

Face Off

PCA + Logistic Regression
vs. FaceNet — which wins?

APPROACH A · BASELINE

PCA + Logistic Regression

150 components · Grayscale 64×64 · Multinomial LR

— VS —

APPROACH B · DEEP LEARNING

FaceNet

InceptionResNetV1 · 512-d embeddings · MTCNN

DATASET

LFW · 13,233 images · 5,749 identities

Edward
Sanjay
Murabet
Priyanshi

Today's Agenda

- 01** Introduction
- 02** Dataset
- 03** EDA
- 04** Methodology A: PCA
- 05** Methodology B: FaceNet
- 06** Results: Closed-set [FaceNet]
- 07** Results: Open-set [FaceNet]
- 08** Verification
- 09** Summary
- 10** Discussion and Conclusion

Introduction

Why Facial Recognition Matters

Facial recognition is proliferating across law enforcement, security, and consumer applications — yet rigorous, accessible evaluations of its accuracy remain limited.

CLASSICAL ML

PCA + Logistic Regression

- Reduce pixel data via PCA (150 components, whiten=True)
- Retains 90.44% of variance — top eigenfaces
- Multinomial logistic regression, L-BFGS solver
- No GPU required · 1.8s training · 0.002ms inference

DEEP LEARNING

FaceNet (InceptionResNetV1)

- MTCNN face detection & alignment to 160×160px
- Pretrained on VGGFace2 — robust to pose, lighting
- 512-dimensional L2-normalized embedding per face
- Cosine distance threshold for identity decision

Core Question: Is the added complexity of deep learning worth the accuracy gain, or can a classical pipeline remain competitive across 600+ identity classes?



DATASET DESCRIPTION

Labeled Faces in the Wild (LFW)

13,233

TOTAL IMAGES

5,749

TOTAL IDENTITIES

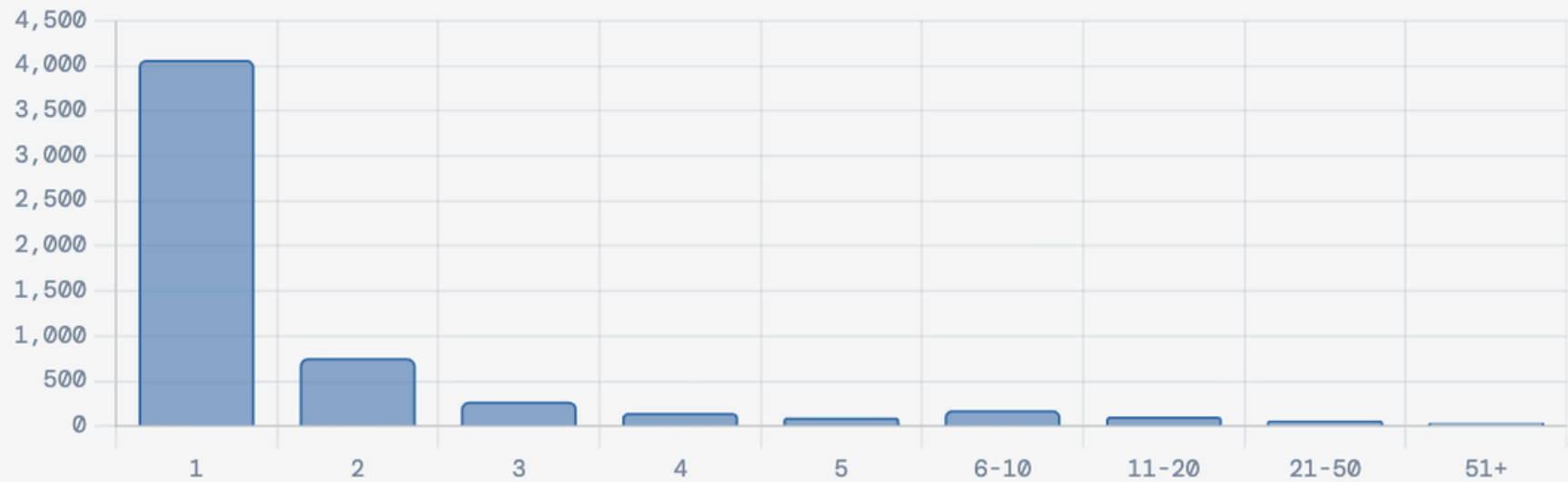
610

FILTERED IDS (≥ 4)

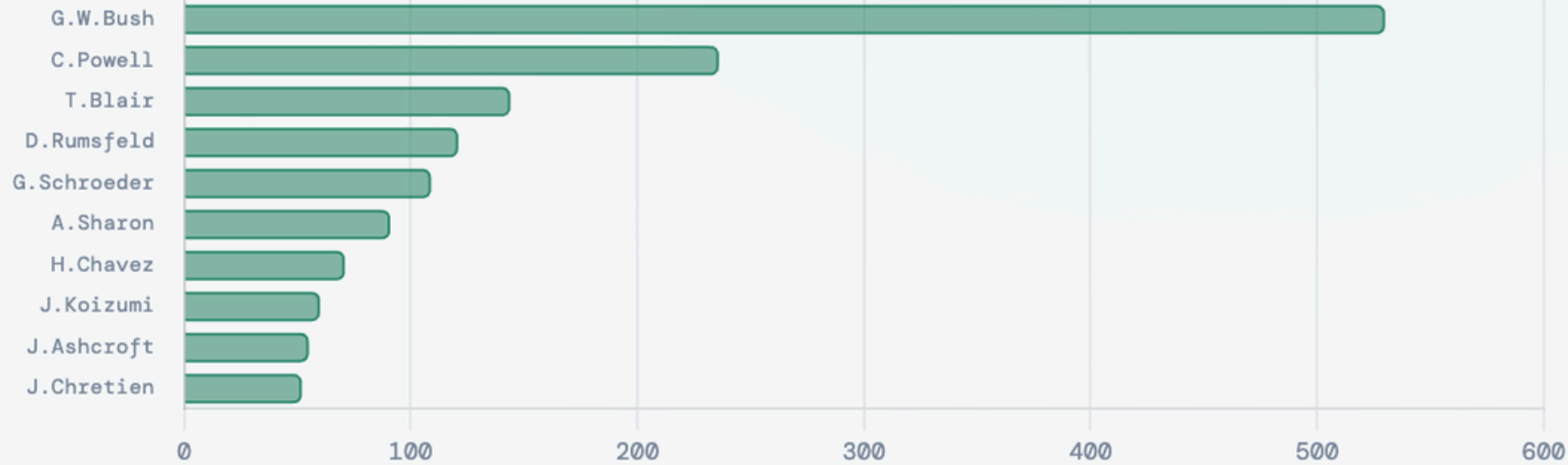
6,733

FILTERED IMAGES

IDENTITY DISTRIBUTION (IMAGES PER PERSON)



TOP 10 MOST-PHOTOGRAPHED IDENTITIES



ALIGNMENT

Funneled (pre-aligned)

SPLIT

80% train / 20% test · stratified

FILTER

≥ 4 images per identity

VARIATION

Lighting · Pose · Expression · Background

Key Observations

Heavy Long-Tail Distribution

Median images/identity: 1, mean: 2.3, max: 530 (G.W. Bush). Highly imbalanced.

Filtering Strategy

After ≥ 4 images/identity: retain 610 identities and 6,733 images for meaningful classification.

Class Imbalance Challenge

G.W. Bush: 530 images vs many with exactly 4. Severe imbalance disadvantages both models for underrepresented classes.

IMAGE COUNT DISTRIBUTION (FILTERED SUBSET)



610

IDENTITIES

90.44%

PCA VARIANCE

5,386

TRAIN IMAGES

PCA + Logistic Regression

PROCESSING PIPELINE



1 · Preprocessing

Grayscale → 64×64 → flatten 4,096-d → normalize [0,1]

2 · Dimensionality Reduction

PCA(n_components=150, whiten=True)

Retains **90.44%** variance.

3 · Classification

Multinomial logistic regression, L-BFGS, max 2,000 iterations, n_jobs=-1

4 · Evaluation Protocols

Closed-set: 80/20 stratified split · 610 classes

Open-set: 80% known / 20% unknown · Euclidean gallery · threshold sweep

0.35s

PCA FIT

0.002ms

PER-IMAGE

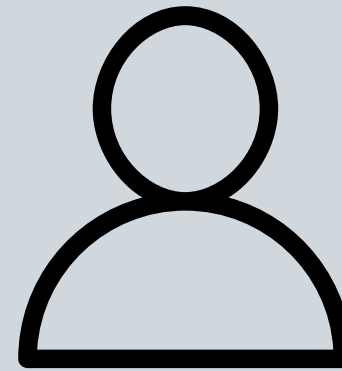
1.45s

LR TRAIN

150

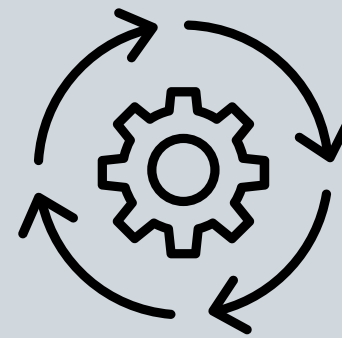
COMPONENTS

Face Identification with FaceNet



TASK

- “Who is this person?”



Process

- If you have a gallery of known identities, you can:
- Compute embeddings for each person in your gallery using FaceNet



Result

- For a new image, find the closest embedding (using cosine distance)
- Return the identity if distance $<$ threshold, otherwise “unknown”

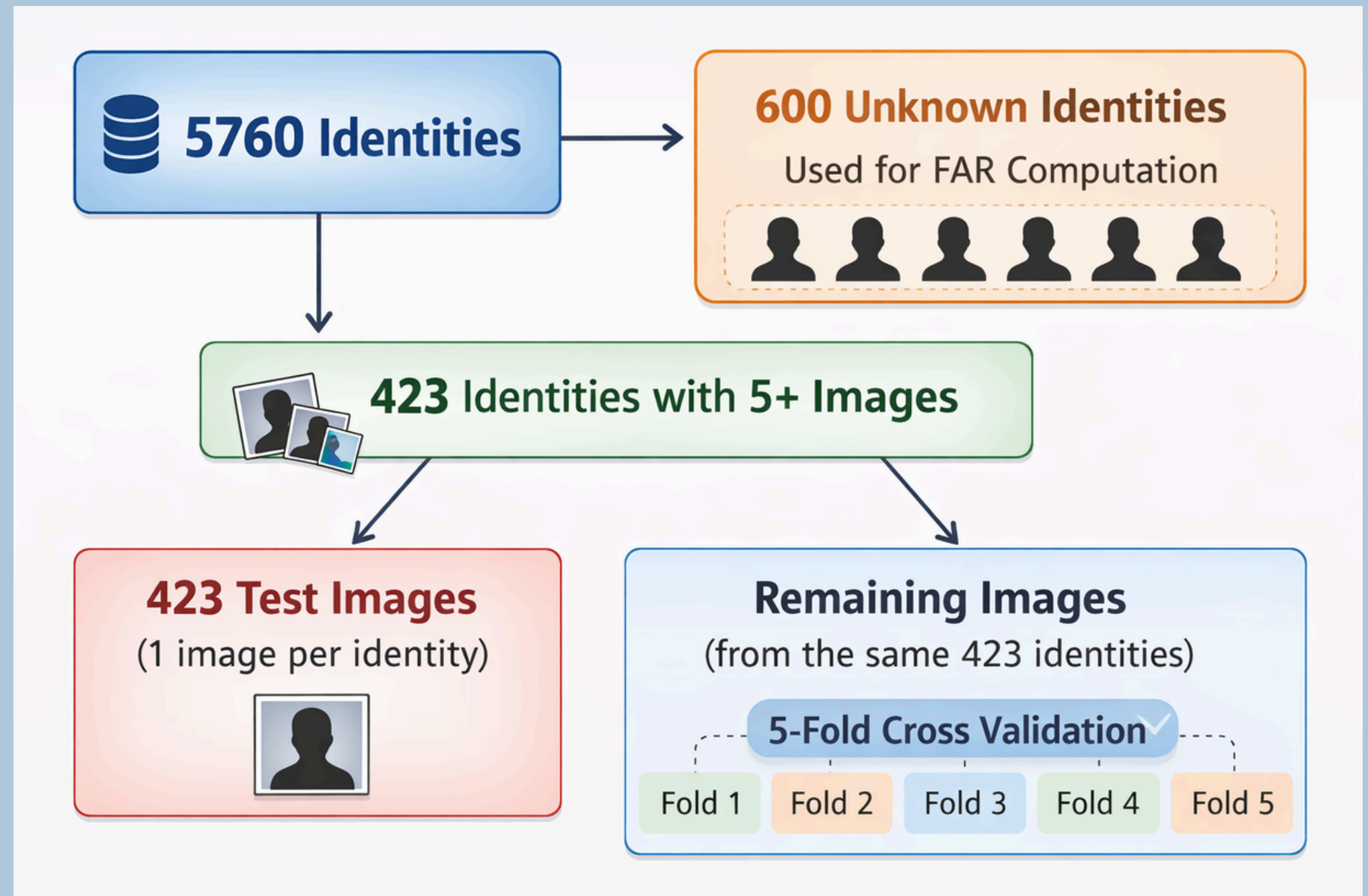
Threshold Tuning

- When selecting the optimal threshold, here are some metrics we look at:

Metric	Meaning	Formula	Focus
TIR (True Identification Rate)	Known user correctly identified as their true identity	Correct identifications / Total known probes	Accuracy / Usability
FIR (False Identification Rate)	Known user incorrectly identified as another enrolled user	Known probes assigned the wrong identity / Total known probes	Accuracy / Security
FAR (False Acceptance Rate)	Unknown user incorrectly accepted as a known user (impostor accepted)	Unknown probes accepted as known / Total unknown probes	Security

DataSet

- of course the more images per identity - the better so system can capture variation (in things like pose, expression, and lighting) and generalize well. This is why we filtered out identities with less than 5 images in our gallery.



Results

- 0.25 to 0.4 range is appropriate for L2-normalized FaceNet embeddings in open-set recognition.
- Based on this table, we selected a threshold of 0.31 to balance between security and usability.
- Based on the specific application, you may choose to select a different threshold

```
----- FINAL HELD-OUT TEST RESULTS -----  
Threshold: 0.31  
Known test identities: 423  
Unknown test identities: 600  
TIR: 0.9362  
FIR: 0.0071  
FRR: 0.0567  
FAR: 0.0300
```

Threshold	TIR	FIR	FAR
0.250	0.8794	0.0024	0.0087
0.255	0.8879	0.0028	0.0090
0.260	0.8960	0.0028	0.0107
0.265	0.9002	0.0033	0.0130
0.270	0.9083	0.0043	0.0147
0.275	0.9116	0.0047	0.0153
0.280	0.9191	0.0047	0.0153
0.285	0.9248	0.0047	0.0167
0.290	0.9343	0.0047	0.0193
0.295	0.9362	0.0047	0.0220
0.300	0.9371	0.0047	0.0247
0.305	0.9404	0.0047	0.0270
0.310	0.9452	0.0047	0.0283
0.315	0.9461	0.0047	0.0307
0.320	0.9466	0.0057	0.0347
0.325	0.9470	0.0057	0.0373
0.330	0.9475	0.0066	0.0427
0.335	0.9494	0.0066	0.0497
0.340	0.9499	0.0066	0.0533
0.345	0.9508	0.0066	0.0567
0.350	0.9532	0.0066	0.0687
0.355	0.9560	0.0066	0.0757
0.360	0.9570	0.0071	0.0837
0.365	0.9570	0.0076	0.0943
0.370	0.9574	0.0076	0.1097
0.375	0.9593	0.0090	0.1193
0.380	0.9626	0.0095	0.1303
0.385	0.9626	0.0104	0.1430
0.390	0.9626	0.0104	0.1607
0.395	0.9631	0.0104	0.1780
0.400	0.9655	0.0104	0.1947

FaceNet Pipeline

PROCESSING PIPELINE



1 · Detection & Alignment

MTCNN detects face, aligns eyes horizontally, crops & resizes to 160×160 with 20px margin

2 · Embedding Extraction

InceptionResNetV1 on VGGFace2 → 512-d unit-length vector (mathematical fingerprint)

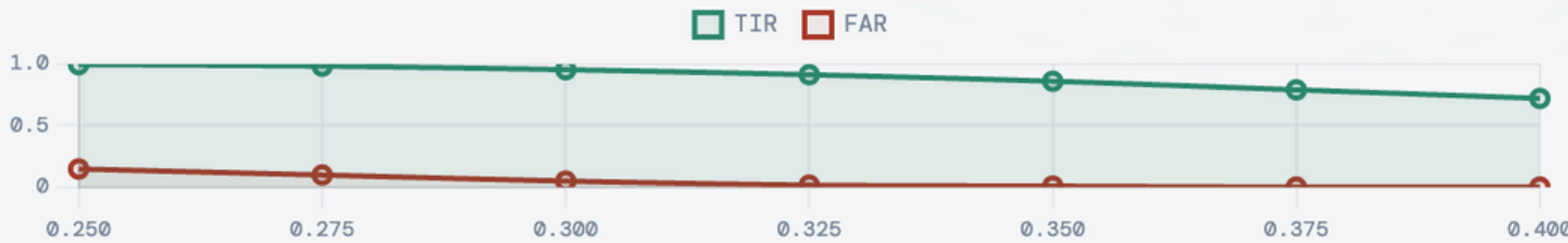
3 · Gallery Construction

80/20 identity split · avg centroid per known identity (L2-norm) · 1 probe reserved per identity

4 · Cosine Distance Decision

Accept if $\text{dist} < \theta$, else "unknown". Optimal: $\theta=0.30$ (open-set), $\theta=0.56-0.58$ (verification).

FACENET TIR & FAR VS THRESHOLD



$\theta = 0.30$
OPEN-SET OPT.

$\theta = 0.56-0.58$
VERIFICATION OPT.

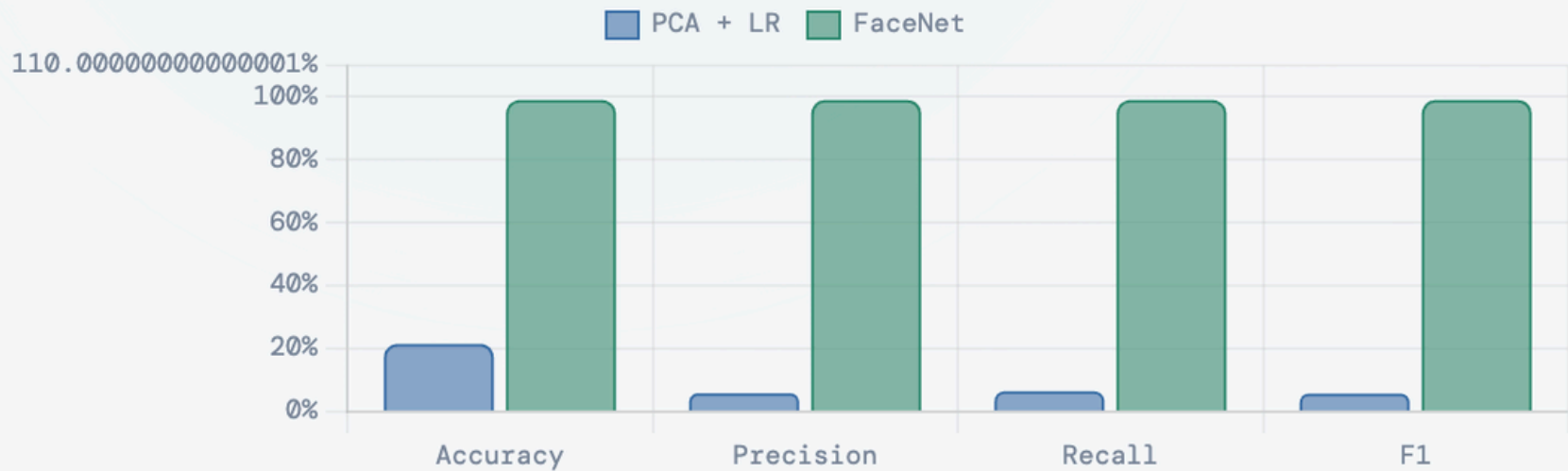
~45ms
PER-IMAGE

RESULTS · CLOSED-SET CLASSIFICATION

Closed-Set Performance

All test identities seen during training · 610 classes · Accuracy, Precision, Recall, F1 (macro avg)

METRIC COMPARISON — PCA+LR VS FACENET



SIDE-BY-SIDE METRIC VIEW

Accuracy		21.53%	vs	99%
Precision		5.89%	vs	99%
Recall		6.46%	vs	99%
F1 Score		5.74%	vs	99%

PCA + LR

21.53%

CLOSED-SET ACC.

Bush F1: 0.72, Koizumi: 0.96

FACENET

99%

CLOSED-SET ACC.

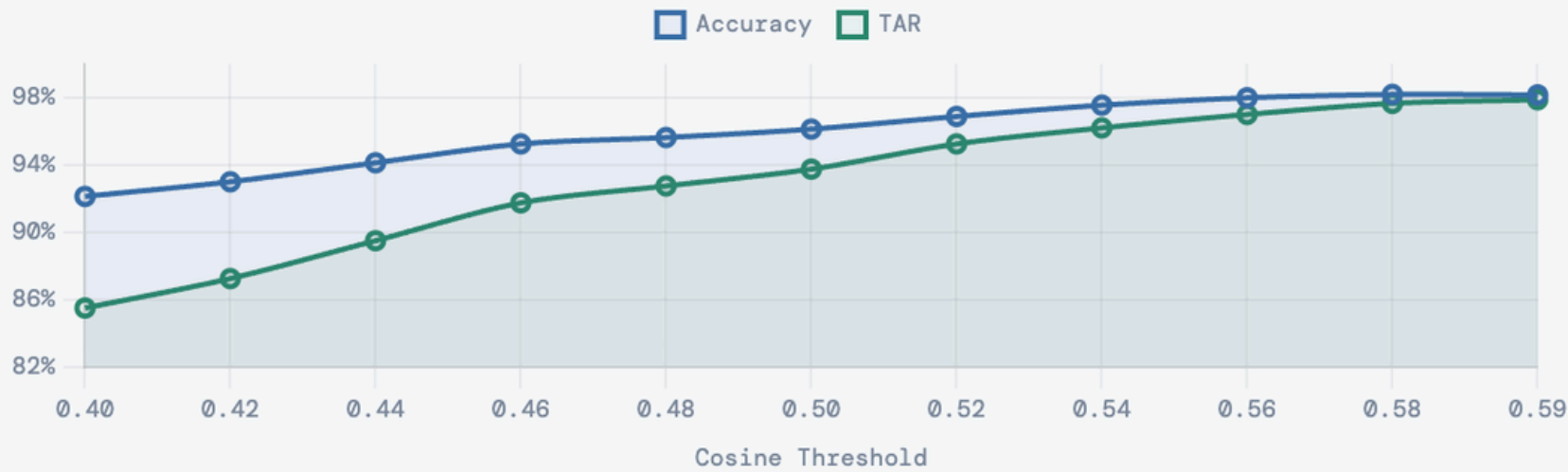
Gap: +77.47%

RESULTS · FACE VERIFICATION

FaceNet Verification Performance

Task: "Same person or different?" · Input: 2 images · Metric: TAR, FAR, Accuracy

VERIFICATION ACCURACY & TAR VS COSINE THRESHOLD



Optimal Range: $\theta = 0.56-0.58$

97.8%
ACCURACY

97.3%
TAR

~1.1%
FAR

VERIFICATION THRESHOLD TABLE (SELECTED)

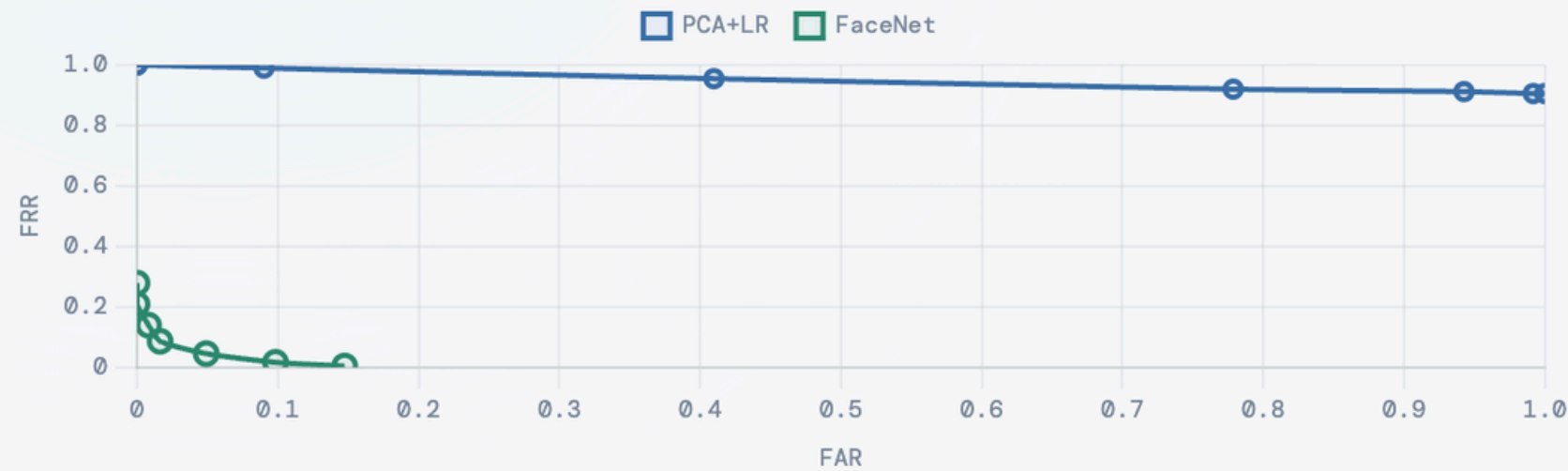
θ	ACCURACY	TAR	FAR
0.40	0.9213	0.8550	0.0125
0.50	0.9613	0.9375	0.0150
0.56	0.9799	0.9700	0.0103
0.58	0.9819	0.9766	0.0127
0.59	0.9816	0.9787	0.0154

RESULTS · OPEN-SET IDENTIFICATION

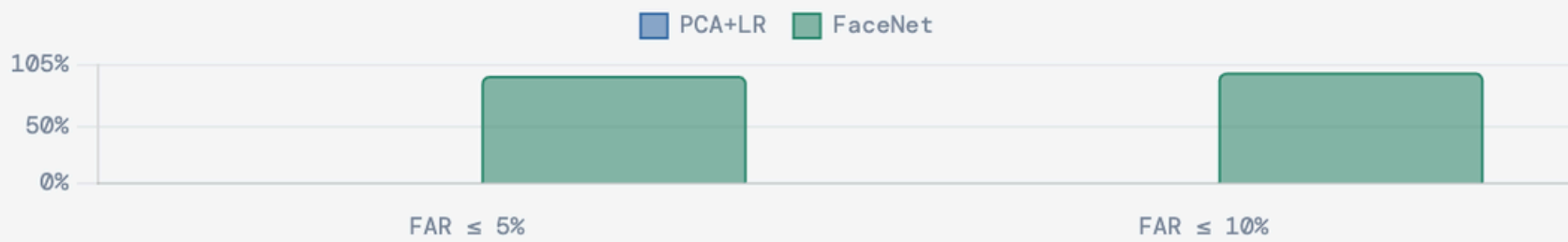
Open-Set Evaluation

80% known / 20% unknown identities · Models must reject unknowns · TIR, FRR, FAR threshold sweep

DET CURVE — FAR VS FRR



TIR AT MATCHED FAR LEVELS



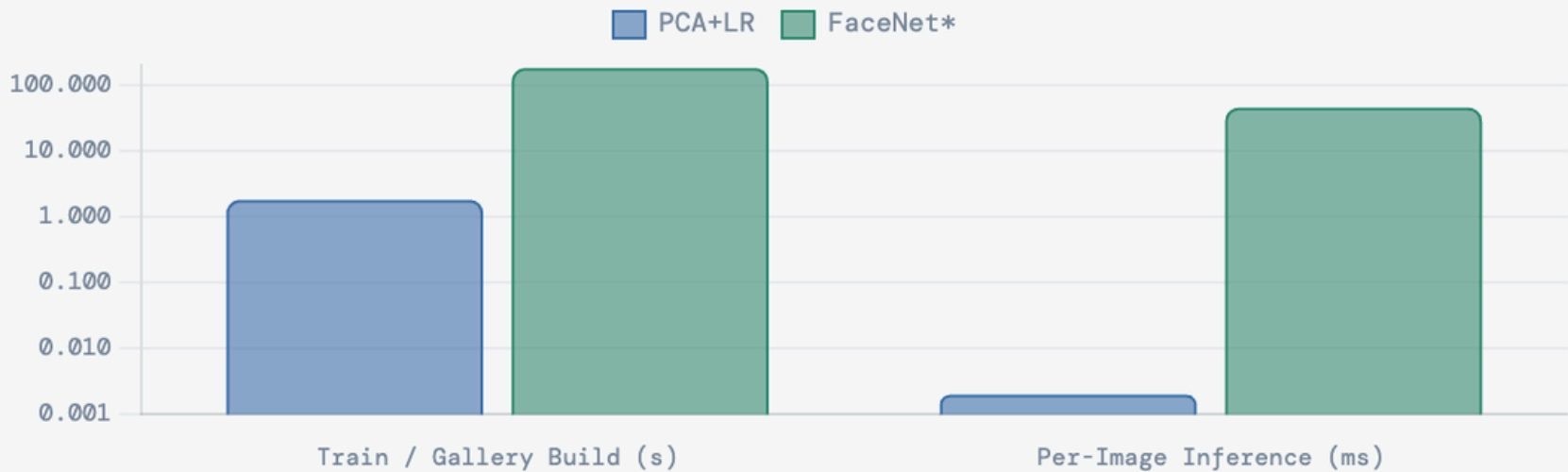
FACENET THRESHOLD SWEEP (COSINE DIST)

θ	TIR	FRR	FAR
0.250	0.9959	0.0041	0.1475
0.300	0.9549	0.0451	0.0492
0.325	0.9139	0.0861	0.0164
0.375	0.7910	0.2090	0.0000

RESULTS · HEAD-TO-HEAD SUMMARY

Complete Comparison

TRAINING & INFERENCE TIMING (* ESTIMATED FOR FACENET)



FULL METRIC COMPARISON

METRIC	PCA+LR	FACENET	GAP
Closed-Set Acc	21.53%	99%	+77.47%
Precision (macro)	5.89%	99%	+93.11%
Recall (macro)	6.46%	99%	+92.54%
F1 (macro)	5.74%	99%	+93.26%
TIR @FAR≤5%	0.00%	95.49%	+95.49%
TIR @FAR≤10%	1.02%	98.36%	+97.34%
Train Time	1.8s	~180s*	100×
Per-Image Infer.	0.002ms	~45ms*	22,500×

DISCUSSION

Why FaceNet Dominates

PCA + LR · LIMITATION

Operates on **raw pixel variance**. Top 150 principal components capture global illumination patterns & head pose — features varying *within* identities too. Logistic regression tries to carve linear boundaries across 600+ classes in this noisy 150-d space. Feature space is simply not discriminative enough at scale.

FACENET · ADVANTAGE

Learns a **512-d embedding space** via triplet/contrastive loss so same-identity faces cluster tightly and different-identity faces separate maximally. Creates **semantically meaningful distances** that pixel-level methods fundamentally cannot achieve. Pretrained on millions of diverse faces.

PCA+LR USE CASES

- Small closed sets (<50 IDs)
- Educational or prototype use
- No GPU, low-power devices
- Ultra-fast inference needed

FACENET USE CASES

- Access control & surveillance
- Open-set recognition
- Large-scale identity databases
- Production deployments

SHARED LIMITATIONS

- Severe class imbalance in LFW
- Single train/test split
- No demographic diversity eval
- News-photo bias in LFW

Potential Improvements

PCA + LR SIDE

- Try 250+ PCA components for additional discriminative variance
- Replace LR with SVM (RBF kernel) — handles nonlinear boundaries
- Histogram equalization for illumination normalization
- Apply LDA after PCA for inter-class separability

FACENET SIDE

- Fine-tune on LFW data for domain adaptation
- ArcFace loss for tighter, more discriminative embeddings
- Multi-embedding gallery for better variation coverage
- Ensemble with other face models for robustness

BOTH MODELS

- Data augmentation: flip, rotations, brightness shifts
- k-fold cross-validation instead of single split
- SMOTE or weighted loss for class imbalance
- Evaluate on age- and ethnicity-diverse benchmarks

Key Insight: The evaluation protocol matters enormously — closed-set accuracy alone is insufficient. Open-set and verification metrics reveal true production readiness.

CONCLUSION

Answering Our Research Questions

Q1 Can we get good results using PCA + LR?
Partially. Works for top-represented classes (Bush F1:0.72, Koizumi F1:0.96) but drops to **21.53%** overall across 600+ classes. Useful baseline only.

Q2 How much better is FaceNet?
Dramatically. FaceNet: **99%** closed-set acc vs 21.53%. At FAR<5%, FaceNet: **95.49%** TIR vs PCA+LR: **0.00%**. Gap consistent across all metrics.

Q3 Is the accuracy gain worth the extra complexity?
Yes — for real-world use. Modern frameworks (facenet-pytorch, MTCNN) abstract away complexity. PCA+LR retains value for education and tiny closed sets only.

KEY TAKEAWAY 1

Classical methods reach their limits quickly in high-class face recognition

KEY TAKEAWAY 2

Learned embeddings fundamentally solve the representation problem PCA cannot

KEY TAKEAWAY 3

Evaluation protocol matters — closed-set accuracy alone is insufficient

**Thank
You** for your time
and attention