口罩人臉識別

生機碩二 R09631031 賴怡穎

ABSTRACT

在 Covid 19 的肆虐下，與口罩共生成為新的生活模式，在帶著口罩的情況下，人臉辨識的準確度下降，為解決此問題，本研究開發一口罩人臉識別系統，使用以有的人臉資料庫人臉圖片用Dlib進行特徵點辨識，額外生成一遮罩，增加訓練過程中眼部的訓練權重。在臉部校正過後，透過在人臉辨識模型Google Facenet 中加入Spatial Attention 機制，加重於眼部特徵的識別，進而增加在臉部區域產生部分阻擋後，人臉辨識的正確率。

LITERATURE REVIEW

在臉部特徵擷取方面，2001由 Paul Viola 與 Micheal Jones [1]提出使用Haar-like小波特徵和積分圖方法進行人臉檢測，並使用AdaBoost訓練出強分類器區別人臉與非人臉，最後將分類器進行集連，提高準確率，然而Haar演算法對於光線與形狀的變形容易導致誤判，因此在之後Dlib演算法提出後，Dlib逐漸變成常用方法。在Dlib演算法中，以方向梯度直方圖（HOG）作為特徵擷取，可有效降低光源與形狀的影像，並透過取得臉部68個特徵點後，可使用支援向量機(Support Vector Machine)卷積神經網路（Convolutional Neural Network, CNN）作為人臉識別分類，實際檢測效果以使用深度學習網路[2]獲得的效果最佳。而另一方面，隨著深度網路的興起，Kaipeng Zhang 與 Zhanpeng Zhang 所提出的MTCNN[3]採用級聯CNN結構，透過將圖片以下採樣（Down-Sampling）的方式建構出影像金字塔（ Image Pyramid），並透過三個網路的級聯，提高臉部候選區域的篩選效率，最後模型輸出人臉的區域與五個特徵點（雙眼，鼻，左右嘴角）的位置。

除了MTCNN外，由Google所提出的FaceNet[5]為有效解決在不同照片中形變導致歐式距離的不同，其模型採用 Triple Loss將所有訓練照片映射到高維空間，並選擇錨點樣本，與錨點樣本屬於同類的正確樣本，與錨點樣本分數不同類的錯誤樣本，透過與錨點樣本相差最小的的錯誤臉孔，和與錨點樣本相差最大的正確臉孔進行特徵訓練，迭代拉近相差大的正確臉孔與錨點距離，拉遠相差小的錯誤臉孔與錨點距離，進而獲得最佳的辨識結果，其訓練結果在 LFW 人臉資料庫以 99.63% 的最佳成績刷新了記錄，由於其易於理解的演算原理以及應用方便，使得 Facenet 在眾多競爭者中（如 DeepFace、DeepID、Face++…等）異軍突起，成為目前最流行的臉部識別技術。

而除了 Facenet 外，基於 one-stage 的人臉檢測網路 RetinaFace[5] 與其前身 InsightFace透過影像獲取不同大小的圖像特徵，配合SSH演算法中的Context Modeling，並且改良了損失函數，其損失函數透過手動標註人臉上的五個特徵點，利用GCN將二維人臉特徵點映射到三維模型上，最後以Mesh Decoder解碼回二維，比較編碼前後的特徵點距離，形成新的Dense Regression作為圖像誤差分類。在實作方面RetinaFace是目前開源系統中於Wider Face資料及準確率最高的，也是目前已知的最強臉部特徵辨識模型。

MATERIAL AND METHOD

本研究在圖片前處理方面，分為兩個部分：

1. 臉部特徵點辨識與形狀校正

透過使用Dlib獲得臉部68個特徵點，在獲取資料集中的臉部輪廓後，框選出臉部的定界匡，並透過特徵點的幾何距離進行臉孔形狀校正，使臉部眼睛校正為水平線條，同時將偏移的五官校正為正臉形狀。

特徵點選取方面使用了Dlib的預訓練模型shape_predictor_68_face_landmarks.dat，Dlib是一個開源的C++函式庫，其主要應用於線性代數與機器學習方面，而此預訓練模型會回傳臉部的68個特徵點，其中包含臉部邊界與五官輪廓，如下方圖一所示，而回傳的資料點，可以進行臉部輪廓的修正，其修正方法為：

1.1. 取得左右兩眼的中心點

1.2. 依據此兩點形成的直線，計算與水平線的角度

1.3. 依雙眼間距計算旋轉後的縮放比例，在本此實驗種使用的縮放後大小為200

1.4. 取得雙眼的中間點，此點亦作為臉部旋轉的中心點

1.5. 獲得轉換矩陣M，並使用轉換矩陣獲得修正後臉部輪廓

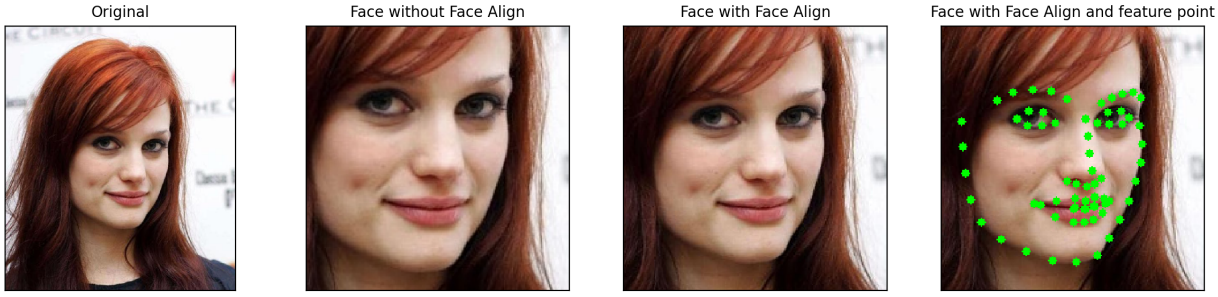


圖1.左1是原始人臉圖片。左2不經過臉部輪廓的修正，直接進行Dlib輪廓選取後獲得的人臉特徵，可以看出雖然有抓取到臉部輪廓，但有部分臉部線條被裁切，同時臉部位置不在整張圖片的中心。右2是經過臉部輪廓的修正後的臉部圖片，可以看到整張臉部線條都有包含在圖片之中。右1是在輪廓的修正後標註的臉部特徵點位置。

2. 臉部遮罩的建立

在獲得臉部特徵點後，選取眼睛與眉毛的特徵點，作為定位點，繪製包含臉部上半部輪廓至眼鼻的遮罩，此遮罩可用於之後模型學習時，作為注意力學習機制的學習範圍，遮罩樣子如圖2所示。在訓練時，會將遮罩與臉部圖片一起送入模型中進行訓練。

圖2.左1是原始人臉圖片。左2經過選擇後，範圍僅包含上半臉的注意力遮罩。

在模型選擇上，由於RetinaFace的模型複雜度較高，固本實驗使用擴充效果較佳的Facenet作為訓練基底架構，Facenet的整體架構如下圖3所示，其模型內容主要包含五個部分：Batch Input, Deep Network Architecture, L2 normalization, Embedding, Triplet Loss。

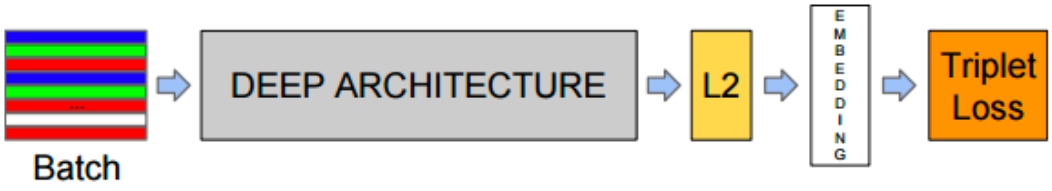


圖3. Facenet 整體架構，圖片來自Reference[5]中的Facenet論文。

2.1. Batch Input

本實驗訓練資料來自三個開源資料集，包含VGGFace2，LFW(Labeled Faces in the Wild)，與Real-World-Masked-Face-Dataset，其中Real-World-Masked-Face-Dataset不用於訓練與驗證，僅作為最後測試用途，VGGFace2用於訓練，LFW則用於驗證。

在將資料集中的圖片先經過第一階段的清洗，刪除掉長寬少於250 pixel的圖片後，進行前面說過的資料前處理，並根據資料類別進行錨點/正確臉孔/錯誤臉孔的資料配對，最後生成10萬筆的配對資料，以Batch size為 30 作為模型的input。

2.2. Network Architecture

本實驗的模型架構使用Resnet 34，並透過加上Spatial Attention[6] 機制，作為整體模型架構，Spatial Attention的架構如下圖4所示，透過將資料不同channel的資料進行數值加總後，轉為一維數據後，進行softmax以提出每個數據點的權重，最後轉回二維數據，與原始資料進行卷積。

在模型中加入此架構的原因在於在神經網路中使用全局平均池化Global average pooling (GAP)雖可有效聚焦圖片中對於特徵點的有效訊息，但也會造成部分訊息的丟失，因此在池化層前面加入Spatial Attention可有效改善此問題。

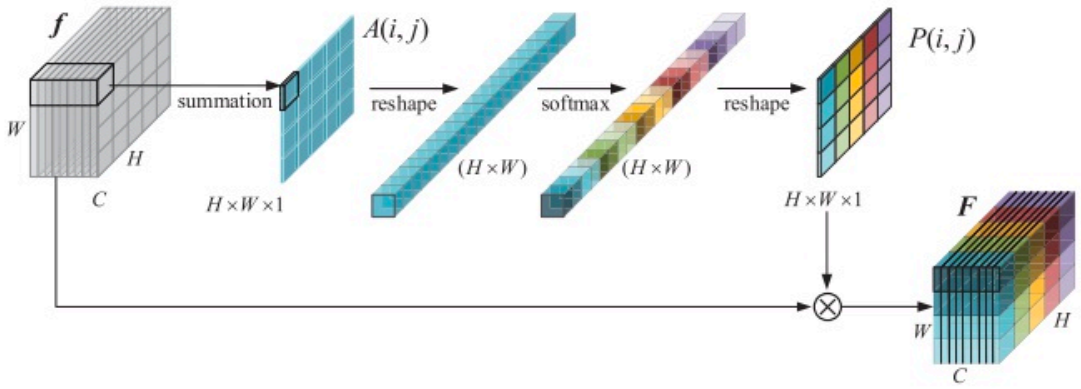


圖4. Spatial Attention 整體架構，圖片來自Reference[6]中的論文。

在 Batch Input 時，由於圖片會與其遮罩一起送入模型中，因此遮罩中未被遮住的上半臉特徵會在Spatial Attention層中加重其權重，進而增加眼部特徵的訓練，用以達到在較好的訓練效果。

2.3. L2 normalization

透過L2 normalization 對資料進行歸一化

2.4. Embedding

生成 output 向量特徵

2.5. Triplet Loss :損失函數

$$L = max(d(a,p) - d(a,n) + margin, 0) \qquad (1)$$

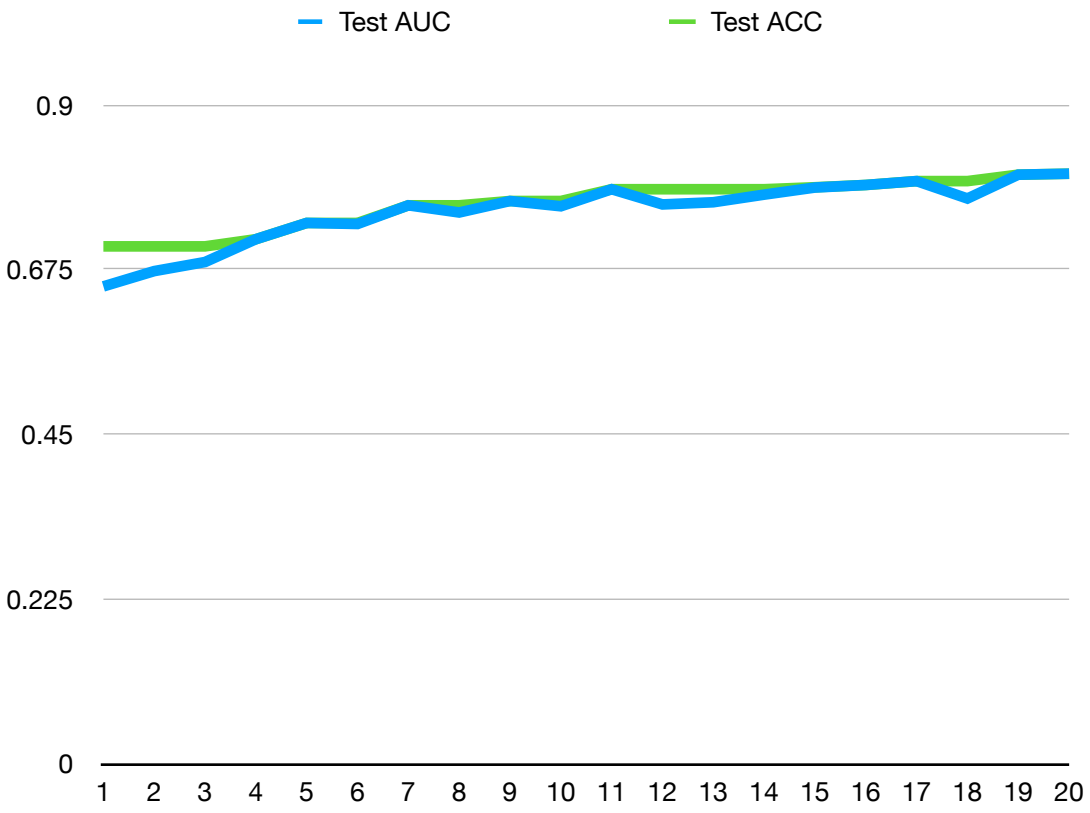
其中a為錨點，p與 a 是同一類別的樣本，n與 a 是不同類別的樣本，d為歐式距離，即透過拉近(a,p)的距離，拉遠(a,n)的距離來獲得最小的誤差。

RESULT AND DISCUSSION

因模型還沒有完全訓練完，因此這個部分先將我之後要討論的內容逐點敘述，前處理的結果已寫在MATRIAL裡面，所以這邊不會再提到

1. 模型訓練結果

模型的訓練過程中測試資料的準確度如下圖所示，由於訓練過程中Loss 的數據會根據動態Margin的設定有大小的差異，因此圖表中以AUC曲線與分辨得準確率ACC作為評斷標準。可以看到模型訓練過程中準確度有持續在上升，並且在訓練到20個epoch的時候接近收斂，故在增加臉部遮罩後，對於人臉辨識依舊有一定程度的準確率。



2. 不同模型比較

這裡會與 Resnet 34 (無Attention) 與 Inception-ResNet (無Attention）與我建立出的Model做準確度比較。

3. 辨識錯誤的圖像

這裡會拿辨識失誤的圖像與辨識正確的圖像做比較，看錯誤的原因是什麼，並且配合Attention的熱點圖來作為判斷。

REFERENCE

1. **D.N. C., G. A., M. R. FACE DETECTION USING A BOOSTED CASCADE OF FEATURES USING OPENCV. IN: VENUGOPAL K.R., PATNAIK L.M. (EDS) WIRELESS NETWORKS**

AND COMPUTATIONAL INTELLIGENCE. ICIP 2012. COMMUNICATIONS IN COMPUTER AND INFORMATION SCIENCE, VOL 292. SPRINGER, BERLIN, HEIDELBERG.

2. **DENG, H.; FENG, Z.; QIAN, G.; LV, X.; LI, H.; LI, G. MFCOSFACE: A MASKED-FACE RECOGNITION ALGORITHM BASED ON LARGE MARGIN COSINE LOSS. *APPL. SCI.* 2021, *11*, 7310. HTTPS://DOI.ORG/10.3390/APP11167310**

3. **KAIPENG ZHANG., ZHANPENG ZHANG., ZHIFENG LI, YU QIAO. JOINT FACE DETECTION AND ALIGNMENT USING MULTI-TASK CASCADED CONVOLUTIONAL NETWORKS. INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS (IEEE). 2016, 11, 1499–1503.**

4. **SCHROFF, FLORIAN & KALENICHENKO, DMITRY & PHILBIN, JAMES. (2015). FACENET: A UNIFIED EMBEDDING FOR FACE RECOGNITION AND CLUSTERING. 815-823. 10.1109/CVPR.2015.7298682.**

5. **J. DENG, J. GUO, E. VERVERAS, I. KOTSIA AND S. ZAFEIRIOU, "RETINAFACE: SINGLE-SHOT MULTI-LEVEL FACE LOCALISATION IN THE WILD," 2020 IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2020, PP. 5202-5211, DOI: 10.1109/CVPR42600.2020.00525.**

6. **WANG, H., FAN, Y., WANG, Z., JIAO, L., & SCHIELE, B. (2018). PARAMETER-FREE SPATIAL ATTENTION NETWORK FOR PERSON RE-IDENTIFICATION. *ARXIV, ABS/ 1811.12150*.**

7.