

# 基於圖片識別人物表情實驗報告

## 一、摘要

本實驗聚焦於利用卷積神經網絡（CNN）對源自 Kaggle 的 FER2013 數據集進行深度挖掘，旨在實現精准的圖片人物表情識別。模型在驗證集上準確率約處於 57% - 58% 區間，後續將基於這些結果，深入剖析現有困境，圍繞實驗全過程展開詳細闡述，探尋改進方向。

## 二、引言

### 2.1 研究背景

在當今數字化浪潮下，人物表情識別技術在人機交互、安防監控、數字娛樂等諸多領域展現出巨大的應用潛力。智能客服借助表情識別可精准感知用戶情緒，及時優化服務策略；安防系統能通過分析人群表情提前預警潛在風險。據市場研究機構預測，全球表情識別市場規模將以迅猛態勢增長，彰顯出該技術的迫切需求與廣闊前景。

### 2.2 目的

本次實驗意在充分發揮 CNN 的圖像特徵自動提取優勢，鑽研 FER2013 數據集，構建泛化性強的圖片人物表情識別模型。

### 三、相關理論與技術基礎

#### 3.1 人物表情識別原理

人類面部表情受大腦神經調控，引發面部肌肉協同運動，遵循面部動作編碼系統（FACS）規律。例如，快樂時嘴角上揚、眼部收縮，悲傷則伴隨嘴角下撇、眉頭緊鎖等特徵。

#### 3.2 基於 CNN 的表情識別流程

原始圖像輸入後，卷積層憑借滑動卷積核捕捉人臉輪廓、五官細節等局部特徵；池化層通過下採樣降低特徵圖分辨率，保留關鍵特徵同時減輕計算負擔；全連接層整合全部，映射至表情分類空間輸出預測結果。相較傳統手工特徵提取，CNN 自動學習特徵的能力大幅提升效率與準確性。

#### 3.3 涉及的算法與技術

實驗選用優化後的 AlexNet 變體作為 CNN 模型，其結構精細調整。含 5 個卷積層，前兩層以  $3 \times 3$  小卷積基精細提取面部細節，後續層依序增大卷積核尺寸捕捉特徵，通道數從 64 逐步遞增至 256。

訓練優化層面，歸一化技術穩定每一層輸入數據分布，加速收斂、規避梯度異常；Dropout 層依 0.4 概率隨機斷開神經元連接，預防過擬合，強化模型泛化性能。

針對 FER2013 數據集，將圖像歸一化至模型要求尺寸，利用數據增強手段豐富訓練樣本多樣性，包括旋轉、平移、縮放、錯切、翻轉等操作，採用最鄰近像素填充法處理變換後的圖像。

## 四、實驗方案設計

### 4.1 數據集選取

FER2013 數據集源於 Kaggle，是表情識別領域經典標桿，涵蓋 7 種基本表情，超 3 萬張訓練圖片，人物來源廣泛，種族、年齡、性別多樣，代表性極強，且已精准標注表情類別。

### 4.2 實驗流程與步驟

實驗涵蓋數據加載、預處理、模型構建與訓練、模型評估環節，詳細步驟如下：

- 數據加載：運用 TensorFlow 內置函數讀取 FER2013 數據集至內存。  
通過 `flow_from_directory` 方法從指定文件夾（`train_folder`、`val_folder`）加載圖像數據，生成訓練生成器 `train_generator` 和驗證生成器 `valid_generator`。
- 預處理：對加載數據進行標準化操作，圖像歸一化、灰度轉換、高斯濾波依次執行，淨化數據提升質量。利用 `ImageDataGenerator` 類進行了

豐富的數據增強操作，如 `rotation_range=40` 表示圖像隨機旋轉角度範圍為 40 度，`width_shift_range=0.2` 和 `height_shift_range=0.2` 是水平和垂直方向的平移範圍，`shear_range=0.2` 為錯切變換範圍，`zoom_range=0.2` 是縮放範圍，`horizontal_flip=True` 進行水平翻轉，這些操作旨在擴充訓練數據，提高模型的泛化能力。

- 模型構建：基於優化 AlexNet 變體搭建 CNN 模型，卷積核大小為 3，填充方式為 `same` 確保圖像尺寸不變，激活函數為 `relu` 引入非線性，接著是三個 `MaxPool2D` 池化層，池化窗口大小為 2 進行下採樣，之後通過 `Flatten` 層展平數據，再連接兩個全連接層，第一個全連接層有 128 個神經元，激活函數為 `relu`，並使用 `Dropout(0.4)` 防止過擬合，最後一個全連接層輸出節點數為 7（對應 7 種表情類別），激活函數為 `softmax` 用於多分類任務。
- 模型訓練：設定初始學習率 0.001，每 10 個 epoch 衰減 0.1，迭代 100 個 epoch，批處理大小 64，啟用 Adam 優化器，依交叉熵損失函數反向傳播更新參數。代碼中使用 `model.compile` 方法編譯模型，指定優化器為 `adam`，損失函數為 `categorical_crossentropy` 適用於多分類問題，評估指標為 `accuracy`，然後通過 `model.fit_generator` 方法利用訓練生成器和驗證生成器進行模型訓練，訓練次數 為 50，`steps_per_epoch`

和 `validation_steps` 分別根據訓練集和驗證集樣本數及批處理大小計算得出，確保每個 `epoch` 訓練和驗證完整遍歷數據集。

- 模型評估：訓練中每 `epoch` 結束在驗證集評估性能，記錄損失與準確率。模型在訓練集上的損失值約為 1.11 - 1.18 之間，準確率在 0.55 - 0.55 左右波動；在驗證集上，損失值約為 1.11，準確率約為 0.57 - 0.57 之間，後續需進一步分析這些初步結果，探尋模型性能提升路徑。

```
Epoch 49/50  
448/448 — 51s 113ms/step - accuracy: 0.5576 - loss: 1.1773 - val_accuracy: 0.5767 - val_loss: 1.1016  
Epoch 50/50  
448/448 — 2s 4ms/step - accuracy: 0.5625 - loss: 1.1188 - val_accuracy: 0.5773 - val_loss: 1.1011
```

## 五、實驗結果與分析

### 5.1 模型訓練結果

根據已有結果，訓練集損失值隨 `epoch` 上升穩步下降，驗證集損失同步下滑，暗示模型學習良好。

為對抗過擬合，結合數據增強手段，如隨機翻轉、旋轉、裁剪擴充訓練數據，驗證集準確率最終穩定在 57% - 58% 左右。

### 5.2 準確率思考

在多次調整參數後準確率仍無法突破 60%，後續抽檢發現，數據集中混入不少非表情圖片。這不僅破壞數據集純淨度，更可能誤導模型學習，嚴

重影響識別準確率。同時，多樣採集場景雖利於泛化訓練，但非表情圖片的存在使得場景噪聲加劇。

## 六、討論與展望

數據集層面，非表情圖片混入致不平衡加劇，關鍵表情樣本學習受干擾，應人工去除或找更完善的數據集。

## 七、結論

回溯實驗全程，從 CNN 選型、FER2013 數據集應用、環境搭建至模型訓練、評估，各環節緊密相扣，雖受非表情圖片困擾，但積累寶貴經驗。當前驗證集約 57% 準確率，不同表情識別有一定穩定性。雖準確率待提升，但收獲了很多經驗，豐富了自身 CNN 應用實踐。