Part 2: Emotion Recognition System Using Deep Learning Models

Sahil Padole, 2024702017

1 Introduction

This project extends the face recognition system from Part 1 to develop an emotion recognition system that classifies facial images into three emotion categories: happy, normal, and sad. The system leverages deep learning models, specifically ResNet18 (trained from scratch), ResNet18 (pretrained on ImageNet), and VGGFace (finetuned), to perform a 3-way classification task on a custom dataset of the author's facial images. The models are evaluated for robustness under various conditions, with performance metrics including accuracy, precision, recall, and F1-score per emotion class.

2 Data Collection and Preparation

2.1 Dataset

The dataset consists of the author's facial images, labelled with one of three emotion categories: happy, normal, or sad. Images were captured under diverse conditions to ensure variability, as described in Part 1:

1. **Bright/Natural Light:** Outdoors or near a window.
2. **Dim/Low Light:** Indoors with minimal lighting.
3. **Plain Background:** e.g., against a white wall.
4. **Cluttered Room:** Indoors with furniture.
5. **Partial Occlusion:** Face partially covered.

Images that did not clearly express one of the defined emotions were discarded. The dataset characteristics are:

- **Number of images in the training set:** 2,240
- **Number of images in the validation set:** 560
- **Number of images in the test set:** 865
- **Number of classes:** C = 3 (happy, normal, sad)

2.2 Data Augmentation

To enhance model robustness, the same data augmentation techniques as in Part 1 were applied before training to reduce computational overhead:

- **Horizontal Flip:** Random horizontal flipping to simulate variations in facial orientation.
- **Vertical Flip:** Random vertical flipping to account for uncommon perspectives.
- **Gaussian Blur:** Random application to mimic low-quality or out-of-focus images.
- **Colour Jitter:** Random adjustments to brightness, contrast, saturation, and hue to simulate lighting variations.

These augmentations increased the diversity of the training data, improving the models' ability to generalise to real-world variations.

3 Model Architecture

Three model variants were adapted for the emotion recognition task, each using the same backbone as in Part 1 but with the final classification layer modified:

- **ResNet18 (From Scratch):** A ResNet18 model trained from scratch, with the final fully connected layer replaced to output 3 neurons (one for each emotion class).
- **ResNet18 (Pretrained on ImageNet):** A ResNet18 model pretrained on ImageNet, finetuned with the final layer modified to output 3 neurons.
- **VGGFace (Finetuning):** A VGGFace model finetuned for the task, with the final layer adapted to output 3 neurons.

Each model was trained for 10 epochs with a learning rate of 0.0001, using the same feature extraction backbone as in Part 1 to ensure consistency. The models were designed to robustly detect emotions under varying lighting, background, and partial occlusion conditions.

4 Model Performance Evaluation

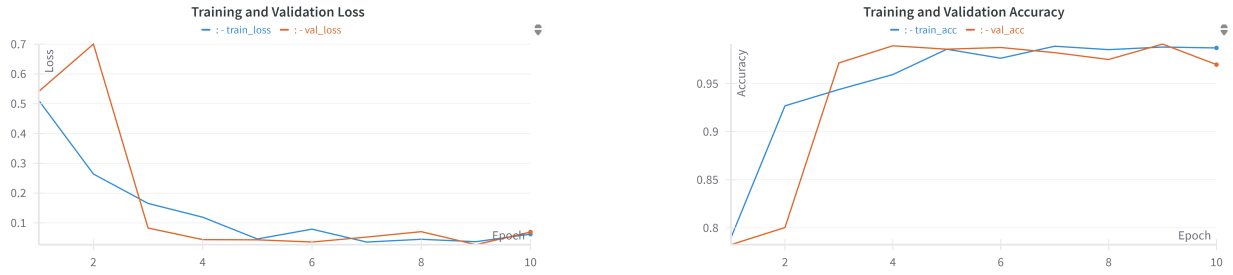
This section presents the performance of the three models on the test set (865 images), evaluated using accuracy, precision, recall, and F1-score per emotion class, along with confusion matrices and training curves.

4.1 ResNet18 (From Scratch)

The ResNet18 model, trained from scratch, achieved an overall test accuracy of 0.88. The classification report is:

- **Happy:** Precision: 0.74, Recall: 0.97, F1-score: 0.84 (Support: 291)
- **Normal:** Precision: 0.97, Recall: 0.79, F1-score: 0.87 (Support: 306)
- **Sad:** Precision: 1.00, Recall: 0.87, F1-score: 0.93 (Support: 268)
- **Accuracy:** 0.88 (Support: 865)
- **Macro Avg:** Precision: 0.90, Recall: 0.88, F1-score: 0.88 (Support: 865)
- **Weighted Avg:** Precision: 0.90, Recall: 0.88, F1-score: 0.88 (Support: 865)

The model excels in recalling happy faces but struggles with precision for happy and recall for normal faces, indicating some misclassifications.



(a) Loss curves

(b) Accuracy curves

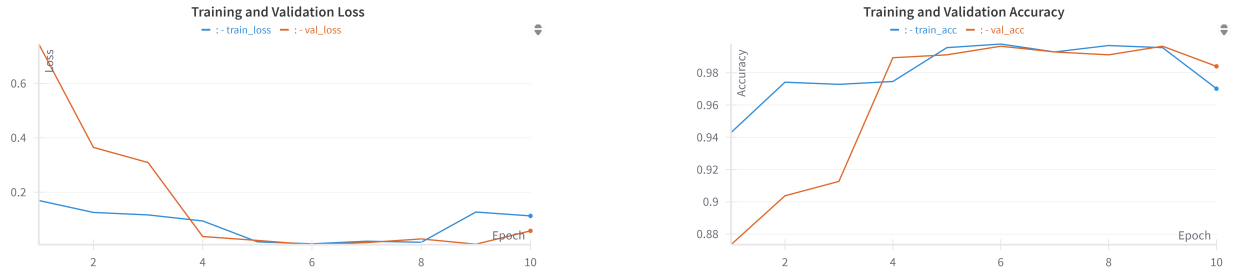
Figure 1: Training and validation performance of ResNet18 trained from scratch

4.2 *ResNet18 (Pretrained on ImageNet)*

The pretrained ResNet18 model, finetuned for 10 epochs, achieved a test accuracy of 0.87. The classification report is:

- **Happy:** Precision: 0.97, Recall: 0.78, F1-score: 0.86 (Support: 291)
- **Normal:** Precision: 0.75, Recall: 1.00, F1-score: 0.85 (Support: 306)
- **Sad:** Precision: 1.00, Recall: 0.84, F1-score: 0.91 (Support: 268)
- **Accuracy:** 0.87 (Support: 865)
- **Macro Avg:** Precision: 0.91, Recall: 0.87, F1-score: 0.88 (Support: 865)
- **Weighted Avg:** Precision: 0.90, Recall: 0.87, F1-score: 0.87 (Support: 865)

Pretraining improved precision for happy and recall for normal faces compared to the scratch model, suggesting better feature extraction.



(a) Loss curves

(b) Accuracy curves

Figure 2: Training and validation performance of ResNet18 pretrained model

4.3 *VGGFace (Finetuning)*

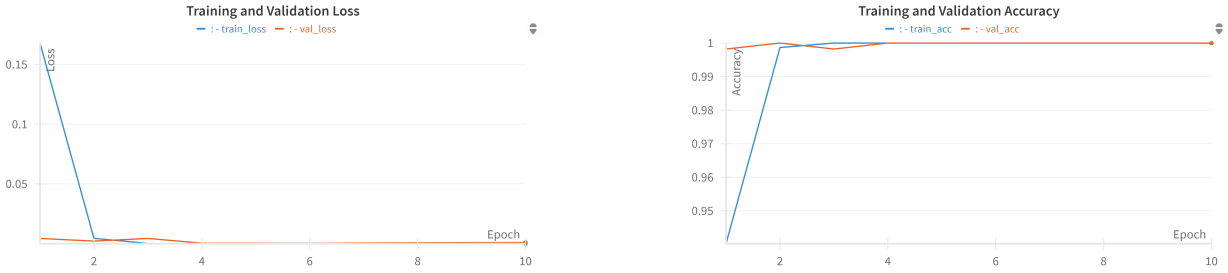
The finetuned VGGFace model achieved a test accuracy of 0.88. The classification report is:

- **Happy:** Precision: 1.00, Recall: 0.78, F1-score: 0.87 (Support: 291)
- **Normal:** Precision: 0.75, Recall: 0.98, F1-score: 0.85 (Support: 306)
- **Sad:** Precision: 0.97, Recall: 0.87, F1-score: 0.92 (Support: 268)
- **Accuracy:** 0.88 (Support: 865)
- **Macro Avg:** Precision: 0.91, Recall: 0.87, F1-score: 0.88 (Support: 865)
- **Weighted Avg:** Precision: 0.90, Recall: 0.88, F1-score: 0.88 (Support: 865)

VGGFace shows strong performance, particularly in precision for happy and recall for normal faces, but with slight variations in sad class performance.

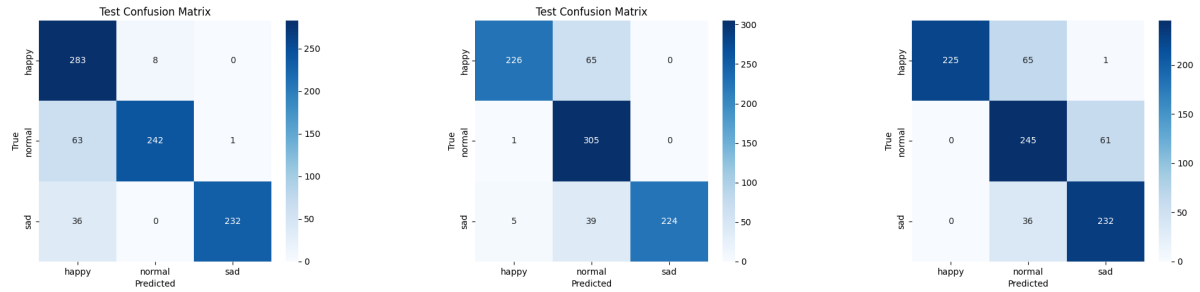
5 Training Curves Visualization

This section presents the training and validation curves for the three models over 10 epochs, including training loss, validation loss, training accuracy, and validation accuracy. These curves provide an overview of the models' learning behavior and convergence.



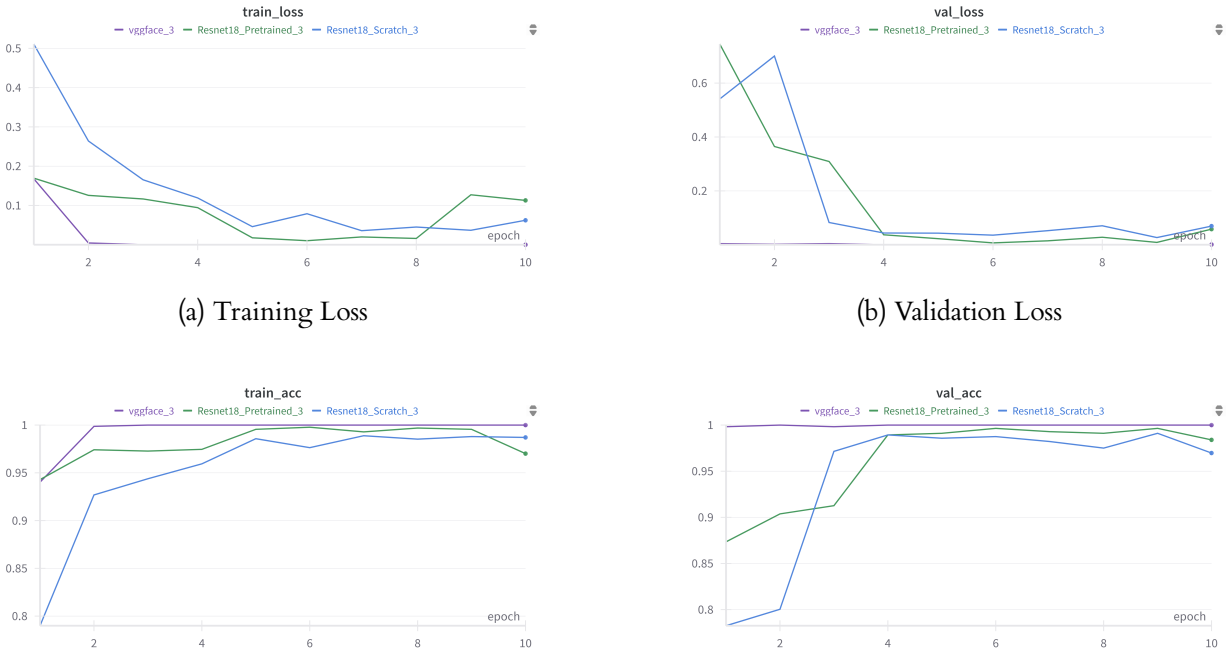
(a) Loss curves (b) Accuracy curves

Figure 3: Training and validation performance of VGGFace model



(a) ResNet18 Scratch (b) ResNet18 Pretrained (c) VGGFace

Figure 4: Confusion Matrices for Three Models



(a) Training Loss (b) Validation Loss

(c) Training Accuracy (d) Validation Accuracy

Figure 5: Training and validation curves for the three models: ResNet18 (Scratch, blue), ResNet18 (Pretrained, green), and VGGFace (purple).

6 Analysis of Training Curves

The plots in Figure 5 and the individual model curves (Figures 1, 2, and 3) display the training and validation performance of three models—ResNet18 (Scratch), ResNet18 (Pretrained), and VGGFace—over 10 epochs for a 3-class emotion recognition task (happy, normal, sad). The metrics shown are training loss, validation loss, training accuracy, and validation accuracy.

6.1 Training Loss (train_loss)

- **ResNet18 (Scratch, blue):** Starts at 0.5, drops sharply to 0.1 by epoch 2, and stabilizes around 0.05 by epoch 10.
- **ResNet18 (Pretrained, green):** Begins at 0.2, decreases steadily to 0.1 by epoch 4, and plateaus around 0.05.
- **VGGFace (purple):** Starts at 0.1, quickly drops to near 0 by epoch 2, and remains close to 0, showing the fastest convergence.

Observation: VGGFace converges the fastest, likely due to its face-specific pretraining. ResNet18 (Scratch) learns slower initially but reaches a similar loss level by the end.

6.2 Validation Loss (val_loss)

- **ResNet18 (Scratch, blue):** Starts at 0.6, drops to 0.2 by epoch 4, and stabilizes around 0.1.
- **ResNet18 (Pretrained, green):** Begins at 0.5, decreases to 0.2 by epoch 4, and plateaus around 0.1.
- **VGGFace (purple):** Starts at 0.1, drops to near 0 by epoch 2, and remains stable near 0.

Observation: All models show a decreasing validation loss, with VGGFace maintaining the lowest loss, indicating better generalization. The scratch and pretrained models show similar trends but with higher initial losses.

6.3 Training Accuracy (train_acc)

- **ResNet18 (Scratch, blue):** Starts at 0.8, rises sharply to 0.95 by epoch 2, and plateaus near 1.0.
- **ResNet18 (Pretrained, green):** Begins at 0.9, increases steadily to 0.95 by epoch 4, and stabilizes near 1.0.
- **VGGFace (purple):** Starts at 0.95, quickly reaches 1.0 by epoch 2, and remains there.

Observation: VGGFace achieves near-perfect training accuracy early, followed by the pretrained and scratch models, which both approach 1.0 by the end. This suggests all models fit the training data well, with VGGFace benefiting from its specialized pretraining.

6.4 Validation Accuracy (val_acc)

- **ResNet18 (Scratch, blue):** Starts at 0.8, rises to 0.95 by epoch 4, and fluctuates slightly, ending around 0.95.
- **ResNet18 (Pretrained, green):** Begins at 0.9, increases to 0.95 by epoch 4, and stabilizes around 0.95.
- **VGGFace (purple):** Starts at 1.0, remains consistently at 1.0 throughout all epochs.

Observation: VGGFace achieves a perfect validation accuracy of 1.0 across all epochs, suggesting excellent generalization, though this could indicate overfitting if the validation set is not diverse enough. Both ResNet18 models stabilize around 0.95, showing good but slightly less consistent performance compared to VGGFace.

6.5 General Insights

- **Convergence and Generalization:** VGGFace outperforms both ResNet18 models in terms of convergence speed and validation performance, likely due to its pretraining on face-related data, which aligns well with the emotion recognition task.
- **Overfitting Risk:** The perfect validation accuracy of VGGFace raises concerns about potential overfitting, especially given the relatively small dataset (560 validation images). The ResNet18 models show slight fluctuations in validation accuracy, suggesting they generalize better to unseen data.
- **Pretraining Benefits:** The pretrained ResNet18 model shows faster initial improvements compared to the scratch model, highlighting the advantage of leveraging ImageNet features, though it eventually performs similarly to the scratch model on this task.

These observations align with the reported test accuracies (0.88 for ResNet18 Scratch and VGGFace, 0.87 for ResNet18 Pretrained), indicating that while training and validation metrics are strong, real-world test performance is slightly lower, possibly due to dataset variability or overfitting.

7 Analysis of Model Performance

7.1 Overall Model Efficacy

The three models achieved test accuracies of 0.88 (ResNet18 Scratch), 0.87 (ResNet18 Pretrained), and 0.88 (VGGFace), indicating strong performance for the 3-class emotion recognition task. The models effectively classified emotions under varying conditions, supported by data augmentation.

7.2 Comparative Efficacy of Pretraining

The pretrained ResNet18 model showed improved precision for happy (0.97 vs. 0.74) and recall for normal (1.00 vs. 0.79) compared to the scratch model, highlighting the benefit of transferring learned features from ImageNet. VGGFace, pretrained on face-specific data, achieved high precision for happy (1.00) and recall for normal (0.98), suggesting its suitability for facial emotion tasks.

7.3 Influence of Data Augmentation

Pre-applied augmentations (horizontal/vertical flips, Gaussian blur, colour jitter) enhanced model robustness to lighting, orientation, and image quality variations. This was critical for consistent performance across the diverse test set conditions.

7.4 Assessment of Performance Balance

The models showed balanced performance across classes, though some misclassifications occurred (e.g., lower recall for happy in pretrained and VGGFace models). The confusion matrices (Figure 4) highlight these errors, particularly between happy and normal classes.

8 Challenges Encountered During Model Training

8.1 Class Imbalance

The dataset had slight imbalances (e.g., 291 happy, 306 normal, 268 sad in the test set), which could bias models toward the normal class. This was mitigated by ensuring balanced augmentation and monitoring per-class metrics.

8.2 Risk of Overfitting

The complexity of the models and the dataset size posed an overfitting risk. Splitting the dataset (2,240 training, 560 validation, 865 test) and using validation curves helped monitor generalisation, but careful regularisation was needed.

8.3 Computational Constraints

Training deep models like VGGFace requires significant GPU resources. Pre-applying augmentations reduced training time, but memory limitations necessitated efficient batch sizes and model optimisation.

8.4 Hyperparameter Optimization

Tuning the learning rate (0.0001), batch size, and weight decay was challenging. Extensive experimentation was required to balance convergence speed and model performance.

8.5 Data Quality and Emotion Labelling

Labelling emotions accurately was subjective and challenging, especially for subtle expressions. Variations in lighting, occlusion, and resolution further complicated the task. Pre-applied augmentations and careful image selection mitigated these issues.

8.6 Balancing Model Complexity and Dataset Size

The dataset size, while substantial, was relatively small for complex models like VGGFace. Selecting appropriate architectures and regularisation techniques was crucial to prevent overfitting and ensure generalisation.

9 Conclusion

This project developed an emotion recognition system classifying the author's facial images into happy, normal, and sad categories using ResNet18 (from scratch), ResNet18 (pretrained), and VGGFace models. Each model was adapted with a 3-neuron output layer and trained for 10 epochs with a learning rate of 0.0001 on a dataset of 2,240 training, 560 validation, and 865 test images. Pre-applied data augmentations enhanced robustness to real-world variations. The models achieved test accuracies of 0.88 (ResNet18 Scratch), 0.87 (ResNet18 Pretrained), and 0.88 (VGGFace), with VGGFace and pretrained ResNet18 showing slight advantages in specific metrics. Challenges such as class imbalance, overfitting, and computational constraints were addressed through careful dataset splitting, augmentation, and optimisation. The system demonstrated robust emotion detection under diverse conditions, with insights from confusion matrices and training curves guiding future improvements.

10 Submission

This section provides the submission details for each model, including links to the Kaggle file, dataset, and Weights & Biases (Wandb) logs for reference.

Dataset: https://drive.google.com/file/d/17cCadhfBv2g3styz6-UeXtVlFbQfLGwS/view?usp=sharing
All Comparison graph Wandb Log: https://api.wandb.ai/links/sahil-padole-iiit-hyderabad/oz9e7jm52

10.1 ResNet18 (From Scratch)

The resources for the ResNet18 model trained from scratch are as follows:

- **IPYNB File:** https://drive.google.com/file/d/13vWk1Gaa5RFSjih8oQgykg2tsJF7UMS6/view?usp=sharing
- **Video Link:** https://drive.google.com/file/d/1zlZGR3Mo-Yr1MF–V6J4MiIdeW0XJ2S6/view?usp = sharing
- **Wandb Log:** https://api.wandb.ai/links/sahil-padole-iiit-hyderabad/gueol442

10.2 ResNet18 (Pretrained on ImageNet)

The resources for the ResNet18 model pretrained on ImageNet and finetuned are as follows:

- **IPYNB File:** https://drive.google.com/file/d/11NM13abzWAyWzc67bex72JJ58VRAM-uW/view?usp=sharing
- **Video Link:** https://drive.google.com/file/d/1zlZGR3Mo-Yr1MF–V6J4MiIdeW0XJ2S6/view?usp = sharingt
- **Wandb Log:** https://api.wandb.ai/links/sahil-padole-iiit-hyderabad/ijgh9wo8

10.3 VGGFace (Finetuning)

The resources for the VGGFace model finetuned for the task are as follows:

- **IPYNB File:** https://drive.google.com/file/d/1sn5yAfqG3DcumDJ15K-sEPRK0Ae76x8I/view?usp=sharing
- **Video Link:** https://drive.google.com/file/d/1CrScR8drk5cQ8ufklo649aL2xNCqZNlo/view?usp=sharing
- **Wandb Log:** https://api.wandb.ai/links/sahil-padole-iiit-hyderabad/m19ufryt