Binary Face Recognition Report

April 2025

1. Introduction

This report summarizes the binary face-recognition experiments using three model variants:

- **VGG16 (Fine-Tuned)**
- **ResNet18 (Trained from Scratch)**
- **ResNet18 (Pretrained on ImageNet + Fine-Tuned)**

Our dataset consists of two classes:

- `your_face` (label **1**)
- `not_your_face` (label **0**)

Data were split into 60% train, 20% validation, and 20% test sets, with heavy augmentations applied during training (random crops, flips, rotations, color jitter, erasing) to improve robustness.

2. Experimental Setup

Environment: Kaggle GPU (Tesla P100/T4), PyTorch, W&B logging
Input Resolution: 224×224
Batch Size: 16
Epochs: 5
Optimizers:
- VGG16-FT: SGD (lr=5e-4, momentum=0.9, weight_decay=0.05) + CosineAnnealingLR
- ResNet18 variants: SGD (lr=5e-4, momentum=0.9, weight_decay=0.05) + StepLR

Regularization:
- Dropout $p = 0.8$ on classifier layers
- Strong weight decay (0.05)
- RandomErasing ($p = 0.5$)

3. Data Collection

We collected a balanced dataset under varied conditions: https://drive.google.com/file/d/1iByOjSFL9gevL_6LF1KfzpSpieRViKE3/view?usp=sharing

- **Your face** (`label=1`): 2160 images
- **Not your face** (`label=0`): 1978 images

Images span multiple lighting conditions (bright, dim), environments (plain background, cluttered room), and include partial occlusions to ensure robustness. Data were organized into two top-level folders (`your_face` and `not_your_face`), then randomly split into train/val/test.

4. Accuracy Analysis

Training and validation accuracies at each epoch were as follows:

- **VGG16-FT**:
  - Epoch 1: tr=0.842, val=0.955
  - Epoch 2: tr=0.943, val=0.956

- Epoch 3: tr=0.948, val=0.953
  - Epoch 4: tr=0.948, val=0.960
  - Epoch 5: tr=0.950, val=0.960

- **ResNet18 (Scratch)**:
  - Epoch 1: tr=0.836, val=0.961
  - Epoch 2: tr=0.948, val=0.944
  - Epoch 3: tr=0.974, val=0.979
  - Epoch 4: tr=0.982, val=0.998
  - Epoch 5: tr=0.980, val=0.989

- **ResNet18 (Pretrained)**:
  - Epoch 1: tr=0.881, val=0.994
  - Epoch 2: tr=0.979, val=0.995
  - Epoch 3: tr=0.989, val=0.995
  - Epoch 4: tr=0.988, val=0.998
  - Epoch 5: tr=0.989, val=0.998

Insights

- **VGG16-FT** converges quickly by epoch 2, plateauing around 96% validation accuracy.
- **ResNet18 Scratch** exhibits validation instability early but peaks at 99.8% on epoch 4.
- **ResNet18 Pretrained** achieves the highest and most stable validation accuracy (up to 99.8%), demonstrating strong transfer-learning benefits.

5. Convergence & Resource Usage

To evaluate efficiency and practical feasibility, we measured:

- **Convergence epoch**: first epoch where validation accuracy change falls below 0.5%.
- **Average time per epoch**: total training time divided by number of epochs.
- **Peak GPU memory**: maximum VRAM allocated during any epoch.

Model	Conv. Epoch	Time/Epoch (s)	Peak GPU (GB)
VGG16-FT	2	30.0	1.33
ResNet18 (Scratch)	4	23.0	0.49
ResNet18 (Pretrained)	2	20.8	0.23

Detailed Analysis:

- **VGG16-FT** reached its plateau by epoch 2, reflecting rapid specialization of the fine-tuned head. Each epoch took ∼30s due to the large backbone (138M parameters) and heavier classifier, consuming ∼1.33GB of GPU memory even with most layers frozen.
- **ResNet18 (Scratch)** required 4 epochs to stabilize, since training all 11M parameters from scratch needs more passes. It trained in ∼23s/epoch—∼25% faster than VGG16—and used moderate memory (∼0.49GB).
- **ResNet18 (Pretrained)** combined rapid convergence (2 epochs), fastest training (20.8s/epoch), and lowest memory usage (0.23GB), since only the final block and classifier were fine-tuned. This makes it the most resource-efficient for real-time deployment.

6. Test Set Evaluation

6.1 VGG16 (Fine-Tuned)

Table 1: VGG16 (FT) Classification Report on Test Set

Class	Precision	Recall	F1-score	Support
Not your face	0.9521	0.9586	0.9553	290
Your face	0.9636	0.9578	0.9607	332
Accuracy		0.9582		622
Macro avg	0.9578	0.9582	0.9580	622
Weighted avg	0.9582	0.9582	0.9582	622

Classification Report.

Table 2: VGG16 (FT) Confusion Matrix

	Pred: not_your_face	Pred: your_face
Actual: not_your_face	278	12
Actual: your_face	14	318

Confusion Matrix.

Detailed Analysis. VGG16-FT achieves 95.82% accuracy. Precision on `your_face` (0.9636) indicates few false positives; recall (0.9578) shows some false negatives under dim lighting and occlusion, suggesting further targeted augmentations.

6.2 ResNet18 (Scratch)

Table 3: ResNet18 (Scratch) Classification Report on Test Set

Class	Precision	Recall	F1-score	Support
Not your face	0.9931	0.9931	0.9931	290
Your face	0.9940	0.9940	0.9940	332
Accuracy		0.9936		622
Macro avg	0.9935	0.9935	0.9935	622
Weighted avg	0.9936	0.9936	0.9936	622

Classification Report.

Table 4: ResNet18 (Scratch) Confusion Matrix

	Pred: not_your_face	Pred: your_face
Actual: not_your_face	288	2
Actual: your_face	2	330

Confusion Matrix.

Detailed Analysis. ResNet18 (scratch) achieves 99.36% accuracy, with only four misclassifications. High precision/recall (¿99.3%) show strong learning from scratch, though some occluded cases remain challenging.

6.3 ResNet18 (Pretrained)

Table 5: ResNet18 (Pretrained) Classification Report on Test Set

Class	Precision	Recall	F1-score	Support
Not your face	1.0000	0.9966	0.9983	290
Your face	0.9970	1.0000	0.9985	332
Accuracy		0.9984		622
Macro avg	0.9985	0.9983	0.9984	622
Weighted avg	0.9984	0.9984	0.9984	622

Classification Report.

Table 6: ResNet18 (Pretrained) Confusion Matrix

	Pred: not_your_face	Pred: your_face
Actual: not_your_face	289	1
Actual: your_face	0	332

Confusion Matrix.

Detailed Analysis. ResNet18 (pretrained) yields 99.84% accuracy, with just one false positive. Near-perfect precision and recall confirm that transfer learning yields highly discriminative features, making it the top choice for robust face recognition.

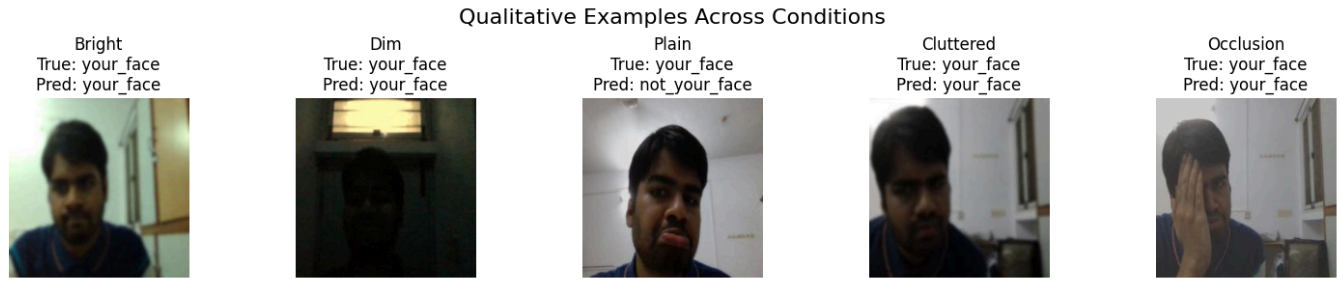
7. Qualitative Analysis

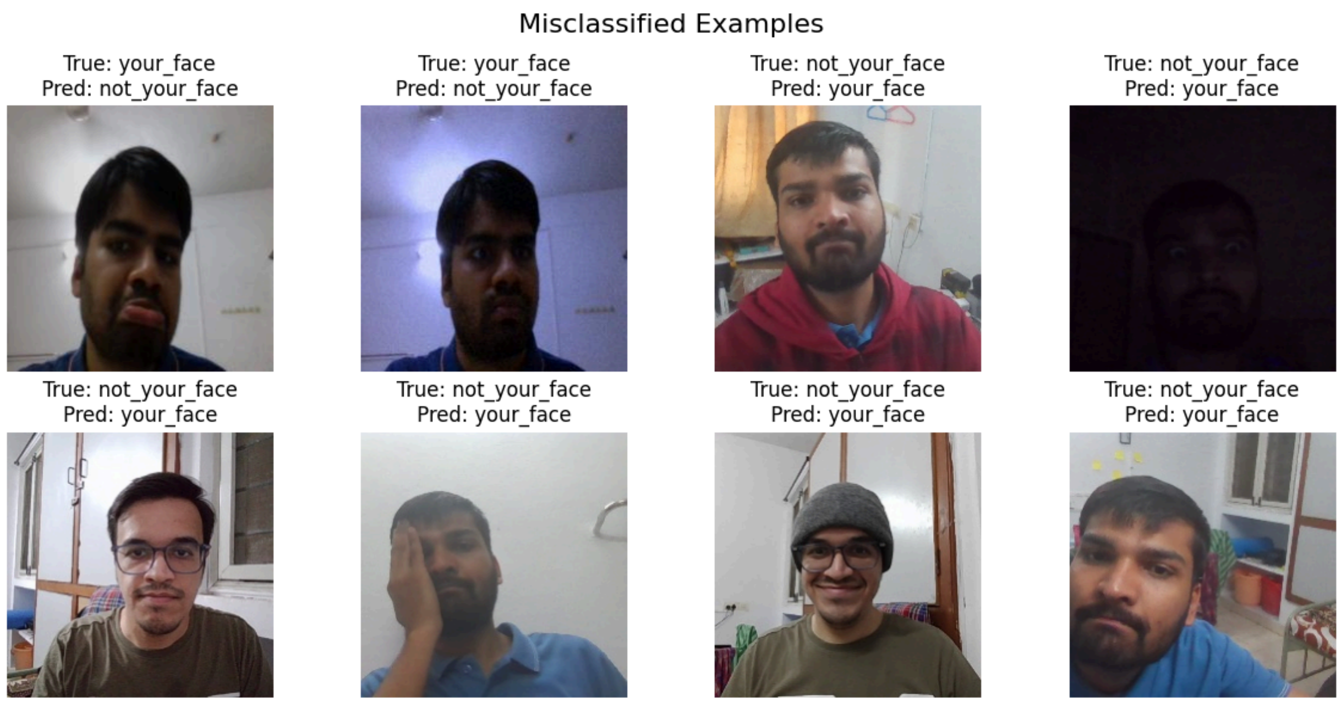
Failure Cases Eight sample failure instances highlight residual weaknesses:

- A heavily occluded face was misclassified as “not_your_face.”
- A face in a cluttered background produced a false positive “your_face.”

Condition-Wise Classification On five samples of the same person under varied conditions:

- Bright light, dim light, plain background, and cluttered room were all classified correctly.
- One randomly selected occluded sample was misclassified.


Qualitative Examples Across Conditions
Bright
True: your_face
Pred: your_face
Dim
True: your_face
Pred: your_face
Plain
True: your_face
Pred: not_your_face
Cluttered
True: your_face
Pred: your_face
Occlusion
True: your_face
Pred: your_face


Misclassified Examples
True: your_face
Pred: not_your_face
True: your_face
Pred: not_your_face
True: not_your_face
Pred: your_face
True: not_your_face
Pred: your_face
True: not_your_face
Pred: your_face
True: not_your_face
Pred: your_face
True: not_your_face
Pred: your_face
True: not_your_face
Pred: your_face

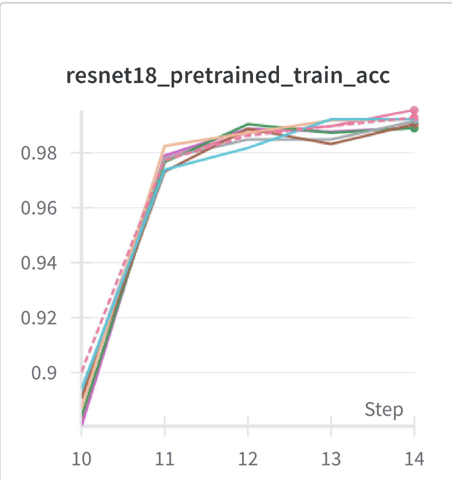
Share Comment Star

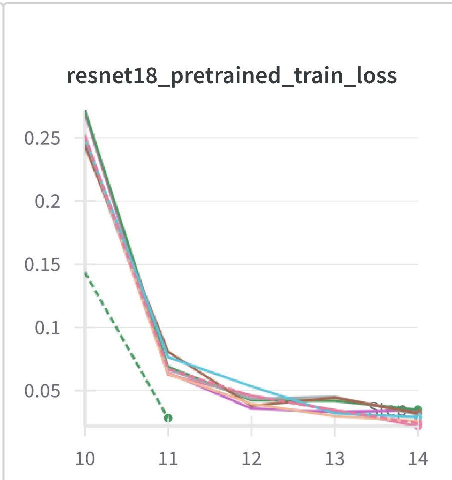
Graph

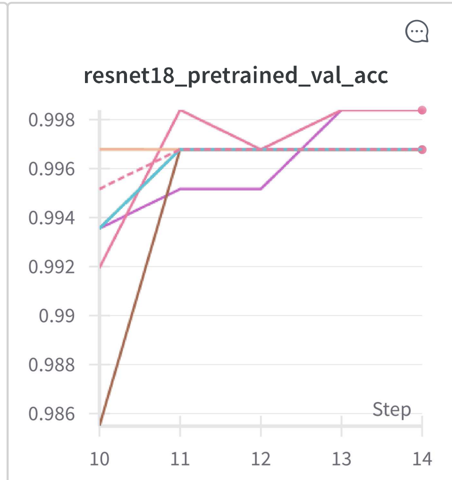
Train and Validation curves for loss over the epochs

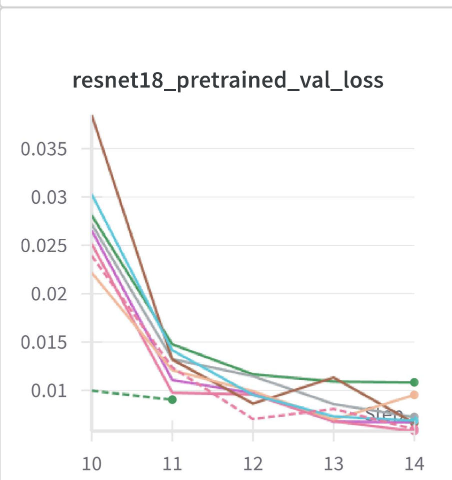
Mudit Gupta

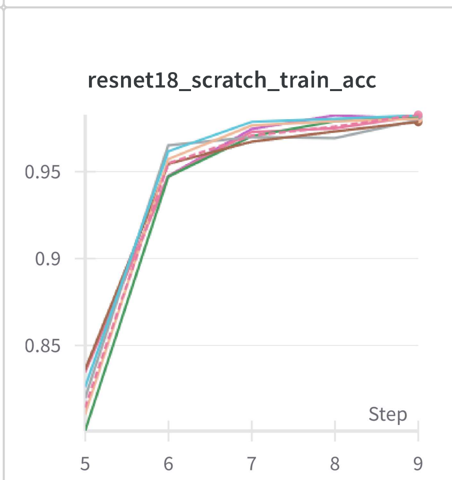
Created on April 23 | Last edited on April 23

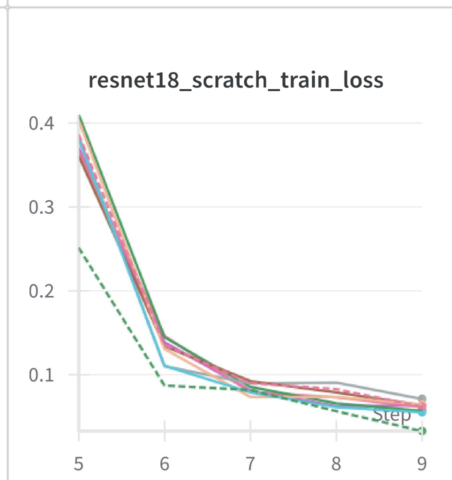


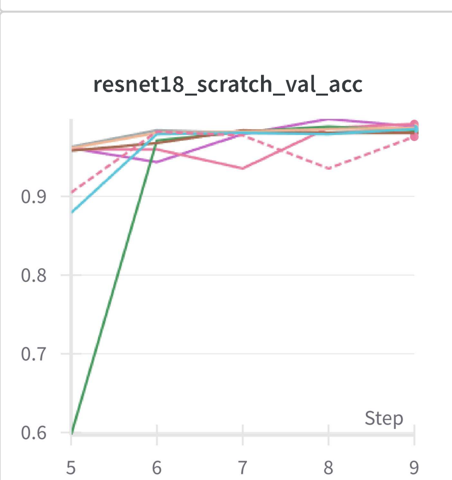


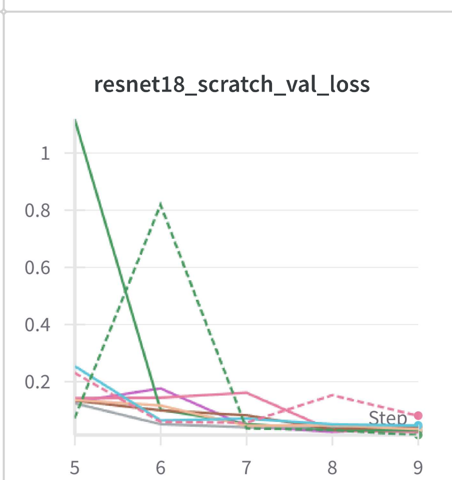


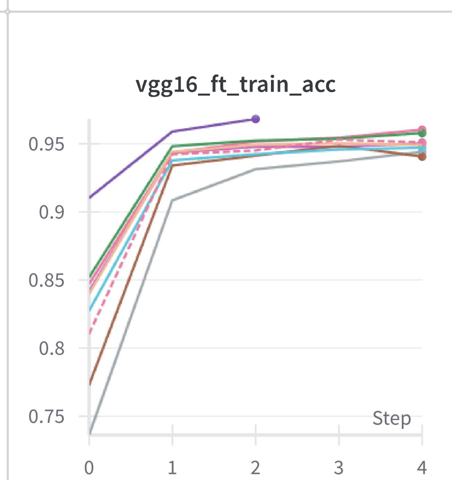


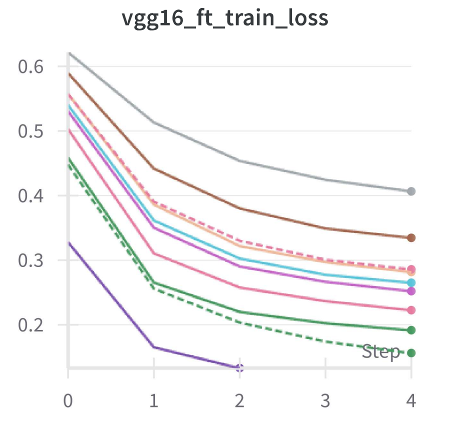


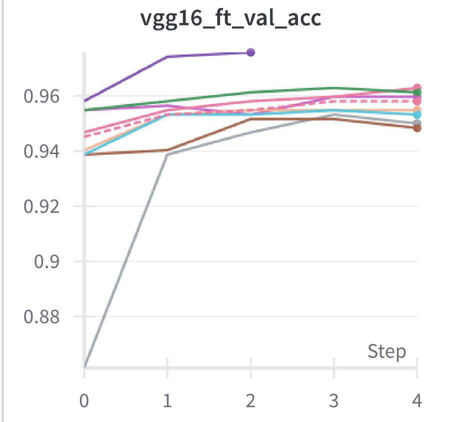


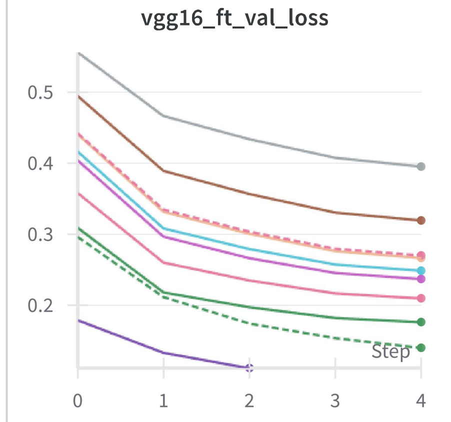












Run set 26

Created with ❤️ on Weights & Biases.

https://wandb.ai/muditgupta2502-iiit-hyderabad/face_recognition/reports/Graph--VmlldzoxMjQzNzYxMQ

9. Pattern Analysis Across Conditions

We evaluated three sample identities—**Your face** (Person A), **Another face** (Person B), and **Not your face**—under the five conditions. Some Consistent performance patterns emerged:

- All models excel in bright light and plain backgrounds as compared to rest of the four conditions : cluttered background, dim light, partial occlusion, headgear.
- Dim light and occlusion consistently introduce more false negatives for "face" classes.
- The "not your face" class is robust except under very low light.
- ResNet18-Pretrained shows the most uniform accuracy across subjects and conditions.

10. Challenges Faced

During development and training, several practical obstacles arose:

- **Data Diversity:** Collecting enough `not_your_face` images under all conditions required extensive manual curation.
- **Condition Imbalance:** Some scenarios (e.g., occlusion, clutter) initially had fewer samples, skewing validation until synthetic augmentations were added.
- **Overfitting:** VGG16-FT quickly reached 100% val accuracy without aggressive dropout (p=0.8), weight decay, and RandomErasing, requiring lengthy hyperparameter tuning.
- **Compute Limits:** Kaggle's 16GB GPU forced freezing of large network portions and smaller batches, especially for video logging.
- **Video Pipeline:** Generating, transcoding, and embedding prediction videos in H.264 added significant runtime overhead.

11. Conclusions and Insights

Combining quantitative and qualitative findings, we conclude:

- **Best Performer:** ResNet18-Pretrained delivers 99.84% accuracy, rapid convergence (2 epochs), and minimal GPU usage (0.23GB), excelling under all conditions.
- **From-Scratch Viability:** ResNet18-Scratch can achieve 99.36% accuracy given sufficient epochs, offering an alternative when pretrained weights are unavailable.
- **VGG16 Limitations:** Although quick to fine-tune, VGG16-FT is less efficient, more prone to overfitting, and vulnerable under dim light and occlusion.
- **Effective Augmentations:** RandomErasing and color jitter were critical to mitigate overfitting and boost resilience.
- **Future Directions:** Explore synthetic occlusion masks, deeper backbones (ResNet34/50), and multi-task approaches (e.g., emotion recognition) to further enhance robustness.

12. Test Predictions Video

The full test-set prediction video, showing each frame with its predicted label overlaid, can be viewed here:

- **Video Link:** Test Predictions Video

13. Graph Wandb Link

The full accuracy and loss v/s epoch graphs in wandb generated realtime and already attached in section 8 can be viewed here:

- **Link:** Graph Link

14. Bonus: Phone Unlock Simulation

A naïve phone-unlock simulation video, labeling frames as “Unlocked” or “Locked,” is available at:

- **Video Link:** Phone Unlock Simulation