

期末專題

壹、研究背景：

近年來，隨著卷積神經網路的發展，這些模型已經能有效的完成影像分類的任務，然而訓練一個高效的 CNN 模型往往需要龐大的資料集和長久的訓練時間，並不容易開發。因為從零開始訓練的成本過高，遷移學習成為一個解決方案，利用在 ImageNet 上預先訓練好的模型，將其應用於特定領域或是較小規模的任務中，可以節省訓練時間並且提升模型效能。

貳、研究動機與目的：

市面上已經存在多個 CNN 模型，每個模型的設計理念都不同，我將透過遷移學習的方式，將模型應用在不同的資料集上，比較不同模型對於相同資料集在情緒辨識上的表現

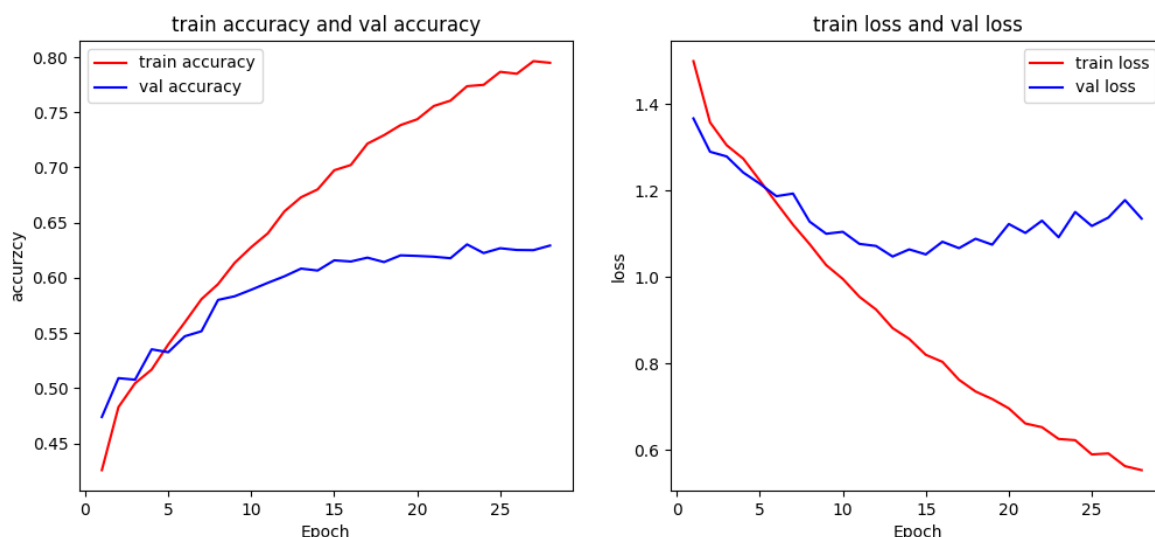
參、研究方法：

每個模型先從 ImageNet 上進行遷移學習並且對 FERplus 資料集進行訓練，將訓練好的模型遷移到 RAF-DB 資料集，再將利用 RAF-DB 訓練的模型遷移到 JAFFE 資料集，最後將用 JAFFE 訓練好的模型遷移到 CK+ 資料集做最後的訓練，完成後總共會有 12 個模型。訓練時為了避免過擬合發生，在訓練初期會凍結大部分預訓練層，隨著訓練解凍更多預訓練層並降低學習率。最後將訓練過程的 accuracy 和 loss 變化用折線圖呈現，再利用 Grad-CAM 和混淆矩陣視覺化呈現模型在測試集上的表現。

肆、實驗結果：

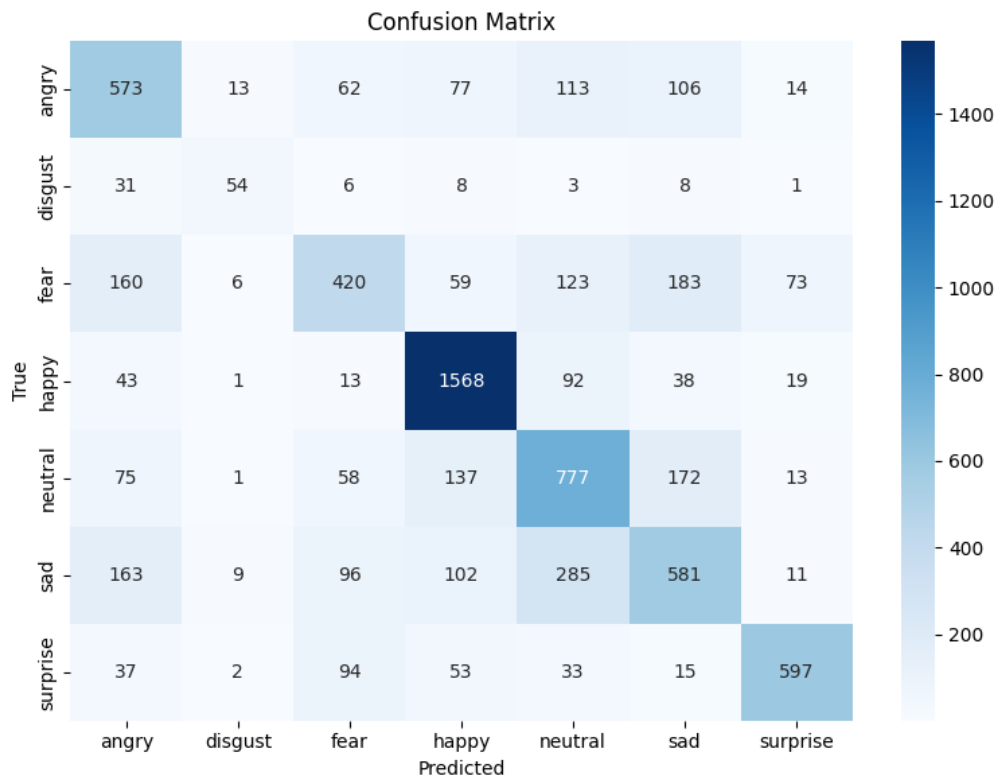
一、VGG19 建立適用於 FERplus 訓練集的情緒辨識模型：

1. 訓練集和驗證集的 accuracy 和 loss 折線圖：



模型在訓練集上的表現較穩定，accuracy 穩定上升，loss 穩定下降，但驗證集的 accuracy 只有在初期到 15 個 epoch 時穩定上升，15 後趨於平緩，且 loss 在 15 個 epoch 後有震盪且上升的現象，模型有過擬合現象發生。

2. 測試集的混淆矩陣：



3. 整體結果分析：

Test Accuracy: 0.6367

Classification Report:

	precision	recall	f1-score	support
angry	0.53	0.60	0.56	958
disgust	0.63	0.49	0.55	111
fear	0.56	0.41	0.47	1024
happy	0.78	0.88	0.83	1774
neutral	0.54	0.63	0.58	1233
sad	0.53	0.47	0.49	1247
surprise	0.82	0.72	0.77	831
accuracy			0.64	7178
macro avg	0.63	0.60	0.61	7178
weighted avg	0.63	0.64	0.63	7178

表中可以看出 happy 和 surprise 的表現最好，而 fear、sad、disgust 的辨識效果差，其中 disgust 可能是因為屬於此類別的太少，造成模型沒有學到很多關於 disgust 的特徵，而 angry 和 neutral 的表現尚可。

4. Grad-CAM 熱圖分析：

True: disgust, Pred: sad
Input Image



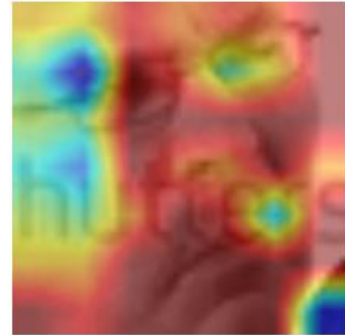
Grad-CAM



True: disgust, Pred: angry
Input Image



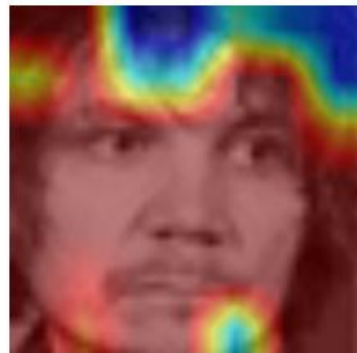
Grad-CAM



True: sad, Pred: neutral
Input Image



Grad-CAM



True: surprise, Pred: happy
Input Image



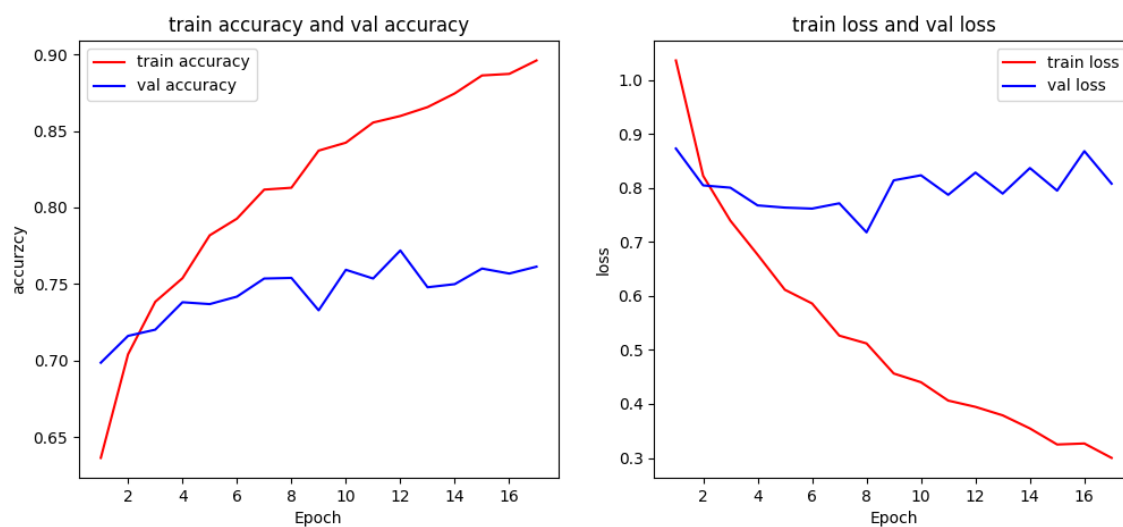
Grad-CAM



這裡找出幾個錯誤的例子分析，對於 FERplus 資料集而言，有些資料的圖片，連我都很難分辨出真正的表情為何，有些表情和其他表情過於相似，例如右下角的 surprise，以及左上角的 disgust，其實也不太能說模型預測出錯，因為這些表情和模型的預測結果也十分相似，但左上角的 disgust 模型就沒有學到 disgust 的吐舌頭的特徵，也印證了，disgust 資料數量少，模型沒有學到關於真正學到關於 disgust 的特徵。

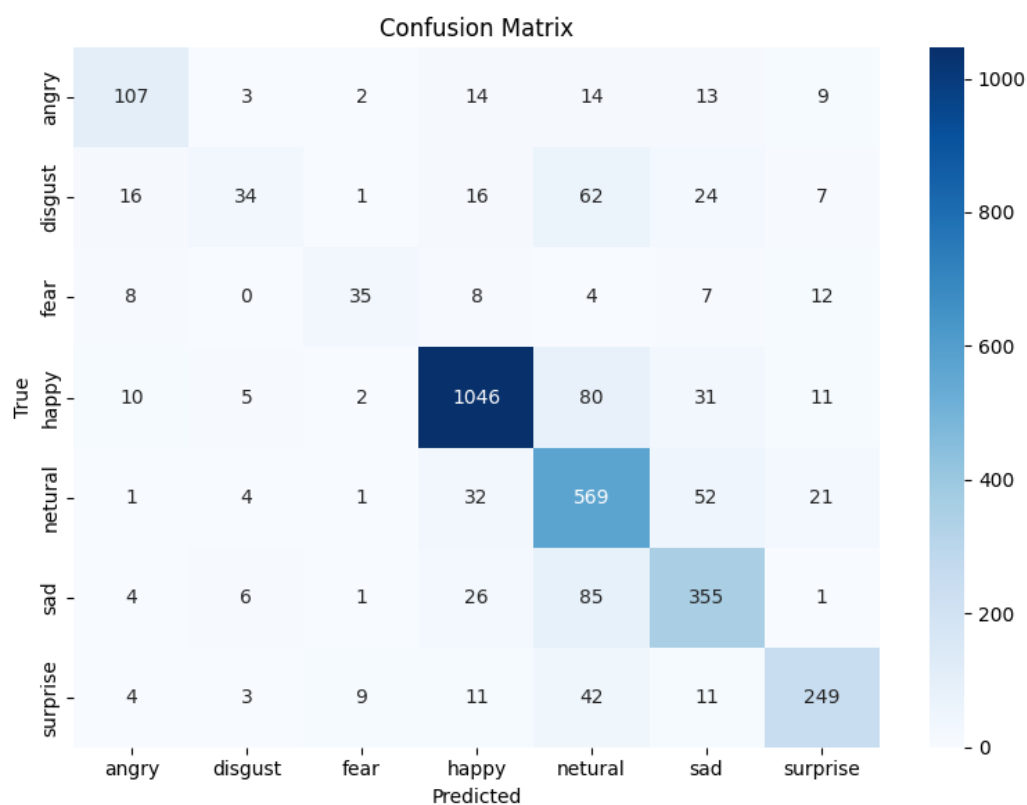
二、VGG19 建立適用於 RAF-DB 訓練集的情緒辨識模型：

1. 訓練集和驗證集的 accuracy 和 loss 折線圖：



訓練集和驗證集的準確率都有進一步上升，但驗證集的 loss 依然出現震盪的現象，也同樣有過擬合的現象發生。

2. 測試集的混淆矩陣：



3. 整體結果分析：

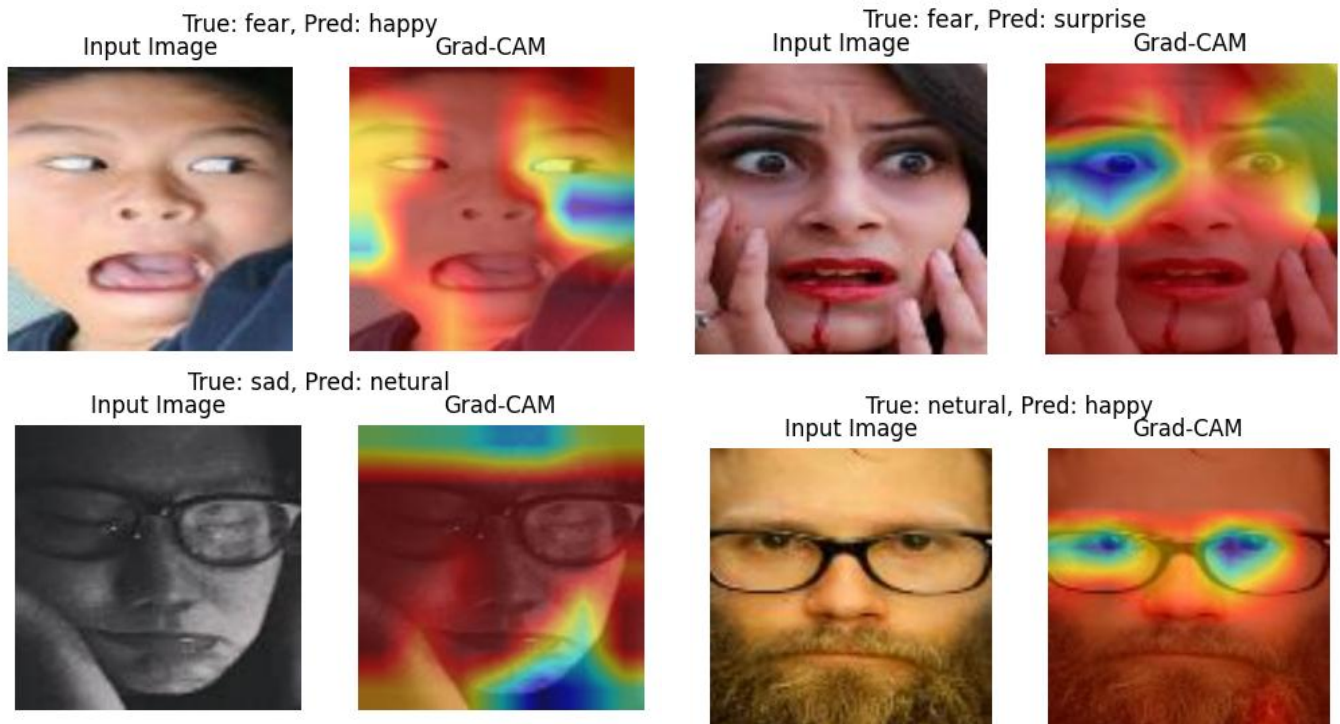
Test Accuracy: 0.7806

Classification Report:

	precision	recall	f1-score	support
angry	0.71	0.66	0.69	162
disgust	0.62	0.21	0.32	160
fear	0.69	0.47	0.56	74
happy	0.91	0.88	0.89	1185
netural	0.66	0.84	0.74	680
sad	0.72	0.74	0.73	478
surprise	0.80	0.76	0.78	329
accuracy			0.78	3068
macro avg	0.73	0.65	0.67	3068
weighted avg	0.78	0.78	0.77	3068

happy 的 f1-score 從 0.83 變成 0.89 代表模型表現非常穩定且準確，表示模型對這類情緒學得很好、泛化能力強，sad 的 recall 和 F1-score 明顯改善，模型現在學會辨識 sad 特徵，neutral 的 recall 和 F1-score 也有改善，模型現在也學會辨識 neutral 特徵，而 disgust 在此資料集的資料也很少，進一步造成 disgust 的表現下降。

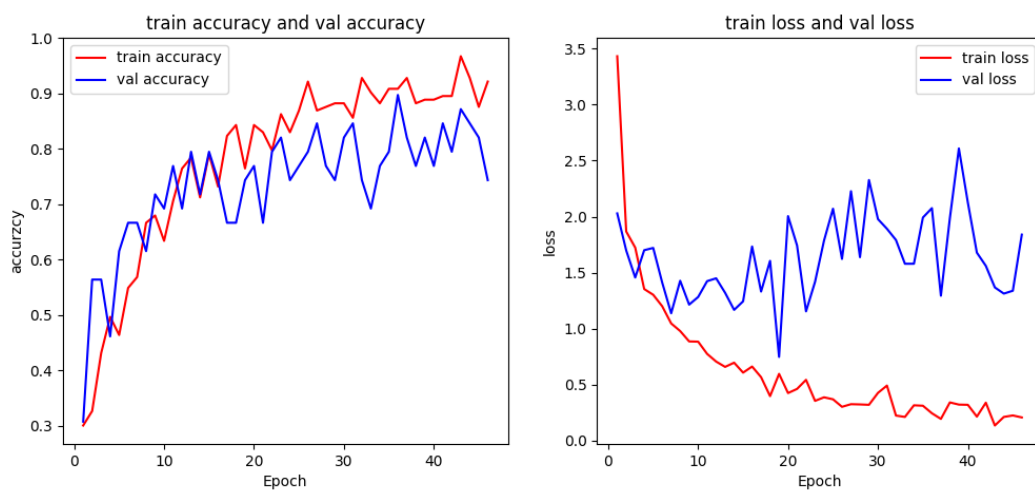
4. Grad-CAM 熱圖分析：



雖然模型學會辨識 sad 和 netural 特徵，但還是有錯誤的識別狀況，而模型還是沒有辦法準確做出 fear 的識別。

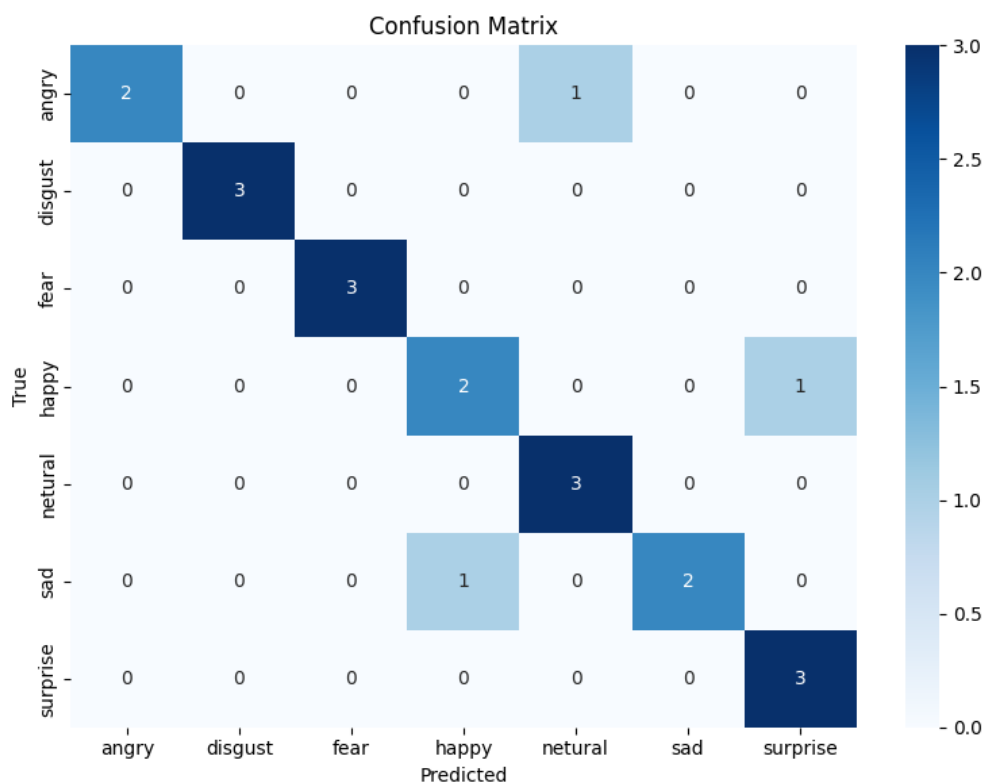
三、VGG19 建立適用於 JAFFE 訓練集的情緒辨識模型：

1. 訓練集和驗證集的 accuracy 和 loss 折線圖：



模型 accuracy 和 loss 都有明顯震盪現象，推測是資料集中資料少，造成模型的 accuracy 和 loss 震盪

2. 測試集的混淆矩陣：



3. 整體結果分析：

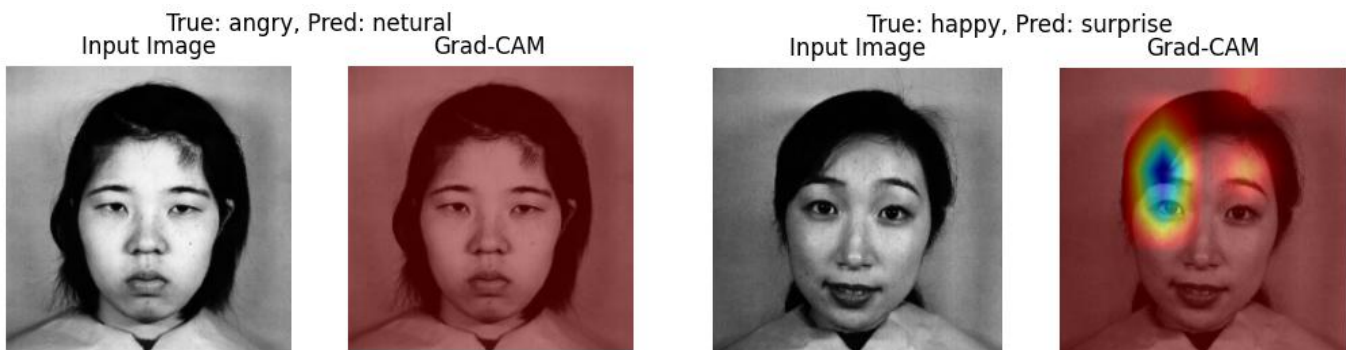
Test Accuracy: 0.8571

Classification Report:

Classification Report:				
	precision	recall	f1-score	support
angry	1.00	0.67	0.80	3
disgust	1.00	1.00	1.00	3
fear	1.00	1.00	1.00	3
happy	0.67	0.67	0.67	3
netural	0.75	1.00	0.86	3
sad	1.00	0.67	0.80	3
surprise	0.75	1.00	0.86	3
accuracy			0.86	21
macro avg	0.88	0.86	0.85	21
weighted avg	0.88	0.86	0.85	21

模型在某些特徵下可以正確分類，但在前面表現好的某些特徵的表現卻下降，我推測是因為測試集大小過小，而特徵的誤判現象依然存在，這樣一來造成，模型只要誤判一個類別的特徵，這個類別的準確率看起來就很低，但實際上整體準確率還是有上升。

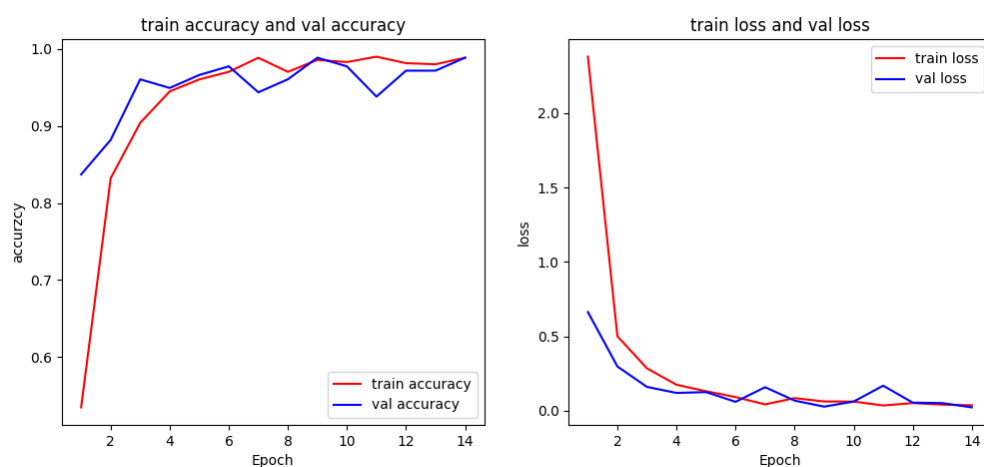
4. Grad-CAM 熱圖分析：



特定幾個特徵的預測錯誤造成該標籤的準確率下降。

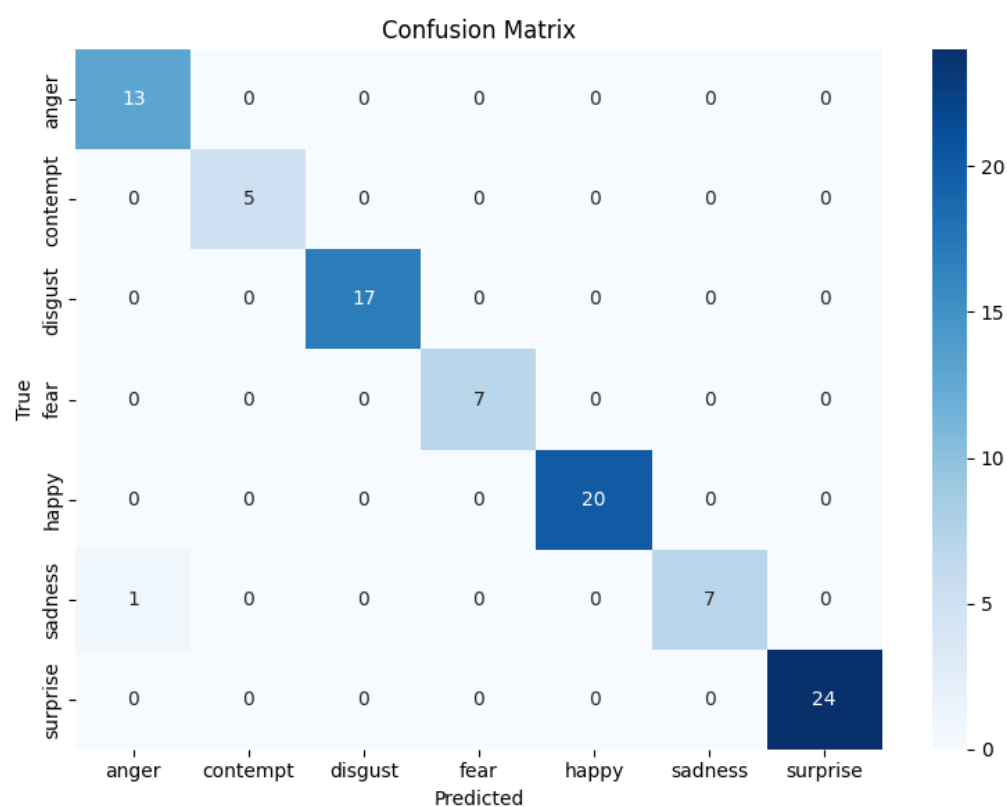
四、VGG19 建立適用於 CK+訓練集的情緒辨識模型：

1. 訓練集和驗證集的 accuracy 和 loss 折線圖：



訓練集和驗證集的表現良好，準確率最後高達 100%，也沒有過擬合現象發生

2. 測試集的混淆矩陣：



3. 整體結果分析：

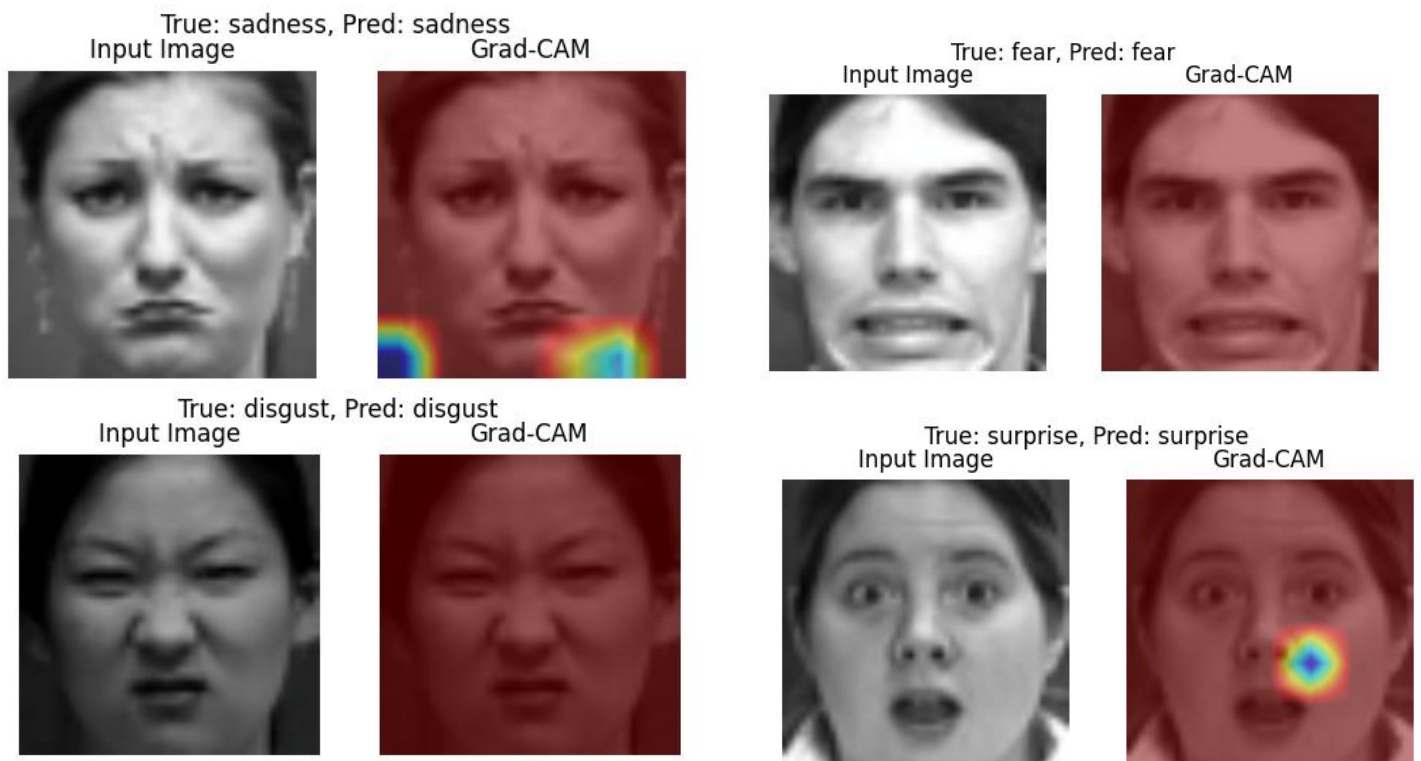
Test Accuracy: 0.9894

Classification Report:

	precision	recall	f1-score	support
anger	0.93	1.00	0.96	13
contempt	1.00	1.00	1.00	5
disgust	1.00	1.00	1.00	17
fear	1.00	1.00	1.00	7
happy	1.00	1.00	1.00	20
sadness	1.00	0.88	0.93	8
surprise	1.00	1.00	1.00	24
accuracy			0.99	94
macro avg	0.99	0.98	0.99	94
weighted avg	0.99	0.99	0.99	94

模型幾乎全部預測正確。

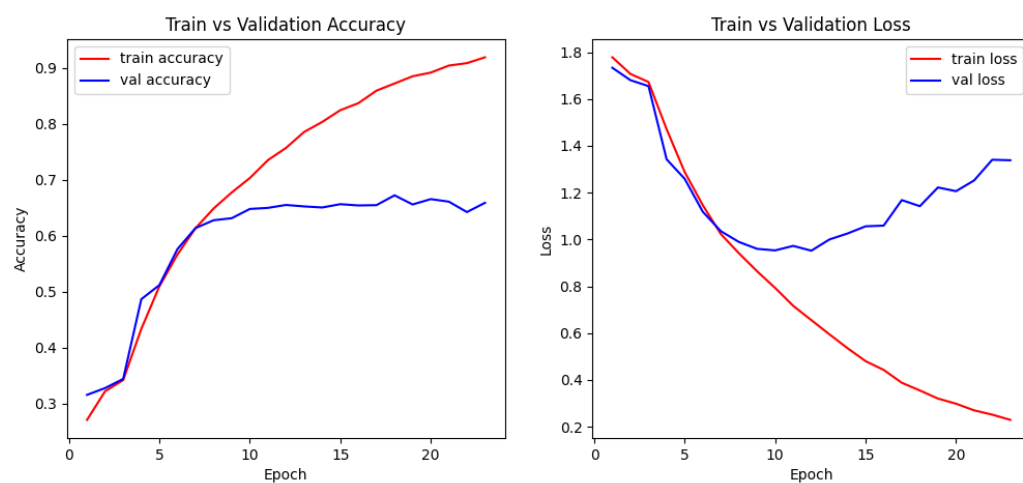
4. Grad-CAM 熱圖分析：



所有特徵幾乎都正確分類。

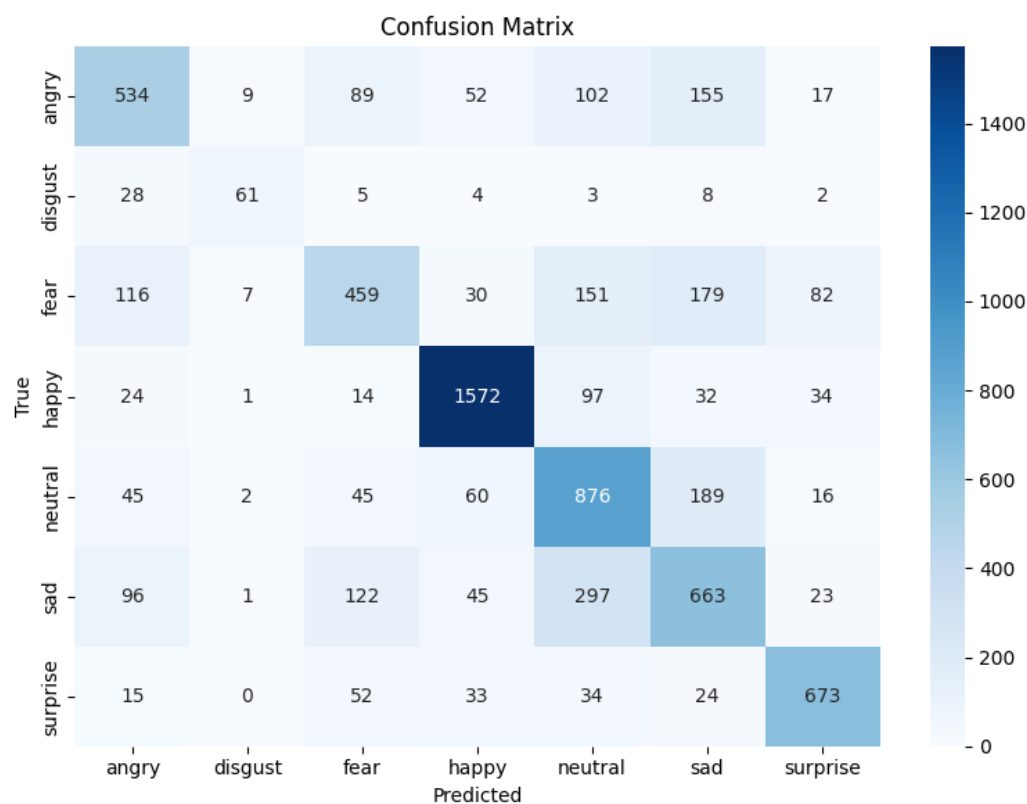
五、ResNet50 建立適用於 FERplus 訓練集的情緒辨識模型：

1. 訓練集和驗證集的 accuracy 和 loss 折線圖：



模型的過擬合現象更加明顯，驗證集 loss 在後面的 epoch 不減反增。

2. 測試集的混淆矩陣：



3. 整體結果分析：

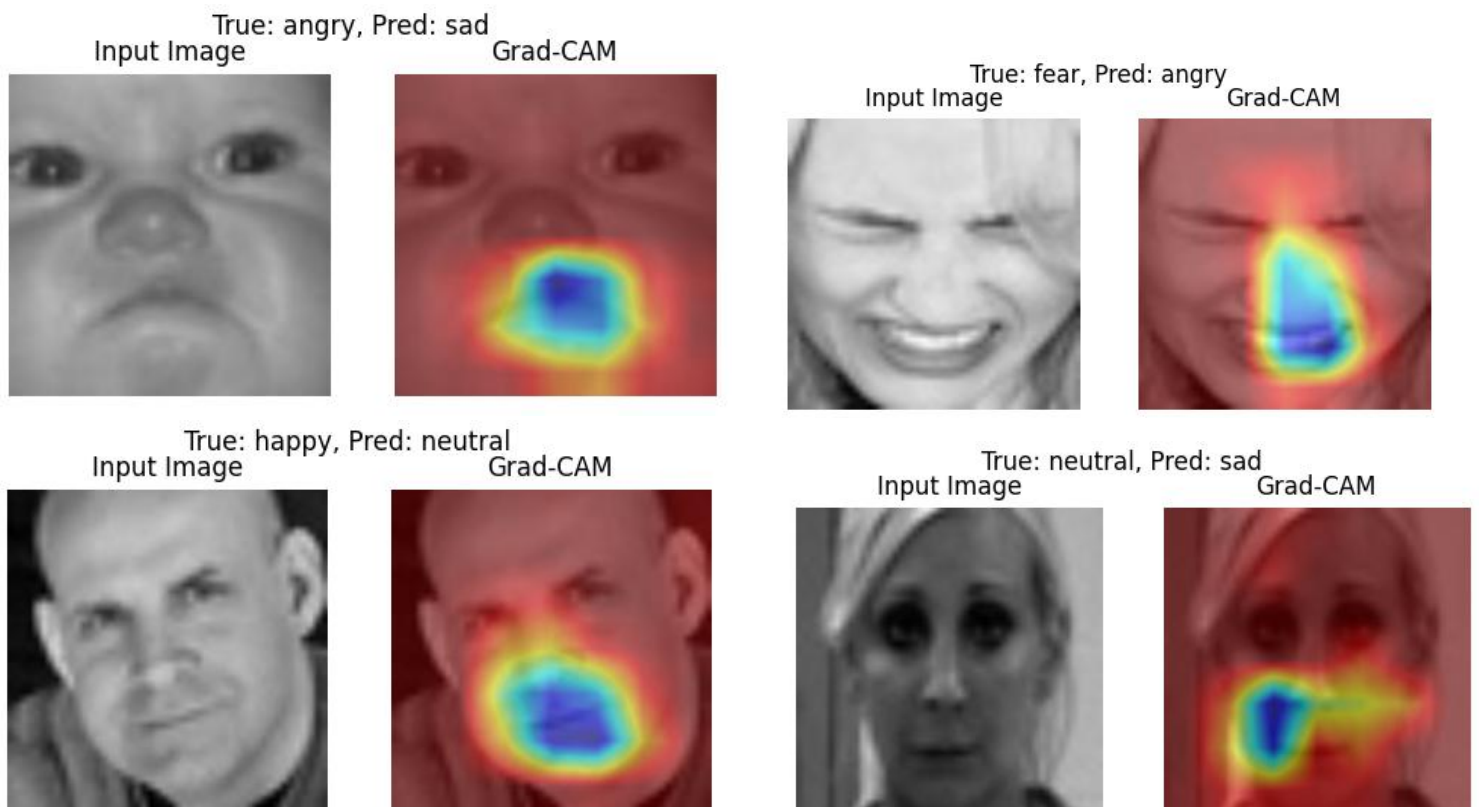
Test Accuracy: 0.6740

Classification Report:

	precision	recall	f1-score	support
angry	0.62	0.56	0.59	958
disgust	0.75	0.55	0.64	111
fear	0.58	0.45	0.51	1024
happy	0.88	0.89	0.88	1774
neutral	0.56	0.71	0.63	1233
sad	0.53	0.53	0.53	1247
surprise	0.79	0.81	0.80	831
accuracy			0.67	7178
macro avg	0.67	0.64	0.65	7178
weighted avg	0.67	0.67	0.67	7178

整體的準確率達到 0.67，除了 fear 的特徵表現較差之外，其餘的特徵分辨和 VGG19 相比都表現不錯。

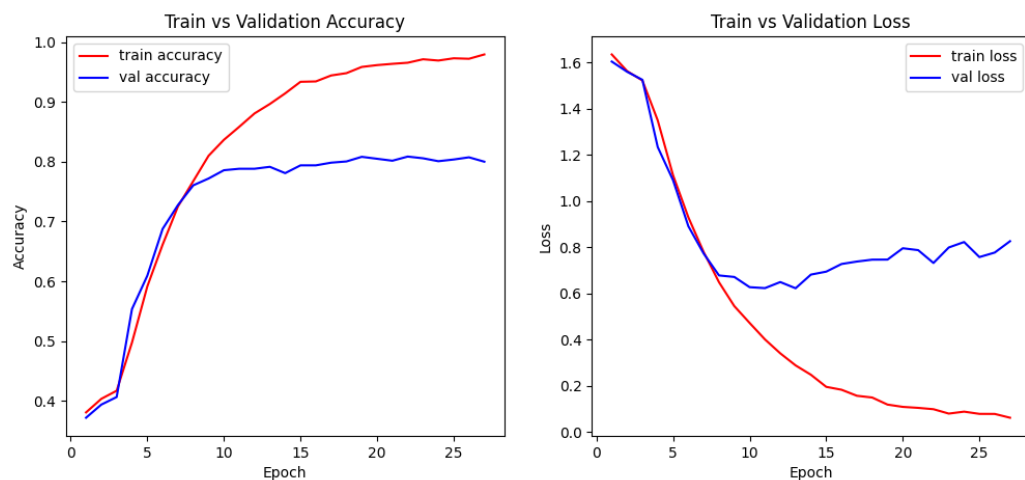
4. Grad-CAM 熱圖分析：



模型整體情緒分辨表現不錯，但有些相似的表情還是無法做到完美分辨。

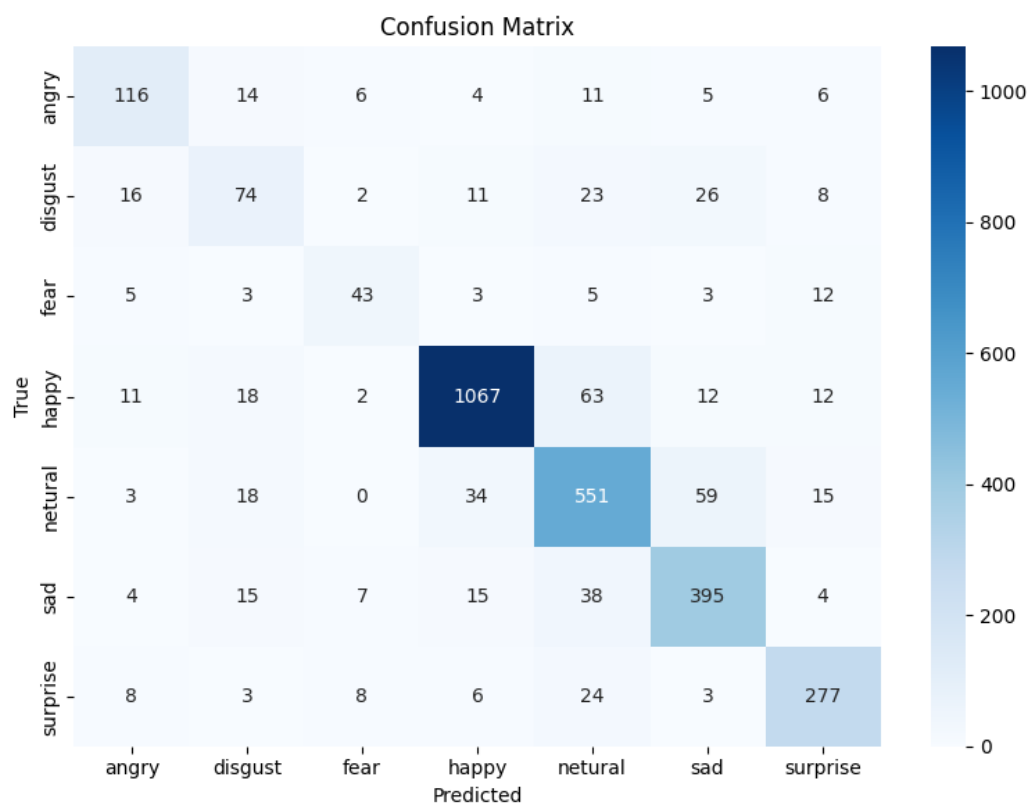
六、ResNet50 建立適用於 RAF-DB 訓練集的情緒辨識模型：

1. 訓練集和驗證集的 accuracy 和 loss 折線圖：



雖然模型整體的準確率有上升，但依然有過擬合現象。

2. 測試集的混淆矩陣：



3. 整體結果分析：

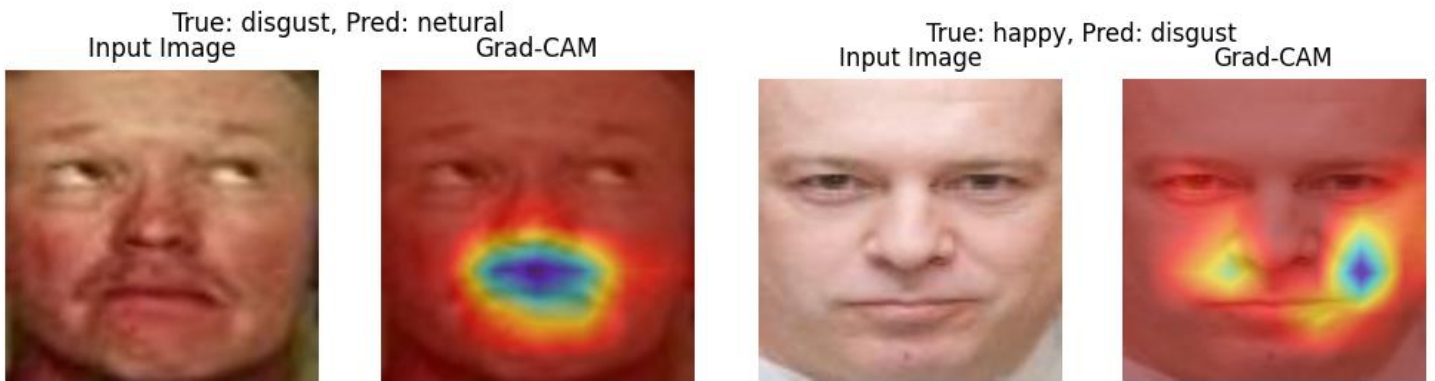
Test Accuracy: 0.8224

Classification Report:

	precision	recall	f1-score	support
angry	0.71	0.72	0.71	162
disgust	0.51	0.46	0.49	160
fear	0.63	0.58	0.61	74
happy	0.94	0.90	0.92	1185
netural	0.77	0.81	0.79	680
sad	0.79	0.83	0.81	478
surprise	0.83	0.84	0.84	329
accuracy			0.82	3068
macro avg	0.74	0.73	0.74	3068
weighted avg	0.82	0.82	0.82	3068

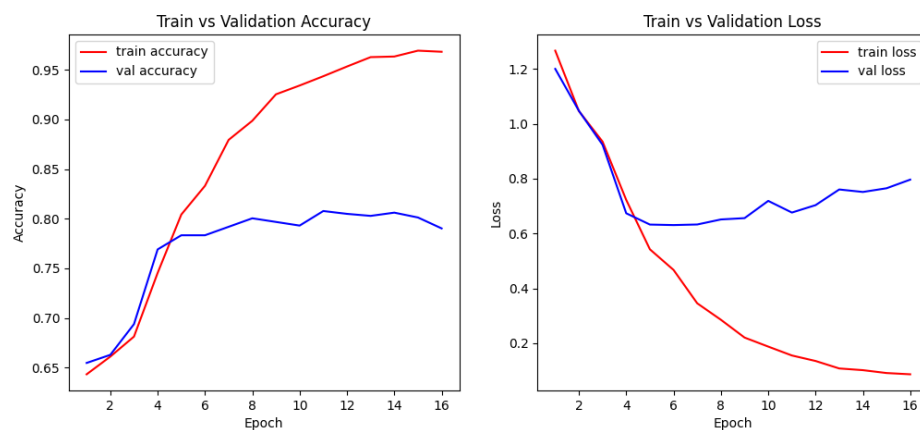
除了 disgust 可能因為訓練資料量不多，表現不是很好之外，上一個資料集表現差的 fear 特徵在這個資料集中學到如何有效分辨，模型的 fear 分辨表現提高。

4. Grad-CAM 熱圖分析：



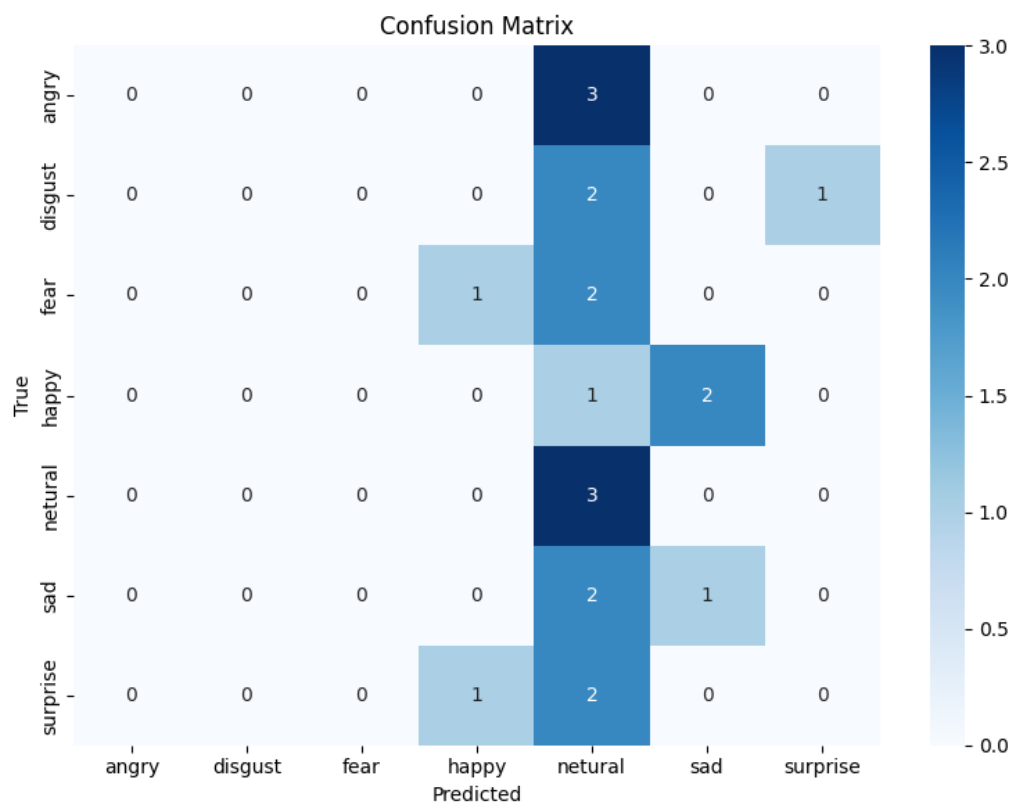
七、ResNet50 建立適用於 JAFFE 訓練集的情緒辨識模型：

1. 訓練集和驗證集的 accuracy 和 loss 折線圖：



雖然有過擬合現象，但模型表現平穩。

2. 測試集的混淆矩陣：



3. 整體結果分析：

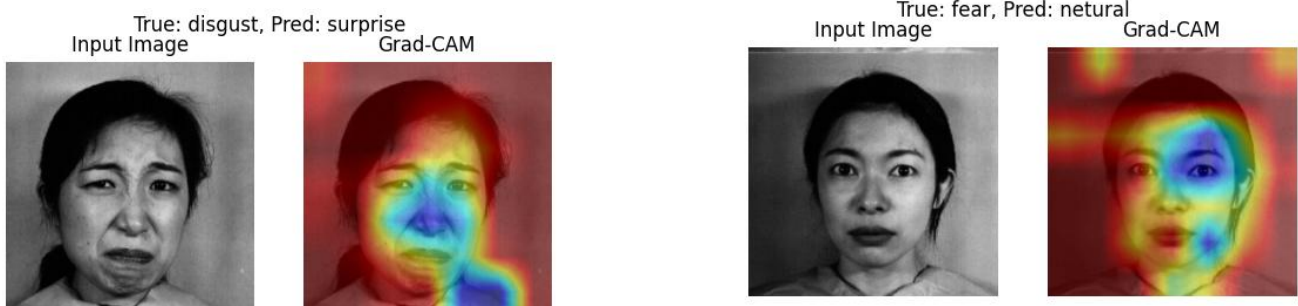
Test Accuracy: 0.1905

Classification Report:

precision	recall	f1-score	support	
angry	0.00	0.00	0.00	3
disgust	0.00	0.00	0.00	3
fear	0.00	0.00	0.00	3
happy	0.00	0.00	0.00	3
netural	0.20	1.00	0.33	3
sad	0.33	0.33	0.33	3
surprise	0.00	0.00	0.00	3
accuracy			0.19	21
macro avg	0.08	0.19	0.10	21
weighted avg	0.08	0.19	0.10	21

雖然模型在訓練時表現平穩，但事實上是模型並沒有在此資料及學到任何特徵的分辨，推測是前面過擬合的現象加上測試集過小造成的結果。

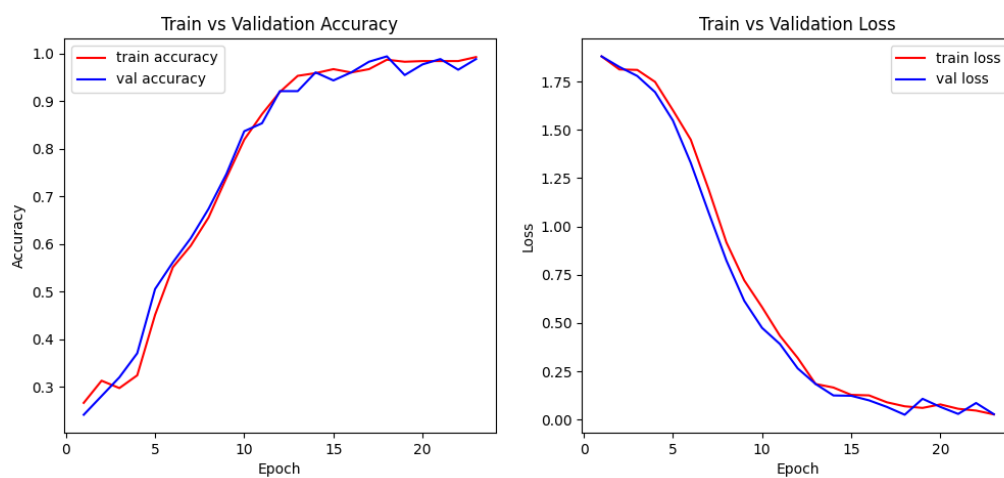
4. Grad-CAM 熱圖分析：



幾乎所有的結果都預測錯誤。

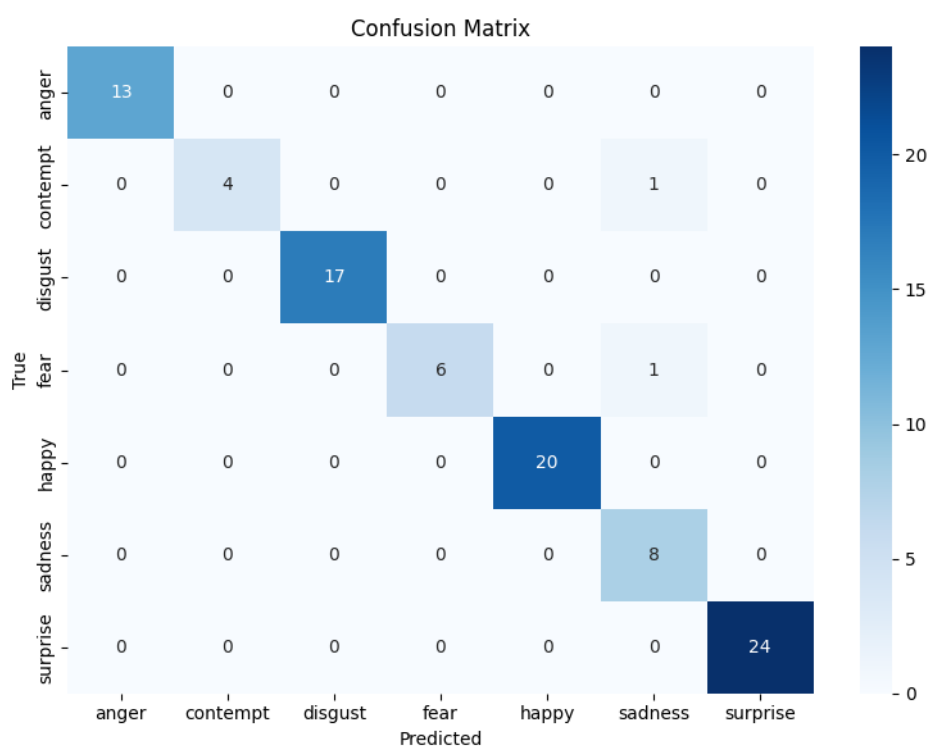
八、ResNet50 建立適用於 CK+訓練集的情緒辨識模型：

1. 訓練集和驗證集的 accuracy 和 loss 折線圖：



模型訓練狀況良好，無過擬合現象。

2. 測試集的混淆矩陣：



3. 整體結果分析：

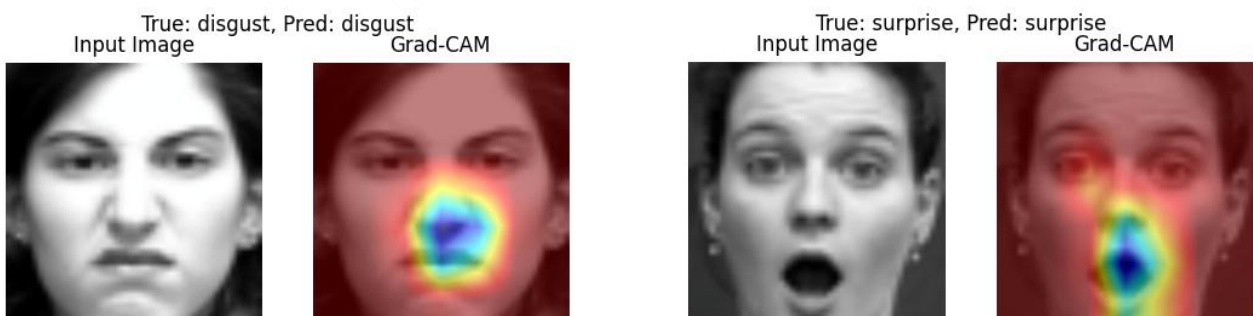
Test Accuracy: 0.9787

Classification Report:

	precision	recall	f1-score	support
anger	1.00	1.00	1.00	13
contempt	1.00	0.80	0.89	5
disgust	1.00	1.00	1.00	17
fear	1.00	0.86	0.92	7
happy	1.00	1.00	1.00	20
sadness	0.80	1.00	0.89	8
surprise	1.00	1.00	1.00	24
accuracy			0.98	94
macro avg	0.97	0.95	0.96	94
weighted avg	0.98	0.98	0.98	94

幾乎所有特徵都能有效分辨。

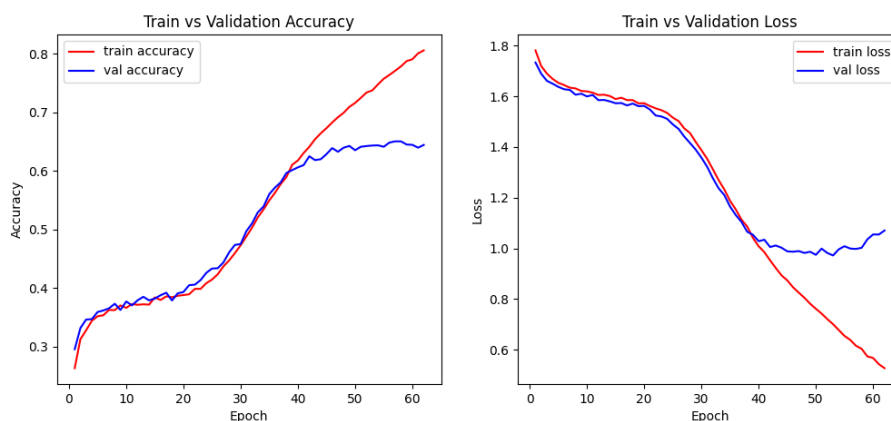
4. Grad-CAM 熱圖分析：



幾乎所有特徵都正確分辨

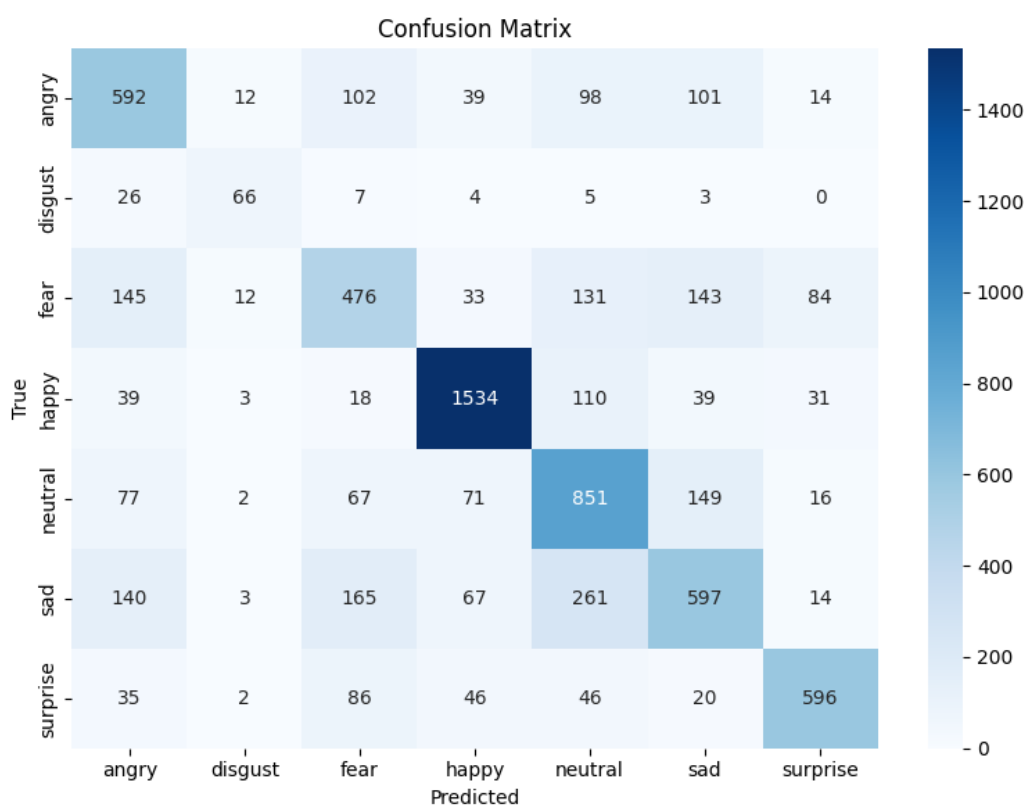
九、MobileNetV2 建立適用於 FERplus 訓練集的情緒辨識模型：

1. 訓練集和驗證集的 accuracy 和 loss 折線圖：



模型訓練表現還不錯，雖然後期有過擬合現象發生，但是被 early stopping 中止。

2. 測試集的混淆矩陣：



3. 整體結果分析：

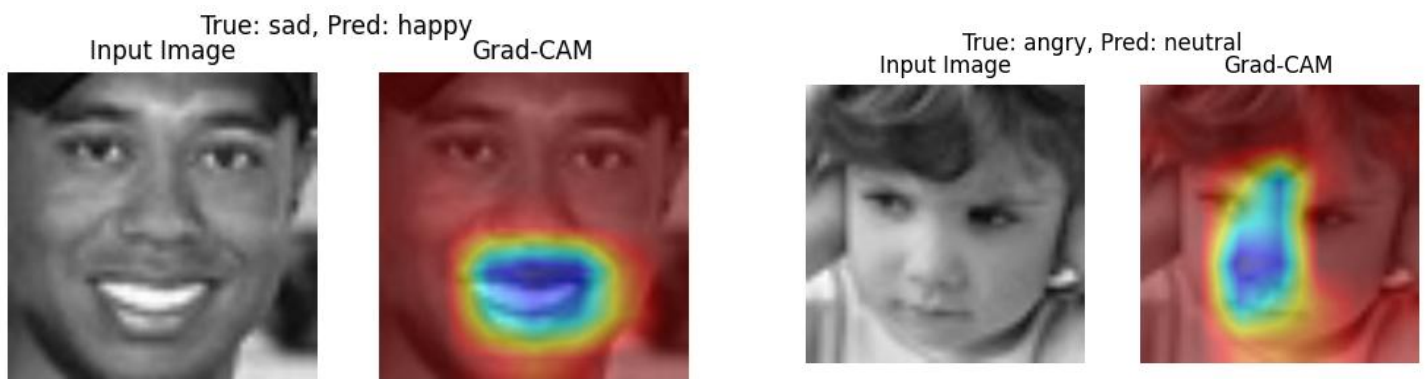
Test Accuracy: 0.6565

Classification Report :

	precision	recall	f1-score	support
angry	0.56	0.62	0.59	958
disgust	0.66	0.59	0.63	111
fear	0.52	0.46	0.49	1024
happy	0.86	0.86	0.86	1774
neutral	0.57	0.69	0.62	1233
sad	0.57	0.48	0.52	1247
surprise	0.79	0.72	0.75	831
accuracy			0.66	7178
macro avg	0.65	0.63	0.64	7178
weighted avg	0.66	0.66	0.65	7178

除了 fear 和 sad 其他特徵分辨表現都不錯。

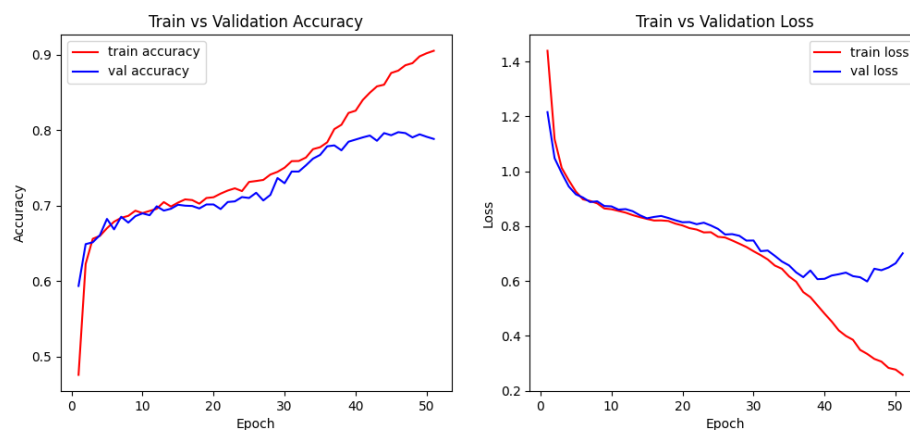
4. Grad-CAM 熱圖分析：



除了一些相似的表情特徵，模型都能正確分辨。

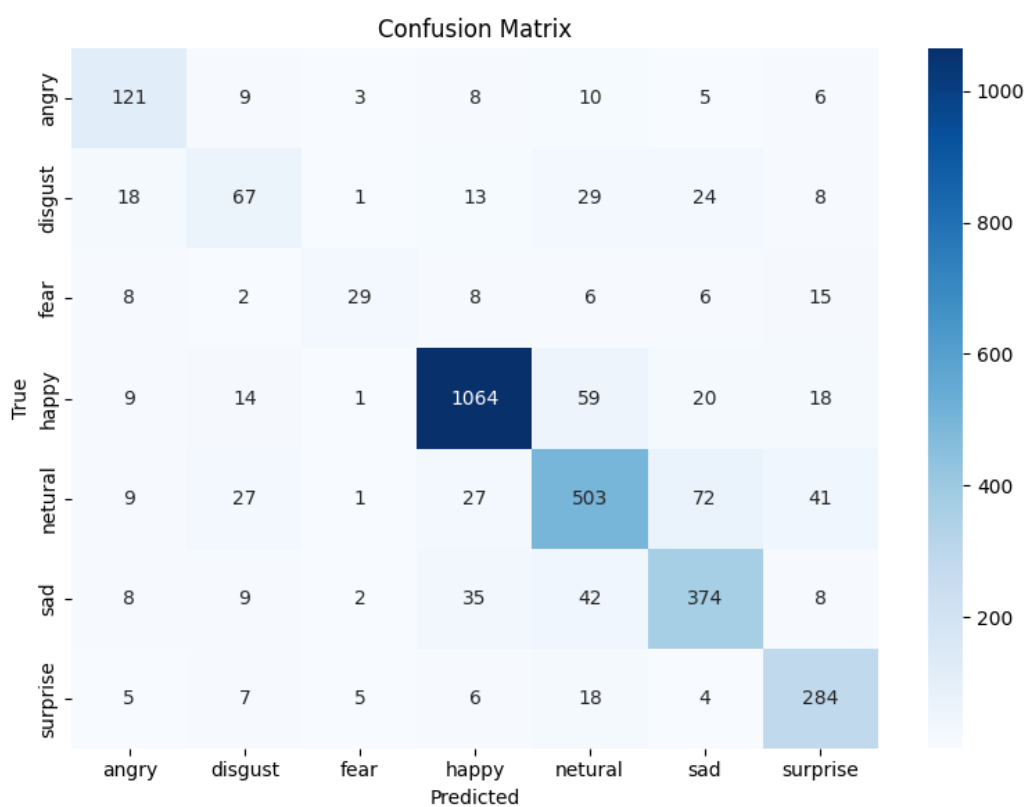
十、MobileNetV2 建立適用於 RAF-DB 訓練集的情緒辨識模型：

1. 訓練集和驗證集的 accuracy 和 loss 折線圖：



訓練表現不錯，整體準確率上升，雖然過擬合現象依然在最後發生。

2. 測試集混淆矩陣：



3. 整體結果分析：

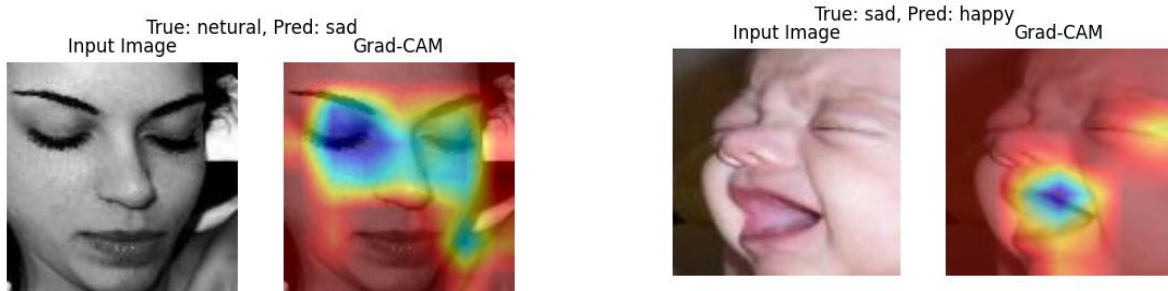
Test Accuracy: 0.7960

Classification Report:

	precision	recall	f1-score	support
angry	0.68	0.75	0.71	162
disgust	0.50	0.42	0.45	160
fear	0.69	0.39	0.50	74
happy	0.92	0.90	0.91	1185
netural	0.75	0.74	0.75	680
sad	0.74	0.78	0.76	478
surprise	0.75	0.86	0.80	329
accuracy			0.80	3068
macro avg	0.72	0.69	0.70	3068
weighted avg	0.80	0.80	0.79	3068

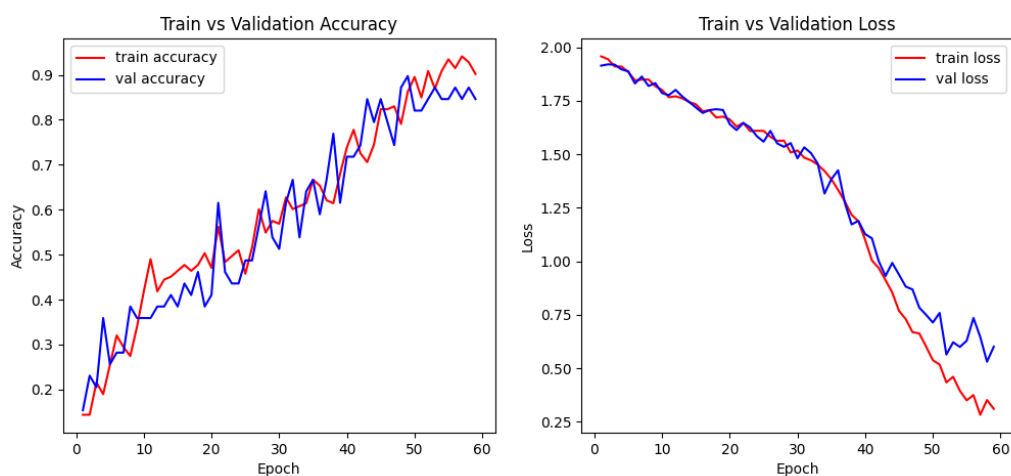
disgust 和 fear 分辨力差，推測是因為資料量少，模型無法有效學習這兩項特徵。

4. Grad-CAM 熱圖分析：



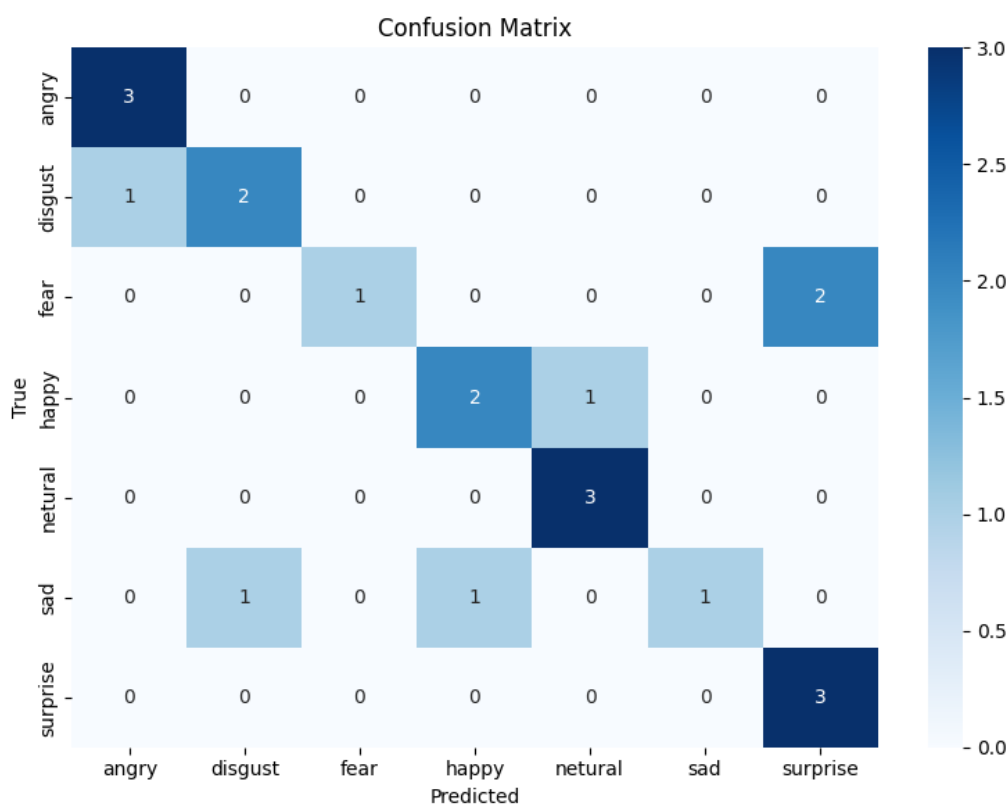
十一、 MobileNetV2 建立適用於 JAFFE 訓練集的情緒辨識模型：

1. 訓練集和驗證集的 accuracy 和 loss 折線圖：



模型整體訓練情況良好，雖然有明顯的 accuracy 和 loss 震盪問題，但整體的 accuracy 穩定上升，loss 穩定下降，過擬合的現象也不明顯。

2. 測試集混淆矩陣：



3. 整體結果分析：

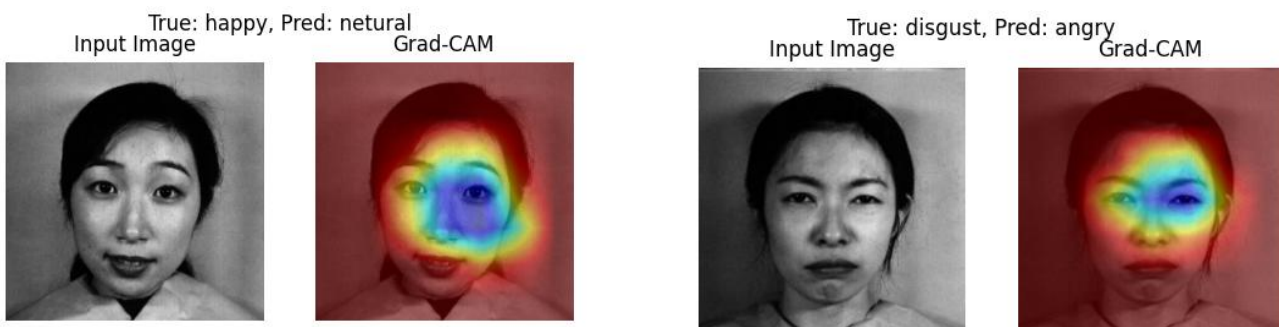
Test Accuracy: 0.7143

Classification Report:

	precision	recall	f1-score	support
angry	0.75	1.00	0.86	3
disgust	0.67	0.67	0.67	3
fear	1.00	0.33	0.50	3
happy	0.67	0.67	0.67	3
netural	0.75	1.00	0.86	3
sad	1.00	0.33	0.50	3
surprise	0.60	1.00	0.75	3
accuracy			0.71	21
macro avg	0.78	0.71	0.69	21
weighted avg	0.78	0.71	0.69	21

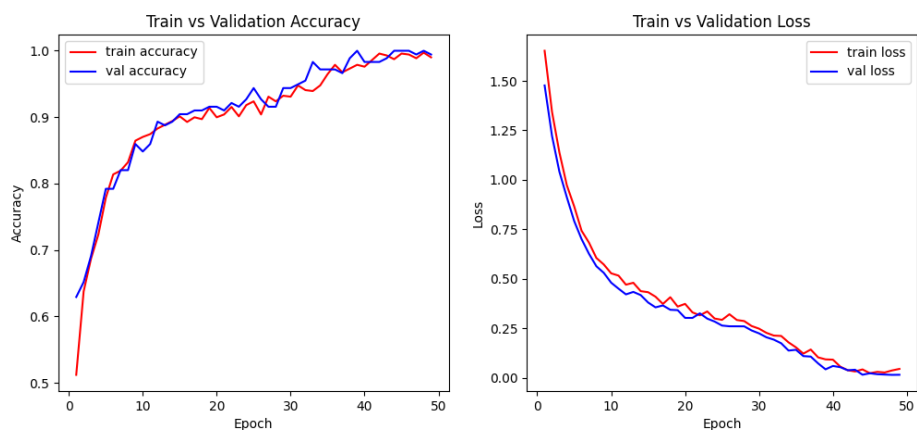
雖然模型在訓練時表現良好，但並沒有很好的泛化到測試集中，測試集的表現和訓練時有明顯差距，推測是因為資料集小，模型辦法有效泛化。

4. Grad-CAM 熱圖分析：



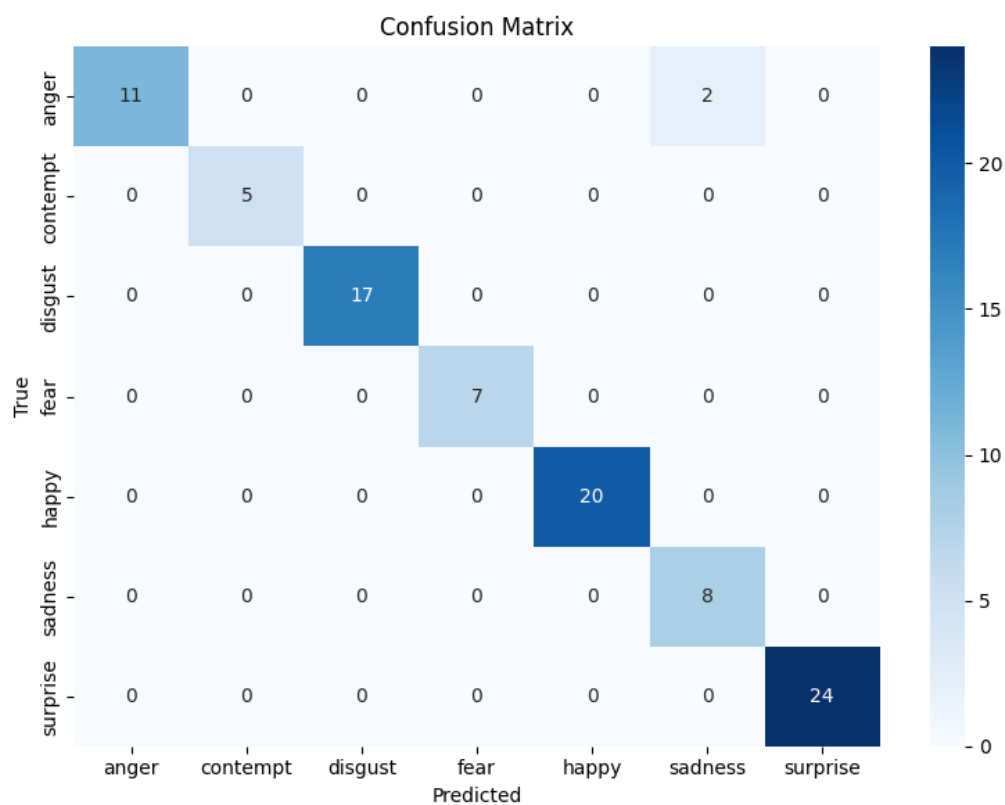
十二、 MobileNetV2 建立適用於 CK+訓練集的情緒辨識模型：

1. 訓練集和驗證集的 accuracy 和 loss 折線圖：



模型訓練時表現良好，沒有過擬合現象，accuracy 也穩定上升。

2. 測試集混淆矩陣：



3. 整體結果分析：

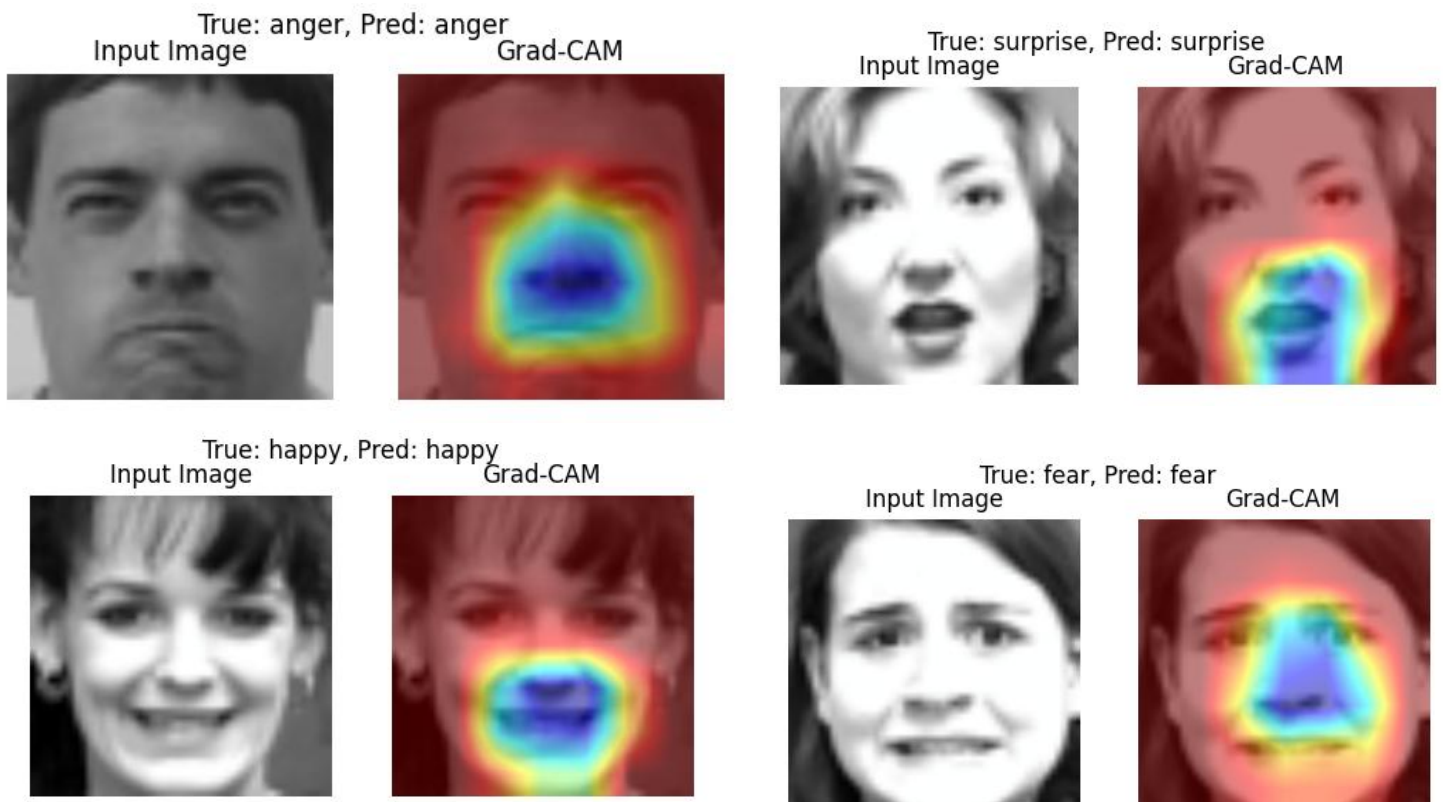
Test Accuracy: 0.9787

Classification Report:

	precision	recall	f1-score	support
anger	1.00	0.85	0.92	13
contempt	1.00	1.00	1.00	5
disgust	1.00	1.00	1.00	17
fear	1.00	1.00	1.00	7
happy	1.00	1.00	1.00	20
sadness	0.80	1.00	0.89	8
surprise	1.00	1.00	1.00	24
accuracy			0.98	94
macro avg	0.97	0.98	0.97	94
weighted avg	0.98	0.98	0.98	94

模型對大部分特徵都能有效分辨。

4. Grad-CAM 熱圖分析：



對於大部分的特徵都可以正確分類。

伍、總結：

VGG19 在訓練過程花的時間比其他兩個模型多上不少，且他在每個資料集的測試集總體表現並不算最好，但隨著每一次的遷移學習，VGG19 在每一個資料集的表現都是穩定上升，最後到 CK+ 資料集能達到 0.98 左右的準確率。ResNet50 雖然在 Ferplus2013 和 RAF-DB 的表現都領先其他兩個模型，但是可能因為過擬合的影響，在 JAFFE 資料集的表現十分不理想，最後到了 CK+ 的表現反而不如 VGG19。MobileNetV2 相較於其他兩個模型較輕量，訓練速度最快，整體表現略遜於 VGG19。

在三個模型中，VGG19 對每個資料集的表現應該最好，但我認為 ResNet50 如果正確的改善了過擬合的問題，能夠有效泛化到每一個資料集，它的表現應該能領先其他兩個模型，MobileNet50V2 相較於其他兩個模型較為輕量，訓練速度快。儘管輕量，但實際表現並沒有到非常差，在部分的資料集還是有 7 成將近 8 成的準確率。