八、深度學習即時辨識，增加判斷是否戴口罩。

五、結論

本專題我們分別使用特徵學習與深度學習建立了即時 AI 人臉辨識系統，雖然在模型的建立上需要花較多時間，但是有了模型後便能對影像作快速即時的辨識。使用 HOG features 建立特徵學習模型雖然只需要兩分鐘左右，相較於深度學習建立的模型需要一小時以上花費時間較少，但是深度學習模型得辨識率比特徵學習的模型好很多。建立深度學習的訓練集不能只使用單純臉部的照片，訓練集若過於單一會造成模型過擬合，需增加訓練集多樣性或使用資料擴增來防止過擬合，經優化後的深度學習模型辨識率可達到 100%。由於時間的關係我們只辨識了三張人臉，若能夠蒐集更多的訓練集，我們便可以建立辨識更多人臉的模型，未來可以應用在門禁系統、身分核對，判斷是否有戴口罩的功能也能應用在防疫措施上。

六、參考文獻

- [1] Brunette, E. S., Flemmer, R. C., & Flemmer, C. L. (2009, February). A review of artificial intelligence. In *2009 4th International Conference on Autonomous Robots and Agents* (pp. 385-392). Ieee.
- [2] Trigueros, D. S., Meng, L., & Hartnett, M. (2018). Face recognition: From

traditional to deep learning methods. *arXiv preprint arXiv:1811.00116*.

- [3] Kang, Suk-Ju. "Multi-user identification-based eye-tracking algorithm using position estimation." *Sensors* 17.1 (2017): 41.
- [4] Kobayashi, Takuya, Akinori Hidaka, and Takio Kurita. "Selection of histograms of oriented gradients features for pedestrian detection." *International conference on neural information processing*. Springer, Berlin, Heidelberg, 2007.
- [5] Shrestha, Ajay, and Ausif Mahmood. "Review of deep learning algorithms and architectures." *IEEE Access* 7 (2019): 53040-53065.
- [6] Lemley, Joseph, Shabab Bazrafkan, and Peter Corcoran. "Transfer Learning of Temporal Information for Driver Action Classification." *MAICS*. 2017.
- [7] Han, Xiaobing, et al. "Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification." *Remote Sensing* 9.8 (2017): 848.
- [8] Mukti, Ishrat Zahan, and Dipayan Biswas. "Transfer learning based plant diseases detection using ResNet50." *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*. IEEE, 2019.
- [9] Theckedath, Dhananjay, and R. R. Sedamkar. "Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks." *SN Computer Science* 1.2 (2020): 1-7.

