

111 學年度專題報告競賽

題目：基於人臉辨識的點名系統

系所班別：統計學系

姓名學號：410874246 (王廷宇)

410878013 (王宣儒)

410878015 (吳宥融)

410878064 (黃博宣)

410878065 (林明宏)

410878066 (葉育豪)

報告日期：2023 / 06 / 09

指導教授：汪群超 教授

專題競賽報告分工表

學號	姓名	分工內容
410874246	王廷宇	人臉辨識系統模型製作
410878013	王宣儒	主講人、資料蒐集
410878015	吳宥融	簡報製作、主講人、資料蒐集
410878064	黃博宣	人臉辨識系統模型製作
410878065	林明宏	人臉辨識系統模型製作
410878066	葉育豪	組長、書面報告撰寫、外部設備、簡報製作

摘要

在這篇研究文章中，我們探討了利用人臉辨識技術來加速教師點名的速度。我們希望解決傳統點名方法所需的長時間以及其效率低下的問題，期望透過應用人臉辨識的技術，使點名過程變得更加迅速且準確。為此，我們打算設計一個基於人臉辨識的點名系統。

在分析方法部分，我們分為人臉辨識和人臉偵測兩個部分。在人臉辨識方面，本研究首先選擇使用與機器學習相關的統計方法，包括 K-近鄰算法(K-Nearest Neighbors, KNN)、支持向量機(Support Vector Machines, SVM)、隨機森林分類器(Random Forest Classifier, RFC)，以及深度學習的方法包括 Facenet 和 ArcFace 的深度殘差網路(ResNet)進行人臉辨識。人臉偵測方面，我們實驗了哈爾特徵(Haar features)、主成分分析(Principal Component Analysis, PCA)、線性判別分析(Linear Discriminant Analysis, LDA)、局部二值模式(Local Binary Patterns, LBP)、方向梯度直方圖(Histogram of Oriented Gradients, HOG)和單一尺度特徵金字塔(Single Shot MultiBox Detector, SSD)進行人臉偵測。

在經過測試之後，Facenet 的實驗中達到了約 90%的準確率，而 ArcFace 的實驗準確率約 85%。在人臉偵測方面，我們發現 HOG 和 SSD 能最準確地偵測到人臉，而 HOG 能更快速地偵測到人臉，且能夠擷取到極端的人臉，這更符合我們點名系統的需求。因此，我們最終選擇使用 Facenet 和 HOG 來設計一個人臉辨識點名系統。該系統將使用手機鏡頭來進行人臉辨識。另外，我們還在此人臉辨識系統中加入了語音功能，教師只需要使用手機掃描學生的臉部，學生則能透過語音獲知自己是否被點名，系統就能自動識別出學生的身份，並記錄他們的出席情況。

目錄

第一章 緒論	1
1.1 前言	1
1.2 研究動機與目的	1
1.2.1 研究動機	1
1.2.2 研究目的	2
第二章 文獻探討	2
2.1 人臉偵測	2
2.2 人臉辨識	3
第三章 分析方法	7
3.1 資料收集與處理	7
3.2 統計方法人臉辨識模型	8
3.3 卷積神經網路模型	8
3.3.1 CNN	8
3.3.2 Arcface 模型實作	10
3.3.3 Facenet 模型實作	10
第四章 點名系統與實驗設備	12
第五章 結果分析與討論	13
第六章 結論	16
參考文獻	16

圖目錄

圖 2-1 損失函數 "Triplet Loss" 示意圖，出自 Florian.S 等人 (2015)	5
圖 2-2 Facenet 的網路架構出自 Florian.S 等人 (2015).....	5
圖 2-3 ArcFace 計算在 2D 投影下的對比示意圖，出自 Deng, J.等人 (2018) [12]	6
圖 3-1 卷積神經網絡運作原理，出自 Florian.S 等人 (2015)	9
圖 3-2 GoogLeNet 整體架構，出自 C Szegedy 等人 (2015)	11
圖 3-3 神經網路的主幹，出自 C Szegedy 等人 (2015)	11
圖 5-1 FaceNet 不同閾值之準確率折線圖	15

表目錄

表 5-1 ArcFace 不同方法之準確率情形	14
表 5-2 FaceNet 不同閾值之準確率情形	15

第一章 緒論

1.1 前言

人臉辨識一直以來都是計算機視覺和模式識別領域的重要研究方向之一。隨著科技的發展，人臉辨識技術廣泛應用於安全監控和身份驗證等領域。準確且高效的人臉辨識系統對於現代社會的生活和安全至關重要。

本文旨在設計和實現一個基於 **Facenet** 和 **ArcFace** 的人臉辨識點名系統，將深入研究 **Facenet** 和 **ArcFace** 的原理和技術細節，包括訓練過程和特徵表示。接著，本文將探討數據集的蒐集、選擇和預處理方法，以確保訓練和測試數據的質量，並設計並完成一個完整的人臉辨識點名系統，包括人臉偵測、特徵提取和人臉匹配等關鍵模型。最後，本文將對所提出的人臉辨識點名系統進行評估和測試，並與其他的方法進行比較。

本文旨在實現一個高準確性、高效率且實時性的人臉辨識點名系統，以滿足在課堂中的人臉辨識點名需求。

1.2 研究動機與目的

1.2.1 研究動機

傳統的課堂點名方式耗時且容易產生錯誤和冒名頂替行為，這些問題對課堂教學效率和課堂管理帶來挑戰。為了解決這些問題，本文提出利用人臉辨識技術進行點名，這種方法不僅提高了點名效率，也增強了點名的準確性和減少冒名頂替行為。

人臉辨識系統利用深度學習模型識別學生的獨特面部特徵，實現在各種情況下的準確人臉辨識。此外，由於每個人的面部特徵都是獨一無二的，人臉辨識系統能有效防止冒名頂替的情況，提高點名記錄的真實性和效率。因此，研究和開發一個準確、高效且可靠的人臉辨識點名系統具有重要意

義，並有望在實際應用中產生廣泛影響。本研究旨在探索此技術的可能性，並嘗試實現其在課堂點名中的應用。

1.2.2 研究目的

本研究旨在開發基於人臉辨識的點名系統，以提高課堂點名的效率。希望設計一個高效的人臉辨識系統，能快速辨識並驗證學生身份，解決傳統點名方式的困難和限制。透過引入先進的人臉辨識模型，提升辨識準確率並確保點名的準確性和可靠性。本文將建立完整的點名系統，包括人臉圖像擷取、特徵提取、身份驗證和點名記錄等功能。最終我目標是將研究成果應用於課堂中，提高課堂點名效率，並探索其他潛在應用領域，如企業的考勤管理和會議出席確認，將網路上資源整理合併成一個完整的產品，應用在上述的各個場景中。

第二章 文獻探討

人臉辨識的研究最早於 1960 年代，而隨著時間的演進，人們也開發出許多不同的方法進行人臉辨識的相關研究，而人臉辨識主要又分為人臉辨識以及人臉偵測，本文將會分成這兩個面向進行相關文獻的探討。

2.1 人臉偵測

人隨著人臉辨識技術的進步，人臉偵測的準確性已成為其發展的重要前提。為此，學者們提出了多種方法以改善人臉偵測的準確性。在 20 世紀末，Linear Discriminant Analysis (LDA) 成為人臉偵測技術的主流方法。作為一種監督式學習的降維技術，LDA 旨在最大化類別間的變異，同時最小化類別內的變異，進而提高識別的準確性【1】。

進入 21 世紀，人臉偵測技術迎來了新的發展。Paul Viola 和 Michael Jones (2001) 提出了 Haar 特徵，一種基於人眼對亮度變化的敏感度的新型人臉偵測方法【2】。Ojala 等人 (2002) 提出了 Local Binary Patterns (LBP)，一種有效的紋理特徵描述符【3】。Navneet Dalal 和 Bill Triggs (2005) 提出了 Histogram of Oriented Gradients (HOG)，一種用於對象識別的特徵描述子，其在人臉偵測和識別上展現出良好的效果【4】。

更進一步，Wei Liu (2016) 等人提出了 Single Shot MultiBox Detector (SSD)，一種在準確性和速度之間取得良好平衡的物件檢測演算法【5】。此算法的主要特點是只需要一次前向傳播即可完成所有的預測，因此在實時物件檢測任務中，SSD 展現出高效率，並能有效識別出圖片中的人臉位置。

隨著時間的演進以及機器學習技術的發展，人臉偵測的準確性逐漸提高。Adrian Rosebrock (2021) 在論壇中比較了 Haar cascades、SSD、HOG 以及 CNN 臉部檢測器的性能【6】。他認為，SSD 和 HOG 是目前較好的臉部偵測器，因為它們在速度和準確性上達到了良好的平衡，並可以在現代的筆記本電腦和桌面電腦上實時運行。

2.2 人臉辨識

人臉辨識的發展也隨著時代的變化有了許多不同面向的應用，Matthew Turk 和 Alex Pentland (1991) 使用基於主成分分析的 Eigenface 進行人臉辨識【7】，這種方法將臉部圖像分解為一小組特徵圖像，這些特徵圖像可以被視為臉部圖像的主要成分。識別是通過將新的圖像投影到由 eigenfaces 構成的子空間（也被稱為“臉部空間”）中，然後通過比較其在臉部空間中的位置與已知個體的位置來進行分類。這種方法具有速度和簡單性、學習能力以及對臉部圖像的小變化或漸變變化的不敏感性等優點。不過此種方法很容易被噪聲外在因素影響，例如，在變化的光照條件下，未知率提高到 19%，對於方向變化，未知率為 39%，對於大小變化，未知率為 60%。當將未知率設定為 20% 時，正確識別率分別為 100%，94%，和 74%。這種方法的準確性會隨著閾值的降低而提高，但這也意味著有更多的圖像會被識別為未知。

進入 21 世紀，學者們就已經引進了 K-近鄰演算法（K-Nearest Neighbors，K-NN）這一結合機器學習和統計方法的工具來進行人臉辨識。KNN 方法會將偵測到的人臉分類到與其 k 個最近鄰居中最常見的類別。其核心是透過比較人臉特徵向量的距離以實現辨識的目的。Ni Kadek Ayu Wirdiani（2019）使用了 K-Nearest Neighbor（KNN）方法進行臉部辨識的訓練和測試【8】。在訓練階段，他們使用了 790 張圖像，這些圖像包含 158 個類別，每個類別有三張訓練圖像和兩張測試圖像。他們的訓練和測試過程分為兩個部分，第一部分使用 KNN 方法，準確率為 84.85 %。另外，他們使用了不同的 k 值進行參數調整，並觀察其對準確率的影響。當 k = 1 時，準確率為 81%；當 k = 2 時，準確率為 53 %；當 k = 3 時，準確率為 47 %。

與此同時，學者們也採用支持向量機（Support Vector Machines，SVM）來進行人臉辨識。SVM 是一種強大的分類和回歸分析工具，它的目標是尋找一個超平面，以最大化不同類別之間的邊界。在人臉識別中，SVM 可用於對人臉特徵向量進行分類。Phillips, P. J.（1998）使用 SVM 進行人臉辨識，準確率達到了 77 % - 78 %。【9】而隨機森林分類器（Random Forest Classifier，RFC）也被用於人臉辨識。RFC 運行多個決策樹，並採取多數投票結果。在人臉辨識中，RFC 可對特徵進行分類，並考慮多種可能的特徵組合。Emir Kremic 和 Abdulhamit Subasi（2015）使用隨機森林進行人臉辨識，準確率達到了 97.17 %。【10】

2014 年 Facebook 的 AI 團隊 Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. 研發 Deepface 人臉辨識系統。該系統使用深度學習技術，設計了一種深度神經網路架構，將人臉識別問題轉變為一個分類問題，大幅提高了人臉識別的準確性。【10】

Google 於 2015 年 CVPR 發表 Facenet 演算法【11】，其中使用 LFW（Labeled Faces in the Wild）資料集，該資料集包含有 5749 個人臉圖片，總圖片個數超過一萬三千多張，是一個相當龐大的資料集。最後 Facenet 演算法在 LFW 人臉資料庫以 99.63 % 的最佳成績刷新了記錄，有別於當時主流的人臉辨識方式是針對不同人物去提取特徵與編碼，再取得特徵後最後一層接上 Softmax 做分類，而 Facenet 的訓練目標則是訓練一個高維度變換空間，訓練完成後它可以直接將人臉圖像映射至歐氏空間，透過比對兩個人臉圖像映射（embedding）的歐幾里德距離能直接反應

出人臉間的相似度，對於曾未出現在資料集中的人臉也能藉此透過映射轉換後取得人臉特徵，並可接上任意分類器做分類，像是常見的 KNN、SVM 等分類方式。Facenet 演算法中使用的損失函數並非是常見的 "cross - entropy"，取而代之的是 "Triplet Loss"，也是此方法最大的特色，核心概念是讓同一個人的臉在映射後的特徵其距離越小越好，而不同人物的臉部特徵則是越遠越好，圖 2.1 為論文提供的示意圖，Anchor 代表欲辨識的目標人臉，Positive 代表和 Anchor 為同一人的另一張臉部圖像，反之 Negative 與 Anchor 為不同人的臉部圖像。尚未經過學習時，原空間的 Anchor 與 Negative 可能距離會比較相近，但經由訓練之後讓 Anchor 與 Positive 的距離拉近，而 Anchor 與 Negative 的距離變遠，而本文接下來會使用非線性函數（Embedding）做為轉換原始人臉圖像的方式[參照圖 2.1]。

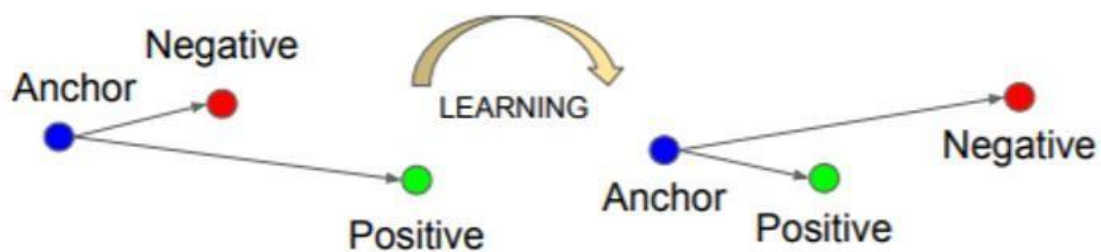


圖 2-1 損失函數 "Triplet Loss" 示意圖，出自 Florian.S 等人 (2015)

Facenet 的網路架構，首先將資料集分成幾組批量 (batch)，實驗中對於訓練資料進行取樣，讓每個批量中的身份大約包含 40 張臉孔，此外，也對 Negative 隨機取樣並新增到每個 batch 中，後續接上一個深度學習框架，以及 L2 正規化並以目標函數 triplet loss 做訓練 [參照圖 2.2]。



圖 2-2 Facenet 的網路架構出自 Florian.S 等人 (2015)

Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2018) 發表 Deep Residual Learning for Image Recognition 【12】，內文介紹 ArcFace 是一種基於角度餘弦補釘 (Angular

Cosine Margin) 損失函數的人臉辨識方法，提升人臉辨識的準確性和適應性。其中使用了 ImageNet 數據集，該數據集超過 1400 萬張圖片，這些圖片都被分類並標註了相應的物體標籤，也是專為訓練深度學習模型而設計。ArcFace 演算法在 ImageNet 的測試集上達到了僅 3.57 % 的錯誤率，這是一個非常高的準確率。

ArcFace 的運作原理包括以下幾個步驟：首先，利用深度學習模型（如 ResNet）對人臉圖像進行特徵提取，將人臉圖像映射到高維特徵空間，其中每個特徵向量代表一個人臉的特徵表示。然後，為了增強特徵的鑒別性，ArcFace 會對特徵向量進行正規化，使其長度為 1，將特徵表示映射到高維球面上。最後，ArcFace 引入了角度餘弦補釘損失函數，這是一種針對人臉辨識任務設計的特殊損失函數。它通過最大化正確類別的得分與其他類別的得分之間的差距，使得同一個人臉的特徵表示更加接近，不同人臉的特徵表示更加分散。ArcFace 在特徵空間中定義了一個球面中心，並將每個類別的特徵表示點與該球面中心相連。

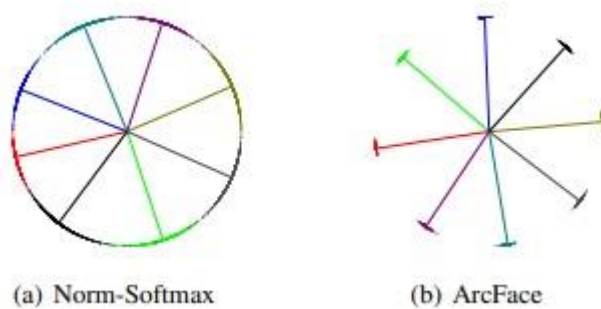


圖 2-3 ArcFace 計算在 2D 投影下的對比示意圖，出自 Deng, J.等人 (2018) [12]

第三章 分析方法

3.1 資料收集與處理

在這篇論文中，集中探討如何使用基於 FaceNet 和 ArcFace 的深度學習技術來進行人臉識別。為了確保本文的模型具有足夠的數據來學習並達到準確的辨識結果，需要進行一系列的資料收集與處理。以下將詳細描述如何收集數據，以及這些數據如何被處理以優化人臉識別系統。

首先，本文挑選了 32 位自願參與的受試者，這些受試者包括學校同學和組員的家人。另外，本文確保這些受試者在年齡、性別等方面有一定的多樣性，以便讓模型可以在各種條件下都能有良好的識別效果。對於每一位受試者，本文收集了 20 張照片作為訓練數據。這 20 張照片包括四種不同的情緒表現（喜怒哀樂各一張）以及臉部的 8 個不同角度各兩張。這種資料收集方式讓模型有機會學習到人臉在不同情緒和角度下的變化，訓練集總共 640 張照片（參見附錄 1.1）。另外，本文同時也收集了 32 位受試者的 15 張隨機照片當作測試集，測試集總共 480 張（參見附錄 2）。

在收集完所有的照片後，本文進行了一系列的前處理步驟。首先，將每一位受試者的照片分類並存放至一個資料夾下的獨立的子資料夾中，並給予子資料夾的編號作為這位受試者的標籤（label）。這種方式使得在訓練模型時可以輕鬆地識別出每一張照片對應到的受試者。

接下來，把所有的照片轉換為灰階圖像（參見附錄 1.1）。儘管這種轉換會使圖像失去一些顏色資訊，但是灰階圖像的計算量只佔原圖 RGB 三通道的三分之一，可以增加人臉偵測抓取人臉的速度。

先前文獻探討的部分提到使用 SSD 和 HOG 人臉偵測較為優秀，不過礙於硬體設備的問題選擇使用 Histogram of Oriented Gradients（HOG）的方法對灰度圖像進行特徵提取。HOG 擅長捕捉影像的紋理信息，且對影像的平移和縮放具有優良的穩定性。利用 HOG 的手法，本文對 640 張照片進行了人臉檢測。然而，由於某些照片的拍攝角度不佳，本文實際上只在 636 張照片中偵測到了人臉。對於這 4 張

偵測失敗的照片，如果將它們納入訓練模型，可能會降低辨識的成功率。因此，本文手動移除這 4 張照片，確保訓練集僅包含 636 張人臉照片 (參見附錄

1.1)。

綜上所述，以上資料收集和處理流程取得一個高質量的訓練集，並且對訓練集進行了有效的前處理，以便本文後續的模型可以更好地學習和識別人臉。

3.2 統計方法人臉辨識模型

本文在剛開始探索人臉識別系統時，最先使用 KNN、SVM 和 RFC 統計方法來進行人臉辨識模型的設計，透過這些統計方法進行人臉辨識模型在實驗過後都能達到高達 9 成的準確率，不過當在實際真人測試的時候卻相當不準，準確率約落在 4 成，本文我現這些方法將會把每張測試圖片做分類，因此很常出現長得相似的人而造成辨識錯誤的問題，如果出現資料集內未收集的臉，也會強制分類，且易受到光線及角度變化的影響，這三種方法也就不適合使用於點名系統的人臉辨識模型。於是本文將研究的方向轉向卷積神經網路模型。

3.3 卷積神經網路模型

3.3.1 CNN

本文使用卷積神經網路 (Convolutional Neural Network, CNN) 來實現人臉辨識的功能。CNN 是一種深度學習模型，特別適合處理圖像識別任務，其運作原理如下：

- 卷積層 (Convolutional Layer)：卷積層是 CNN 的核心組成部分，它通過在圖像上滑動一個稱為卷積核的小窗口，計算出每個區域的特徵。卷積核會根據圖像的局部特徵對其進行過濾和提取，產生一個特徵圖 (Feature Map)。通過多個卷積層的堆疊，可以捕捉圖像的不同層次的特徵。

- 激活函數 (Activation Function)：在每個卷積層的特徵圖上應用激活函數，以引入非線性和表達能力，它們可以增強網絡的非線性建模能力。
- 池化層 (Pooling Layer)：池化層用於對特徵圖進行降維和抽取主要特徵。最常見的池化操作是最大池化，它選擇每個區域中的最大值作為池化後的特徵。這樣可以減少特徵的空間尺寸，同時保留重要的特徵信息。
- 全連接層 (Fully Connected Layer)：在經過多個卷積層和池化層之後，將特徵圖展平成一維向量，然後將其輸入到全連接層中。全連接層通常由多個神經元組成，每個神經元都與前一層的所有神經元相連。這些神經元可以學習到不同的特徵組合，並進行最終的分類和識別。
- 損失函數 (Loss Function)：在訓練過程中，需要定義一個損失函數來衡量模型的預測結果和真實標籤之間的差異，進而進行模型的參數優化和更新。對於人臉辨識任務，常見的損失函數包括交叉熵損失和三元組損失。交叉熵損失用於多類別分類，而三元組損失則用於學習如何將同一個人的特徵表示盡量拉近，不同人的特徵表示盡量推遠。

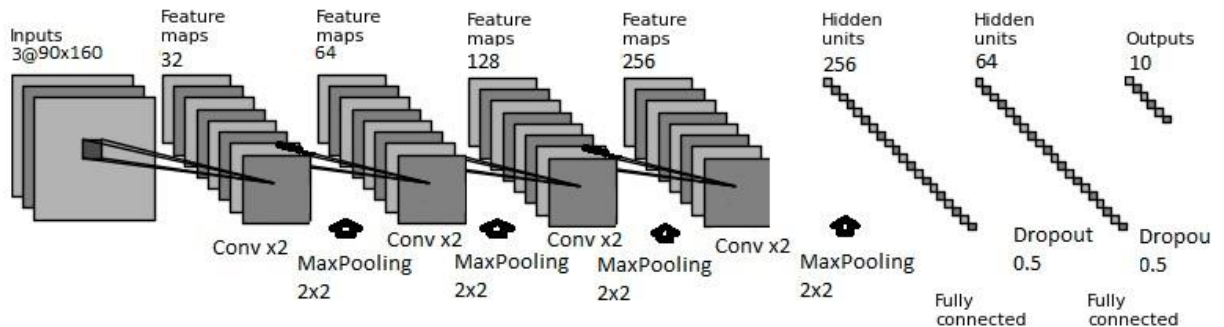


圖 3-1 卷積神經網絡運作原理，出自 Florian.S 等人 (2015)

綜合來說，基於 CNN 的人臉辨識模型具有高辨識率。透過深度學習，它能夠學習到更高層次的抽象特徵，並更好地捕捉人臉的細微差異。此外，CNN 模型具有非線性建模能力和適應性，能夠應對光線、角度和尺度變化等困難。通過共享特徵表示，它能夠實現更準確的人臉辨識。基於這些優勢，基於 CNN 的人臉辨識系統在準確性和適應性方面取得了顯著提升，為人臉辨識技術的發展和應用提供了強大支持。

3.3.2 Arcface 模型實作

DeepFace 是一個以深度學習為基礎的面部識別與分析程式庫。它基於深度神經網路的技術，提供了一個簡易而又強大的人臉識別工具集。

本文運用其中的“find”函式來進行人臉特徵的識別與分析。該功能的運作方式是：輸入一個含有眾多照片的資料夾，然後該功能會分析資料夾下的每一張照片，並把每張照片的人臉特徵輸出成一個特徵向量。值得注意的是，DeepFace 提供多種將照片轉換成特徵向量的方式，本文選擇使用其中的 ArcFace 方法，因為在本文欲達到的需求和使用情境下，ArcFace 提供了相對其他方法更好的結果。

在轉換完特徵向量後，會輸入一張新的照片，然後 DeepFace 將計算這張新照片與資料夾中每張照片的餘弦距離。find 函式中，也可以選擇使用不同的距離計算方式，如歐氏距離等，但考慮到本文選擇使用的 ArcFace 模型是與角度有關的，所以本文決定使用餘弦距離。

在計算完所有餘弦距離後，本文將會挑選餘弦距離最近的前 5 張照片來進行進一步的分析與判斷。

3.3.3 Facenet 模型實作

為實現人臉辨識之目的，本文嘗試採用 Facenet 作為人臉辨識系統的架構之一，透過 Keras_facenet 套件將神經網絡模型匯入，Keras 是一個開源神經網路庫，基於 Python 深度學習的程式庫，在 Keras 下的 Facenet 是基於 MS-Celeb-1M 資料庫預訓練而成，MS-Celeb-1M 是微軟於 2016 年 6 月所公開發佈的人臉資料庫，包含 100 萬個名人總共約 800 多萬張人臉影像。

Facenet 模型的主幹可以採取不同的 CNN 架構，論文中採用的 NN2 架構是基於 Inception v1 (GoogLeNet) 作為基礎建立的深度學習框架，論文中提到它與其它常見的網路架構相比準確率較為出色，它的計算量和參數也比 AlexNet 、

VGGNet 或其他經典神經網路少了許多，而本專題 Facenet 模型使用的主幹則為 Inception-ResNet V1，它是由上述提及之 Inception 模組與 ResNet 結合後的網路架構，下圖三為它的整體架構，圖四則為神經網路的主幹。

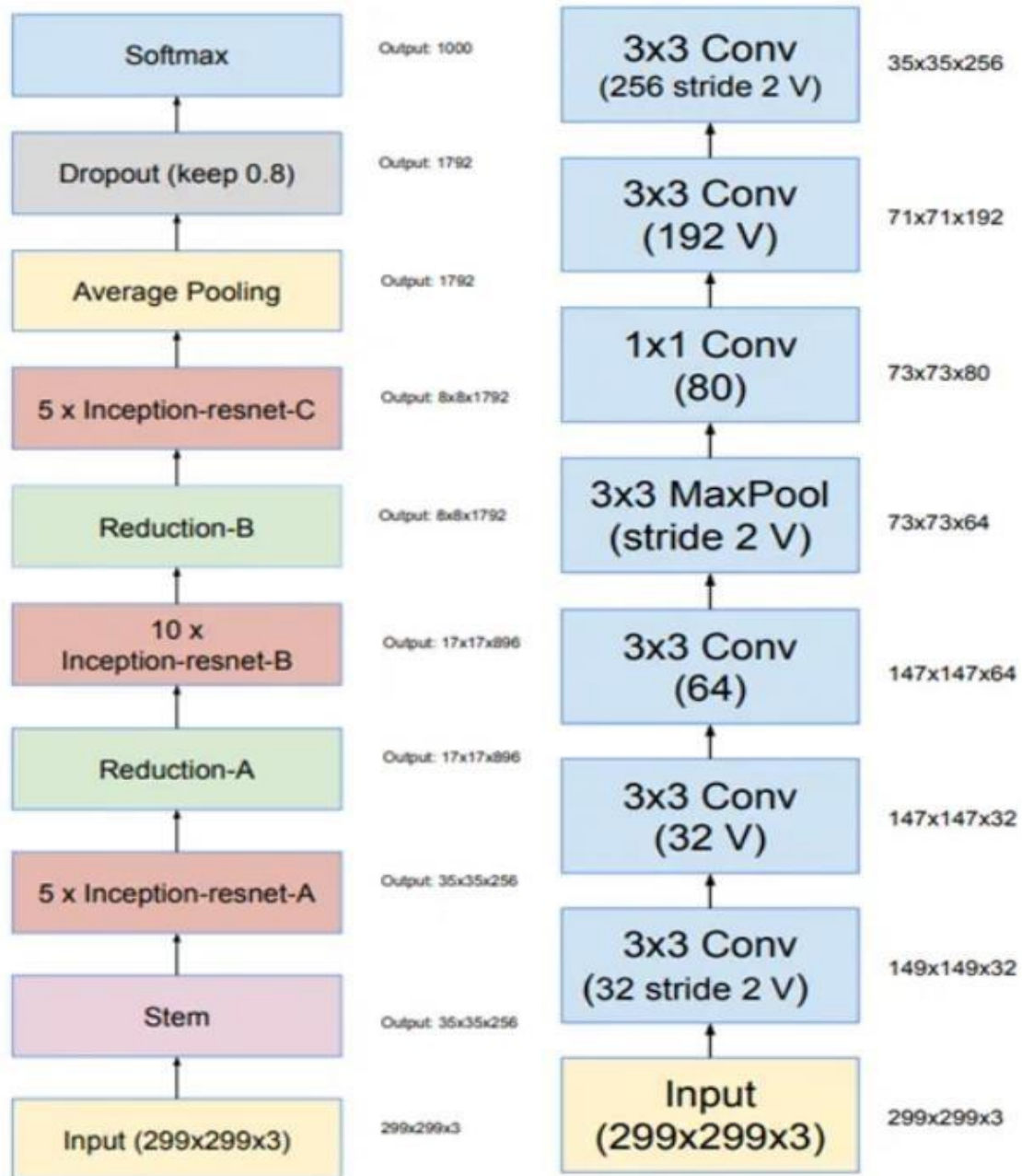


圖 3-2GoogLeNet 整體架構，出自 C Szegedy 等人 (2015)

圖 3-3 神經網路的主幹，出自 C Szegedy 等人 (2015)

如同文獻探討中所提，Facenet 演算法旨在訓練一個高維度轉換空間，期望訓

練後的模型能將輸入的人臉資料藉由映射至歐氏空間，並取得該人物臉部之特徵值，接著依序將訓練集之人臉資料輸入模型，取得訓練集人臉的特徵值，並儲存為一個 Python 字典格式，以供後續測試和現場人臉辨識。

經過整理與刪減後的訓練集內有 32 人共 636 張的人臉照片（參見附錄 1.1），映射至歐氏空間後每張圖片能取得一串 512 個變量的人臉特徵值，最後訓練集的形狀為 636×512 ，測試集則為 478×512 ，擁有訓練集和測試集的特徵值後即可進行分類，分類方式可採取常見的機器學習演算法，像 KNN 或是 SVM。

雖然測試集中的人物皆與訓練集內的人物相同，但在現場人臉辨識點名下，無法保證會不會有非課堂的同學出現，若採用機器學習分類，機器可能會將非這堂課的同學強迫進行標籤，又或者班上有兩位同學的長相極為相似，可能會造成分類上的判斷錯誤，為避免上述情況發生，本文嘗試以計算歐氏距離的方式來替輸入之照片分類，與訓練集內的照片一一比對，若兩者間的距離最小則代表兩人臉的特徵最相近，即返回訓練集內該人物的標籤作為輸出預測，但這樣依然無法避免非這堂課的同學出現的狀況，因此我們會設定一個閾值，若輸入照片與訓練集內照片之最小距離沒有超過該閾值，則會判定為 "???"，代表無法辨識該人物。

第四章 點名系統與實驗設備

在人臉辨識與點名系統研究中，影像擷取和實驗設備的部分也是十分關鍵。接下來，將會詳細介紹本文在這兩個方面所做的工作。

首先是影像擷取。本文使用了 Histogram of Oriented Gradients（HOG）方法來進行人臉偵測，這種方法通常能夠準確地偵測出影像中的人臉區域。然而，在本文的訓練集中，有 4 張照片的人臉區域並未被 HOG 成功識別出來。這可能是由於影像的質量、光線條件、或是人的姿態與臉部表情等因素所引起的。

接下來是實驗設備。主要使用筆記型電腦來進行人臉辨識與點名系統的開發和實現。這種選擇主要基於其便攜性和足夠的計算能力。可以在任何時間、任何地點進行開發工作，並且能夠順利地執行所需的深度學習模型。此外，本文也將手機鏡頭納入了點名系統。利用手機 APP(IP Webcam)，連接手機鏡頭，可以更靈活地調整

取樣角度，並且能夠得到更清晰更多面向的人臉影像，這有助於提升人臉識別準確度。

最後，本文我系統中也加入了語音點名功能。當系統通過攝像鏡頭確認了一個人臉辨識成功後，會自動唸出該學生的姓名。可以提供一種視覺以外的確認方式，進一步提高點名的互動性。

本文會將人臉辨識的結果儲存為一個串列，主要用來記錄有哪些學生出席，最後以一個簡易的網頁來呈現學生的出席狀況，主要使用 **Flask** 撰寫網頁，它是一個使用 **Python** 編寫的輕量級 **Web** 應用框架，網頁上的表單會列出所有課堂成員，透過人臉點名系統得到到課學生之串列後，再透過 **Python** 程式碼把結果匯入並結合 **Flask** 的網頁架構，即可將學生的出席狀況呈現在網頁上。

綜上所述，本文使用的影像擷取和實驗設備上的設定都是為了確保人臉辨識與點名系統能夠在各種情況下都能獲得最佳的表現。

第五章 結果分析與討論

上述研究方法中，本文透過蒐集到的人臉訓練集進行完整的訓練，接著將透過先前資料處理中提到測試集照片進行人臉辨識的實驗。

在 **Arcface** 方面，本文將假設測試集的不同成功標準，以及設置不同的餘弦距離去實驗哪種假設會得到最高的準確率和最低的失敗率，而如果為偵測到人臉時，會顯示「??？」，以下是各個方法的實驗結果。

採用距離最近的 照片張數	採用照片中需符 合條件張數	餘弦距離 設定條件	準確率	失敗率	未偵測到
1	1	最近	91%	9%	0%
1	1	< 0.5	87.87%	5.65%	6.48%
5	4	< 0.5	85.98%	3.77%	10.25%
5	4	< 0.55	85.77%	2.72%	11.51%
5	4	< 0.6	80.75%	2.09%	17.16%
5	3	< 0.55	83.89%	4.6%	11.51%

表 5-1 ArcFace 不同方法之準確率情形

為了提高辨識的準確率，預將失敗率控制在 3% 以下，才能確保不會太常發生偵測錯誤的問題發生，在這個前提之下，可以看出找距離最近的前 5 張照片，有 4 張是同一人，且餘弦距離設置在小於 0.5 時，會有最大的成功率。另外，也從實驗中發現，若是受試人的測試集有戴眼鏡，但在順練集中未戴眼鏡也會造成人臉辨識錯誤的狀況發生。

而在 Facenet 方面，依照研究方法中所提到的分類方式機器學習演算法，KNN 的準確率達到 97.90%，而 SVM 的準確率更高達 99.79%。不過這兩種方法，並不會考慮到無法辨識的情況，無法保證會不會有非課堂的同學出現，很可能會造成分類上的判斷錯誤，於是還是需要使用設定閾值的方式來實驗哪種準確率最高，而準確率、失敗率以及為偵測到的比率定義與 Arcface 相同。

閾值 (Threshold)	成功率	失敗率	未偵測到
0.80	95.40%	0.21%	4.39%
0.70	95.40%	0.21%	4.39%
0.65	87.03%	0.21%	12.76%
0.60	82.01%	0.00%	17.99%

表 5-2 FaceNet 不同閾值之準確率情形

結合實驗結果可以產出下圖 5.1 閾值設定之準確率折線圖，透過圖形能發現閾值 (threshold) 設定的越小，為偵測該人臉的比例會越高，但相對的可以很好地避免辨識錯誤的情況發生，因此本文願意犧牲一點成功辨識的比例來取代錯誤率的發生，在全部的測試集中發現僅僅只有一張照片辨識錯誤，因此閾值設定從 1 到 0.60 內的錯誤率皆相同，最終當閾值小至 0.60 的時候就沒有任何照片辨識錯誤，經過反覆的測試本文決定將現場人臉辨識的閾值設定在 0.60，不僅擁有高準確率且可以避免辨識錯誤，無法辨識的比例也還在可接受的範圍，但上述閾值設定可能還是無法避免兩個長相極為相似的同學被判定為同一人物之情形發生。

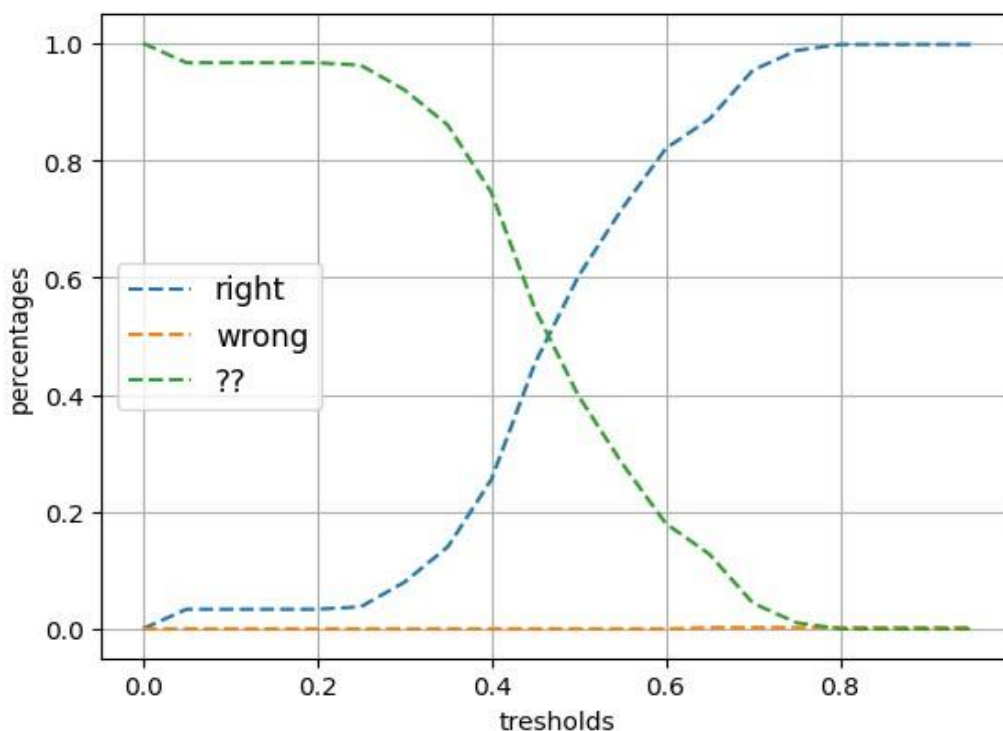


圖 5-1 FaceNet 不同閾值之準確率折線圖

總和上述的實驗結果，可以看出 Facenet 的失敗率明顯比 Arcface 來得低，另外，準確率方面，Facenet 也明顯比 Arcface 來得高，因此本文最終將選用 Facenet 作為人臉辨識系統的人臉辨識模型。

第六章 結論

在這次的專題研究中，本文採用了 Facenet 進行人臉辨識，並使用 HOG 進行人臉偵測。這兩種技術的結合，使該系統能夠在一定程度上實現人臉辨識點名。

然而，本文的點名系統還存在一些問題。一個可能的問題是，人臉辨識可以用照片來進行，這對於點名系統來說可能會有漏洞。例如，一個學生可以使用他人的照片來冒充他人出席課程。為了解決這個問題，之後需要引入活體辨識的技術。活體辨識是一種可以判斷被攝像頭捕捉到的人臉是否為真實人臉的技術。它可以防止使用照片、錄像或者 3D 模型等方式進行欺詐。結合活體辨識和人臉辨識，可以建立一個更加安全和可靠的點名系統。

未來，致力於將活體辨識技術與人臉辨識點名系統結合起來，並將使我們的系統更加完善，能夠更好地滿足實際應用的需求。

參考文獻

- [1] Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 711-720.
- [2] Paul, V., Michael, J. (2001) Rapid Object Detection using a Boosted Cascade of Simple Features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001 (Vol. 1, pp. I-I)*. IEEE.
- [3] Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971-987.
- [4] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 886-893.

- [5] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. European Conference on Computer Vision (ECCV), 9905, 21-37.
- [5] Rosebrock, A. (2021). Face detection tips, suggestions, and best practices. Retrieved from <https://pyimagesearch.com/2021/04/26/face-detection-tips-suggestions-and-bestpractices/>
- [6] Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 71-86.
- [7] Wirdiani, N. K. A., Hridayami, P., Widiari, N. P. A., Rismawan, K. D., Candradinata, P. B., & Jayantha, I. P. D. (2019). Face Identification Based on KNearest Neighbor. *Scientific Journal of Informatics*.
- [8] Phillips, P. J. (1998). Support Vector Machines Applied to Face Recognition.
- [9] Kremic, E., & Subasi, A. (2015). Performance of Random Forest and SVM in Face Recognition. Faculty of Engineering and Information Technologies, International Burch University, Bosnia and Herzegovina.
- [10] Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1701-1708).
- [11] Florian.S, Dmitry.K, James.P(2015),FaceNet: A Unified Embedding for Face Recognition and Clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [12] Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2018). ArcFace: Additive Angular Margin Loss for Deep Face Recognition.
- [13] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, InceptionResNet and the Impact of Residual Connections on Learning.

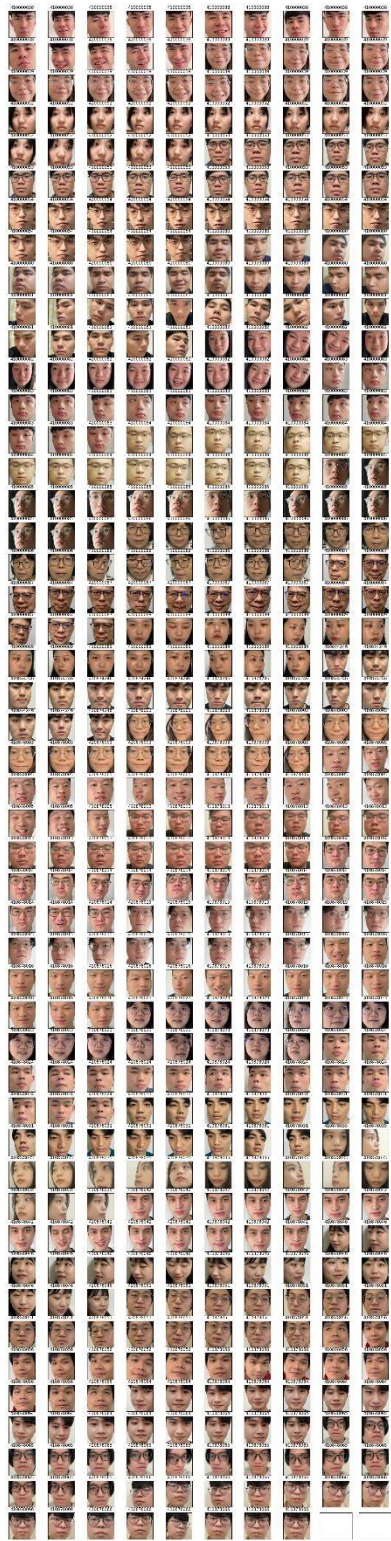
(附錄 1.1) - 彩色訓練集



(附錄 1.2) - 黑白訓練集



(附錄 2.1) - 彩色測試集



(附錄 2.2) - 黑白測試集

