**Facial and gender classification of stickers using CNN**

陳柏勳-輔仁大學資訊工程學系

林博文-輔仁大學資訊工程學系

殷珮珊-輔仁大學資訊工程學系

**I.** **介紹**

近年來隨著Alpha-go的崛起，人工智慧等相關議題也順勢跟著水漲船高，而與之相關的類神經及機器學習更是密不可分，涉及到的應用層面也非常廣泛，從最基本的影像辨識、語音辨識，甚至已被廣泛使用的Siri，到較為複雜且多方面的資料探勘(Data Mining)，不只將學術界的發展帶向另一個高度，同時對於物質生活帶來很大的轉變。

其中圖像辨識更是與我們息息相關，像是醫學的腫瘤分析、偵測，或是台灣警方已經使用許久的人臉辨識等，都為我們的生活帶來很大的轉變。因此本次的主題選擇著重在圖像辨識上，並使用CNN(Convolutional Neural Networks)作為辨識圖像的model，是一種前饋神經網絡，它的人工神經元可以針對一部分覆蓋範圍內的周圍單元，對於大型圖像處理有良好的表現。

對於人臉辨識來說，最為基本的就是性別辨識，也是許多進階應用的基礎，並以此為特徵，作為其他分類器的分類條件之一。因此如何精確的識別男女， 是本次著重的要點。

**II.** **方法**

我們將方法分為三個部分，第一個部分為特徵選擇，此部分將介紹本文的特徵是依據什麼條件挑選，第二部分介紹圖片的處理方法，第三部分則為介紹CNN的架構。

1. 特徵選擇  
    圖片的挑選如圖一、圖二。

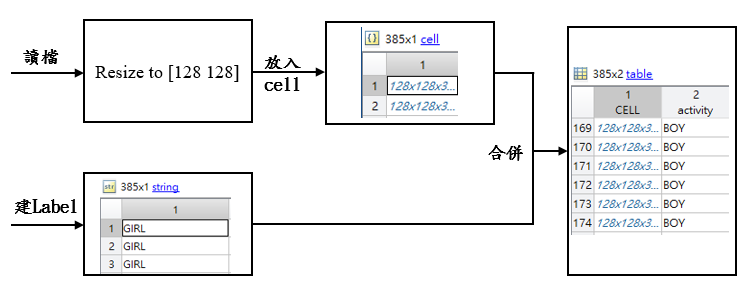
圖一、男



|  圖二、女



選擇圖片時大致可以依照五個條件做選擇：  
 1. 臉部位置：臉部的位置非常重要，若挑選圖片時臉部位置不統一，CNN在訓練時就無法成功抓到臉部的位置，導致最終訓練失敗，因此圖片選擇臉部為放置在中央的圖片為主。  
 2. 背景：圖片背景盡量選擇素色且並無太多雜物。  
 3. 髮型：男性通常為短髮或平頭，而女性則須多挑選短髮、長髮，選取短髮女的原因是為了防止最後的model在做預測時把短髮女性的圖片誤斷為男性。  
 4. 修圖：有修圖的照片，臉部會顯得特別平滑與白皙且五官不清晰，model容易誤判為女性，所以挑選時盡量以修圖幅度較低的為主。  
 5. 角度：臉部的角度主要以正面照為基準，若有側面、歪臉及歪頭則會被排除，由於相片角度的差異會影響識別，如眼睛的大小、形狀或因為角度而有所不同，因此以證件照的需求為主。

1. 圖片處理

圖三、圖片處理流程圖

本文程式使用matlab撰寫，因為每張圖片大小不一定相同，所以讀檔後須將每張圖片重新調整成128\*128的大小，接著放入cell中，而Label則為cell中圖片的正確配對，最後將cell與Label合併，合併後出來的Data便可以直接經由CNN訓練成model。

1. CNN架構

CNN設計的目標就是用來處理以多陣列型態表達的資料，如以RGB三通道表達的彩色圖片。CNN和普通神經網路之間的一個實質差別在於，CNN是對原始圖像直接做操作，而傳統神經網路是人為的先對影像提取特徵(例如灰階化，二值化)才做操作。

本研究使用卷積神經網絡（以下簡稱CNN）架構做為辨識模型。其主要分為卷積層、線性整流層、池化層、全連接層，再進行損失函數以及分群，圖四為分層架構，以下分述各層作用。

1. 卷積層（Convolutional Layer）

相較於傳統型深度學習網路，捲積層能保留圖像的空間排列並取得局部圖像作為輸入特徵。要計算特徵和圖片局部的相符程度，其透過一個指定尺寸的window，由上而下依序滑動取得圖像中各局部特徵作為下一層的輸入，我們給定3個隨機3\*3大小的卷積核，利用平移兩兩做乘積，製作出篩選過後的圖，也就是新的二維矩陣。

1. 線性整流層（ReLU）

線性整流層為此層神經的激勵函數，其可以增強判定函數和整個神經網絡的非線性特性，而本身並不會改變卷積層。

使用ReLU有幾個特性，相較於sigmoid 和 tanh 函數，ReLU可有效解決梯度消失問題；其次，ReLU會使部分神經元的輸出為0，可以讓神經網路變得稀疏，緩解過度擬合的問題；且ReLU計算簡單，只需將圖片上的所有負數轉為 0。

1. 池化層（Pooling Layer）

池化層的目的為簡化卷積層的訓練輸出數據。它實際上是一種降採樣的形式，利用降採樣來將圖片縮小，可以減少後續所需的參數，以增加效率；且可降低過度擬合的情況。常見的形式有三種：最大化（Max-Pooling）、平均化（Mean-Pooling）、隨機（Stochastic-Pooling）…等。

本模型採用Max-Pooling做降採樣。

1. 全連接層（Full connected layer）

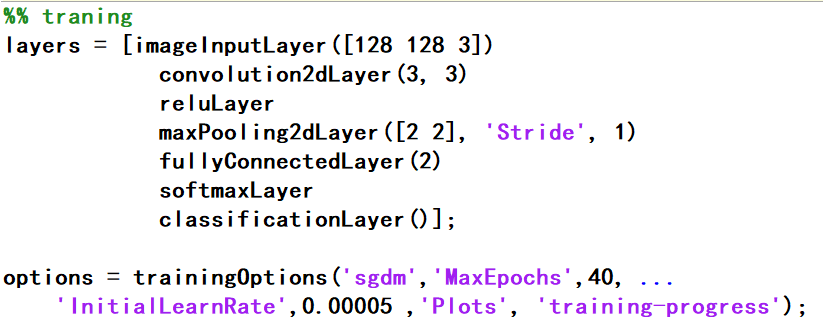
全連接層指的就是一般的神經網路，將之前的結果平坦化之後接到最基本的神經網絡。其會集合高階層中篩選過的圖片，並將這些特徵資訊轉化為票數。

在此我們將群分為兩類，分別為男性以及女性。

1. 損失函數層（Loss Layer）

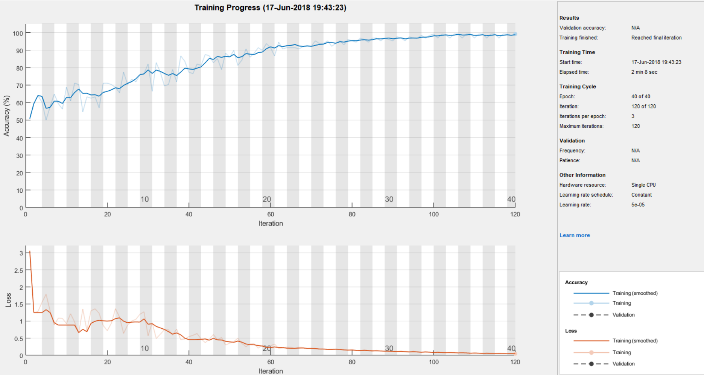
損失函數用於決定訓練過程如何來「懲罰」網絡的預測結果和真實結果之間的差異，各種不同的損失函數適用於不同類型的任務，Sigmoid交叉熵損失函數常常用於多個獨立的二分類問題；歐幾里德損失函數常常用於結果取值範圍為任意實數的問題。

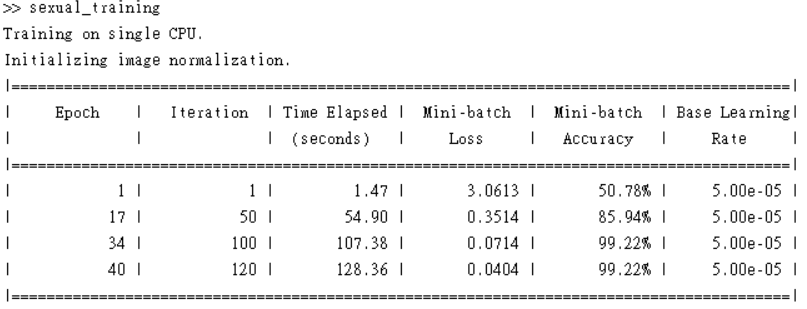
本模型使用的Softmax交叉熵損失函數，被用於在K個類別中選出一個。



圖四、CNN七層

**III.** **結果與分析**

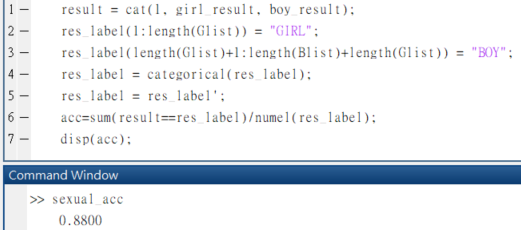




圖五、Training Progress

Training時的第一筆資料是隨機挑選，而第一筆資料會影響到training時每一筆資料的權重，且程式會依據每筆一資料算出來的權重的不同去平衡Mini-batch Accuracy和Mini-batch Loss。

在訓練過40個世代(epcho)，Mini-batch Accuracy也從50.78%提升到99.22%，Mini-batch Loss則降低到0.0404，而這兩個值對於最後分類的結果也有參考性，若Mini-batch Accuracy不夠高或Mini-batch Loss過高，都會造成最終結果不正確甚至是Undefined。



圖六、精確度計算

將拿來做Testing的圖片執行完分類後，我們將分類的結果與正確的答案拿來進行精確度的運算，得到88%的正確率，雖然這結果並沒有達到理想的精確度，但也表示了我們在挑選特徵的方向和CNN架構上並無太大的錯誤。

**IV. 結論**

以目前的結果而言仍舊無法達到非常高的精確率，產生誤差的原因，或許可以透過改善樣本的多樣性，藉以降低不同膚色、髮型、臉型所造成的差異性，或是改以RCNN、Fast-RCNN、Faster-RCNN等變形來作為演算法，也可能更為適合用來實作性別辨識。

除了依據臉部來辨識，未來也希望能夠透過身材、肌肉分布或衣服穿著等身體特徵作為分類標準，不再拘束於五官或髮型，並能夠將精確度提高，且不再因男女相似的肉眼特徵發生預測錯誤。

**V. 參考文獻**

1.google圖片

2.卷積神經網路wiki.-https://zh.wikipedia.org/wiki/%E5%8D%B7%E7%A7%AF%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C

3. Unsplash - https://unsplash.com/