Application of Deep Neural Networks in Cat and Dog Recognition

深度神經網路應用於貓狗辨識

專題領域: 系統領域

組 別: B361

指導教授:鄭桂忠 教授

組員姓名:劉羿彣、魏宜汝

報告摘要

本專題著重於應用深度神經網路於貓狗和貓狗品種的識別，旨在展現人工智慧在圖像識別領域的實際應用。我們利用 Kaggle 提供的貓狗數據集，在 Colaboratory 平台上建立神經網路模型並進行訓練。

這項研究的動機來自於當前圖像辨識技術的快速發展，以及深度學習模型在這個領域的卓越應用。隨著大數據的蓬勃發展和硬體運算能力的提升，圖像辨識已成為人工智慧領域中一個極具前景且引人注目的研究方向。透過這項研究，我們期望能更深入地理解圖像辨識技術，同時為實際應用領域提供有價值的貢獻，並推動該技術在社會中更廣泛的應用。

在研究中，我們不僅對原始數據進行預處理，還討論了數據集的大小對於預測正確性的影響。我們發現手動裁剪數據集能顯著提高模型的學習性能，且擴大數據集有助於減少過度擬合的現象，進一步提升預測的準確性。此外，我們採用了多種數據增強技巧，同時紀錄其他訓練參數的設置對模型訓練的影響，以找出最佳的預測正確性。

我們的目標是透過評估訓練過程中的預測正確性和損失值的變化，來建立一個高效且準確的貓狗辨識深度神經網路模型。本專題將有助於深化對神經網路在實際應用中性能的理解，同時提供實用的模型優化經驗。

本文

一. 研究主題：深度神經網路應用於貓狗辨識

二. 研究背景和動機：

神經網路的發展對現代科技產生了深遠的影響，並為未來的人工智慧研究和應用奠定了堅實的基礎。在神經網路中，含有多個神經元，每個神經元代表一個計算單元，這些神經元透過不同的權重彼此連接。神經網路通過調整這些權重以學習和適應不同的任務，使其具有辨識目標、做出預測，甚至在未見過的數據上進行泛化的能力。

在應用方面，神經網路在影像辨識、語音辨識、自然語言處理等領域取得了顯著的成就，其靈活性和適應性使得它成為解決複雜問題和執行智能任務的強大工具。

這項研究的動機源於當前圖像辨識技術的快速發展，以及深度學習模型在這個領域的卓越應用。隨著大數據的蓬勃發展和硬體運算能力的提升，圖像辨識已成為人工智慧領域中一個極具前景且引人注目的研究方向。透過這項研究，我們期望更深入地理解圖像辨識技術，同時為實際應用領域提供有價值的貢獻，並促進該技術在社會中更廣泛的應用。

三. 研究目的：

我們的目標是透過評估訓練過程中的預測正確性和損失值的變化，來建立一個高效且準確的貓狗辨識深度神經網路模型。此外，我們的研究還將著重於數據集的預處理和擴充，並探討如何建構神經網路來降低模型的過度擬合。

初步目標是建立一個能有效識別這些動物的神經網絡，隨後將功能擴展至精準辨認各種品種，以助於動物福利和預防疾病。該系統還計劃應用於動物收容所、獸醫診所及生物多樣性研究，提供數據支持和分析。

四. 研究過程：

(一)挑選貓狗數據集：

我們的數據集來自於 Kaggle，其中包含了 15 個狗的品種和 8 個貓的品種，每個品種都包含了 170 張相應品種的圖像，這些圖像格式為.jpg 或.jpeg。由於我們的研究主題是貓狗辨識，我們從 Kaggle 的數據集中挑選出 6 個狗的品種和 6 個貓的品種，每個品種均包含 168 張相應品種的圖像，成為我們最終的原始數據集。

由於辨識貓狗為二分法預測，我們希望每種動物類別（貓和狗）的數據量均佔總數據集的 50%，這種平衡的數據結構對於我們的研究具有以下優點：

首先，這種平衡確保了數據集中不會出現對某一類別的偏見，這有助於減少模型在訓練過程中過度偏向於學習任何一個類別的風險。其次，均衡的數據集有助於提升模型的泛化能力，使其能夠更有效地學習和識別不同類別的特徵，這在面對新的未知數據時特別重要。這樣的均衡分佈也有助於減少過度擬合的現象，因為模型不會過分依賴於某一類別的特定特徵，從而提高其在實際應用中的表現。

均等分佈的數據還能提供更準確的性能評估。在測試階段，模型需要在兩個相同規模的類別上展示相似的準確率，這樣的設置有助於更公平、全面地評估模型的性能。最後，這種數據結構還可以提高訓練效率，因為數據分佈均衡，模型在學習過程中可以更均等地從每個類別中獲取信息，這不僅加快了模型收斂的速度，還提高了結果的一致性。

我們選擇的數據集結構將有助於我們建立一個均衡、高效且具有強大泛化能力的深度神經網路模型。

(二)數據集預處理：

我們希望數據集的圖片能夠更加地單純，避免任何不必要的干擾對訓練過程造成影響，所以我們對圖片進行了手動的預處理。我們認為狗與貓的面部特徵是區分它們的關鍵，所以我們刻意裁去了圖片中多餘的背景以及貓狗的身體其他部分，並以方形來裁剪每一張貓狗圖片，最終只保留了它們的完整頭像。此專注於面部的裁剪方式，旨在幫助模型辨識重點區域，從而提高模型的預測正確性。

同時，我們也對數據集進行了嚴格的質量控制。我們刪除了那些含有多隻狗或多隻貓的圖片，以確保每張圖片只含有單隻貓狗的頭像。另外，我們還排除了那些貓狗的假圖片以及過度暗淡或過度明亮的圖片。這些做法是為了確保圖片質量的一致性，並為深度神經網路模型提供最佳的學習素材。

經過這一系列細心的預處理後，我們得到了一個高質量、專注於面部特徵的數據集。這個經過優化的數據集包含了 6 個狗的品種和 6 個貓的品種，每個品種均包含了 168 張經過挑選和預處理的相應品種圖像。此精心整理和優化的數據集為為深度神經網路模型提供了最佳的學習素材，從而為未來建立高效且準確的貓狗辨識深度神經網路模型奠定了關鍵的第一步。

(三)針對數據集探討訓練結果：

在我們的研究中，我們發現對數據集進行不同的預處理，會對最終的訓練結果有著顯著的影響。我們首先探討了使用未裁減的原始數據集（包含 6 個狗的品種和 6 個貓的品種，每個品種包含 112 張相應品種的圖像）進行訓練的情況。使用未裁減的原始數據集進行模型的訓練，我們發現訓練出的預測正確性相對較低，且過度擬合的現象相當明顯。這表示原始數據集中含有過多不必要的信息，如背景或其他不相關的元素，這些可能都會讓模型產生混淆，導致準

確率降低和過度擬合問題的出現，這也是我們決定手動裁剪圖片以確保每張圖片只含有單隻貓狗的頭像，從而增加圖片質量的一致性。

相較之下，當我們使用經過裁剪後的數據集（同樣包含 6 個狗的品種和 6 個貓的品種，每個品種均包含 112 張圖片）進行訓練時，結果得到了顯著的改善。這些經過裁剪的圖片因只包含了貓狗的主要區別特徵：它們的面部，所以模型的預測正確性有了大大的提升，同時也改善了過度擬合的現象。

以上訓練結果展現了簡化數據集並專注於關鍵特徵識別的重要性。為了探討數據集總量對過度擬合的影響，我們準備了三份含有不同圖片總量的數據集。每份都包含了 6 個狗的品種和 6 個貓的品種，但每個品種的圖片數量分別為 56 張、112 張和 168 張，其相關的訓練結果比較會在後面的篇章做更深入的討論。

(四)神經網路模型的建立與訓練：

運用 Colaboratory 來進行神經網路模型的建立與訓練。

1. 建立訓練、驗證和測試集：在 Colab 環境中導入數據集，原先按照 70%／15%／15%的比例劃分成訓練、驗證和測試集，但在經過多次嘗試後，發現按照 70%／12%／18%的比例劃分成訓練、驗證和測試集，會達到更好的預測正確性。
2. 數據增強以及數據生成器：利用 TensorFlow 和 Keras 庫提供的功能進行數據增強，這包括對圖像進行旋轉、平移、隨機縮放、水平翻轉等操作。然後使用數據生成器根據指定的目錄結構讀取圖像，同時讀取相應的標籤。我們設定輸入神經網路模型的圖像大小為32 × 32像素，每批次處理的圖像數量為 32 張，以便進行後續的二元分類任務。
3. 神經網路模型的建立：建立的模型包括多個卷積層，每一層都負責學習圖像中的不同特徵。在每個卷積層之後都接了一個最大池化層，以降低特徵圖的尺寸，進一步縮小模型的參數量。此外，我們為了防止模型的過度擬合，引入了 Dropout 層。最後，使用全連接層進行高級特徵的學習，並使用 sigmoid 激活函數來進行二元分類的輸出。在訓練過程中，我們使用了 Adam 優化器來有效地最小化損失函數，以提升模型的訓練效率。
4. 神經網路模型的訓練：在這個階段，我們開始訓練模型，並在每個訓練循環結束後進行驗證。訓練過程中，模型會根據訓練數據逐步地調整權重，以提高模型在訓練數據上的性能。最後，我們會保存訓練和驗證的損失值以及正確性。
5. 評估模型性能：透過將模型應用在測試集上，獲得測試損失值和正確性，再利用 Matplotlib 庫進行視覺化，呈現訓練過程中損失值和正確性的變化趨勢。這將有助於我們深入了解模型的性能表現以及在訓練過程中的動態變化。

五. 研究比較與結果：

我們選取 6 個狗的品種和 6 個貓的品種，每個品種包含 168 張相應品種的圖像，總共有$168 \times 6 \times 2 = 2016$張圖片，再以手動裁剪圖片的方式，來確保每張圖片只含有單隻貓狗的頭像，進而增加圖片質量的一致性。以下數據集 A、數據集 B、數據集皆為裁剪後的圖片。

數據集 A：共有 2016 張照片，貓與狗各佔 50%。

數據集 B：在每個品種中留下 2／3 的照片，共有 1344 張圖片。

數據集 C：在每個品種中留下 1／3 的照片，共有 672 張圖片。

(一)裁剪圖片對訓練結果的影響：

我們以數據集 B 為基礎，在數據集 B 上進行裁剪圖片前後的訓練結果比較，發現裁剪圖片後的模型預測正確率為 92.06%，速度為 1.077 秒／步，預測損失率為 0.2621；而裁剪圖片前的模型預測正確率為 73.02%，速度為 3.419 秒／步，預測損失率為 0.6017。這表示裁剪圖片可以顯著提高模型的計算效率和準確性。

(二)數據集總量對訓練結果的影響：

數據集 A、B、C 的正確率分別為 95.70%、92.06%、87.12%，這顯示出數據總量越多，模型的預測正確率越高，過度擬合的問題下降。

(三)Dropout 率對訓練結果的影響：

在不同 Dropout 率下，數據集 A、B、C 的預測正確率各有差異。特別是在數據集 B 和 C 上，0.2 的 Dropout 率對提升預測正確率非常有效。但數據集 A 在無 Dropout 下的性能最好。

藉由預測正確率和預測損失值的比較，我們可以看到 Dropout 在不同的數據集上對於提高模型預測正確性和減少預測損失值的效果不同。對於較大的數據集，無 Dropout 時預測正確率最高和預測損失值最低，但輕微的 Dropout 也能維持較高的預測正確率和較低的預測損失值，從而提供更好的泛化能力。對於中等和較小的數據集，使用 Dropout 能夠達到提高模型預測正確性和減少預測損失值，這表明這些數據集更容易受到過擬合現象的影響，特別是在數據集較小時，模型會出現較嚴重的過擬合現象，更需要透過引入 Dropout 來有效地緩解這一問題，從而使模型在面對新的、未見過的數據時表現得更加穩健。

(四)迭代次數（Epochs）對訓練結果的影響

模型在不同訓練迭代次數下的表現顯示，在 10 次迭代時，模型有較高的預測正確率但也有較高的預測損失值。當增加至 20 次迭代後，預測正確率提升，預測損失值降低，顯示出更好的數據擬合。但在 30 次迭代時，模型的預測正確率和預測損失值與 10 次迭代相似，暗示過度擬合的現象可能發生，因為性能沒有進一步改善。這表明 20 次迭代是這些設置中的最優選擇，提供了最佳的預測正確率和預測損失值，而較無過度擬合的跡象。

(五)數據集切割比例(訓練集%／驗證集%／測試集%)對訓練結果的影響：

比較 70%／15%／15%與 70%／12%／18%的數據集切割比例，後者在訓練和測試中顯示出更高預測正確率和更低的預測損失值，表明其模型性能較優。

(六)學習率（Learning Rate）對訓練結果的影響：

學習率 = 0.001 時提供了最高預測正確率和最低預測損失值，顯示出適中學習率有利於模型有效學習並達到最佳性能。

(七)圖片像素尺寸(pixel × pixel）對訓練結果的影響：

圖片像素尺寸 = 32 × 32的模型有最高預測正確率和最低預測損失值，表示較小的圖片像素尺寸已足夠捕捉重要特徵，且計算成本更低。

六. 總結

圖 1 Training and Validation Loss

Training and Validation Loss 顯示了訓練集（藍色線條）和驗證集（橙色線條）的損失值。損失值是量化預測值和實際值差異的指標：損失值越低，模型的預測越準確。一開始，訓練和驗證損失都迅速下降，這在訓練初期是典型的現象。然而，隨著訓練持續進行，損失值繼續下降，而驗證集損失值在降到一個點後開始出現波動，這可能是過擬合的現象，表示模型過度學習了訓練數據，因此在驗證集上的表現較在訓練集上的表現不佳。

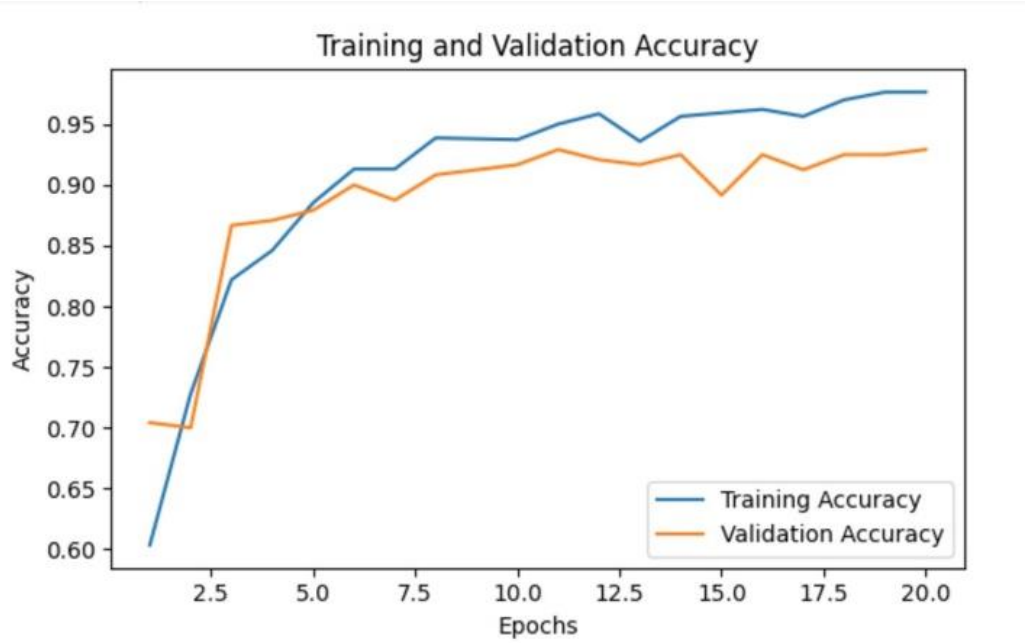

圖 2 Training and Validation Accuracy

Training and Validation Accuracy 顯示了訓練集（藍色線條）和驗證集（橙色線條）的正確性。正確性隨著時間的增長而提高，這是理想的狀況。其中，訓練集正確性最終超過了驗證集正確性，並持續改進，而驗證集正確度則趨於平穩甚至略有下降，這表示模型可能失去了對新未見數據的泛化能力。

由上述兩張圖來看，可以發現模型在訓練集上的性能隨時間改善，但在驗證集上的性能改善隨時間增長而改善較少，表示有過度擬合的現象出現。為了防止過度擬合，可能還需近一步的調整。

七. 延伸研究(貓與狗品種的辨識)：

在進一步的研究中，我們不僅實現了基本的貓狗辨識，還進行了品種辨識。透過優化神經網絡架構，我們將數據集重新分配，原先的 70%／15%／15% 調整為 65%／20%／15%，以平衡數據集並減少過擬合，提升泛化能力。特別是在進行品種辨識時，這種平衡尤為重要。 為了應對更複雜的多類別分類任務，我們增加了代表性，提高了圖片像素尺寸至128 × 128，並將學習率調低至 0.0001，訓練週期增至 30 次。這些調整帶來了顯著成果，貓品種辨識的預測正確率達到了 0.6218，而狗品種的預測正確率達到 0.7628。

貓與狗品種分類的比較顯示，狗品種的辨識表現更佳，可能是因為狗之間的差異更為顯著。而貓品種的較低辨識率及較高損失值反映出貓品種間差異較小，辨識難度較高。訓練時間上，貓品種的分類速度略快於狗，這可能與數據集特性或模型對不同品種的適應性有關，在貓品種分類上可能需要額外的模型優化或數據調整以提高準確率。

八. 心得

我在整個研究過程中，我們從數據處理、模型構建到觀測訓練結果中學到了許多有價值的知識和經驗。首先，我們明白了從 Kaggle 等平台獲取豐富的數據集的重要性。在圖像預處理階段，我們深刻了解到對圖像進行裁剪和篩選的重要性，這對提高訓練準確性具有決定性的影響。其次，透過使用 TensorFlow 和 Keras 構建和訓練神經網絡模型，我們對卷積層、最大池化、dropout 等技術有了更深入的理解。

在最後的比較過程中，我們發現了一些關鍵的因素會對模型性能產生影響。圖像預處理、數據集大小、dropout 率、訓練迭代次數、數據集拆分比例、學習率以及圖像像素大小都是需要透過訓練結果的比對來找出最合適的設置。

這個研究讓我們更深入地理解了深度學習在圖像識別中的應用，並提高了我們的實際操作能力。

九. 參考文獻

Efficient Processing of Deep Neural Networks: A Tutorial and Survey, Vivienne Sze, Senior Member IEEE, Yu-Hsin Chen, Student Member IEEE, Tien-Ju Yang, Student Member IEEE, and Joel S. Emer, Fellow IEEE, Vol. 105, No. 12, December 2017

十. 團隊合作

應用發想與實作：劉羿彣、魏宜汝
設計神經網絡架構：劉羿彣、魏宜汝
成果報告書編寫與海報設計：劉羿彣、魏宜汝