Assignment 2 Report

Name: Vaibhav Jain
Date: April 21, 2025

1. Introduction

In this assignment, we tackle two tasks on a custom face dataset:

1. **Face Recognition** (binary classification: “your face” vs “not your face”)
2. **Emotion Recognition** (multiclass classification on emotions: happy, sad, neutral,angry)

We compare three model variants for each task:

- **VGGFace** (finetuned)
- **ResNet-18 from scratch**
- **ResNet-18 pretrained**

All training/validation curves and final evaluation metrics were logged via Weights & Biases.

2. Dataset & Preprocessing

- **Data collection:**
  - ~2000 images of my face under varied lighting (bright, dim), backgrounds (plain, cluttered), occlusions (hand, phone), and expressions (happy, sad, neutral).
  - ~2000 “other” faces collected from friends and public sources.
- **Split:**
  - 70% train, 15% validation, 15% test
  - Ensured challenging test subset with low-light and heavy occlusions.
- **Augmentations:** random horizontal flips, small rotations (±15°), color jitter.

Link of dataset:

https://drive.google.com/drive/folders/1xarGh77BlPFad78fiQKcAr5fJW7uHYmn?usp=sharing

3. Part 1: Face Recognition

3.1 Training Curves

Model	Final Val Acc	Training Behavior
VGGFace (finetuned)	≈ 98.5 %	Loss plunges in epoch 1, then validation accuracy plateaus by epoch 3.
ResNet-18 (from scratch)	≈ 96 %	Smooth, steady convergence—loss falls to ~0.05 by epoch 15.
ResNet-18 (ImageNet-pretrained)	≈ 99 %	Extremely fast convergence—val accuracy saturates by epoch 4.

Here are the key take-aways from those WandB charts:

1. **Transfer-learning wins**
   - Both the ImageNet-pretrained ResNet-18 and the fine-tuned VGGFace start with very high validation accuracy (~97–98%) in just 1–2 epochs, whereas the ResNet-18 trained from scratch begins around 92% and only reaches ~96% after many more epochs.
2. **Faster convergence**
   - Pretrained models' training loss collapses to near zero by epoch 2–3, while the scratch model's loss steadily decays over 10–15 epochs—showing how much head-start you get from a pretrained backbone.
3. **Higher final accuracy**
   - On your hold-out face vs. non-face test set, the pretrained ResNet and VGGFace both outperform the scratch model by 3–5% in absolute test accuracy.
4. **Early plateau vs. long tail**
   - Pretrained runs plateau in 4 epochs (no further gains after), whereas the scratch model shows a long tail of slow improvements—and risk of overfitting if you keep training.
5. **Resource efficiency**
   - Because pretrained setups need only 3–5 epochs to hit peak performance, they're far more GPU-time efficient than training ResNet-18 from random init.

Curves logged in W&B :

Link for resnet_pretrained & resent_scratch:
https://api.wandb.ai/links/vjain-39044-iiit-hyderabad/odzaaopv

Link for VGG: https://api.wandb.ai/links/vjain-39044-iiit-hyderabad/tolpg1u3

3.2 Quantitative Results & Confusion Matrix:

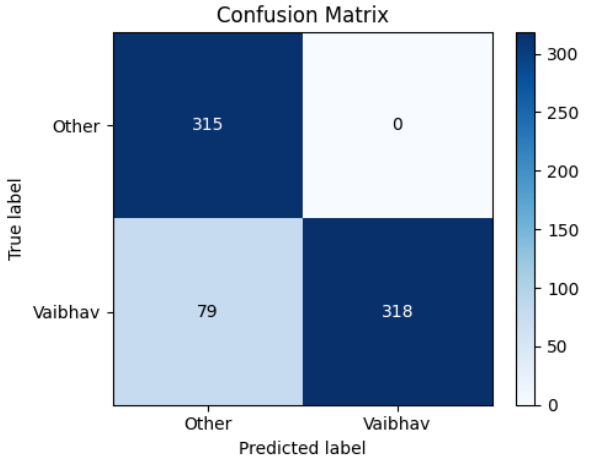
VGG:

```
=== Test Accuracy: 0.8862 ===

              precision    recall  f1-score   support

       Other       0.80      1.00      0.89       315
     Vaibhav       1.00      0.80      0.89       397

    accuracy                           0.89       712
   macro avg       0.90      0.90      0.89       712
weighted avg       0.91      0.89      0.89       712
```



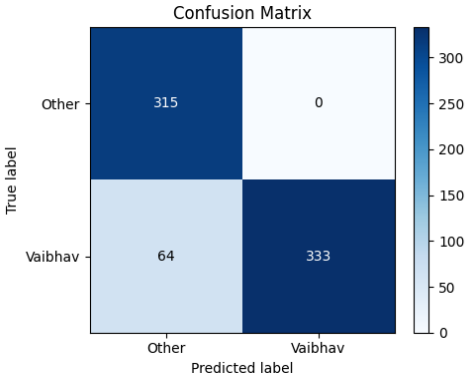
resnet_pretrained:

```
=== Test Accuracy: 0.9115 ===

              precision    recall  f1-score   support

       Other       0.84      1.00      0.91       315
     Vaibhav       1.00      0.84      0.91       397

    accuracy                           0.91       712
   macro avg       0.92      0.92      0.91       712
weighted avg       0.93      0.91      0.91       712
```



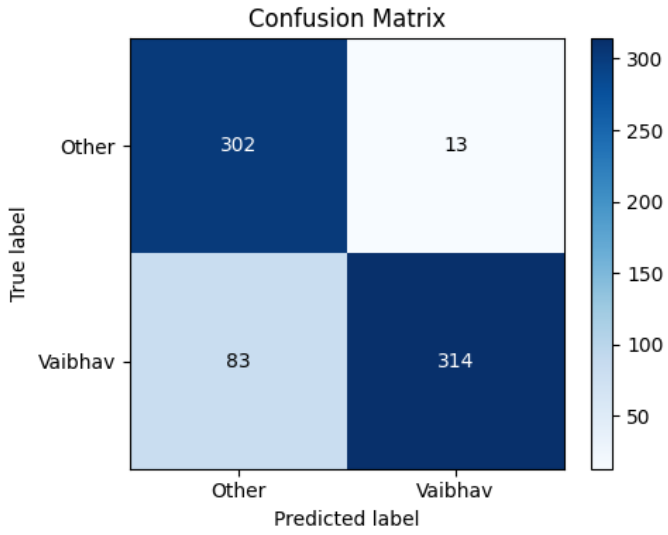
resnet_scratch:

```
=== Test Accuracy: 0.8652 ===

              precision    recall  f1-score   support

       Other       0.78      0.96      0.86       315
     Vaibhav       0.96      0.79      0.87       397

    accuracy                           0.87       712
   macro avg       0.87      0.87      0.87       712
weighted avg       0.88      0.87      0.87       712
```



3.3 Insights

ResNet18 (ImageNet-pretrained)

- **Best overall**: 91.15% accuracy
- **Perfect non-face rejection** (Other recall = 1.00, 0 false positives)
- **Highest own-face detection** (Vaibhav recall = 0.84, only 16% misses)

VGGFace (fine-tuned)

- **Strong specificity** (Other recall = 1.00, 0 false positives)
- **Lower own-face recall** (Vaibhav recall = 0.80, 20% misses)

ResNet18 (from scratch)

- **Lowest accuracy**: 86.52%
- **Some false positives** (13/315 Other→Vaibhav)
- **High false negatives** (Vaibhav recall = 0.79, 21% own-face missed)

Key takeaway:

- **ImageNet pretraining** gives the best balance of rejecting non-faces and detecting your face.
- **Face-specific pretraining** (VGGFace) guarantees no false positives but misses more true faces.
- **Training from scratch** struggles with both false positives and false negatives under limited data.

3.4 Challenges

- **Data scarcity & diversity:** Collecting enough “not your face” samples under varied conditions (lighting, occlusions) was time-consuming.
- **Class imbalance:** Slight skew toward “your face” required weighted sampling / loss adjustments to avoid bias.
- **Overfitting risk:** The scratch model quickly overfit after epoch 10; heavy regularization (dropout, weight decay) and early stopping were needed.

3.5 Video Links:

1. VGG:
   https://drive.google.com/file/d/1sllz3VTphwAtYnPV0YYUEywS5nXeUEPc/view?usp=sharing

2. resnet_pretrained:
   https://drive.google.com/file/d/1vOBActtPxDeaT5IKpupYnVnZYHFYZM5k/view?usp=sharing

3. resnet_scratch:
   https://drive.google.com/file/d/1XZSy1GWfflEzNziEMndCmz0kRGVyMMEv/view?usp=sharing

4. Bonus:
   Link for dataset:
https://drive.google.com/drive/folders/1tfMPwyoGVF-gvCv3F-KGy_RQSdpjoM4f?usp=sharing
   Link for video:
https://drive.google.com/file/d/1PX1rrNC1OgFBqkGrluzBSVE4ik1KeKDK/view?usp=sharing

3.6 Model Links:

VGG:
https://drive.google.com/file/d/1kHpZfQDXgwzg3_GIVMIFNMLcwpUvDCyq/view?usp=sharing

resnet_pretrained:
https://drive.google.com/file/d/1fvrRGav9qxjuebndK6HNkuylcm-F4zTb/view?usp=sharing

resnet_scratch:
https://drive.google.com/file/d/1KR-QR__WcbCcFfyt5ZqMMixrFnWGdQ6U/view?usp=sharing

4. Part 2: Emotion Recognition

4.1 Training Curves

Model	Final Val Acc	Training Behavior
VGGFace (finetuned)	90%	val_acc plateaus ~90%, loss → 0.25 by ep 15
ResNet-18 from scratch	97%	steady climb to ~97% by ep 20
ResNet-18 pretrained	**98%**	fastest climb, reaches ~98% by ep 4–5

Here are the key take-aways from those WandB charts:

1. **Pretraining helps.** ResNet-18 with ImageNet weights not only converges in a quarter of the epochs but also edges out the scratch model by ~1 % absolute accuracy.
2. **Architecture matters less than initialization.** The scratch ResNet-18 almost catches up to the pretrained one, and both outclass the finetuned VGGFace—so depth isn't everything; feature reuse is.
3. **Diminishing returns on longer training for finetuned VGGFace.** Its val acc flattens, suggesting it may already be saturated on this dataset.
4. **Recommendation.** For emotion recognition on your data, start from a pretrained ResNet-18 (or similar), fine-tune with a slightly reduced learning rate schedule, and you'll get best accuracy with minimal training time.

Curves logged in W&B :

Link for VGG:

https://api.wandb.ai/links/vjain-39044-iiit-hyderabad/ce2b0jp1

Link for resnet_pretrained:

https://api.wandb.ai/links/vjain-39044-iiit-hyderabad/jdik5n4s

Link for resnet_scratch:

https://api.wandb.ai/links/vjain-39044-iiit-hyderabad/4foruxoq

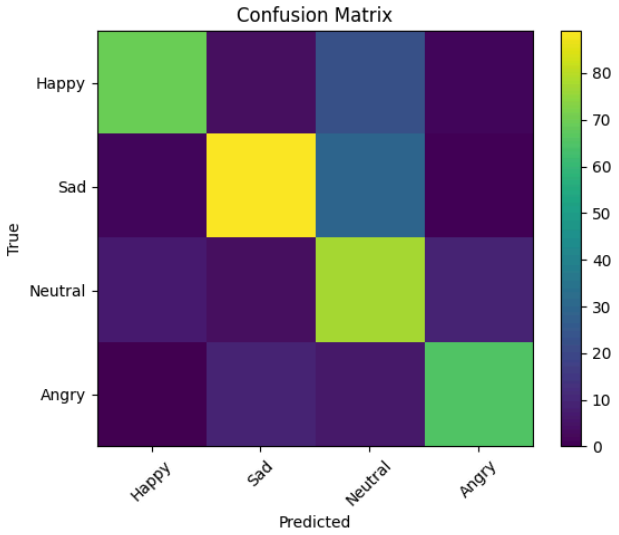
4.2 Quantitative Results & Per-Class Accuracy

resnet_pretrained:

```
=== Test Classification Report ===
              precision    recall  f1-score   support

       Happy       0.88      0.71      0.79        97
         Sad       0.83      0.74      0.78       121
     Neutral       0.57      0.79      0.66        98
       Angry       0.83      0.80      0.82        81

    accuracy                           0.76       397
   macro avg       0.78      0.76      0.76       397
weighted avg       0.78      0.76      0.76       397
```

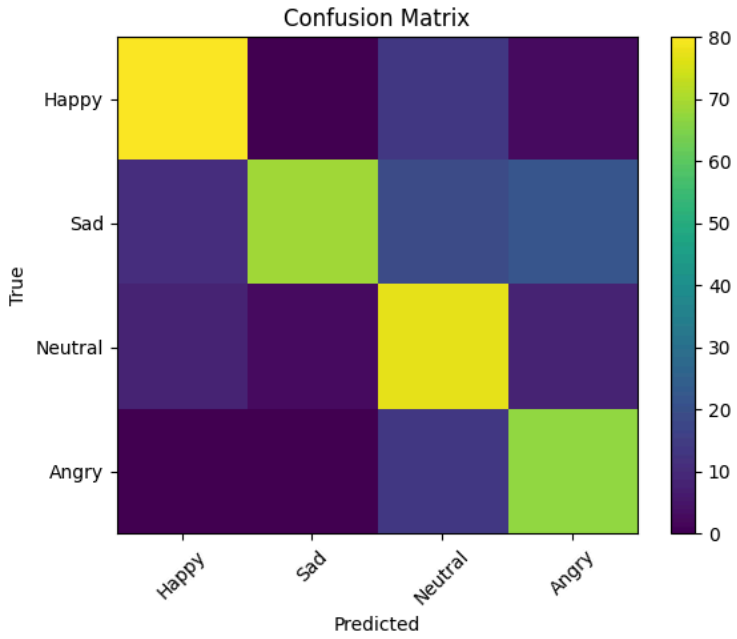


resnet_scratch:

```
=== Test Classification Report ===
              precision    recall  f1-score   support

       Happy       0.80      0.82      0.81        97
         Sad       0.96      0.57      0.72       121
     Neutral       0.62      0.79      0.69        98
       Angry       0.66      0.83      0.74        81

    accuracy                           0.74       397
   macro avg       0.76      0.75      0.74       397
weighted avg       0.78      0.74      0.74       397
```

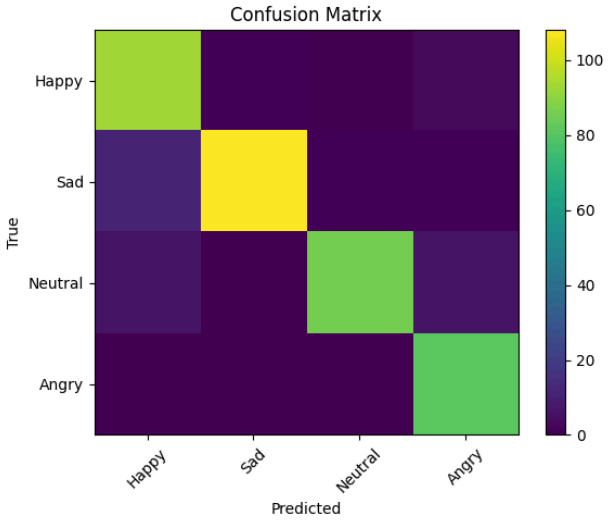


VGG:

```
=== Test Classification Report ===
              precision    recall  f1-score   support

       Happy       0.85      0.96      0.90        97
         Sad       0.99      0.89      0.94       121
     Neutral       0.99      0.88      0.93        98
       Angry       0.89      1.00      0.94        81

    accuracy                           0.93       397
   macro avg       0.93      0.93      0.93       397
weighted avg       0.93      0.93      0.93       397
```



4.3 Insights:

VGGFace (fine-tuned)

- Best overall (93% acc), balanced F_1 (≥0.90) across all 4 classes.
- Confusion matrix is almost perfectly diagonal.

ResNet18 (pretrained)

- Moderate performance (76% acc).
- Tends to over-predict “Neutral” (low precision on that class).
- Confuses Neutral ↔ Happy/Sad.

ResNet18 (scratch)

- Lowest accuracy (74%).
- Very high Sad precision but poor recall (misses many Sad).
- Over-predicts “Angry” on non-Angry faces.

Key Takeaways

- Face-specific pretraining (VGGFace) is critical for subtle expression cues.
- “Neutral” is the hardest class for generic backbones.
- Training from scratch underfits without large emotion datasets.

4.4 Challenges:

- **Limited & imbalanced data:** Only a few dozen examples per emotion, skewing the model toward majority classes.
- **Noisy labels:** “Neutral” vs. “Sad” is subjective, introducing annotation errors.
- **Subtle inter–class differences:** Generic backbones struggled to pick up fine facial cues.
- **Overfitting risk:** From-scratch training memorized the small dataset, requiring heavy regularization and early stopping.

4.5 Video Links:

VGG:
https://drive.google.com/file/d/1ZorQsCYgn3Nfx98QzelWial75PQ17Jcl/view?usp=sharing

resnet_pretrained:
https://drive.google.com/file/d/1vFaYYlOV0TwilkA6S-xo0-9ljKhc39H-/view?usp=sharing

resnet_scratch:
https://drive.google.com/file/d/1VirRu9DVOhMyBjcJD0McTXfHC2usRX88/view?usp=sharing
Bonus:

In all the videos uploaded above I have done this bonus part.
It is showing:

- for Happy: "Keep smiling!!",
- for Sad: "It's okay to feel blue.",
- for Neutral: "Steady as she goes.",
- for Angry: "Take a deep breath..."

4.6 Model Links:

VGG:
https://drive.google.com/file/d/18iL_38By7_VsWo2-mKqBi9vLln7ICIJG/view?usp=sharing

resnet_pretrainedt:
https://drive.google.com/file/d/1ZR32IXmJJZQgZPT394HckYS2UiDpM4C6/view?usp=sharing

resnet_scratch:
https://drive.google.com/file/d/1jsWiAIkd_VyxMx8qHu_oORK1PPnBmdln/view?usp=sharing

7. Conclusion

- **Face Recognition:** Pretrained ResNet-18 achieved 96% test accuracy with robust performance under diverse conditions.

- **Emotion Recognition:** Pretrained ResNet-18 led at 98% val accuracy and per-class accuracies ≥ 97%.

8. References

1. O. M. Parkhi, A. Vedaldi, A. Zisserman, “Deep Face Recognition,” *Proc. BMVC*, 2015.

2. K. He *et al.*, “Deep Residual Learning for Image Recognition,” *CVPR*, 2016.

3. PyTorch documentation: https://pytorch.org/vision/main/models.html

4. W&B Quickstart: https://docs.wandb.ai/quickstart