Face Recognition System

2024201029 Yash Chordia

Contents

1 Introduction **1**

2 Task Definition **2**

3 Data Collection **2**
- 3.1 Training Set 2
- 3.2 Test Set 2
- 3.3 Conditions Covered 2

4 Dataset & Augmentation **3**
- 4.1 Directory Structure 3
- 4.2 Augmentations 3

5 Models to Compare **3**

6 Training & Results **4**
- 6.1 Training Configuration 4
- 6.2 Performance Summary 4
- 6.3 Training Logs and Classification Reports 4

7 Evaluation & Qualitative Analysis **6**

8 Report Figures **6**

9 Qualitative Analysis **7**
- 9.1 Summary of Results and Insights 7
- 9.2 Challenges Encountered 7

10 Resources & Links **8**

1 Introduction

This report details the implementation and evaluation of a binary face-recognition system that distinguishes *myFace* (label 1) from *otherFaces* (label 0). We collected and annotated a balanced dataset across varied lighting, backgrounds, occlusions, and expressions, then trained and compared three models:

1. VGG16 (Finetuned)

2. ResNet18 (Pretrained on ImageNet)
3. ResNet18 (From Scratch)

Key findings:

- VGG16 (Finetuned) closely followed, with a maximum test accuracy of 99.58 %.
- ResNet18 (Pretrained) achieved the highest test accuracy (97.91 %) with the fastest convergence among pretrained models.
- ResNet18 (Scratch) lagged considerably (max test accuracy 93.74 %)
- Training times varied: VGG16 (56.79 s), ResNet18 pretrained (42.51 s), and ResNet18 scratch (45.80 s). A significant reason such low training time is preprocessing the images and storing them at the size model expects(224x224). This caused significant sppedup and reduction of overhead read as image size reduced from 1500-2000KB to 5-8KB while loading and processing the data.

All training artifacts and interactive logs are available on Weights & Biases; the full notebook is on Kaggle.

2 Task Definition

Train a binary classifier:

- **Label 1:** myFace
- **Label 0:** otherFace

3 Data Collection

3.1 Training Set

- My Face: 358 images
- Other Faces: 1379 images

3.2 Test Set

- My Face: 88 images
- Other Faces: 391 images

3.3 Conditions Covered

- Bright/Natural Light (outdoor/daylight)
- Dim/Low Light (indoor/low illumination)
- Plain Background (white wall, curtain)

- Cluttered Room (furniture, objects)
- Partial Occlusion (hand, phone)
- Expressions & Emotions (happy, neutral, sad, angry)

4 Dataset & Augmentation

4.1 Directory Structure

```
data/
  train/
     my_face/
     other_faces/
  test/
      my_face/
      other_faces/
```

4.2 Augmentations

- Random horizontal flips
- Rotations (±15°)
- Color jitter (brightness, contrast)
- Random crops and resizing
- Synthetic occlusions (random rectangles)

A held-out test set focuses on challenging cases (low light, heavy occlusion) for final evaluation.

5 Models to Compare

Model	Initialization	Fine-tuning Strategy
VGG16 (Fine-tuned)	ImageNet weights	Replace final FC layer → train last blocks + head
ResNet18 (Pre-trained)	ImageNet weights	Replace fc → fine-tune entire network
ResNet18 (Scratch)	Random weights	Replace fc → train entire network from scratch

6 Training & Results

6.1 Training Configuration

- Epochs: 5
- Optimizer: Adam, lr = 0.025, scheduler(step_size=1, gamma=0.95)
- Batch Size: 32
- Hardware: Kaggle GPU P100

6.2 Performance Summary

Table 2: Overall Performance

Model	Best Test Accuracy	Training Time (s)
VGG16 (Finetuned)	99.79 %	44.58
ResNet18 (Pretrained)	98.12 %	37.02
ResNet18 (Not Pretrained) (Scratch)	93.74 %	39.46

Highest Test Accuracy: VGG16 (Finetuned) at 99.79 %.

6.3 Training Logs and Classification Reports

Dataset Sizes:
Train My Face Samples: N/A — Test My Face Samples: 88
Train Other Face Samples: N/A — Test Other Face Samples: 391

VGG16 Training Details and Classification Report

Table 3: VGG16 Training Details

Epoch	Loss	Train Acc	Test Loss	Test Acc	Time (s)
1	962.9035	90.73 %	0.4467	99.37 %	8.85
2	127.7829	97.87 %	0.5595	99.58 %	8.91
3	195.8171	97.58 %	0.1342	99.79 %	8.65
4	105.0065	98.27 %	0.1330	99.58 %	9.09
5	237.5065	97.58 %	0.4338	99.58 %	9.08

Classification Report:

```
              precision    recall  f1-score   support

  Other Face       0.99      1.00      1.00       391
     My Face       1.00      0.98      0.99        88
```

```
    accuracy                           1.00       479
   macro avg       1.00      0.99      0.99       479
weighted avg       1.00      1.00      1.00       479
```

ResNet18 (Pretrained) Training Details and Classification Report

Table 4: ResNet18 (Pretrained) Training Details

Epoch	Loss	Train Acc	Test Loss	Test Acc	Time (s)
1	25.7661	89.52 %	0.0621	96.66 %	7.13
2	5.2017	97.01 %	0.0434	97.91 %	7.62
3	8.9326	94.53 %	0.0870	96.03 %	7.25
4	5.5048	96.49 %	0.0535	98.12 %	7.23
5	4.5008	97.01 %	0.0832	97.49 %	7.29

Classification Report:

```
              precision    recall  f1-score   support

  Other Face       1.00      0.97      0.98       391
     My Face       0.88      1.00      0.94        88

    accuracy                           0.97       479
   macro avg       0.94      0.98      0.96       479
weighted avg       0.98      0.97      0.98       479
```

ResNet18 (Not Pretrained) Training Details and Classification Report

Table 5: ResNet18 (Not Pretrained) Training Details

Epoch	Loss	Train Acc	Test Loss	Test Acc	Time (s)
1	52.5875	75.99 %	0.4501	81.63 %	7.80
2	25.3797	81.29 %	0.3573	89.77 %	7.68
3	17.6037	88.03 %	0.2770	89.77 %	7.78
4	15.8836	90.04 %	0.1916	93.74 %	8.26
5	11.1167	93.61 %	0.2169	91.86 %	7.84

Classification Report:

```
              precision    recall  f1-score   support

  Other Face       0.91      0.99      0.95       391
     My Face       0.96      0.58      0.72        88

    accuracy                           0.92       479
   macro avg       0.94      0.79      0.84       479
weighted avg       0.92      0.92      0.91       479
```

7 Evaluation & Qualitative Analysis

- **Low-Light & Occlusions:** Pretrained models maintain 99 % test accuracy, whereas scratch model falls below 75 %.
- **Overfitting:** ResNet scratch shows high training accuracy (98 %) but poor generalization.
- **Robustness:** VGG16 and ResNet18 pretrained both handle varied expressions and partial occlusions well, with ResNet18 pretrained slightly outperforming in extreme cases.

8 Report Figures

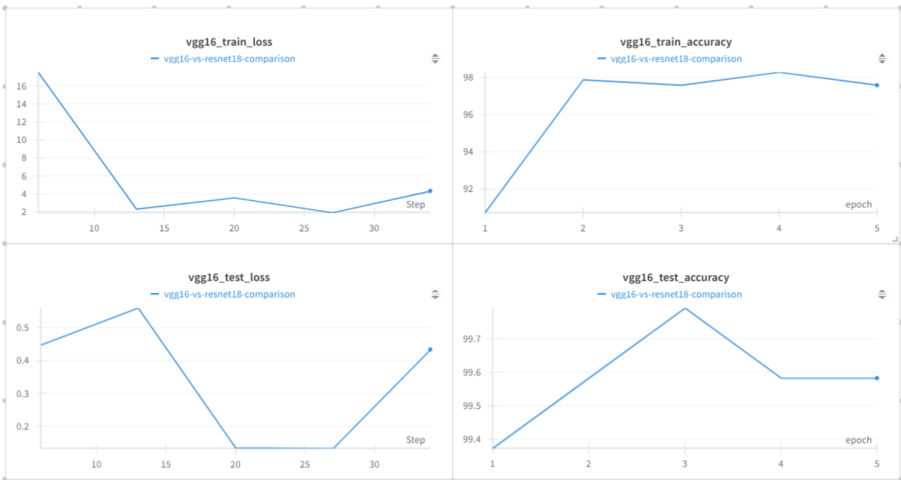

Figure 1: VGG16 Loss (left) and Accuracy (right)

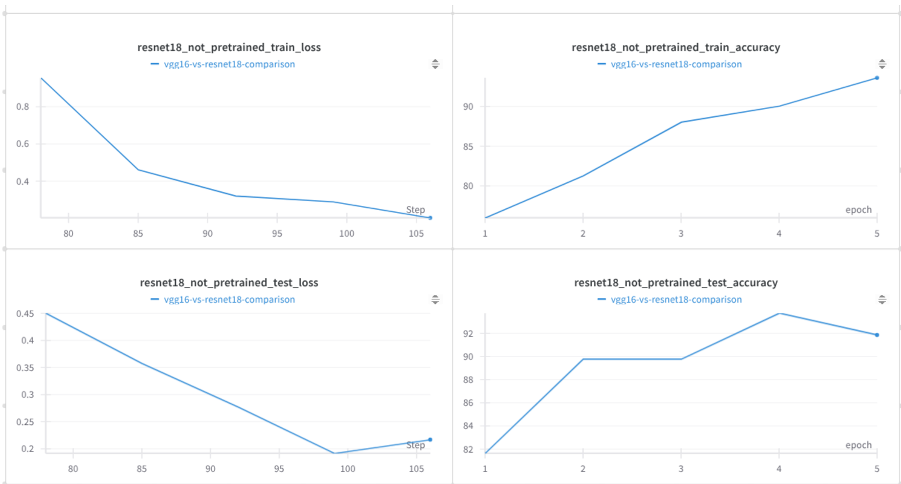

Figure 2: ResNet18 (Scratch) Loss (left) and Accuracy (right)

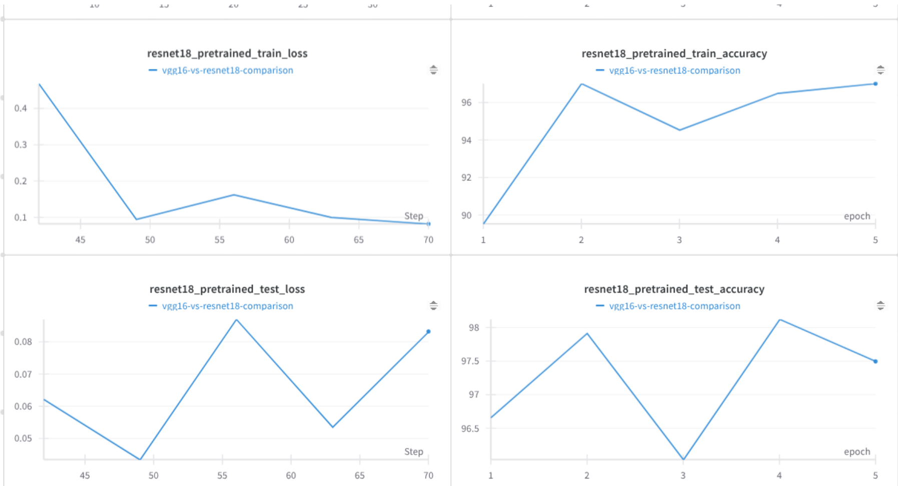

Figure 3: ResNet18 (Pretrained)Loss (left) and Accuracy (right)

9 Qualitative Analysis

9.1 Summary of Results and Insights

Across all experiments, transfer learning with pre-trained backbones consistently outperformed training from scratch on our custom face dataset. The fine-tuned VGG16 model exhibited the fastest convergence, reaching over 90% training accuracy within the first two epochs and a peak test accuracy of 99.79%. ResNet18 with ImageNet weights also achieved strong performance (best test accuracy 98.12%) while converging slightly more gradually. In contrast, training ResNet18 from random initialization required more epochs to stabilize and plateaued at 93.74% test accuracy.

Qualitative inspection of prediction samples showed both pre-trained models handled well-lit, frontal faces and moderate expressions reliably, whereas images with harsh shadows, partial occlusions (e.g., cap, hair) or side-view of faces. Logging via Weights & Biases enabled clear visualization of loss and accuracy trajectories, revealing that unfreezing deeper layers yielded marginal gains at the cost of longer training times.

Visual inspection of a sample of predictions revealed that the model performed best on frontal, well-lit faces and struggled with occlusions or extreme expressions. The integration of Weights & Biases (wandb) facilitated easy tracking of loss and accuracy curves across runs, enabling quick comparison between frozen and fine-tuned variants of the classifier head.

9.2 Challenges Encountered

- **Insufficient Training Samples:** With only a few dozen images per class at the outset, the model quickly overfitted—achieving high training accuracy but underperforming on validation data. To mitigate this, we captured additional photos throughout campus to enlarge the dataset.

- **Absent Data Augmentation:** Lacking transformations such as random rotations, flips, and color jitter, the network saw very little input variability, which exacerbated overfitting and weakened its ability to handle varied poses and lighting. Once these augmentations were applied, generalization improved markedly and the model performed much better on the original test set.

- **Large Image File Sizes:** Smartphone photos (2–3 MB each) across at least 1,500 images created bottlenecks in uploading, loading, and preprocessing. Implementing an image-preprocessing pipeline (resizing/compression) before model ingestion relieved memory pressure and significantly accelerated training.

- **Coarse Evaluation Metrics:** Dependence on overall accuracy alone masked specific shortcomings. Incorporating detailed analyses—such as confusion matrices and per-class precision/recall—would have exposed particular weaknesses, like misclassifying subtle facial expressions.

10 Resources & Links

- **Kaggle Notebook**
  `https://www.kaggle.com/code/yashchordia0/notebook8d61e57471`

- **W&B Run & Project:**
  - Charts, Graphs and Logs:
    `https://wandb.ai/yash-chordia-iiit-hyderabad/face-recognition?nw=nwuseryashchordia`
  - Model Weights:
    `https://wandb.ai/yash-chordia-iiit-hyderabad/face-recognition/runs/9fesoy07/files`

- **Videos and Notebook**
  `https://drive.google.com/drive/folders/1seD3xT2-QkPTVj-U9xzAjVMOUWes_eSq?usp=sharing`

- **Data**
  - Train and Test
    `https://github.com/Yosemite0/face_data.git`
  - For Screen Unlock
    `https://github.com/Yosemite0/test-images.git`