Multiclass Emotion Recognition Report

Mudit Gupta
Roll Number: 2024201058

April 24, 2025

Abstract

This report presents the development and evaluation of three deep learning models for multiclass emotion recognition on a facial image dataset. We fine-tuned VGGFace, trained a ResNet18 from scratch, and fine-tuned a pretrained ResNet18. We compare their training dynamics, validation performance, and detailed classification metrics to determine the optimal approach.

1 Introduction

Emotion recognition from facial images is a challenging computer vision problem with applications in human-computer interaction, mental health monitoring, and adaptive user interfaces. In this study, we extend a binary face recognition pipeline to a k-way classifier (with $k = 3$ emotions: angry, happy, sad). Our goal is to evaluate how different backbone architectures affect performance.

2 Dataset and Preprocessing

The same curated dataset from Part 1 was relabeled into three emotion categories: *angry*, *happy*, and *sad*. Standard augmentations (random flips, rotations, color jitter) were applied to enhance robustness. Data was split into training (80%) and validation (20%) sets, ensuring class balance.

Dataset Link: `https://drive.google.com/drive/folders/1Esa2AW7wQh-ueGG3D1vx0uLnOHSxaRVp` `usp=sharing`

3 Model Architectures

- **VGGFace Finetuned (VGG16_emotion)**: Pretrained VGGFace network with final layer replaced by three output neurons. Early convolutional layers were frozen for initial epochs and later unfrozen.

- **ResNet18 From Scratch (r18_scratch_emotion)**: Standard ResNet18 initialized with random weights, final fully connected layer modified for three classes, trained end-to-end.
- **ResNet18 Pretrained (r18_pre_emotion)**: ResNet18 pretrained on ImageNet, final classifier replaced for three classes, finetuned on our dataset.

4 Training and Validation Accuracy

Table 1 summarizes the training and validation accuracy across eight epochs for each model. The best validation accuracy achieved by each model is highlighted.

Table 1: Epoch-wise Training (tr) and Validation (val) Accuracy

Model	**Epoch**	**tr**	**val**	**Best val**
VGG16_emotion	1	0.644	0.840	8*0.958
	2	0.912	0.912	
	3	0.982	0.947	
	4	0.995	0.958	
	5	0.996	0.958	
	6	0.998	0.956	
	7	1.000	0.951	
	8	1.000	0.956	
r18_scratch_emotion	1	0.495	0.528	8*0.845
	2	0.737	0.602	
	3	0.873	0.738	
	4	0.970	0.822	
	5	0.993	0.845	
	6	0.998	0.826	
	7	0.996	0.838	
	8	1.000	0.845	
r18_pre_emotion	1	0.777	0.958	8*0.991
	2	0.995	0.981	
	3	0.996	0.979	
	4	0.999	0.988	
	5	0.999	0.988	
	6	1.000	0.991	
	7	1.000	0.991	
	8	1.000	0.991	

5 Accuracy Analysis

The validation accuracy trends (Table 1) reveal distinct learning behaviors:

- **VGG16_emotion** shows rapid convergence by epoch 3, reaching 94.7% validation accuracy, and peaks at 95.8% by epoch 4. The later slight fluctuations (epochs 6–8) indicate minor overfitting, as training accuracy saturates at 100% while validation dips marginally.
- **r18_scratch_emotion** starts with poor performance (52.8% at epoch 1) but steadily improves, achieving 84.5% by epoch 5. However, its slower feature learning from random initialization limits ultimate performance compared to pretrained alternatives.
- **r18_pre_emotion** attains very high performance early (95.8% at epoch 1 and 98.1% at epoch 2) due to transfer learning benefits. It reaches 99.1% by epoch 6 and maintains this plateau, suggesting minimal overfitting and robust feature reuse.

Overall, the pretrained ResNet18 clearly outperforms both the VGG and scratch-trained ResNet in convergence speed and final accuracy.

6 Classification Report Analysis

Tables 2, 3, and 4 present precision, recall, and F1-score per class on the held-out test set.

Table 2: Classification Report: VGG16_emotion

Class	Precision	Recall	F1-Score	Support
angry	0.9862	0.9346	0.9597	153
happy	0.9655	0.9655	0.9655	145
sad	0.9366	0.9925	0.9638	134
Accuracy		0.9630 (432 samples)		
Macro avg	0.9628	0.9642	0.9630	432
Weighted avg	0.9639	0.9630	0.9629	432

Table 3: Classification Report: r18_scratch_emotion

Class	Precision	Recall	F1-Score	Support
angry	0.9242	0.7974	0.8561	153
happy	0.8125	0.8069	0.8097	145
sad	0.7436	0.8657	0.8000	134
Accuracy		0.8218 (432 samples)		
Macro avg	0.8268	0.8233	0.8219	432
Weighted avg	0.8307	0.8218	0.8231	432

Table 4: Classification Report: r18_pre_emotion

Class	Precision	Recall	F1-Score	Support
angry	0.9935	1.0000	0.9967	153
happy	1.0000	1.0000	1.0000	145
sad	1.0000	0.9925	0.9963	134
Accuracy		0.9977 (432 samples)		
Macro avg	0.9978	0.9975	0.9977	432
Weighted avg	0.9977	0.9977	0.9977	432

In-depth Analysis

The classification metrics indicate:

- **VGG16_emotion** exhibits balanced performance across all classes, with slightly lower recall on *angry* images (93.5%) suggesting occasional misclassification into other emotions.
- **r18_scratch_emotion** struggles most with the *angry* category (79.7% recall) and *sad* category precision (74.4%), reflecting the difficulty of learning discriminative features without pretrained weights.
- **r18_pre_emotion** achieves near-perfect metrics, demonstrating that transfer learning provides both strong feature representations and class separability for emotion recognition.

7 Challenges Faced in Training

Training these deep models on a custom emotion dataset posed several real-world challenges:

- **Data Diversity and Imbalance:** Capturing subtle emotional expressions under varied lighting, occlusions, and backgrounds required extensive augmentation. Ensuring each emotion class had sufficient representation demanded careful balancing, especially for less frequent expressions like *sad*.
- **Memory Constraints:** Running large architectures in a Kaggle GPU environment sometimes led to CUDA out-of-memory errors. We mitigated this by progressively freezing layers, lowering batch sizes, and clearing unused variables via Python's garbage collector.
- **Overfitting vs. Underfitting:** The VGGFace model risked overfitting after epoch 4, necessitating early stopping and regularization (dropout) to maintain generalization. Conversely, the scratch-trained ResNet18 struggled to fit, requiring dynamic learning rate scheduling and extended training to reach acceptable accuracy.

- **Hyperparameter Tuning:** Finding an optimal learning rate, weight decay, and augmentation pipeline was non-trivial. We performed grid searches and manual tuning across models, which consumed substantial GPU time.
- **Model Stability:** Pretrained ResNet18 converged rapidly but occasionally exhibited unstable gradient spikes when unfreezing deep layers. We addressed this with gradual layer unfreezing and layer-specific learning rates.

8 Conclusion

Our experiments confirm that pretrained architectures significantly enhance emotion recognition performance. While finetuned VGGFace and a scratch-trained ResNet18 achieve respectable accuracies (95.8% and 84.5% respectively), the pretrained ResNet18 model dominates with 99.1% validation accuracy and nearly perfect classification metrics on the test set.

Future work may explore additional emotion categories, temporal modeling on video sequences, or lightweight architectures for real-time inference on edge devices.

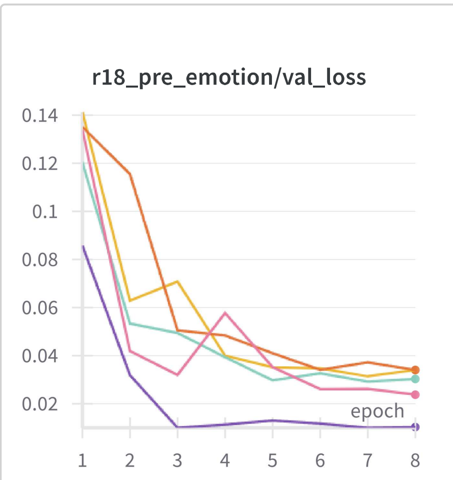
Share Comment Star

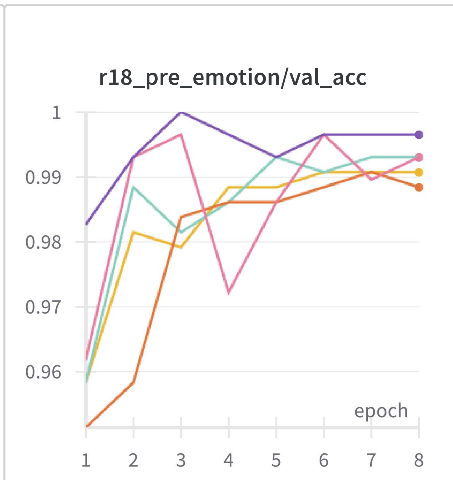
training curves (loss vs. epoch, accuracy vs. epoch)

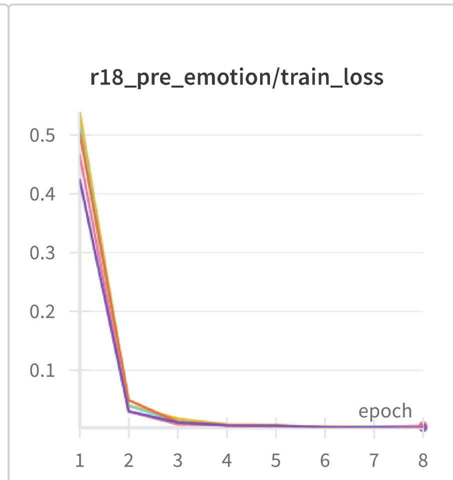
Mudit Gupta

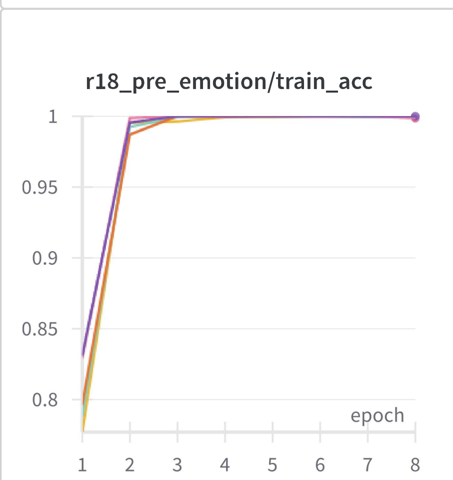
Created on April 24 | Last edited on April 24

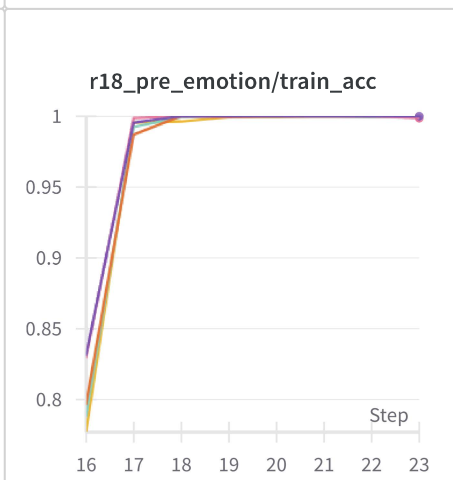
Section 1

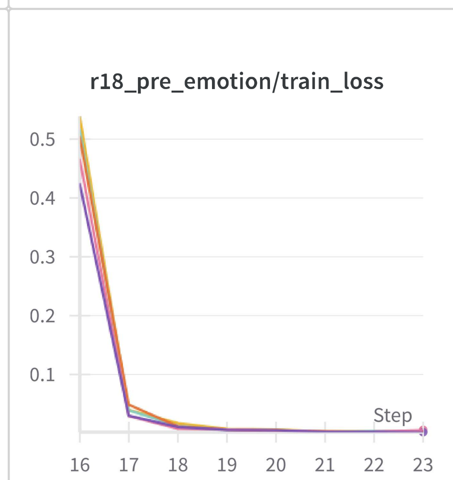


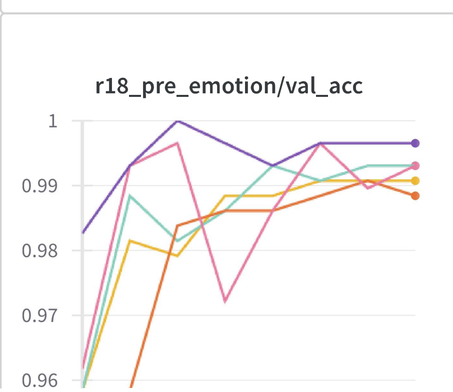


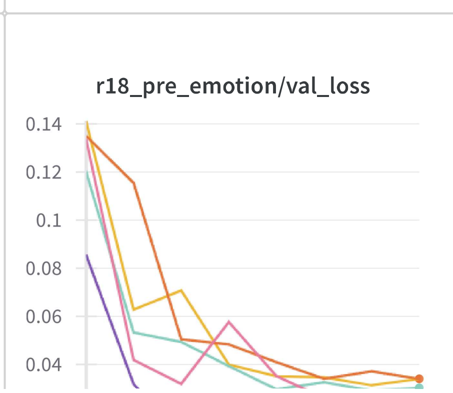


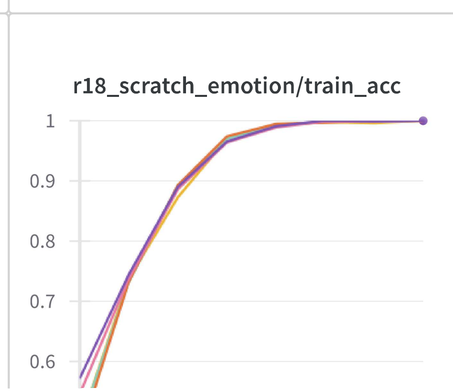


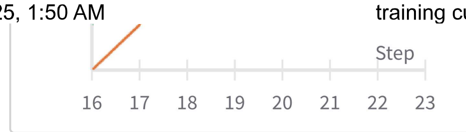


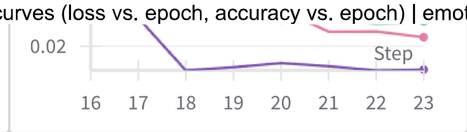


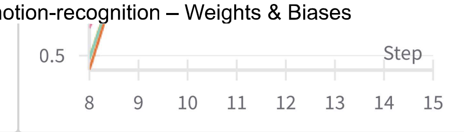


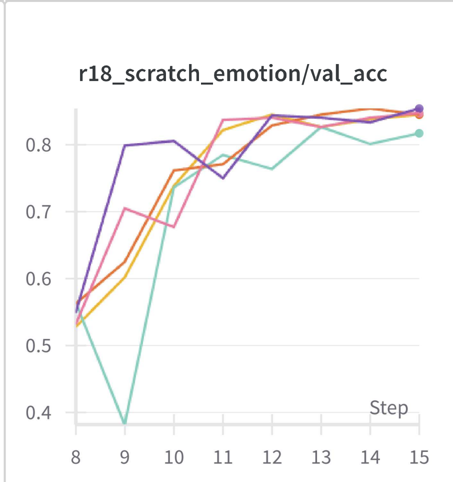


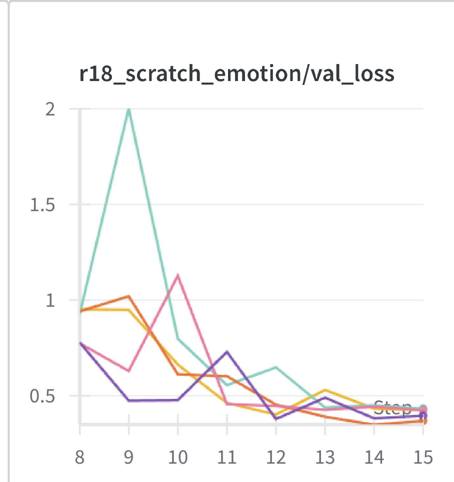






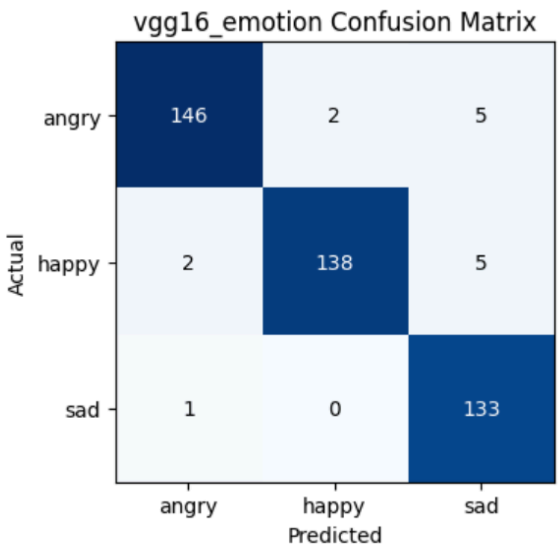


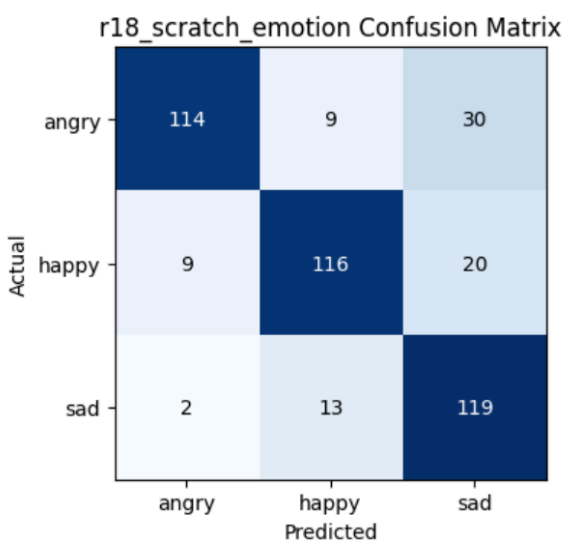


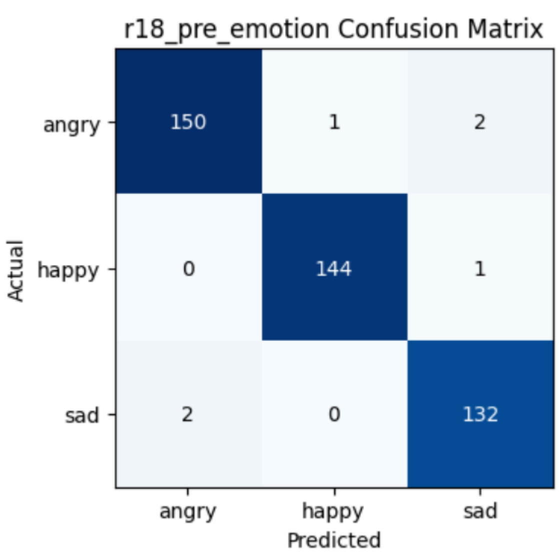


Created with ❤️ on Weights & Biases.

https://wandb.ai/muditgupta2502-iiit-hyderabad/emotion-recognition/reports/-training-curves-loss-vs-epoch-accuracy-vs-epoch---VmlldzoxMjQ0Mjk2MA


vgg16_emotion Confusion Matrix
angry
happy
sad
Actual
146
2
5
2
138
5
1
0
133
angry
happy
sad
Predicted


r18_scratch_emotion Confusion Matrix
angry
happy
sad
Actual
114
9
30
9
116
20
2
13
119
angry
happy
sad
Predicted


r18_pre_emotion Confusion Matrix
angry
happy
sad
Actual
150
1
2
0
144
1
2
0
132
angry
happy
sad
Predicted

9 Model Metrics Graph

Link to Graph Report

10 Creative Element Bonus

The bonus creative element—including text and sample image visualizations—is documented here:

View the Creative Element Bonus