

機器學習技術 Lab5

0116F137 陳廣能

實驗內容描述

這次實驗讓我有機會從零開始建構一個完整的人臉識別系統，說實話，一開始我對於要整合這麼多不同的深度學習模型感到有些緊張。整個系統的設計思路是分階段進行的：先用MTCNN來找到人臉在哪裡，然後用FaceNet提取每張臉的特徵，最後訓練一個分類器來識別是誰。

MTCNN這個模型真的很有趣，它不像我之前接觸過的那些單一功能的模型。它實際上是把三個神經網路串在一起使用 - P-Net、R-Net和O-Net。我覺得這個設計很聰明，因為它模擬了人類識別人臉的過程：先大概看看哪裡可能有臉，然後仔細看看是不是真的是臉，最後精確定位五官位置。P-Net負責快速掃描整張圖片找出可能的人臉區域，R-Net會進一步篩選掉那些明顯不是人臉的區域，O-Net則做最終的精確定位。實際操作時發現，這種階段性的處理方式確實比一步到位的方法效果要好很多。

在特徵提取這部分，我原本以為FaceNet會很複雜，但實際使用起來發現概念其實不難理解。它的厲害之處在於能把一張160x160的人臉圖片壓縮成512個數字，而這512個數字卻能很好地代表這個人的臉部特徵。更神奇的是，如果兩個人長得像，他們對應的這512個數字也會比較接近；如果長得不像，數字就會差很多。這讓我想到了一種很直觀的比喻：就像是給每張臉都分配了一個在512維空間中的座標位置。

資料集的部分我選用了Kaggle上的5-celebrity-faces-dataset，包含了五位名人的照片。雖然資料集不大，但對於理解整個流程來說已經足夠了。我在FaceNet提取的特徵基礎上搭建了一個相對簡單的分類器，就是一個全連接層加上softmax輸出。選擇這樣簡單的架構主要是因為FaceNet已經做了很好的特徵提取工作，後面的分類器不需要太複雜。

實驗結果

實驗結果打從一開始並不是一帆風順的。MTCNN在人臉偵測方面的表現確實讓我印象深刻，即使是一些光線不太好或者角度比較偏的照片，它也能準確找到人臉位置。我特別記得有張照片裡有三個人，背景還很雜亂，但

MTCNN不僅找到了所有的臉，還準確標出了眼睛、鼻子、嘴巴的位置。這讓我第一次真切地感受到深度學習在計算機視覺方面的威力。

FaceNet的特徵提取效果也超出了我的預期。剛開始我對於這512個數字能代表什麼還有些懷疑，但當我把提取出來的特徵做可視化分析時，發現同一個人的不同照片對應的特徵向量確實聚集在一起，而不同人的特徵向量則分布在不同的區域。這種空間分布的規律性讓我對深度學習的特徵學習能力有了更深的認識。

我還嘗試了用聚類算法對提取的人臉特徵進行分組，結果還蠻有意思的。用Agglomerative Clustering把偵測到的人臉分成三組，大部分情況下同一個人的臉都被分到了同一組。雖然偶爾會有分錯的情況，但整體準確率還是讓我滿意的。這個無監督的聚類結果讓我看到了這個系統在實際應用中的潛力，比如可以用來自動整理照片。

模型訓練的過程比我想象的要順利。30個epoch下來，訓練準確率和驗證準確率都在穩步提升，最終達到了相當不錯的水準。我特別關注了loss曲線的變化，沒有出現過擬合的情況，這讓我對模型的泛化能力比較有信心。不過說實話，看著那些數字一點點變好，心裡還是挺有成就感的。

最讓我興奮的是最後的測試階段。把訓練好的模型載入後，對一些新的照片進行識別，系統不僅能正確識別出是哪個名人，還會給出一個信心度分數。看到一個完整的pipeline從人臉偵測到身份識別完全自動化地運行，真的很有成就感。雖然偶爾會有識別錯誤的情況，但考慮到資料集的規模和模型的複雜度，這個結果已經很不錯了。

結果討論與實驗心得

回想整個實驗過程，最讓我頭疼的問題出現在模型整合階段。我原本計劃使用某個特定的FaceNet實現，結果下載時出了問題，檔案總是損壞。沒辦法，只好改用keras_facenet這個套件。本來以為只是換個套件而已，結果發現問題遠比想象的複雜。首先是embedding的維度不對，原來預期的128維變成了512維，這直接導致後面的分類器輸入維度不匹配。然後是API調用方式也不一樣，從原來的.predict()變成了.embeddings()。甚至連輸入圖片的格式要求都有差異。

這個過程真的讓我體會到了實際開發和理論學習的差別。在課堂上或者教程裡，一切都顯得那麼順理成章，但實際動手時總會遇到各種意想不到的問題。不過也正是通過解決這些問題，我學會了如何閱讀錯誤信息、查找文檔、

在Stack Overflow上尋找答案。現在回頭看，這些"意外"反而成了很寶貴的學習經歷。

除了技術層面的挑戰，這次實驗也讓我思考了一些更深層的問題。人臉識別技術確實很厲害，但它也帶來了隱私和倫理方面的擔憂。在做這個實驗的過程中，我不禁想，如果這種技術被用在不當的地方會怎樣？比如未經同意的監控、針對特定群體的歧視等等。我覺得作為技術開發者，我們有責任考慮自己開發的技術可能帶來的社會影響。

從純粹的學習角度來說，這次實驗讓我對深度學習有了更加深入和實際的理解。以前我對神經網路的理解更多停留在理論層面，知道反向傳播、知道各種激活函數，但缺乏對整個系統如何運作的直觀感受。通過這次實驗，我真正體會到了深度學習的魅力：它不僅僅是一堆數學計算，而是能夠學習和理解複雜模式的智能系統。

最後，我覺得這次實驗最大的收穫不是學會了如何使用某個具體的模型或者工具，而是培養了面對複雜技術問題時的思維方式。當遇到問題時，如何分析問題的原因，如何尋找解決方案，如何驗證方案的有效性 - 這些思考過程對我來說比任何具體的技術知識都更加寶貴。我相信這些經驗會在我未來的學習和工作中發揮重要作用。

參考資源：

<https://youtu.be/bG2tNYs7gw8?si=KSG-YMH-NIVqc8XI&t=1388> (原先載入FaceNet的方法行不通，因此改用此影片中的方法，載入keras的keras_facenet)

注：修復後可以正常運行的程式碼會一同繳交，這部分我放在一個**Face_Detection_And_Recognition**的壓縮檔內，它原本為我撰寫時的colab工作區。