

國立雲林科技大學資訊管理系

機器學習 – 專案作業二

Department of Information Management National Yunlin

University of Science & Technology Assignment

CNN 性別分類專案：模型設計、資料增強與遷移學

習實驗

CNN Gender Classification Project: Model Design, Data

Augmentation, and Transfer Learning Experiments

M11323039、M11323021、M11323023、M11323041

蔡昆彥、蔣育承、毛淑娟、林承勳

指導老師：許中川 博士

Advisor: Chung-Chian Hsu, Ph.D.

中華民國114年4月

March 2025

## 摘要

隨著深度學習與卷積神經網路(Convolutional Neural Networks, CNN)技術的快速發展，影像分類已被廣泛應用於人臉辨識等多元應用場景。然而，實務應用上常面臨資料不足、光照與姿態變化、遮擋及模型過擬合等問題，限制了模型效能。本研究旨在探討於性別分類任務中，資料增強、正則化及轉移學習等技術策略對 CNN 模型效能之影響，以提升模型的預測準確率與泛化能力。

研究中採用 Gender Detection & Classification - Face Dataset 作為實驗資料，透過 CNN 架構進行系統性實驗與比較分析。研究結果顯示，資料增強技術能有效提升模型泛化能力，但可能導致訓練穩定性降低。進一步分析發現，在有限資料條件下，結合資料增強與 L2 正則化有助於模型效能改善，而其他正則化技術如 Dropout、Batch Normalization、L1 正則化與 Bagging 則需視資料特性謹慎應用。轉移學習實驗結果則指出，以 Inception-v3 為基礎模型進行微調能顯著提升驗證準確率且具有最佳的泛化穩定性，相較於 VGG-16 和 ResNet50，更適合作為性別分類任務之建議模型架構。

關鍵字：卷積神經網路、性別分類、資料增強、正則化、轉移學習

# 一、緒論

## 1.1 動機

隨著深度學習與 CNN 技術的蓬勃發展，影像分類已廣泛應用於臉部辨識、監控系統等場域(Parkhi et al., 2015; Pawar et al., 2024)。然而在實務應用中，資料不足、光照與姿態變化、遮擋，以及過擬合仍是限制模型效能的重要因素(Ding & Tao, 2016; Sawant & Bhurchandi, 2019; Tan & Triggs, 2010; Zeng et al., 2021)。為突破此瓶頸，近年來學界提出資料增強、正則化與轉移學習等策略，期望在有限資料與硬體條件下，進一步提升 CNN 的預測表現與穩定性(Srivastava et al., 2014; Wang & Perez, 2017; Yosinski et al., 2014)。

本研究以 Gender Detection & Classification – Face Dataset 為實驗資料。該資料集收錄在不同性別、種族以及拍攝背景條件（如室內／戶外、單色背景／複雜場景）下的臉部影像；雖未額外標註年齡、光照或表情變化，仍因跨族群與多元場景的組合而具備足夠變異性，可用來評估各種提升模型泛化能力的方法。

綜合上述背景，如何在具有性別、種族與拍攝背景多樣性的臉部影像資料上，提升 CNN 模型的預測準確率與泛化能力，成為本研究的核心課題。因此，本研究動機歸納為以下三點，並形成三項具體研究問題：

1. 使用資料增強技術，觀察可否提升模型預測績效？
2. 使用資料增強技術搭配正則化技術，例如，Dropout, Batch normalization, L1, L2, Early Stopping, Bagging 等，分析是否可提升模型預測績效？
3. 使用轉移學習，至少分別使用 VGG, Inception-v3, ResNet 當基礎模型，再進行微調訓練。請分析使用轉移學習再微調，是否能提升模型預測績效？

本研究將以 CNN 為核心架構，系統性比較上述策略的單獨與組合效果，期望為性別分類任務提供具實證依據的改進方向與應用參考。

## 1.2 目的

本研究旨在探討於性別分類任務中，採用不同技術策略對 CNN 模型預測效能之影響，藉此提升模型於變異性高之臉部影像資料中的準確率與泛化能力。為達成此目標，具體探討問題如下：

1. 評估資料增強技術在處理性別分類資料集時，對模型預測準確率的影響，分析其是否能有效提升模型對多樣化背景與族群的適應能力。
2. 結合資料增強與正則化技術，如 Dropout、Batch Normalization、L1/L2 正則化、Early Stopping 及 Bagging 等方法，探討其對模型過擬合情形的抑制效果與整體表現之提升程度。
3. 應用轉移學習策略，選用 VGG-16、Inception-v3 及 ResNet50 等預訓練模型作為基礎架構，進行微調訓練，分析其對性別分類任務預測效能的影響與適用性。

透過上述三個方向，本研究期望改善一套具備良好準確率與泛化能力的性別分類模型，並為後續人臉辨識應用提供可行性高且具實用價值的技術參考。

## 二、方法

本研究以 Keras/TensorFlow 框架實作 CNN 性別分類模型，並採用 Gender Detection & Classification – Face Dataset 為資料來源。系統流程涵蓋資料前處理、模型訓練與微調、性能評估與結果可視化。實驗首先比較原始 CNN 模型之效能，進一步導入資料增強技術，以強化泛化能力。隨後結合正則化，評估對預測表現之影響。最後採用轉移學習進行比較，探討其對效能的影響與提升潛力。模型評估採用分類準確率 (Accuracy) 作為主要指標，並以交叉熵作為損失函數進行訓練優化。

### 2.1 架構

本研究採用 Keras/TensorFlow 框架建構 CNN 模型，並使用 Gender Detection & Classification – Face Dataset 資料集，實作性別分類任務。整體系統架構包含資料前處理、模型建構與訓練、微調、性能評估與結果可視化等模組。

### 2.2 執行流程

首先，透過微調現有 CNN 模型並訓練，評估其在指定資料集上的預測績效。接著導入資料增強技術，例如，旋轉、平移與縮放調整等，觀察是否能提升模型泛化能力及準確性。在此基礎上，進一步結合正則化方法，如 Dropout、Batch Normalization、L1/L2 正則化、Early Stopping 與 Bagging 等方法，分析這些技術是否有助於減少過擬合並提升預測表現。此外，亦採用轉移學習方式，以 VGG-16、Inception-v3 與 ResNet50 等主流深度學習模型作為基礎架構進行微調訓練，探討其對模型效能的影響與提升潛力。

## 三、實驗設計

### 3.1 資料集

本研究以 Gender Detection & Classification – Face Dataset 為實驗資料。該資料集包含一系列人物照片，並按「女性」和「男性」的標籤分類存放在資料夾中。每個資料夾包含大量圖像，捕捉了來自不同背景、年齡組和種族的女性和男性個體，以方便性別檢測演算法或模型的訓練和測試。

表 1

資料集特性

特性	描述
資料類型	彩色圖像
訓練樣本數-MEN	110
訓練樣本數-WOMEN	110
測試樣本數-MEN	40
測試樣本數-WOMEN	40

### 3.2 前置處理

在本研究中除了作為比較基準的模型外，為了提升模型對輸入影像的學習效果與泛化能力，針對訓練資料與測試資料分別設計了不同的影像前處理流程。主要包括尺寸調整、資料增強、正規化與格式轉換等步驟，具體如下所述：

- **訓練資料之影像前處理：**

為增加資料多樣性並減少過擬合現象，訓練資料透過下列資料增強與轉換操作：

- 影像尺寸調整：將輸入影像最短邊調整為 256 像素，保留長寬比，確保後續裁切操作有足夠範圍。
- 資料增強：
  - ◆ 隨機尺寸裁切：隨機裁切影像中區域後，縮放為固定大小 (224x224)，有助提升模型對不同視角與構圖的適應能力。
  - ◆ 隨機角度變換：對影像進行隨機角度旋轉 ( $\pm 10^\circ$ )、平移 (最多 10%) 及縮放 (0.9~1.1 倍) 等變形處理，增加模型的穩健性。

- 格式轉換與正規化：
  - ◆ 將影像格式轉換為 Tensor 並轉型為 float32
  - ◆ 依據 ImageNet 預訓練模型之統計值進行標準化 (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225])
- 測試資料之影像前處理：
 

為確保模型評估時輸入資料的一致性與可重現性，測試資料不進行隨機變換，僅保留基本標準化處理：

  - 將影像尺寸直接調整為 (224x224)
  - 將影像格式轉換為 Tensor 並轉型為 float32
  - 套用與訓練資料相同的標準化參數 (mean 與 std)

### 3.3 實驗設計

#### 3.3.1 資料增強

- 隨機尺寸裁切：隨機裁切影像中區域後，縮放為固定大小 (224x224)，有助提升模型對不同視角與構圖的適應能力。
- 隨機角度變換：對影像進行隨機角度旋轉 ( $\pm 10^\circ$ )、平移 (最多 10%) 及縮放 (0.9~1.1 倍) 等變形處理，增加模型的穩健性。

#### 3.3.2 不同正則化方法的比較

本研究基於 TensorFlow 框架構建 CNN 模型，以進行性別分類任務。輸出層設計為兩個類別，分別對應男性與女性，並透過 Sigmoid 函數產生分類機率。隱藏層採用 ReLU 激活函數，損失函數則使用二元交叉熵，參數更新則透過 Adam 優化器進行。此外，為避免過度擬合，本研究實施 Early Stopping 技術，並比較多種正則化方法的效果，包括 Dropout、L1/L2 正則化、Batch Normalization 與 Bagging。

- Dropout：在全連接層加入 Dropout 層，設定隨機丟棄比例為 0.2，以減少神經元共適應性，提升模型泛化能力。
- Batch Normalization：於每一層卷積層後加入操作，規範化每一批輸入的特徵分布，穩定訓練過程並加速收斂。
- L1 正則化：在全連接層的權重引入 L1 正則項，強調稀疏性。
- L2 正則化：在全連接層的權重引入 L2 正則項，抑制過大權重。
- Bagging：建立多個 CNN 子模型，針對原始訓練資料進

行隨機重抽，再透過多模型平均方式整合預測結果，採用閾值 0.5 決定最終分類。

### 3.3.3 轉移學習

本研究基於 TensorFlow 框架，分別構建三種基於預訓練模型的 CNN 架構，進行性別分類任務。三種模型分別為 ResNet50、InceptionV3 與 VGG-16。每個模型皆採用相同的訓練與測試流程，包括兩階段訓練依序為頂層訓練及微調及資料預處理方式。輸出層設計為兩個類別，使用 Sigmoid 函數輸出分類機率。損失函數採用 Binary Cross-Entropy，優化器採用 Adam。

- **VGG-16**：相對於更複雜的 VGG 模型其結構較簡單，易於修改與實作，以連續的 3x3 卷積和最大池化組成。
  - 階段一：凍結所有預訓練層，只訓練最後的全連接層。
  - 階段二：解凍最後 4 層(Block 5)的權重將學習率由 0.001 改為 0.0001 進行微調。
- **Inception-v3**：相較於過去版本多採用了正則化技術，能取得更好的泛化能力。
  - 階段一：凍結所有預訓練層，只訓練最後的全連接層。
  - 階段二：解凍最後 30 層的權重將學習率由 0.001 改為 0.0001 進行微調。
- **ResNet50**：採用殘差學習，解決深層網路中的退化問題，提升訓練穩定性。
  - 階段一：凍結所有預訓練層，只訓練最後的全連接層。
  - 階段二：解凍最後 20 層的權重將學習率由 0.001 改為 0.0001 進行微調。



## 3.4 實驗結果

### 3.4.1 評估資料增強技術對模型預測準確率的影響

表 2

資料增強技術對模型訓練與驗證性能影響比較

		Loss	Accuracy	Val-Loss	Val-Accuracy
Base	最大	0.716	<b>0.714</b>	0.753	0.575
	最小	<b>0.542</b>	0.500	0.689	0.500
	平均	0.651	0.581	0.703	0.538
Augmentation	最大	1.186	0.677	0.738	<b>0.638</b>
	最小	0.630	0.446	<b>0.679</b>	0.500
	平均	0.685	0.563	0.697	0.551

從表 2 可見，引入資料增強後，模型在驗證準確率的最大值與平均值均有提升，但訓練損失與準確率的波動也較大，顯示增強技術雖然提高泛化能力，但可能會有更高的訓練不穩定性。

### 3.4.2 資料增強與正則化對過擬合抑制與模型表現之影響

- Dropout

表 3

資料增強與 Dropout 對模型訓練與驗證性能之比較

		Loss	Accuracy	Val-Loss	Val-Accuracy
Augmentation	最大	1.186	<b>0.677</b>	0.738	<b>0.638</b>
	最小	<b>0.630</b>	0.446	<b>0.679</b>	0.500
	平均	0.685	0.563	0.697	0.551
Augmentation + Dropout	最大	0.754	0.623	0.710	0.550
	最小	0.664	0.514	0.689	0.438
	平均	0.688	0.564	0.694	0.508

從表 3 顯示，Augmentation 透過擴充樣本多樣性，使驗證準確率最大值達63.8%、平均55.1%，顯著提升泛化能力；反觀 Dropout 因資料集樣本數有限，使用 Dropout

反而削弱了模型的學習效果，驗證準確率平均僅50.8%，未能帶來預期的正向改善。

- **Batch Normalization**

表 4

資料增強與 *Batch Normalization* 對模型訓練與驗證性能之比較

		Loss	Accuracy	Val-Loss	Val-Accuracy
Augmentation	最大	1.186	0.677	0.738	<b>0.638</b>
	最小	0.630	0.446	0.679	0.500
	平均	0.685	0.563	0.697	0.551
Augmentation + Batch Normalization	最大	1.185	<b>0.918</b>	1.225	0.575
	最小	<b>0.271</b>	0.546	<b>0.671</b>	0.500
	平均	0.530	0.781	0.750	0.535

由表 4可見，結合 Batch Normalization 後，訓練準確率平均提升至 78.1%，訓練 Loss 平均下降至 0.530；然而驗證準確率反而略降至 53.5%，驗證 Loss 平均上升至 0.750，顯示在資料量有限時，Batch Normalization 雖可加速收斂，卻可能增加過度擬合風險。

- **L1正則化：**

表 5

資料增強與 *L1* 正則化對模型訓練與驗證性能之比較

		Loss	Accuracy	Val-Loss	Val-Accuracy
Augmentation	最大	1.186	<b>0.677</b>	0.738	<b>0.638</b>
	最小	<b>0.630</b>	0.446	<b>0.679</b>	0.500
	平均	0.685	0.563	0.697	0.551
Augmentation + L1	最大	1.043	0.659	0.776	0.613
	最小	0.654	0.459	0.725	0.475
	平均	0.742	0.563	0.748	0.551

由表 5 可見，結合 L1 正則化後，訓練損失略升且準確率無明顯提升；驗證準確率平均維持55.1%，但波動幅度更大，顯示在小型資料集上 L1 正則化的效益有限。

- **L2正則化：**

表 6

資料增強與 L2 正則化對模型訓練與驗證性能之比較

		Loss	Accuracy	Val-Loss	Val-Accuracy
Augmentation	最大	1.186	<b>0.677</b>	0.738	<b>0.638</b>
	最小	<b>0.630</b>	0.446	<b>0.679</b>	0.500
	平均	0.685	0.563	0.697	0.551
Augmentation + L2	最大	1.190	0.614	0.778	0.625
	最小	0.709	0.514	0.729	0.513
	平均	0.757	0.569	0.750	0.569

由表 6 可見，引入 L2 正則化後，訓練損失略有上升且訓練準確率從56.3%小幅提升至56.9%；驗證準確率平均從55.1%提高至56.9%，顯示 L2 正則化在有限資料下能提升模型泛化，但也帶來輕微的訓練誤差增加。

- **Bagging：**

表 7

資料增強與 Bagging 對模型訓練與驗證性能之比較

		Loss	Accuracy	Val-Loss	Val-Accuracy
Augmentation	最大	1.186	0.677	0.738	<b>0.638</b>
	最小	0.630	0.446	<b>0.679</b>	0.500
	平均	0.685	0.563	0.697	0.551
Augmentation + Bagging	最大	0.887	<b>0.996</b>	1.851	0.575
	最小	<b>0.051</b>	0.436	0.690	0.413
	平均	0.469	0.716	1.270	0.494

由表 7 可見，結合 Bagging 後雖提升訓練準確率，但由於資料量有限，重複抽樣易導致子模型過度學習訓練資料特徵，造成集成後模型出現明顯過擬合，驗證損失偏高且準確率下降，顯示 Bagging 在小型資料集下需謹慎使用。

### 3.4.3 轉移學習在性別分類任務之應用與評估

- VGG-16：

表 8

*VGG-16 微調階段對模型訓練與驗證性能之影響比較*

		Loss	Accuracy	Val-Loss	Val-Accuracy
階段一	最大	0.735	0.750	0.751	0.828
	最小	0.598	0.564	0.564	0.375
	平均	0.670	0.639	0.644	0.651
階段二	最大	0.659	<b>0.995</b>	0.682	<b>0.875</b>
	最小	<b>0.048</b>	0.660	<b>0.406</b>	0.563
	平均	0.274	0.895	0.476	0.798
最終結果				0.672	0.600

由表 8 顯示，階段一僅訓練全連接層，特徵提取能力受限，驗證準確率平均僅 65.1%；階段二解凍 Block-5 並降低學習率至 0.0001，能針對高階語意進行細部調整，平均驗證準確率提升至 79.8%，展現深層微調優勢，但最終模型整合時表現反而下滑至 60.0%，可能是因為過度微調或驗證資料特徵偏移導致性能不穩定。

- Inception-v3 :

表 9

*Inception-v3* 微調階段對模型訓練與驗證性能之影響比較

		Loss	Accuracy	Val-Loss	Val-Accuracy
階段一	最大	1.542	0.846	1.002	0.797
	最小	0.359	0.533	<b>0.432</b>	0.484
	平均	0.727	0.708	0.707	0.674
階段二	最大	0.376	<b>1.000</b>	0.588	<b>0.828</b>
	最小	<b>0.006</b>	0.872	0.462	0.766
	平均	0.101	0.971	0.526	0.788
最終結果				0.711	0.763

由表 9 顯示，階段一 僅訓練全連接層，特徵提取能力受限，驗證準確率平均僅達 67.4%；階段二 解凍 Inception-v3 的高層卷積模塊並調降學習率後，進行深層微調，平均驗證準確率顯著提升至 78.8%。儘管微調表現提升明顯，但最終模型在整合階段表現仍下降至 76.3%，可能是因為驗證樣本特徵分布變異造成泛化穩定性不足。

- ResNet50 :

表 10

*ResNet50* 微調階段對模型訓練與驗證性能之影響比較

		Loss	Accuracy	Val-Loss	Val-Accuracy
階段一	最大	0.910	0.532	0.921	0.625
	最小	0.700	0.484	0.655	0.375
	平均	0.767	0.504	0.709	0.516
階段二	最大	0.677	<b>0.957</b>	0.695	<b>0.688</b>
	最小	<b>0.113</b>	0.604	<b>0.647</b>	0.500
	平均	0.332	0.838	0.659	0.617
最終結果				0.672	0.600

由表 10顯示，階段一僅訓練全連接層，驗證準確率平均僅 51.6%；階段二解凍 ResNet50 的深層殘差區塊後進行微調，驗證準確率提升至 61.7%。雖然訓練效果顯著提升，但最終結果僅維持在 60.0%，顯示深層微調需兼顧泛化與穩定性設計。

● 比較三種模型微調後的驗證平均準確率與泛化能力

表 11

三種預訓練模型微調後的驗證平均準確率比較

	VGG-16	Inception-v3	ResNet50
階段一 平均準確率	0.651	<b>0.674</b>	0.516
階段二 平均準確率	<b>0.798</b>	0.788	0.617
最終 準確率	0.600	<b>0.763</b>	0.600

本研究比較三種主流預訓練模型於微調階段的表現差異。由表 11可見，Inception-v3 在微調第二階段的驗證準確率（78.8%）與最終驗證表現（76.3%）均為最佳，顯示其深層架構與多分支特徵提取能力有助於提升泛化效果。VGG-16 雖於微調階段表現優異（79.8%），但最終整合後準確率大幅下降至 60.0%，推測可能因過擬合或驗證資料特徵偏移導致。ResNet50 的訓練穩定性佳，第二階段準確率亦有提升（61.7%），但整體表現仍不如前兩者，可能受限於資料量對殘差架構的充分訓練。

整體而言，Inception-v3 在本資料集與微調策略下最具泛化潛力，為三者中整體性能最穩定的模型。

## 四、結論

本研究透過各種實驗，探討不同策略對 CNN 於性別分類任務之效能影響，針對研究動機提出之三個核心問題，得出以下結論：

### 1. 資料增強對模型預測績效之影響

本研究發現透過資料增強技術，模型的泛化能力與驗證準確率有所提升，但訓練過程的不穩定性也隨之增加。

### 2. 資料增強搭配正則化技術之綜合效益分析

- Dropout：因訓練資料量較小，反而降低模型的預測準確率，無法達到預期泛化效果。
- Batch Normalization：顯著提升訓練準確率，但驗證效果未有明顯改善，甚至略有下降，可能引發輕微過擬合現象。
- L1 正則化：對於本資料集而言效益有限，模型性能無明顯提升，反而增加訓練不穩定性。
- L2 正則化：在有限資料情境中展現適當效益，能小幅提升泛化性能，但也伴隨著訓練誤差增加。
- Bagging：因資料量小易造成過擬合，集成後模型表現差。

整體而言，在有限資料量下，資料增強搭配 L2 正則化效果相對較佳，但模型可能會不太穩定及有過擬合風險，其他正則化技術的使用則需根據資料特性與模型表現再做判斷選擇。

### 3. 轉移學習績效提升效果

本研究評估 VGG-16、Inception-v3 及 ResNet50 三種預訓練模型進行微調後之表現：

- VGG-16：微調階段顯著提升模型預測績效（79.8%），但整體驗證性能穩定性不足，最終準確率僅達 60%，反映出較高的過擬合風險。
- Inception-v3：在微調階段展現最佳泛化效能（78.8%），最終驗證準確率維持於較高水準（76.3%），整體表現最為穩定。
- ResNet50：微調後效能雖有提升，但整體績效低於 Inception-v3，且對資料量敏感，未達理想泛化效果。

透過實驗分析，轉移學習明顯能提升模型於性別分類任務中的預測績效，尤其是 Inception-v3 架構，展現出更佳的穩定性與泛化能力，為性別分類模型之推薦基礎模型架構。

## References

- [1]. Ding, C., & Tao, D. (2016). A comprehensive survey on pose-invariant face recognition. *ACM Transactions on intelligent systems and technology (TIST)*, 7(3), 1-42.
- [2]. Parkhi, O., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. *BMVC 2015-Proceedings of the British Machine Vision Conference 2015*,
- [3]. Pawar, A. B., Shitole, G., Naik, S., Patil, R., & Thakur, Y. (2024). Intelligence Video Surveillance Using Deep Learning. *International Journal of Software Computing and Testing*, 10(1), 9-20p.
- [4]. Sawant, M. M., & Bhurchandi, K. M. (2019). Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging. *Artificial Intelligence Review*, 52, 981-1008.
- [5]. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.
- [6]. Tan, X., & Triggs, B. (2010). Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE transactions on image processing*, 19(6), 1635-1650.
- [7]. Wang, J., & Perez, L. (2017). The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks Vis. Recognit*, 11(2017), 1-8.
- [8]. Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in neural information processing systems*, 27.
- [9]. Zeng, D., Veldhuis, R., & Spreeuwers, L. (2021). A survey of face recognition techniques under occlusion. *IET biometrics*, 10(6), 581-606.