

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO TIỂU LUẬN ĐỒ ÁN

SINH TRẮC HỌC

ĐỀ TÀI:

**NGHIÊN CỨU VÀ XÂY DỰNG HỆ THỐNG NHẬN DIỆN KHUÔN
MẶT SỬ DỤNG GHOSTFACENET VÀ FAISS**

Giảng viên lý thuyết: Thầy Lê Hoàng Thái

Giảng viên thực hành: Dương Thái Bảo

Sinh viên thực hiện:

Trần Đức Thịnh - MSSV: 22120348

Nguyễn Sinh Trục - MSSV: 22120395

Phạm Tuấn Vương - MSSV: 22120446

Thành phố Hồ Chí Minh, tháng 01 năm 2026.

LỜI NÓI ĐẦU

Bài báo cáo **Đồ án - Face recognition** Trước tiên, chúng em xin gửi lời cảm ơn chân thành và sâu sắc đến **Thầy Lê Hoàng Thái** và **Thầy Dương Thái Bảo** đã tận tình giảng dạy và hướng dẫn chúng em trong môn Nhận dạng. Nhờ những bài giảng đầy tâm huyết và sự hỗ trợ tận tình của Thầy, chúng em đã có cơ hội tiếp cận và tìm hiểu sâu hơn về các phương pháp nhận diện khuôn mặt trong video.

Trong quá trình thực hiện đồ án, chúng em đã gặp không ít khó khăn, nhưng nhờ sự chỉ dẫn và bài giảng từ Thầy, chúng em đã có thể hoàn thành bài nghiên cứu này một cách tốt nhất có thể. Những kiến thức mà Thầy truyền đạt không chỉ giúp chúng em hiểu rõ về lý thuyết mà còn giúp chúng em áp dụng vào thực tế một cách hiệu quả. Chúng em kính mong nhận được những đánh giá và nhận xét từ thầy để có thể học hỏi, rút kinh nghiệm và hoàn thiện hơn trong những dự án tương lai.

Cuối cùng, chúng em xin kính chúc hai Thầy luôn dồi dào sức khỏe, nhiều niềm vui và thành công trong sự nghiệp giảng dạy cũng như nghiên cứu khoa học. Kính chúc Thầy luôn giữ vững nhiệt huyết với nghề, tiếp tục truyền cảm hứng và tri thức cho nhiều thế hệ sinh viên sau này. Chúng em xin chân thành cảm ơn và kính chúc Thầy mọi điều tốt đẹp nhất.

MỤC LỤC

I. GIỚI THIỆU	5
II. TỔNG QUAN NHẬN DẠNG KHUÔN MẶT (FACE RECOGNITION).....	5
1. Giới thiệu (Introduction)	5
2. Các kỹ thuật nhận dạng khuôn mặt (Face Recognition Techniques)	6
3. Cơ sở dữ liệu (Databases).....	8
4. Bộ lọc tương quan tiên tiến (Advanced Correlation Filters)	8
5. Tensorfaces	10
6. Mô hình ngoại hình chủ động (AAM) để nhận dạng khuôn mặt	10
7. Siêu độ phân giải khuôn mặt sử dụng LPP	11
8. Kết luận (Conclusion)	11
III. PHƯƠNG PHÁP GHOSTFACENETS CHO BÀI TOÁN NHẬN DẠNG KHUÔN MẶT	11
1. Giới thiệu	11
2. Tổng quan phương pháp	12
3. Kiến trúc và ý tưởng cốt lõi GhostFaceNet	12
4. Huấn luyện và đánh giá thực nghiệm GhostFaceNet	16
5. Tổng kết	18
IV. MÔ HÌNH HÓA KHÔNG GIAN ĐẶC TRƯNG TRONG NHẬN DIỆN KHUÔN MẶT	18
1. Hàm mất mát AdaFace và học biên phân lớp thích nghi.....	18
2. Không gian embedding và độ đo tương đồng.....	19
3. Không gian embedding và độ đo tương đồng.....	19
4. Ngưỡng quyết định và EER (Equal Error Rate)	19
V. QUÁ TRÌNH THỰC NGHIỆM VÀ PHƯƠNG THỨC DEMO ĐỒ ÁN.....	20
1. Hướng tiếp cận	20
2. Quá trình thực nghiệm	21
3. Web face recognition.....	21
VI. KẾT QUẢ	23
1. Kết quả đánh giá mô hình	23
2. Kết quả hệ thống demo.....	24
3. Tổng kết	28
VII. TÀI LIỆU THAM KHẢO	28

Danh mục hình

Hình 1. Sáu vectơ cơ sở đầu tiên của Eigenfaces.	6
Hình 2. Sáu vectơ cơ sở đầu tiên của Fisherfaces.....	6
Hình 3. Mô hình khuôn mặt được xây dựng bởi EBGM	7
Hình 4. Minh họa ảnh trong điều kiện kiểm soát và không kiểm soát của FRGC.	8
Hình 5. Các ví dụ về biến thiên tư thế và ánh sáng trong bộ dữ liệu PIE.....	8
Hình 6. Tổng quan thuật toán KCFA.....	9
Hình 7. Thuật toán CFA.....	9
Hình 8. Các đỉnh tương quan	9
Hình 9. Xây dựng biên quyết định theo mô hình “lớp 2 so với phần còn lại”	10
Hình 10. Tỷ lệ xác thực (Verification Rate – VR)	10
Hình 11. Kết quả siêu độ phân giải khuôn mặt bằng LPP so với các phương pháp khác ..	11
Hình 12. Các feature map của khối đầu tiên VGG16	13
Hình 13. Kiến trúc Ghost bottleneck V1 và V2.....	14
Hình 14. Global Depthwise Convolution (GDC) cải tiến.....	15
Hình 15. ReLU vs. PReLU	15
Hình 16. GhostFaceNetV1-1 và GhostFaceNetV2-1	16
Hình 17. Kiểm tra hai người giống nhau	25
Hình 18. Kiểm tra hai người khác nhau.....	26
Hình 19. Đăng ký hình ảnh	27
Hình 20. Kết quả đăng ký hình ảnh	27
Hình 21. Kiểm tra chức năng search top k.....	28

I. GIỚI THIỆU

1. Đặt vấn đề

Trong kỷ nguyên số, xác thực sinh trắc học đóng vai trò then chốt trong bảo mật và định danh cá nhân. Nhận diện khuôn mặt (Face Recognition) là một trong những phương thức phổ biến nhất nhờ tính tự nhiên và không tiếp xúc. Tuy nhiên, các mô hình Deep Learning hiện đại thường yêu cầu tài nguyên tính toán lớn (như ResNet, Inception), gây khó khăn cho việc triển khai trên các thiết bị biên hoặc hệ thống thời gian thực. Đề án này tập trung nghiên cứu và triển khai hệ thống nhận diện khuôn mặt tối ưu hóa hiệu năng, sử dụng mô hình mạng nơ-ron nhẹ (Lightweight CNN) kết hợp với cơ sở dữ liệu vector tốc độ cao.

2. Mục tiêu đề tài

- Nghiên cứu mô hình GhostFaceNet: Một kiến trúc mạng nhẹ, tối ưu giữa độ chính xác và tốc độ.
- Tìm hiểu thư viện FAISS (Facebook AI Similarity Search) để quản lý và tìm kiếm vector đặc trưng quy mô lớn.
- Xây dựng ứng dụng minh họa trọn vẹn (Full-stack) với khả năng đăng ký, xác thực (1:1) và nhận dạng (1:N) khuôn mặt

3. Phạm vi nghiên cứu

- Input: Ảnh khuôn mặt từ file.
- Model backbone: GhostFaceNet (finetuned với AdaFace)
- Database: Vector database cục bộ sử dụng FAISS.
- Ứng dụng: Web application (ReactJS + FastAPI).

II. TỔNG QUAN NHẬN DẠNG KHUÔN MẶT (FACE RECOGNITION)

1. Giới thiệu (Introduction)

- Các hệ thống Nhận dạng khuôn mặt (Face Recognition) đang có nhu cầu rất lớn nhằm hỗ trợ chống tội phạm và khủng bố, cũng như xác thực người dùng trong kiểm soát truy cập. Tuy nhiên, việc khớp ảnh đầu vào với cơ sở dữ liệu vẫn là một thách thức do sự biến thiên của khuôn mặt dưới các điều kiện vận hành như chiếu sáng, góc quay, biểu cảm, lão hóa và phụ kiện. Hiệu suất thấp trong các kịch bản thực tế ngăn cản việc triển khai rộng rãi, nơi các sai số như Tỷ lệ chấp nhận sai (False Acceptance Rate - FAR) và Tỷ lệ từ chối sai (False Rejection Rate - FRR) cần được tính toán kỹ lưỡng.
- Để cải thiện hiệu suất, nhiều phương pháp tiên tiến đã được đề xuất như Phân tích đặc trưng phụ thuộc lớp hạt nhân (Kernel Class-Dependent Feature Analysis - KCFA), Tensorfaces, các phương pháp học đa tạp (Manifold learning), và các biến thể của Phân tích biệt thức tuyến tính (Linear Discriminant Analysis - LDA). Gần đây, nhận dạng khuôn mặt 3D thu hút sự chú ý nhờ khả năng vượt qua các hạn chế của ảnh 2D về tư thế và ánh sáng, dù yêu cầu sự hợp tác cao hơn từ người dùng. Chương này tập trung chủ yếu vào các phương pháp tiếp cận dựa trên 2D.

2. Các kỹ thuật nhận dạng khuôn mặt (Face Recognition Techniques)

Các thuật toán nhận dạng khuôn mặt có thể chia thành hai nhóm chính: phương pháp dựa trên đặc trưng (Feature-based methods) sử dụng các thuộc tính hình học và phương pháp dựa trên ngoại hình (Appearance-based methods) xem xét các thuộc tính toàn cục của mẫu cường độ ảnh.

- Khuôn mặt riêng (Eigenfaces - PCA)

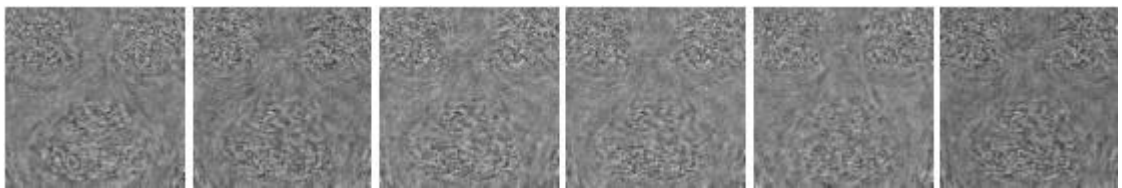
Phương pháp Phân tích thành phần chính (Principal Components Analysis - PCA) tìm kiếm một không gian con tuyến tính có sai số bình phương trung bình tối thiểu để ánh xạ dữ liệu từ không gian N chiều sang không gian đặc trưng M chiều $M \ll N$. Các vector cơ sở PCA tối ưu W được tìm bằng cách cực đại hóa hàm mục tiêu: $W_{PCA} = \arg \max_W |W^T S_T W| = [w_1 w_2 \dots w_m]$. Trong đó S_T là ma trận phân tán tổng thể. PCA tốt cho việc biểu diễn dữ liệu nhưng không nhất thiết tối ưu cho việc phân biệt lớp.



Hình 1. Sáu vector cơ sở đầu tiên của Eigenfaces.

- Phân tích biệt thức tuyến tính (LDA) và Khuôn mặt Fisher (Fisherfaces)

LDA tìm kiếm các hình chiếu giúp phân biệt tốt nhất các lớp khác nhau bằng cách cực đại hóa tỷ lệ giữa ma trận phân tán giữa các lớp S_B và ma trận phân tán trong nội bộ lớp S_W : $W_{LDA} = \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|}$. Để giải quyết vấn đề ma trận S_W bị suy biến, phương pháp Fisherfaces áp dụng PCA trước để giảm chiều, sau đó mới thực hiện LDA trong không gian con PCA: $W_{Fisher} = \arg \max_W \frac{|W^T W_{PCA}^T S_B W_{PCA} W|}{|W^T W_{PCA}^T S_W W_{PCA} W|}$.



Hình 2. Sáu vector cơ sở đầu tiên của Fisherfaces.

- Các biến thể LDA (LDA variants)

Các phương pháp như LDA trực tiếp (Direct LDA - DLDA) và LDA Gram-Schmidt (GSLDA) được thiết kế để khai thác sức mạnh phân biệt trong không gian rỗng (null-space) của ma trận phân tán nội bộ lớp, nơi tỷ lệ Fisher đạt cực đại.

- Phân tích thành phần độc lập (Independent Component Analysis - ICA)

Khác với PCA dựa trên thống kê bậc hai để tìm các đặc trưng không tương quan, ICA tìm kiếm các cơ sở không trực giao sao cho các đặc trưng chuyển đổi độc lập về mặt thống kê, giúp mô hình hóa các mối quan hệ thống kê bậc cao.

- **Phân tích đặc trưng cục bộ (Local Feature Analysis - LFA)**

LFA xây dựng một tập hợp các bộ dò đặc trưng tương quan cục bộ dựa trên sự phân tách không gian con riêng, tạo ra các biểu diễn có khả năng chống lại các thay đổi cục bộ của đối tượng.

- **Khớp đồ thị bó đàn hồi (Elastic Bunch Graph Matching - EBGM)**

EBGM sử dụng các đồ thị hình ảnh để biểu diễn khuôn mặt, trong đó các nút nằm tại các mốc giải phẫu khuôn mặt và sử dụng các hệ số sóng con Gabor (Gabor wavelet) làm đặc trưng cục bộ.



Hình 3. Mô hình khuôn mặt được xây dựng bởi EBGM: (a) đồ thị hình ảnh, (b) các mốc khuôn mặt, (c) đồ thị hình ảnh của ảnh kiểm tra.

- **Mạng thần kinh (Neural Networks - NN) và Máy vector hỗ trợ (Support Vector Machines - SVM)**

NN đòi hỏi khối lượng tính toán lớn khi số lượng người tập huấn tăng lên. Ngược lại, SVM sử dụng "mẹo hạt nhân" (kernel trick) để ánh xạ dữ liệu lên không gian đặc trưng cao chiều, tìm siêu phẳng tối ưu để cực đại hóa biên phân cách, giúp tổng quát hóa tốt với dữ liệu chưa biết.

- **Tensorfaces**

Phương pháp này sử dụng khung đa tuyến tính (multilinear framework) để mô hình hóa các yếu tố biến thiên như danh tính, tư thế, ánh sáng và biểu cảm thông qua Phân tách giá trị đơn lẻ bậc cao (Higher-Order Singular Value Decomposition - HOSVD).

- **Đa tập (Manifolds)**

Các phương pháp học đa tập (như LLE, Laplacianfaces) giả định dữ liệu khuôn mặt nằm trên một đa tập phi tuyến và tập trung vào việc bảo toàn cấu trúc cục bộ thay vì cấu trúc Euclidean toàn cục.

- **Phương pháp hạt nhân (Kernel Methods)**

Bằng cách kết hợp các phương pháp không gian con tuyến tính với mẹo hạt nhân, các thuật toán như KPCA, KDA cho phép biểu diễn các ánh xạ phi tuyến trong không gian cao chiều mà không cần tính toán trực tiếp ánh xạ đặc trưng.

- **Bộ lọc tương quan (Correlation Filters)**

Các bộ lọc tiên tiến như Bộ lọc năng lượng tương quan tối thiểu (Minimum Average Correlation Energy - MACE) được thiết kế để giảm thiểu năng lượng

mặt phẳng tương quan trung bình, tạo ra các đỉnh tương quan nhọn, giúp nhận dạng bền vững với biến thiên ánh sáng và che khuất.

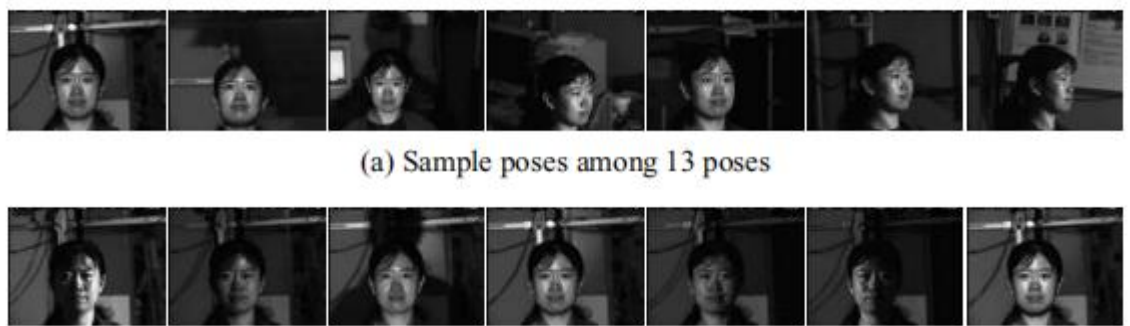
3. Cơ sở dữ liệu (Databases)

Các bộ dữ liệu tiêu chuẩn giúp đánh giá khách quan các thuật toán:

- **FRGC**: Đánh giá hệ thống theo các điều kiện thách thức (kiểm soát và không kiểm soát).



Hình 4. Minh họa ảnh trong điều kiện kiểm soát và không kiểm soát của FRGC.



Hình 5. Các ví dụ về biến thiên tư thế và ánh sáng trong bộ dữ liệu PIE.

- **FERET**: Hỗ trợ phát triển các hệ thống thương mại.
- **PIE (CMU)**: Tập trung vào tư thế (Pose), chiếu sáng (Illumination) và biểu cảm (Expression).
- **AR**: Nổi tiếng với các kịch bản bị che khuất (kính mát, khăn choàng).
- **Yale B**: Chứa các hình ảnh dưới 64 điều kiện chiếu sáng và 9 tư thế.

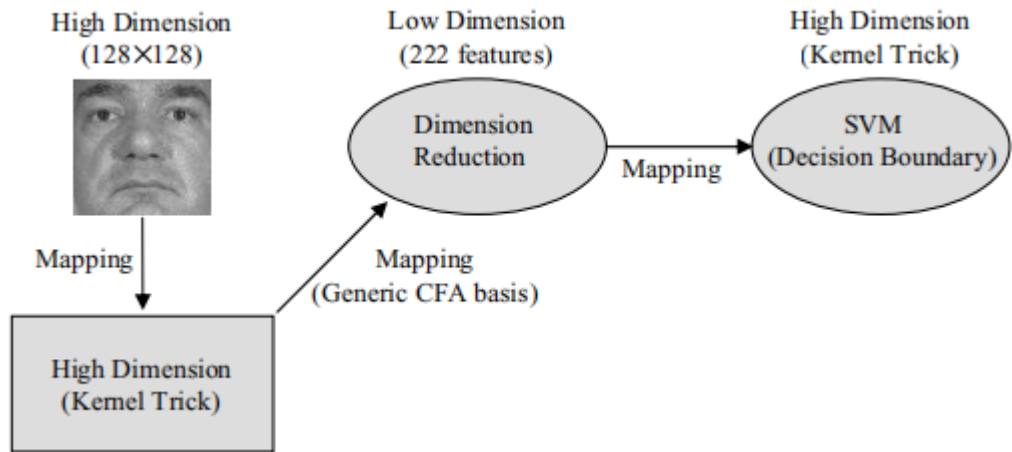
4. Bộ lọc tương quan tiên tiến (Advanced Correlation Filters)

Bộ lọc MACE tối thiểu hóa năng lượng mặt phẳng tương quan trung bình E :

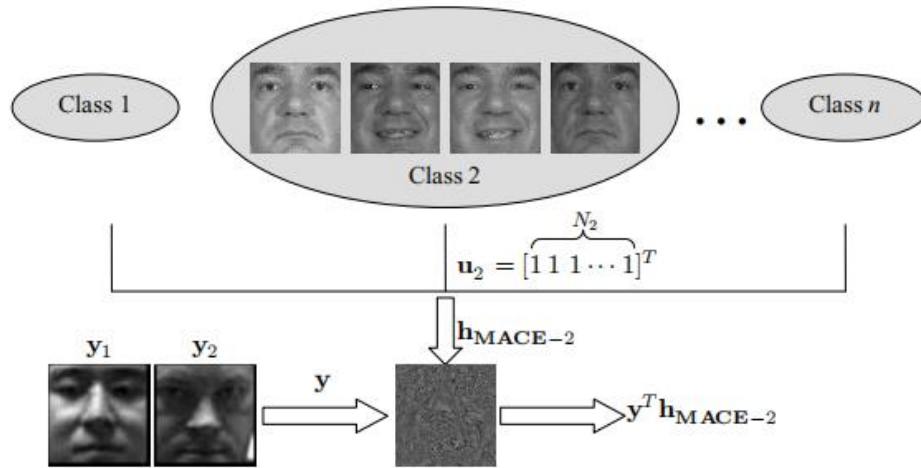
$$E = \sum_{i=1}^N \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} |c_i(x, y)|^2 = h^+ D h$$
 Với ràng buộc đỉnh tương quan tại gốc tọa độ: $X^+ h = c$. Nghiệm đóng cho bộ lọc vector hóa h là: $h = D^{-1} X (X^+ D^{-1} X)^{-1} u$.

- **Phân tích phụ thuộc lớp hạt nhân (KCFA)**

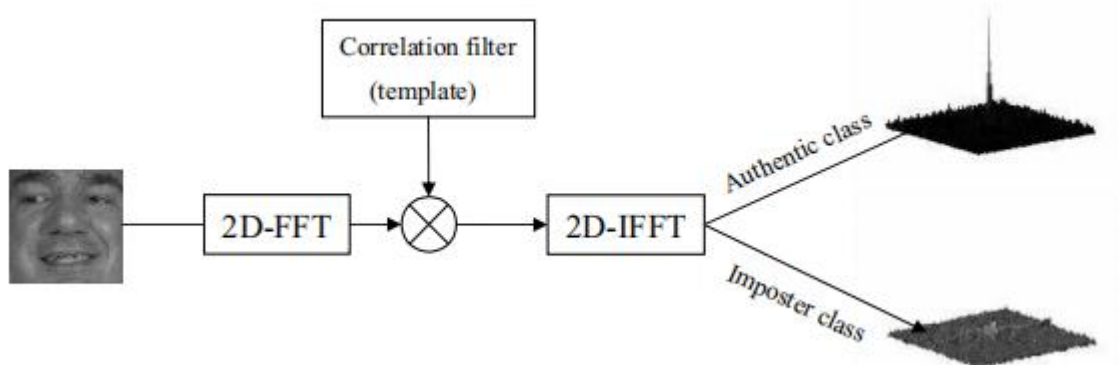
Trong khung làm việc KCFA, một bộ lọc được thiết kế cho mỗi lớp trong tập tập huấn. Ảnh kiểm tra y được đặc trưng bởi các tương quan với n bộ lọc MACE: $c = H^T y = [h_{MACE-1} \dots h_{MACE-n}]^T y$ Đầu ra tương quan hạt nhân được biểu diễn qua hàm hạt nhân K : $\Phi(y) \cdot \Phi(h) = K(y, x'_i) K(x'_i, x'_j)^{-1} u$.



Hình 6. Tổng quan thuật toán KCFA.



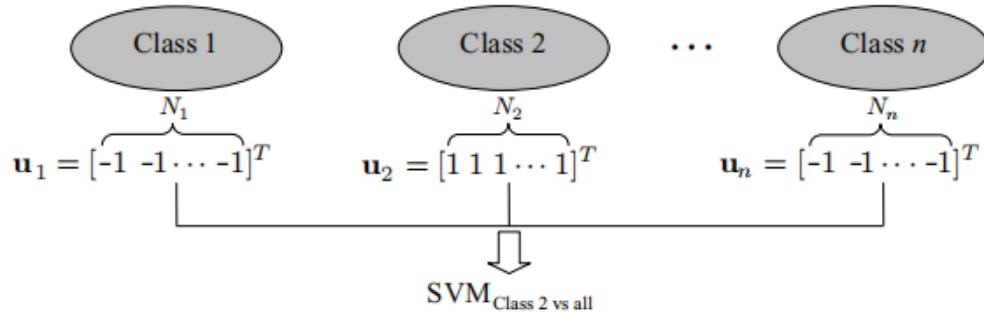
Hình 7. Thuật toán CFA; đáp ứng bộ lọc giữa y_1 và h_{MACE-2} có thể phân biệt rõ ràng so với đáp ứng giữa y_2 và h_{MACE-2} .



Hình 8. Các đỉnh tương quan. Chiều cao của đỉnh biểu thị mức độ tin cậy của phép khớp; do đó, đối với lớp xác thực, một đỉnh cao xuất hiện tại gốc tọa độ với các giá trị tương quan thấp ở phần còn lại của mặt phẳng tương quan. Ngược lại, đối với các lớp giả mạo, không phát hiện được đỉnh tương quan rõ rệt trong mặt phẳng tương quan.

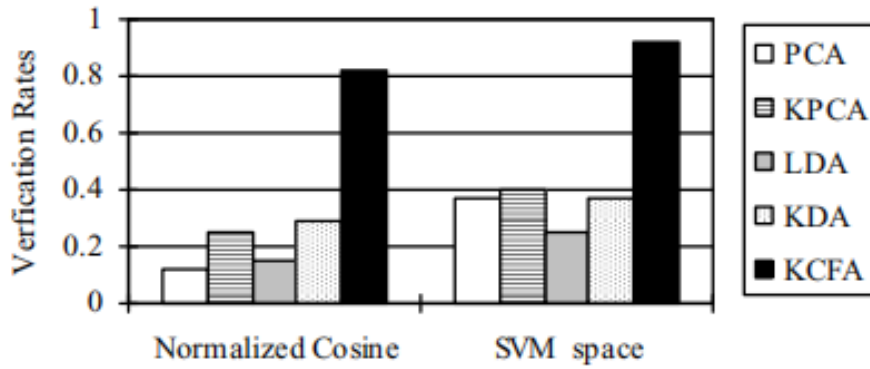
- SVM cho phân loại

Thay vì sử dụng SVM trực tiếp trên pixel ảnh, SVM được huấn luyện trong không gian đặc trưng KCFA (ví dụ: 222 chiều cho dữ liệu FRGC), giúp tăng cường hiệu suất so với các phép đo khoảng cách truyền thống.



Hình 9. Xây dựng biên quyết định theo mô hình “lớp 2 so với phần còn lại”;

u biểu thị vector thông tin gán nhãn lớp và N phụ thuộc vào số lượng ảnh trong mỗi lớp.



Hình 10. Tỷ lệ xác thực (Verification Rate – VR) cho thí nghiệm FRGC 4 với các phương pháp khác nhau, sử dụng khoảng cách cosine chuẩn hóa và không gian SVM (ở mức FAR = 0,1%).

5. Tensorfaces

HOSVD phân tách ma trận tập huấn luyện thành các cơ sở trực giao và giá trị đơn lẻ:

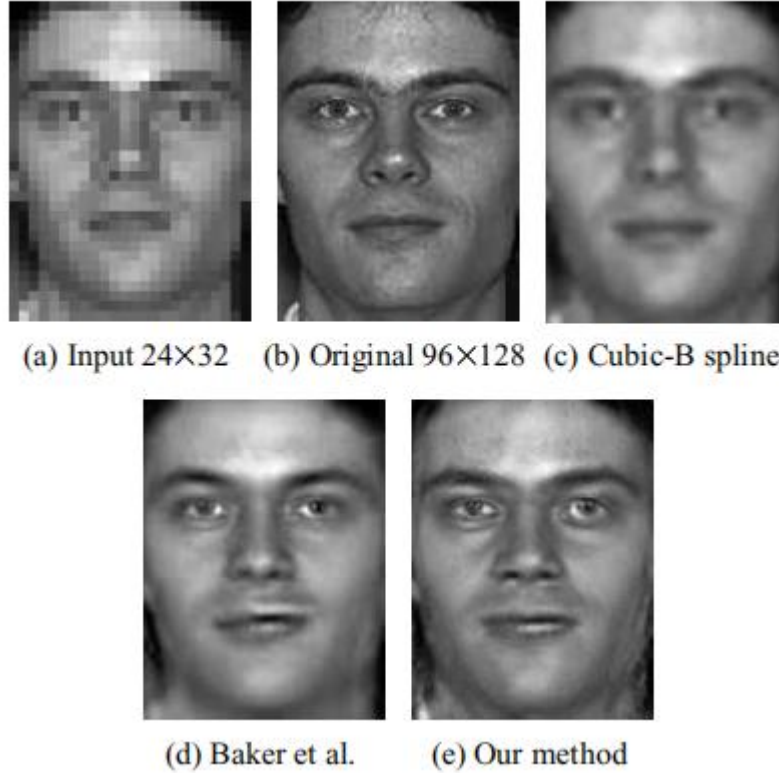
$D_{train} = U_{pixel}Z(U_{people} \otimes U_{light} \otimes U_{pose})^T$ Đối với ảnh kiểm tra chưa biết các yếu tố, việc phân tách được thực hiện thông qua bài toán bình phương tối thiểu có ràng buộc: $\hat{x} = \arg \min_x ||U_{pixel}Zx - d_{test}||^2$ với $||x||^2 = 1$.

6. Mô hình ngoại hình chủ động (AAM) để nhận dạng khuôn mặt

AAM kết hợp mô hình hình dạng (Shape model) và mô hình ngoại hình (Appearance model) bằng PCA. Hình dạng biến dạng được biểu diễn: $s = \bar{s} + V_s p_s$. Ngoại hình (không phụ thuộc hình dạng) được biểu diễn: $A(\bar{s}) = \bar{A}(\bar{s}) + V_A p_A$. AAM được sử dụng để theo dõi các mốc khuôn mặt và hiệu chỉnh tư thế về góc nhìn chính diện (frontal pose), giúp giảm thiểu biến thiên nội lớp.

7. Siêu độ phân giải khuôn mặt sử dụng LPP

Phép chiếu bảo toàn cục bộ (Locality Preserving Projections - LPP) được áp dụng để khôi phục ảnh độ phân giải cao từ ảnh độ phân giải thấp. Ước lượng xác suất hậu nghiệm cực đại (Maximum a posteriori - MAP) được dùng để tìm các hệ số LPP: $y_H^* = (A^T B^T B A + \lambda \Lambda^{-1})^{-1} A^T B^T x_L$.



Hình 11. Kết quả siêu độ phân giải khuôn mặt bằng LPP so với các phương pháp khác.

8. Kết luận (Conclusion)

Các hệ thống nhận dạng khuôn mặt mạnh mẽ phải xử lý được sự biến thiên trong kịch bản thực tế. Thuật toán KCFA cho thấy triển vọng lớn với khả năng giảm chiều hiệu quả, đạt tỷ lệ xác thực 82,4% tại mức 0,1% FAR trên dữ liệu FRGC. Việc sử dụng các kỹ thuật như AAM để hiệu chỉnh tư thế và LPP để tăng độ phân giải là các bước tiền xử lý quan trọng giúp nâng cao độ chính xác của các bộ phân loại truyền thống.

III. PHƯƠNG PHÁP GHOSTFACENETS CHO BÀI TOÁN NHẬN DẠNG KHUÔN MẶT

1. Giới thiệu

Trong những năm gần đây, các thiết bị di động như điện thoại thông minh và máy tính xách tay đã trở thành nền tảng chính để truy cập thông tin trong cuộc sống. Sự phổ biến này đồng thời đặt ra nhiều thách thức về bảo mật thông tin. Trong bối cảnh đó, các bài toán thị giác máy tính như phát hiện, nhận dạng và xác thực khuôn mặt (FD - Face Detection, FR - Face Recognition, FV - Face Verification) ngày càng

được sử dụng rộng rãi như một phương thức xác thực sinh trắc học trên thiết bị di động. Nhờ sự phát triển của học sâu, các mô hình FR/FV hiện nay đã đạt hiệu năng vượt trội so với các phương pháp truyền thống. Tuy nhiên, việc triển khai các mô hình học sâu này trên thiết bị di động vẫn gặp nhiều hạn chế do yêu cầu tài nguyên tính toán lớn và độ trễ thấp.

Để giải quyết vấn đề này, nhóm đề xuất GhostFaceNets (được công bố năm 2023) - một họ kiến trúc nhẹ mới cho bài toán nhận dạng khuôn mặt, với độ phức tạp tính toán rất thấp (chỉ từ 60 đến 275 MFLOPs), đồng thời đạt hiệu năng SOTA trên nhiều bộ benchmark phổ biến.

2. Tổng quan phương pháp

GhostFaceNet là một phương pháp nhận dạng và xác thực khuôn mặt tiên tiến, được thiết kế với mục tiêu đạt độ chính xác cao trong khi vẫn đảm bảo chi phí tính toán thấp, phù hợp cho các thiết bị có tài nguyên hạn chế như điện thoại di động, hệ thống nhúng, Phương pháp này kế thừa ý tưởng cốt lõi của GhostNet: giảm sự dư thừa của feature map thông qua Ghost module, từ đó làm giảm đáng kể số lượng phép toán và tham số mà vẫn giữ được khả năng biểu diễn đặc trưng hiệu quả.

GhostFaceNet được tích hợp nhiều kỹ thuật cải tiến hiện đại nhằm nâng cao hiệu năng nhận dạng, bao gồm:

- Sử dụng GhostNetV1 hoặc GhostNetV2 làm backbone trích xuất đặc trưng.
- Thiết kế lại head của tầng đầu ra, được gọi là Global Depthwise Convolution (GDC) cải tiến cho phép mô hình học được mức độ quan trọng khác nhau của các vùng trên khuôn mặt.
- Thay thế hàm kích hoạt ReLU bằng PReLU để tăng khả năng biểu diễn phi tuyến. Điều chỉnh mô-đun Squeeze-and-Excitation (SE) nhằm nâng cao khả năng phân biệt (discriminative power).
- Sử dụng hàm mất mát ArcFace giúp cải thiện độ gom cụm trong cùng lớp và độ tách biệt giữa các lớp.

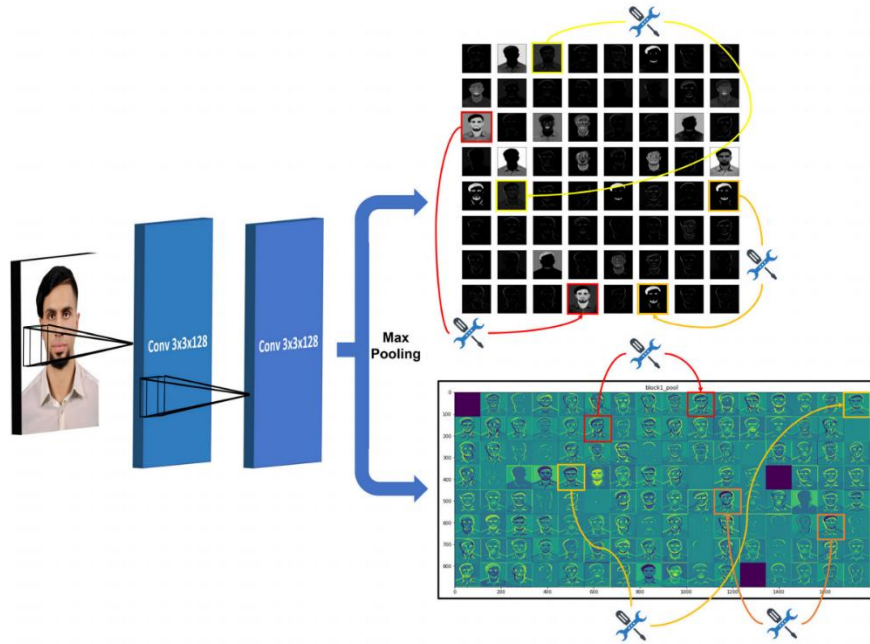
3. Kiến trúc và ý tưởng cốt lõi GhostFaceNet

Nền tảng: Ghost Module và GhostNet

a) Ghost Module và sự dư thừa feature map

GhostFaceNet được xây dựng dựa trên nền tảng GhostNet, trong đó thành phần cốt lõi là Ghost module – một cơ chế nhằm giải quyết vấn đề dư thừa feature map trong các mạng CNN truyền thống.

Trong mạng CNN các tác giả nhận thấy có nhiều feature map rất giống nhau (được gọi là "ghosts").



Hình 12. Các feature map của khối đầu tiên VGG16 cho thấy nhiều bản đồ rất giống nhau (đánh dấu cùng màu)

Do đó thay vì sử dụng toàn bộ convolution chuẩn để tạo ra tất cả feature map, Ghost module tạo ra tất cả feature map này bằng cách:

- Chỉ sử dụng một phần nhỏ các convolution chuẩn để học ra các đặc trưng cốt lõi (intrinsic features).
- Các feature map còn lại được sinh ra từ các đặc trưng cốt lõi thông qua các phép biến đổi tuyến tính chi phí thấp, chủ yếu là depthwise convolution.

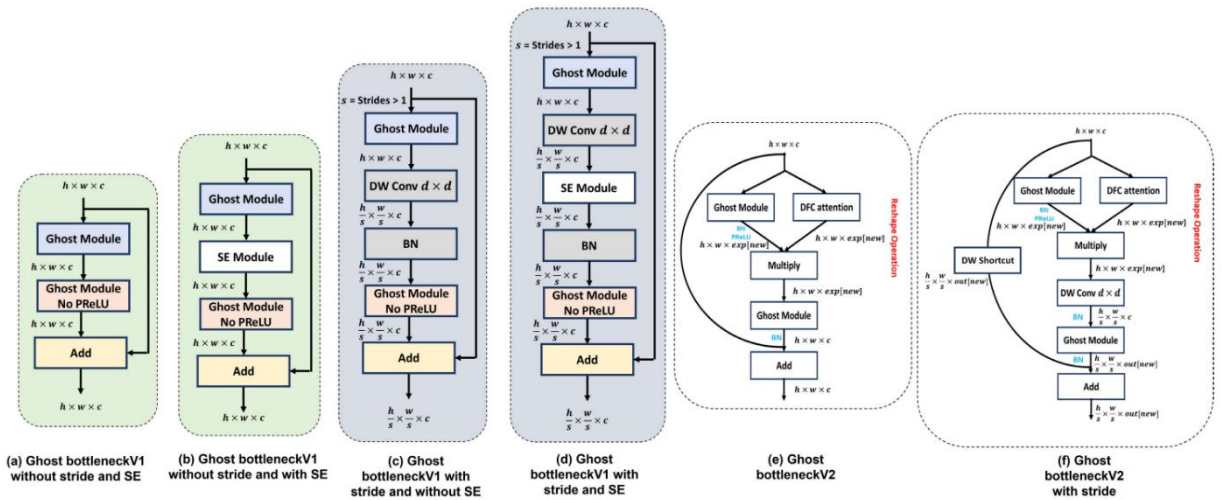
Nhờ cách tiếp cận này, Ghost module tạo ra số lượng feature map tương đương với convolution truyền thống nhưng với số tham số và FLOPs thấp hơn đáng kể, trong khi vẫn duy trì hiệu năng của mô hình.

b) Kiến trúc GhostNetV1 và GhostNetV2

Dựa trên Ghost module, một kiến trúc backbone mới có tên GhostNetV1 đã được đề xuất. GhostNetV1 thực chất là một phiên bản chỉnh sửa của MobileNetV3, trong đó Ghost bottleneckV1 (được cấu thành từ các Ghost module) được sử dụng để thay thế bottleneck gốc. Sau lớp đầu vào (input layer - một lớp convolution thông thường), GhostNetV1 được xây dựng bằng cách xếp chồng các Ghost bottleneckV1, với số channel của tensor tăng dần theo từng giai đoạn.

GhostNetV2 là phiên bản nâng cấp mở rộng Ghost module của GhostNetV1 bằng cách tích hợp khả năng nắm bắt các phụ thuộc dài hạn (long-range dependencies) thông qua một lớp attention mới có tên là nhánh attention DFC (DFC attention branch). Nhánh này sử dụng các phép convolution để tạo ra bản đồ attention có receptive field toàn cục. Trên cơ sở đó, một bottleneck mới được thiết kế với hai nhánh song song: Ghost module và DFC attention module, cùng nhận chung đầu

vào, trích xuất các loại thông tin khác nhau và sau đó được tổng hợp lại để tạo đặc trưng đầu ra giàu thông tin hơn.



Hình 13. Kiến trúc Ghost bottleneck V1 và V2

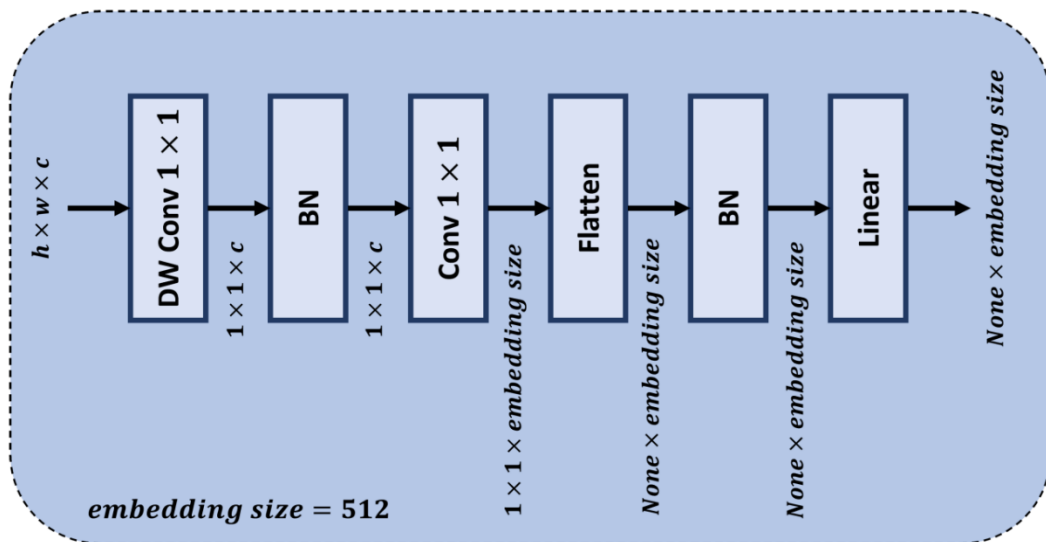
Kiến trúc GhostFaceNet - Những cải tiến mới

Từ các backbone GhostNetV1 và GhostNetV2, nhóm tác giả đề xuất GhostFaceNet – một họ mô hình được tối ưu riêng cho bài toán nhận dạng và xác thực khuôn mặt (FR/FV).

Các cải tiến chính trong GhostFaceNet bao gồm:

- Thiết kế lại lớp đầu ra (Modified GDC Head)

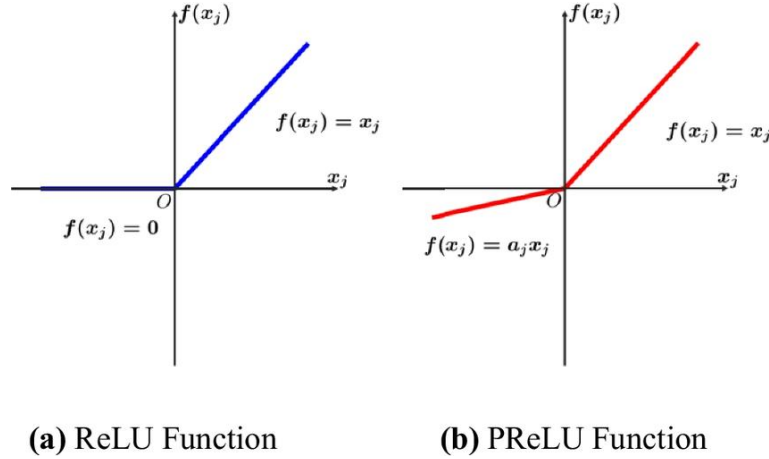
Thay vì sử dụng Global Average Pooling (GAP) GhostFaceNet sử dụng Global Depthwise Convolution (GDC) cải tiến cho phép gán mức độ quan trọng khác nhau cho các vị trí trong feature map (tập trung nhiều hơn vào các vùng đặc trưng quan trọng trên khuôn mặt như mắt, mũi và tai). Từ đó tạo ra vector đặc trưng khuôn mặt có khả năng phân biệt cao hơn, phù hợp với FR/FV.



Hình 14. Global Depthwise Convolution (GDC) cải tiến

- Thay thế hàm kích hoạt ReLU bằng PReLU

PReLU (Parametric Rectified Linear Unit) cho phép các giá trị kích hoạt âm, giúp mạng học được các hàm phi tuyến phức tạp hơn và loại bỏ vấn đề triệt tiêu đạo hàm (vanishing gradient), từ đó nâng cao hiệu suất so với ReLU.



Hình 15. ReLU vs. PReLU

- Điều chỉnh mô-đun Squeeze and Excitation (SE)

Trong GhostFaceNet, các lớp Fully Connected của mô-đun Squeeze-and-Excitation (SE) truyền thống được thay thế bằng convolution 1×1 . Cách thiết kế này giúp tăng cường channel-wise attention với chi phí tính toán gần như không đổi, cho phép mạng tập trung tốt hơn vào các đặc trưng quan trọng và giảm ảnh hưởng của các đặc trưng nhiễu.

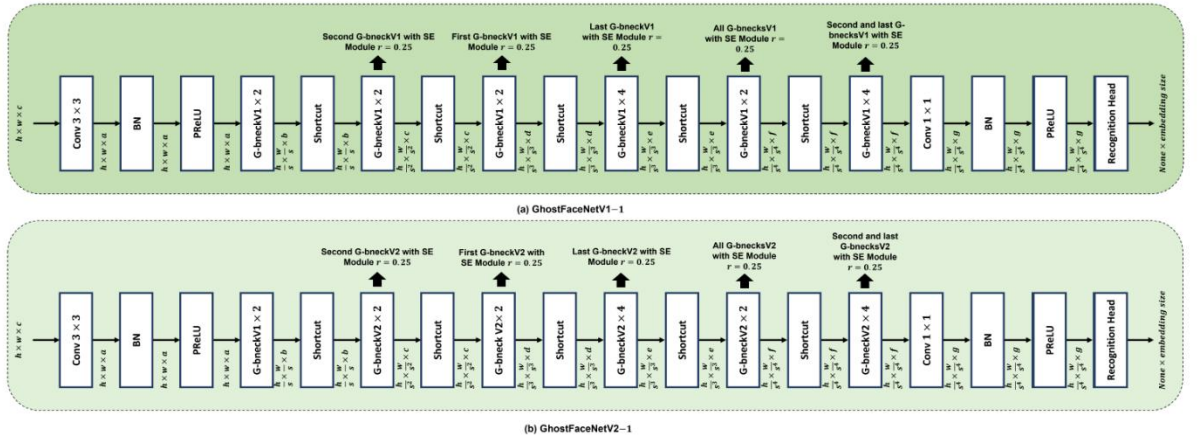
- Hàm mất mát ArcFace

GhostFaceNet sử dụng hàm mất mát ArcFace để huấn luyện vì ArcFace giảm thiểu khoảng cách trong cùng lớp và thể hiện sự phân biệt rõ ràng giữa các lớp vượt trội hơn các hàm khác.

Bên cạnh việc cải tiến kiến trúc, nhóm tác giả còn tạo ra nhiều phiên bản bằng cách tinh chỉnh các siêu tham số như stride và width multiplier thông qua ablation study, từ đó chọn ra mô hình có sự cân bằng tốt nhất giữa độ chính xác và chi phí tính toán.

Model	α	Training Dataset	Complexity FLOPs	# Param s.	LFW (%)	AgeDB_30 (%)	CFP_FP (%)
GhostFace NetV2-2	1	MS1 MV2	47,371,312	4,229,744	95.65	96.212	92.9
GhostFace NetV2-2	1.1	MS1 MV2	56,262,058	5,033,078	99.6	95.683	91.814
GhostFace NetV2-2	1.2	MS1 MV2	66,543,012	5,917,412	99.633	95.9	92.686
GhostFace NetV2-2	1.3	MS1 MV2	76,513,640	6,839,104	99.717	96.55	93.071
GhostFace NetV2-2	1.4	MS1 MV2	88,378,922	7,847,302	99.683	95.983	91.986
GhostFace NetV2-2	1.5	MS1 MV2	100,386,924	8,927,420	99.7	96.483	93.9
GhostFace NetV2-2	1	MS1 MV3	47,371,312	4,229,744	99.667	96.417	94.857
GhostFace NetV2-2	1.1	MS1 MV3	56,262,058	5,033,078	99.667	96.167	92.471
GhostFace NetV2-2	1.2	MS1 MV3	66,543,012	5,917,412	99.683	96.383	94.343
GhostFace NetV2-2	1.3	MS1 MV3	76,513,640	6,839,104	99.683	96.833	94.286
GhostFace NetV2-2	1.4	MS1 MV3	88,378,922	7,847,302	99.683	96.417	94.629
GhostFace NetV2-2	1.5	MS1 MV3	100,386,924	8,927,420	99.667	96.417	94.857

Kiến trúc tổng quan của GhostFaceNet



Hình 16. GhostFaceNetV1-1 (hình a) dùng G-bneckV1 (Ghost bottleneckV1) GhostFaceNetV2-1 (hình b): dùng G-bneckV2 (Ghost bottleneckV2)

4. Huấn luyện và đánh giá thực nghiệm GhostFaceNet

a) Tập dữ liệu huấn luyện (Datasets)

Các mô hình GhostFaceNet được huấn luyện trên hai tập dữ liệu lớn và phổ biến trong lĩnh vực nhận dạng khuôn mặt: MS1MV2 và MS1MV3, là các phiên bản đã được làm sạch của tập MS-Celeb-1M. MS1MV2 bao gồm khoảng 5.8 triệu ảnh khuôn mặt của 85.000 danh tính, trong khi MS1MV3 chứa khoảng 5.1 triệu ảnh của 91.000 danh tính.

b) Các benchmark và tiêu chí đánh giá

Hiệu năng của GhostFaceNet được đánh giá trên nhiều bộ benchmark chuẩn, bao gồm LFW, AgeDB-30, CFP-FP, CFP-FF, CP-LFW, CA-LFW, VGG2-FP, IJB-B, IJB-C và MegaFace. Các thí nghiệm xác thực được đánh giá bằng cách tính Cosine distance giữa các vector embedding.

c) Kết quả thực nghiệm và so sánh với các mô hình SOTA

Kết quả thực nghiệm cho thấy GhostFaceNet đạt hiệu năng vượt trội trên hầu hết các benchmark, trong khi vẫn duy trì độ phức tạp tính toán rất thấp. Các mô hình GhostFaceNet chỉ yêu cầu từ khoảng 60 đến 275 MFLOPs, thấp hơn đáng kể so với nhiều mô hình nhẹ SOTA trước đây như MobileFaceNet hay VarGFaceNet.

Method	#Params. (M)	FLOPs (M)	Model Size (MB)	LFW (%)	CA-LFW (%)	CP-LFW (%)	CFP-FP (%)	CFP-FF (%)	AgeDB _30 (%)	IJB-B (%)	IJB-C (%)
MobileFaceNetV1 [8]	3.4	1100	13.1	99.40	94.47	87.17	95.80	99.50	96.40	92.00	93.90
PocketNetM-256 [11]	1.75	1099.15	-	99.58	95.63	90.03	95.66	-	97.17	90.74	92.70
PocketNetM-128 [11]	1.68	1099.02	-	99.65	95.67	90.00	95.07	-	96.78	90.63	92.63
ShuffleFaceNet 2× [21]	4.5	1050	18	99.62	-	-	97.56	-	97.28	-	-
VarGFaceNet [22]	5	1022	20	99.85	95.15	88.55	98.50	99.50	98.15	92.90	94.70
AirFace [36]	-	1000	-	99.27	-	-	94.11	-	93.25	-	-
MobileFaceNet [8]	2	933.3	4	99.70	95.20	89.22	96.90	99.60	97.60	92.80	94.70
ProxylessFaceNAS [8]	3.2	900	12.5	99.20	92.55	84.17	94.70	98.80	94.40	87.10	89.70
MixFaceNet-M [23]	3.95	626.1	-	99.68	-	-	-	-	97.05	91.55	93.42
ShuffleMixFaceNet-M [23]	3.95	626.1	-	99.60	-	-	-	-	96.98	91.47	91.47
PocketNetS-256 [11]	0.99	587.224	3.9	99.66	95.50	88.93	93.34	-	96.35	89.31	91.33
PocketNetS-128 [11]	0.92	587.11	3.7	99.58	95.48	89.63	94.21	-	96.10	89.44	91.62
ShuffleFaceNet 1.5× [21]	2.6	577.5	10.5	99.67	95.05	88.50	97.26	-	97.32	92.30	94.30
MixFaceNet-S [23]	3.07	451.7	-	99.60	-	-	-	-	96.63	90.17	92.30
ShuffleMixFaceNet-S [23]	3.07	451.7	-	99.58	-	-	-	-	97.05	90.94	93.08
MobileFaceNets [20]	0.99	439.8	8.2	99.55	-	-	-	-	96.07	-	-
ShuffleFaceNet 1× [21]	1.4	275.8	5.6	99.45	-	-	96.04	-	96.33	-	-
GhostFaceNetV2-1 (MS1MV3) (ours)	6.88	272.105	13.743	99.8667*	96.1167*	94.65*	99.33*	99.9143*	98.62*	96.48*	97.75*
GhostFaceNetV2-1 (MS1MV2) (ours)	6.88	272.105	13.743	99.85	96.086	92.93	98.9143	99.84	98.5833	95.745	97.015
GhostFaceNetV1-1 (MS1MV3) (ours)	4.09	215.658	8.17	99.73	95.93	91.93	96.83	99.81	98	93.12	94.94
GhostFaceNetV1-1 (MS1MV2) (ours)	4.09	215.658	8.17	99.77	95.88	90.47	95.64	99.80	97.58	92.19	94.06
MixFaceNet-XS [23]	1.04	161.9	-	99.60	-	-	-	-	95.85	88.48	90.73
ShuffleMixFaceNet-XS [23]	1.04	161.9	-	99.53	-	-	-	-	95.62	87.86	90.43
GhostFaceNetV2-2 (MS1MV3) (ours)	6.84	76.513	13.663	99.683	95.733	90.1667	94.2857	99.6443	96.833	91.889	93.159
GhostFaceNetV2-2 (MS1MV2) (ours)	6.84	76.513	13.663	99.7167	95.7	89.5833	93.07	99.5857	96.55	91.7624	93.0324
ShuffleFaceNet 0.5× [21]	0.5	66.9	1.9	99.23	-	-	92.59	-	93.22	-	-
GhostFaceNetV1-2 (MS1MV3) (ours)	4.06	60.296	8.07	99.68	95.60	90.07	93.31	99.64	96.92	91.25	93.45
GhostFaceNetV1-2 (MS1MV2) (ours)	4.06	60.296	8.07	99.65	95.53	88.57	91.51	99.56	96.18	90.53	92.66

Kết quả thực nghiệm

d) Đánh giá hiệu năng triển khai thực tế

Để đánh giá khả năng ứng dụng thực tế, nhóm tác giả đã sử dụng công cụ TFLite để đánh giá tốc độ suy luận thực tế (real inference speed) của các mô hình GhostFaceNets được đề xuất trên điện thoại di động sử dụng kiến trúc ARM. Kết

qua cho thấy, ở chế độ single-thread, mô hình có kích thước nhỏ (dưới 14 MB) và thời gian suy luận rất ngắn (khoảng 50 ms). Do đó có thể triển khai hiệu quả trong các hệ thống nhận dạng khuôn mặt thời gian thực trên thiết bị di động và hệ thống nhúng.

Model	Inference Time (ms)	Model Size (MB)
GhostFaceNet V1-1	46	8.17
GhostFaceNet V1-2	22	8.07
GhostFaceNet V2-1	52.4	13.706
GhostFaceNet V2-2	28.4	13.606

Kết quả triển khai thực tế

5. Tổng kết

Dựa trên các kết quả thực nghiệm thu được, đặc biệt là trong các bài đánh giá hiệu năng thực tế, có thể khẳng định rằng GhostFaceNet là một phương pháp hiệu quả, tiên tiến và phù hợp cho bài toán nhận dạng và xác thực khuôn mặt trên môi trường tài nguyên hạn chế.

Từ những phân tích và kết quả thực nghiệm trên, nhóm đề xuất lựa chọn GhostFaceNet làm phương pháp triển khai cho hệ thống nhận dạng khuôn mặt trong đồ án. Việc lựa chọn này không chỉ dựa trên hiệu năng cao đạt được trên các benchmark chuẩn, mà còn dựa trên tính khả thi khi triển khai trong thực tế, khả năng mở rộng thông qua điều chỉnh siêu tham số, cũng như sự cân bằng hợp lý giữa độ chính xác, tốc độ và chi phí tài nguyên.

- **Github tác giả:** <https://github.com/HamadYA/GhostFaceNets>

IV. MÔ HÌNH HÓA KHÔNG GIAN ĐẶC TRƯNG TRONG NHẬN DIỆN KHUÔN MẶT

1. Hàm mất mát AdaFace và học biên phân lớp thích nghi

Trong bài toán nhận diện khuôn mặt, mục tiêu của quá trình huấn luyện là học được một ánh xạ

$$f: R^{H \times W \times 3} \rightarrow R^d, \quad d = 512$$

với $d = 512$, sao cho các mẫu thuộc cùng một danh tính có embedding gần nhau, trong khi các mẫu thuộc các danh tính khác nhau được tách biệt rõ ràng. AdaFace là hàm mất mát thuộc nhóm margin-based softmax, mở rộng từ ArcFace và CosFace, trong đó biên phân lớp không cố định mà được điều chỉnh theo chất lượng đặc trưng. Công thức tổng quát có dạng:

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s \cdot \cos(\theta_{y_i} + m_i)}}{e^{s \cdot \cos(\theta_{y_i} + m_i)} + \sum_{j \neq y_i} e^{s \cdot \cos(\theta_j)}}$$

Trong đó: θ_{y_i} là góc giữa embedding của mẫu thứ i và vector trọng số của lớp đúng. m_i là margin thích nghi, phụ thuộc vào chất lượng embedding. s là hệ số scale. AdaFace điều chỉnh m_i dựa trên độ lớn chuẩn hóa của embedding, giúp: Tăng margin cho các mẫu có chất lượng cao (ảnh rõ, ít nhiễu). Giảm margin cho các mẫu có chất lượng thấp, tránh làm mô hình học sai lệch. Nhờ vậy, biên phân tách trong không gian đặc trưng ổn định hơn và khả năng tổng quát hóa được cải thiện.

2. Không gian embedding và độ đo tương đồng

Sau khi qua mạng GhostFaceNet, mỗi khuôn mặt được ánh xạ thành một vector đặc trưng $z \in \mathbb{R}^{512}$, sau đó được chuẩn hóa theo chuẩn L2:

$$\hat{z} = \frac{z}{|z|_2}$$

Việc chuẩn hóa đưa các embedding lên mặt cầu đơn vị, giúp so sánh dựa trên góc giữa các vector. Độ tương đồng Cosine giữa hai khuôn mặt x và y được tính:

$$\text{sim}(x, y) = \cos(\theta) = \frac{\hat{z}_x \cdot \hat{z}_y}{|\hat{z}_x| |\hat{z}_y|}$$

Vì các vector đã được chuẩn hóa, biểu thức rút gọn thành tích vô hướng. Giá trị càng tiến gần 1 thì hai khuôn mặt càng có khả năng thuộc cùng một danh tính.

3. Không gian embedding và độ đo tương đồng

Bài toán tìm kiếm lân cận trong không gian vector và FAISS. Trong bài toán nhận dạng 1:N, ta cần tìm:

$$\text{argmax}_i \text{sim}(q, z_i)$$

với q là embedding truy vấn và $\{z_i\}$ là tập embedding trong cơ sở dữ liệu. Khi số lượng mẫu lớn, việc tính toán tuần tự sẽ có độ phức tạp $O(Nd)$, không phù hợp cho hệ thống thời gian thực. FAISS sử dụng các cấu trúc chỉ mục như IVF, HNSW hoặc PQ để xấp xỉ bài toán Nearest Neighbor, giảm độ phức tạp và cho phép tìm kiếm Top-K nhanh chóng trong không gian chiều cao. FAISS hỗ trợ cả khoảng cách Cosine (thông qua inner product) và L2, đồng thời có khả năng tăng tốc bằng GPU.

4. Ngưỡng quyết định và EER (Equal Error Rate)

Trong bài toán xác thực khuôn mặt (1:1) hoặc nhận dạng (1:N), sau khi tính độ tương đồng Cosine $\text{sim}(x, y) = \cos(\theta)$, hệ thống cần một ngưỡng τ để quyết định hai embedding có thuộc cùng một danh tính hay không:

$$\text{same}(x, y) = 1 \text{ if } \text{sim}(x, y) \geq \tau, 0 \text{ otherwise}$$

Việc lựa chọn ngưỡng τ ảnh hưởng trực tiếp đến hai loại sai số:

- False Accept Rate (FAR): tỉ lệ chấp nhận nhầm hai người khác nhau.
- False Reject Rate (FRR): tỉ lệ từ chối nhầm hai ảnh của cùng một người.

Khi thay đổi τ , FAR và FRR sẽ thay đổi ngược chiều nhau. Một chỉ số thường dùng để đánh giá chất lượng không gian embedding là Equal Error Rate (EER) - giá trị tại đó: $\text{FAR}(\tau_{\text{EER}}) = \text{FRR}(\tau_{\text{EER}})$. Ngưỡng τ_{EER} được xem như ngưỡng cân bằng giữa an toàn (giảm FAR) và tiện dụng (giảm FRR). EER càng nhỏ chứng tỏ: Các embedding cùng danh tính được gom chặt (intra-class compactness cao). Các embedding khác danh tính được tách xa (inter-class separability lớn). Hàm mất mát AdaFace với margin thích nghi giúp phân bố điểm cosine của positive pairs và negative pairs tách biệt rõ hơn, từ đó làm giảm EER.

Trong hệ thống nhận dạng 1:N sử dụng FAISS, sau khi tìm được Top-K vector gần nhất với truy vấn q , ngưỡng τ vẫn được áp dụng:

$$\max_i \text{sim}(q, z_i) \geq \tau \Rightarrow \text{accept}, \quad \max_i \text{sim}(q, z_i) < \tau \Rightarrow \text{unknown}$$

Ngưỡng này thường được chọn dựa trên τ_{EER} hoặc điều chỉnh theo yêu cầu an ninh của bài toán thực tế.

V. QUÁ TRÌNH THỰC NGHIỆM VÀ PHƯƠNG THỨC DEMO ĐỒ ÁN

1. Hướng tiếp cận

- **Mục tiêu:** Nhóm chọn hướng tiếp cận sử dụng pretrained model **GhostFaceNetsV1.3** và thực hiện finetune mô hình trên bộ dataset **WebFace-112x112** và đề xuất hướng đánh giá dựa trên benchmark là các tập LFW (Labeled Faces in the Wild), CFP_FP (Celebrities Frontal-Profile), AgeDB. Bên cạnh đó, VN-Celeb được nhóm đề xuất đánh giá để theo dõi kết quả với của mô hình sau huấn luyện WebFace. Việc finetune trên bộ dữ liệu WebFace mục đích để tăng độ chính xác nhận diện khuôn mặt người châu Á.
- **Cải tiến:** AdaFace là một hàm mất mát thuộc nhóm margin-based softmax, được công bố tại hội nghị Conference on Computer Vision and Pattern Recognition (CVPR). Trong khi các phương pháp trước đó như ArcFace và CosFace sử dụng biên phân lớp (margin) cố định cho toàn bộ tập dữ liệu, giả định rằng mọi ảnh huấn luyện đều có chất lượng tương đương, thì giả định này không còn phù hợp trong các bài toán nhận diện khuôn mặt thực tế, nơi dữ liệu thường chịu ảnh hưởng mạnh của điều kiện chiếu sáng, độ phân giải, góc chụp, che khuất và nhiễu. AdaFace khắc phục hạn chế này bằng cách thiết kế một cơ chế margin thích nghi, trong đó giá trị margin được điều chỉnh động theo chất lượng của từng embedding, ước lượng thông qua độ lớn của vector đặc trưng trước khi chuẩn hóa. Các mẫu có chất lượng cao sẽ được gán margin lớn hơn nhằm tăng cường độ phân tách giữa các lớp, trong khi các mẫu có chất lượng thấp chỉ chịu margin nhỏ hơn để tránh làm sai lệch quá trình tối ưu. Nhờ đó, biên quyết định trong không gian đặc trưng trở nên ổn định, cân bằng giữa các mẫu dễ và khó,

đồng thời cải thiện khả năng tổng quát hóa của mô hình. Vì lý do này, trong đồ án của nhóm, AdaFace được lựa chọn thay thế ArcFace nhằm nâng cao độ chính xác và độ bền vững của hệ thống trong điều kiện dữ liệu không đồng nhất của các bài toán xác thực và nhận dạng khuôn mặt ngoài thực tế.

2. Quá trình thực nghiệm

- Bộ dữ liệu & Tiền xử lý:

- Bộ dữ liệu: WebFace (đã căn chỉnh về kích thước 112x112). Gồm 10572 danh tính và ~500k ảnh.
- Lọc dữ liệu: Các danh tính (identity) có ít hơn 5 hình ảnh sẽ bị loại bỏ và bộ dữ liệu từ 10572 danh tính đã bị loại bỏ 3 danh tính còn 10569 danh tính.
- Tiền xử lý: Giải mã hình ảnh và thay đổi kích thước về 112x112. Chuẩn hóa: $(\text{img} - 127.5) * 0.0078125$ (Đưa về khoảng giá trị $[-1, 1]$).

- Kiến trúc mô hình:

- Backbone: GhostNetV1 (CNN siêu nhẹ).
- Kích thước Embedding: 512 chiều.
- Trọng số Pretrained: Ban đầu tải weight từ GhostFaceNet_W1.3_S1.

- Cấu hình Huấn luyện: Quá trình huấn luyện được chia làm 2 giai đoạn sử dụng hàm mất mát AdaFace. Batch size **128**, learning rate cơ bản 0.00025.

- Hàm mất mát AdaFace tham số: $s=64.0$, $m=0.4$, $h=0.333$, $\text{eps}=1e-3$. Mục đích: Adaptive margin loss giúp xử lý tốt hơn các ảnh chất lượng thấp so với ArcFace.
- Giai đoạn 1: Huấn luyện Head (Đóng băng Backbone), mục tiêu thích nghi lớp phân loại (AdaFace) với bộ dữ liệu mới trong khi giữ nguyên bộ trích xuất đặc trưng (feature extractor). Bộ tối ưu hóa: SGD ($\text{learning_rate}=1e-3$, $\text{momentum}=0.9$). Epochs: 4 Loss: Hội tụ từ **25.37** xuống **8.27**.
- Giai đoạn 2: Tinh chỉnh (Mở đóng băng) mục tiêu tinh chỉnh bộ trích xuất đặc trưng cho miền dữ liệu cụ thể. Cấu hình mở đóng băng (Unfreeze) 30 lớp cuối của backbone. Giảm Learning Rate $0.1 * \text{BASE_LR}$. Epochs 6 Loss hội tụ từ **7.98** xuống **7.12**.

3. Web face recognition

- Tổng quan Hệ thống: Hệ thống được xây dựng nhằm mục đích cung cấp giải pháp xác thực (Verification - 1:1) và nhận diện (Identification - 1:N) khuôn mặt với độ chính xác cao và tốc độ phản hồi nhanh. Hệ thống sử dụng kiến trúc Client-Server, với Backend mạnh mẽ xử lý và Frontend hiện đại cho trải nghiệm người dùng.

- Kiến trúc công nghệ:

- Backend:
 - Framework: FastAPI (Python) - Hiệu năng cao, hỗ trợ bất đồng bộ (async).
 - AI/Deep Learning: TensorFlow 2.15, TF-Keras.
 - Computer Vision: OpenCV, Pillow.

- Face Detection & Alignment: MTCNN (Multi-task Cascaded Convolutional Networks).
- Feature Extraction Model: GhostFaceNet (tối ưu hóa cho thiết bị biên/nhẹ) với hàm loss AdaFace.
- Vector Database: FAISS (Facebook AI Similarity Search) để tìm kiếm vecto tốc độ cao.
- Frontend:
 - Framework: Vite + React (dựa trên cấu trúc thư mục src và vite.config.ts).
 - Language: TypeScript/JavaScript.
- **Chi tiết quy trình xử lý:** Quy trình xử lý một ảnh khuôn mặt đầu vào trải qua các bước sau:
 - Bước 1:
 - Tiền xử Lý (Preprocessing) Face Detection (Phát hiện khuôn mặt): Sử dụng MTCNN để phát hiện vị trí khuôn mặt trong ảnh. Lấy khuôn mặt lớn nhất (nếu có nhiều mặt) dựa trên diện tích bounding box.
 - Face Alignment (Căn chỉnh khuôn mặt): Dựa vào 5 điểm mốc (landmarks): mắt trái, mắt phải, mũi, khóe miệng trái, khóe miệng phải. Sử dụng phép biến đổi Affine (Similarity Transform) để xoay và căn chỉnh khuôn mặt về chuẩn chính diện.
 - Kích thước chuẩn đầu ra: 112x112 pixel.
 - Normalization (Chuẩn hóa): Chuyển đổi không gian màu sang RGB. Chuẩn hóa giá trị pixel về khoảng $[-1, 1]$ theo công thức: $(\text{image} - 127.5) / 128.0$.
 - Bước 2:
 - Trích Xuất Đặc Trưng (Feature Extraction) Mô hình: GhostFaceNet.
 - Input: Tensor (1, 112, 112, 3). Output: Embedding vector kích thước 512 chiều. Vector này sau đó được chuẩn hóa L2 (L2 Normalization) để đưa về nằm trên hypersphere đơn vị.
 - Bước 3:
 - So Khớp (Matching) Xác thực 1:1 (Verification): Tính độ đo Cosine Similarity giữa hai vector đặc trưng. So sánh với ngưỡng (threshold) đã được hiệu chỉnh (ví dụ: 0.25 - 0.35) để kết luận là "Cùng người" hay "Khác người".
 - Nhận diện 1:N (Identification): Sử dụng FAISS Index để tìm k vector trong cơ sở dữ liệu gần nhất với vector truy vấn. Kết quả trả về danh sách các ứng viên có độ tương đồng cao nhất.
- **Các chức năng chính:**

- POST /verify_faces: Nhận 2 ảnh khuôn mặt, trích xuất embedding, tính độ tương đồng Cosine, trả về similarity score, kết luận cùng/người khác (True/False) và thời gian xử lý.
- POST /register_face: Nhận ảnh khuôn mặt kèm thông tin định danh (tên, tuổi, mô tả), trích xuất embedding, lưu vector vào FAISS Index và metadata vào cơ sở dữ liệu, trả về ID người dùng mới.
- POST /recognize_face: Nhận 1 ảnh khuôn mặt, trích xuất embedding, tìm kiếm Top-K láng giềng gần nhất trong FAISS, trả về danh sách các người dùng đã đăng ký có độ tương đồng cao nhất.
- POST /verify_faces_with_db: Nhận 2 ảnh, vừa thực hiện so khớp trực tiếp giữa hai ảnh, vừa kiểm tra mỗi ảnh với cơ sở dữ liệu để xác định chúng có thuộc về người đã đăng ký hay không.
- Quản lý dữ liệu: Vector đặc trưng được lưu trong FAISS Index để tối ưu tốc độ tìm kiếm, trong khi metadata (tên, mô tả, thông tin cá nhân) được lưu riêng và ánh xạ với vector thông qua ID định danh.

VI. KẾT QUẢ

1. Kết quả đánh giá mô hình

- Mô hình được đánh giá trên các tập chuẩn (benchmark) nhận diện khuôn mặt (LFW, CFP_FP, AgeDB) sử dụng độ tương đồng Cosine (Cosine Similarity).

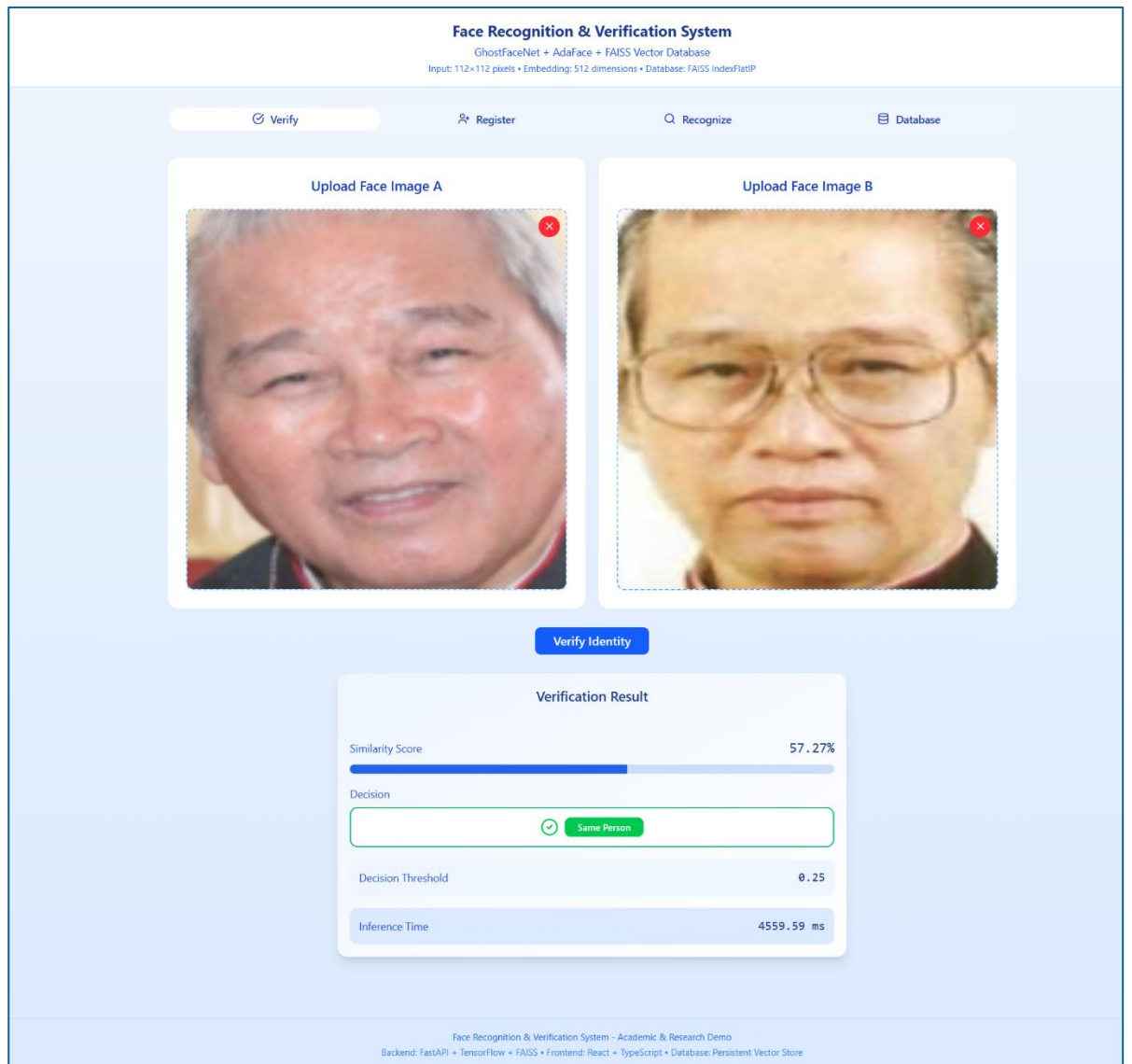
Tập Benchmark	Độ chính xác Pretrained	Độ chính xác Fine-tuned	Thay đổi (Δ)	Ghi chú
LFW	99.60%	99.52%	-0.08%	Giảm nhẹ trên khuôn mặt phổ quát
CFP_FP	95.36%	95.13%	-0.23%	Giảm nhẹ trên góc nghiêng-trực diện
AgeDB	97.37%	97.52%	+0.15%	Cải thiện trên nhận diện bất biến độ tuổi

- **Đánh giá trên tập dữ liệu Việt Nam (VN-Celeb):** Ngoài các tập benchmark quốc tế, mô hình cũng được đánh giá trên tập dữ liệu VN-Celeb để kiểm tra khả năng thích nghi với khuôn mặt người Việt Nam. Số lượng cặp kiểm thử: 2000 cặp (được tạo ngẫu nhiên từ tập VN-Celeb). Kết quả: Pretrained Accuracy: 78.65% (Ngưỡng: 0.140) Fine-tuned Accuracy: 79.90% (Ngưỡng: 0.174) Cải thiện: +1.25% Nhận xét: Việc tinh chỉnh (fine-tuning) đã mang lại sự cải thiện rõ rệt (+1.25%) trên tập dữ liệu người Việt, chứng minh tính hiệu quả của chiến lược huấn luyện trong việc thích nghi miền dữ liệu (domain adaptation).
- **Nhận xét:**

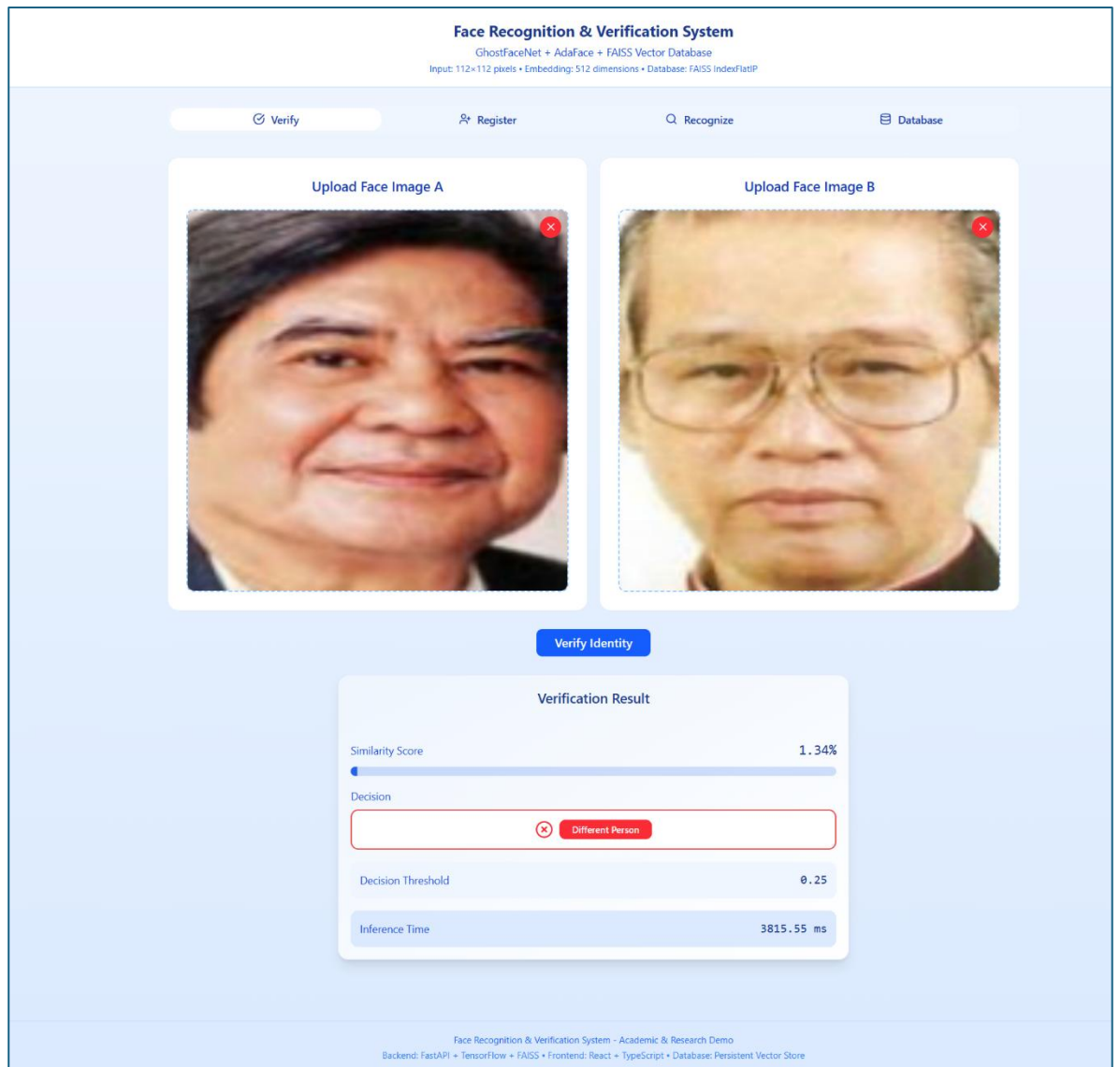
- **Hiệu quả của Tinh chỉnh:** Quá trình tinh chỉnh đã giúp Loss hội tụ ổn định. Việc giảm nhẹ độ chính xác trên LFW/CFP_FP gợi ý rằng quá trình tinh chỉnh trên tập dữ liệu con (có thể là tập "Asian" dựa trên tên file) đã làm mô hình chuyên biệt hóa hơn một chút và lệch khỏi phân phối phổ quát của LFW, hoặc thời gian huấn luyện còn ngắn.
- **Bất biến độ tuổi:** Sự cải thiện trên AgeDB cho thấy mô hình đã học được các đặc trưng tốt hơn để xử lý sự thay đổi do tuổi tác.
- **Hiệu năng:** GhostNetV1 duy trì hiệu quả tính toán cao (~4M tham số), phù hợp cho ứng dụng thời gian thực.
- **Hạn chế và hướng tiếp cận tiếp theo**
 - Trong quá trình thực nghiệm, nhóm chưa thử nghiệm với số epochs lớn hơn cũng như chưa thử nhiều tập hyperparameters nên kết quả thể hiện chưa quá nổi trội.
 - Hướng tiếp cận tiếp theo:
 - Mở rộng Tinh chỉnh: 6 epochs cho Giai đoạn 2 là khá ngắn. Nên kéo dài thời gian huấn luyện kết hợp với lịch trình giảm learning rate (ví dụ: Cosine Decay) để đạt kết quả tốt hơn.
 - Mở đóng băng toàn bộ: Thử nghiệm mở đóng băng nhiều lớp hơn hoặc toàn bộ backbone với learning rate rất nhỏ để huấn luyện lâu hơn.
 - Thử nghiệm với learning rate scheduler (ví dụ: cosine annealing). Thêm data augmentation để tăng tính đa dạng. Thử nghiệm với các bộ hyperparameters.

2. Kết quả hệ thống demo

- Chức năng xác thực 1-1:

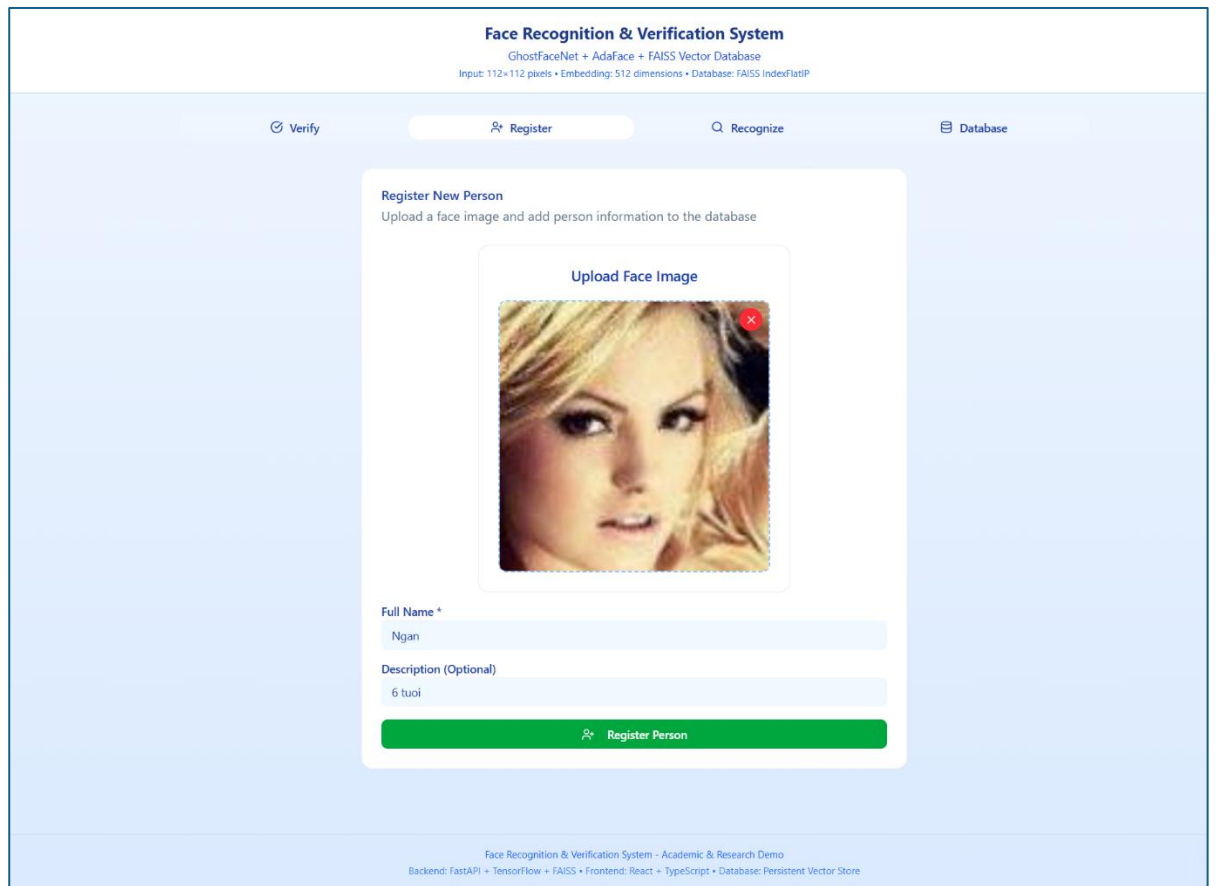


Hình 17. Kiểm tra hai người giống nhau

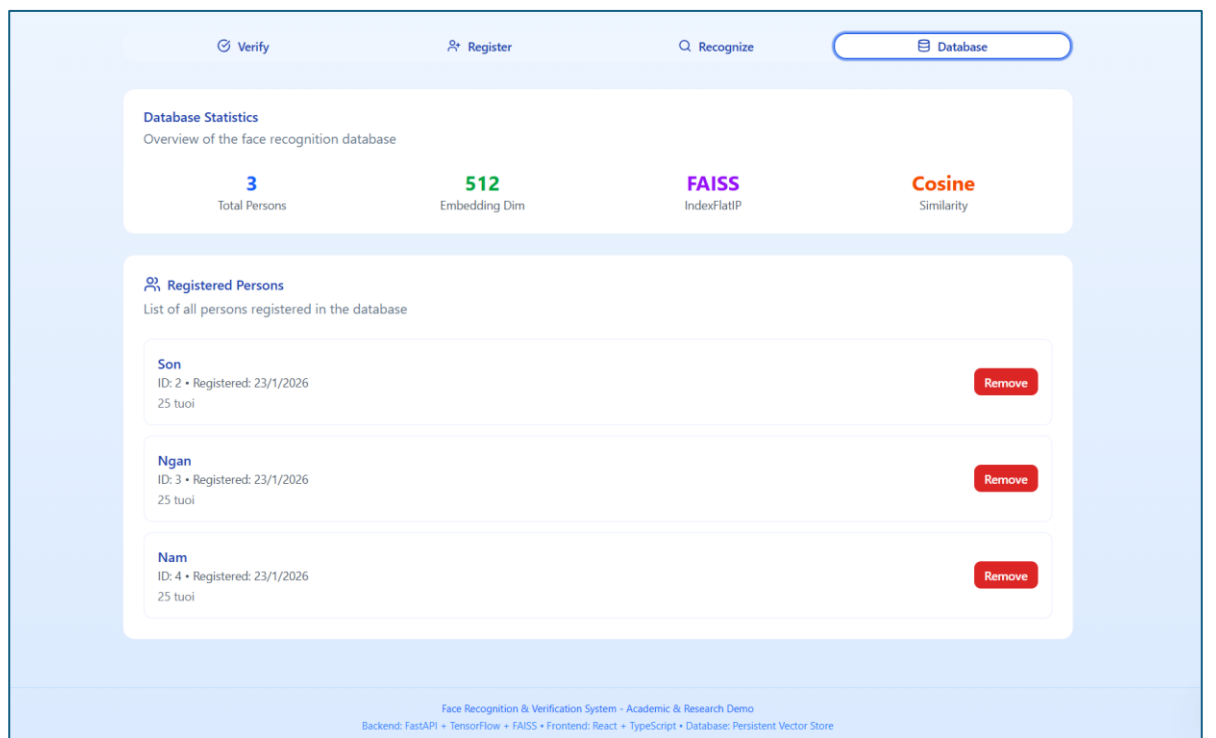


Hình 18. Kiểm tra hai người khác nhau

- Chức năng đăng ký database:

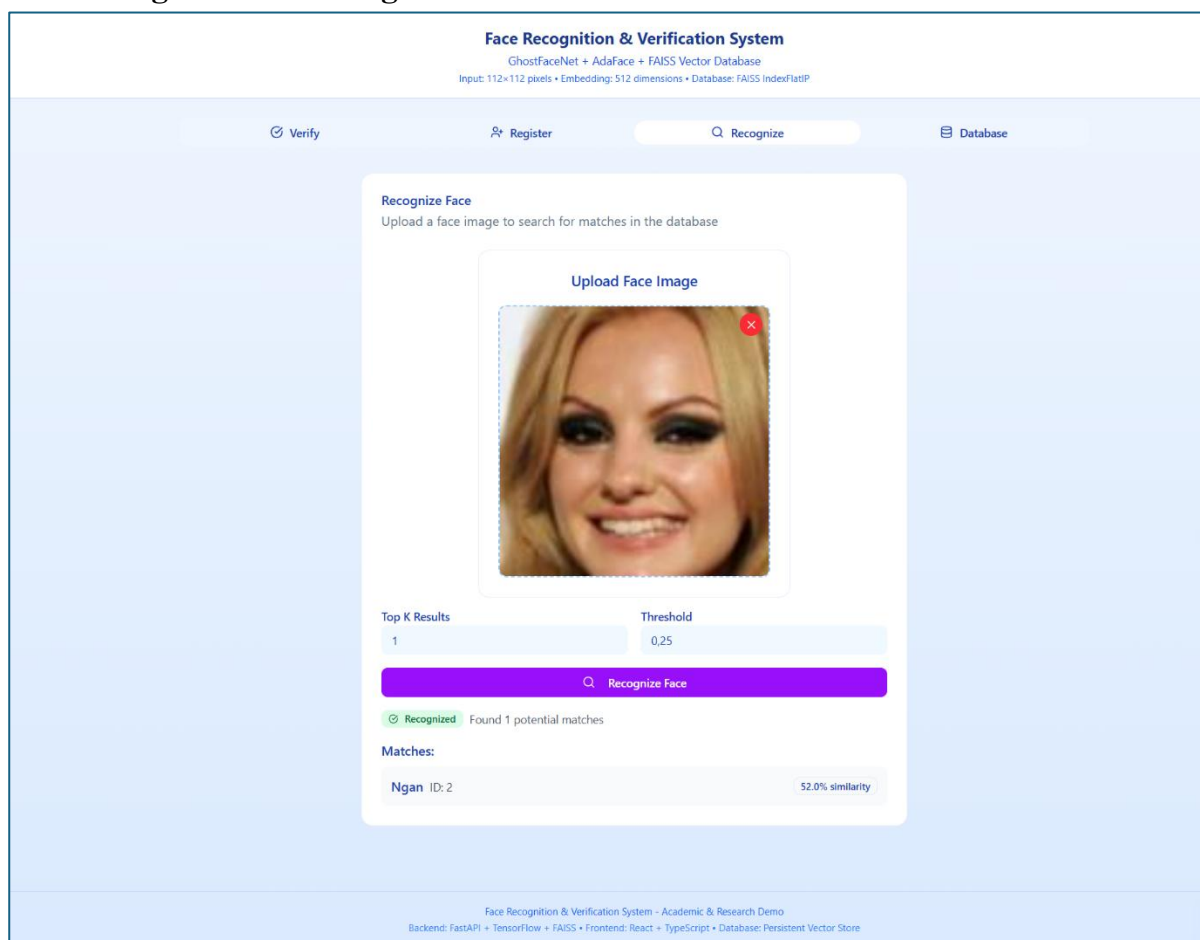


Hình 19. Đăng ký hình ảnh



Hình 20. Kết quả đăng ký hình ảnh

- **Chức năng tìm kiếm trong database:**



Hình 21. Kiểm tra chức năng search top k

3. Tổng kết

Hệ thống đã tích hợp thành công các kỹ thuật tiên tiến trong nhận diện khuôn mặt (GhostFaceNet, MTCNN, FAISS) vào một kiến trúc vi dịch vụ (Microservice-like) với FastAPI. Việc bổ sung bước Alignment và Detection đảm bảo độ chính xác cao ngay cả với ảnh đầu vào chưa qua cắt chỉnh (in-the-wild inputs).

Link github nhóm: [GhostFaceNetv1_adaface](#).

VII. TÀI LIỆU THAM KHẢO

- Repository: [GhostFaceNets](#), [AdaFace](#).
- Paper: [GhostFaceNets: Lightweight Face Recognition Model From Cheap Operations](#), [AdaFace: Quality Adaptive Margin for Face Recognition](#).
- Dataset: [Labelled Faces in the Wild](#), [Celebrities in Frontal-Profile in the Wild](#), [Webface_112*112](#), [VN-celeb: Dữ liệu khuôn mặt người nổi tiếng Việt Nam](#).

