

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Nguyễn Thị Vân

**NGHIÊN CỨU ỨNG DỤNG NHẬN DIỆN KHUÔN MẶT TRONG XÁC
THỰC EKYC TẠI NGÂN HÀNG VPBANK**

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT
(Theo định hướng ứng dụng)

HÀ NỘI – 2025

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Nguyễn Thị Vân

**NGHIÊN CỨU ỨNG DỤNG NHẬN DIỆN KHUÔN MẶT TRONG XÁC
THỰC EKYC TẠI NGÂN HÀNG VPBANK**

Chuyên ngành: Hệ thống thông tin
Mã số: 8.48.01.04

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT
(Theo định hướng ứng dụng)

NGƯỜI HƯỚNG DẪN KHOA HỌC

TS. VŨ VĂN THÒA

HÀ NỘI - 2025

LỜI CAM ĐOAN

Tôi cam đoan rằng công trình nghiên cứu này là thành quả của công sức cá nhân của tôi và không sao chép từ bất kỳ nguồn nào khác. Tất cả thông tin được trình bày trong đề án này đều là sản phẩm của công việc cá nhân hoặc được tổng hợp từ nhiều nguồn tài liệu khác nhau. Mọi tài liệu tham khảo đều được trích dẫn một cách hợp pháp và có nguồn gốc rõ ràng. Các dữ liệu và kết quả được trình bày trong đề án đều là trung thực và chưa từng được công bố trong bất kỳ công trình nghiên cứu nào khác.

Tác giả đề án

Vân

Nguyễn Thị Vân

LỜI CẢM ƠN

Lời đầu tiên, học viên xin gửi lời cảm ơn sâu sắc tới các thầy, cô giáo giảng viên khoa Công nghệ thông tin 1, khoa Đào tạo Sau đại học nói riêng và các thầy, cô giáo giảng viên Học viện Công nghệ Bưu chính Viễn thông nói chung. Trong suốt quá trình học tập tại Học viện, các thầy cô đã chỉ bảo, giảng dạy cho em biết bao kiến thức, kinh nghiệm quý báu để em có hành trang vững bước trong tương lai.

Em cũng xin được gửi lời cảm ơn tới thầy hướng dẫn TS. Vũ Văn Thóa, cảm ơn thầy đã luôn hướng dẫn chỉ bảo tận tình em trong suốt quá trình học tập, nghiên cứu và thực hiện đề án này. Những lời khuyên, sự chỉ bảo của các thầy đã giúp em hoàn thành đề án tốt nghiệp này cũng như có thêm rất nhiều kiến thức, kinh nghiệm trong việc học tập và nghiên cứu.

Dù đã nỗ lực hoàn thành đề án, em hiểu rằng có thể không tránh khỏi những sai sót. Kính mong được thầy cô và các bạn thông cảm và đóng góp ý kiến.

Em xin trân trọng cảm ơn.

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
MỤC LỤC.....	iii
DANH MỤC TỪ VIẾT TẮT.....	iv
DANH SÁCH CÁC BẢNG.....	v
DANH MỤC CÁC HÌNH.....	vi
MỞ ĐẦU.....	1
CHƯƠNG 1. TỔNG QUAN VỀ NHẬN DIỆN KHUÔN MẶT VÀ XÁC THỰC eKYC ...	3
1.1. Giới thiệu bài toán nhận diện khuôn mặt.....	3
1.2. Tổng quan các phương pháp nhận diện khuôn mặt	4
1.3. Ứng dụng của nhận diện khuôn mặt	7
1.4. Quy trình eKYC và ứng dụng nhận diện khuôn mặt	9
1.6. Kết luận chương.....	17
CHƯƠNG 2. NGHIÊN CỨU MÔ HÌNH NHẬN DIỆN KHUÔN MẶT	18
2.1. Khảo sát một số mô hình nhận diện khuôn mặt sử dụng học sâu.....	18
2.2. Các tiêu chí đánh giá mô hình nhận diện khuôn mặt.....	20
2.3. Một số yếu tố trong bài toán phát hiện khuôn mặt	24
2.4. Lựa chọn mô hình nhận diện khuôn mặt cho xác thực eKYC của ngân hàng.....	26
2.5. Kết luận chương.....	42
CHƯƠNG 3: GIẢI PHÁP ỨNG DỤNG NHẬN DIỆN KHUÔN MẶT TRONG XÁC THỰC eKYC TẠI NGÂN HÀNG VPBANK	43
3.1. Giới thiệu chung về eKYC tại ngân hàng VPBank	43
3.2. Giải pháp triển khai nhận diện khuôn mặt trong xác thực eKYC.....	44
3.3. Thử nghiệm và đánh giá	47
3.4. Kết luận chương.....	52
KẾT LUẬN.....	53
DANH MỤC TÀI LIỆU THAM KHẢO.....	55

DANH MỤC TỪ VIẾT TẮT

STT	Từ viết tắt	Tiếng Anh	Tiếng Việt/Giải thích
1	eKYC	Electronic Know Your Customer	Quy trình định danh khách hàng điện tử
2	CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
3	AI	Artificial intelligence	Trí tuệ nhân tạo
4	GAT	Graph Attention Network	Mạng đồ thị chú ý
5	FAR	False Acceptance Rate	Tỷ lệ chấp nhận sai
6	FRR	False Rejection Rate	Tỷ lệ từ chối sai
7	EER	Equal Error Rate	Điểm lỗi cân bằng

DANH SÁCH CÁC BẢNG

Bảng 3. 1: Kết quả thực nghiệm	50
--------------------------------------	----

DANH MỤC CÁC HÌNH

Hình 1. 1: Ví dụ minh họa nhận diện khuôn mặt.....	3
Hình 1. 2: Minh họa một số ứng dụng của nhận diện khuôn mặt	9
Hình 1. 3: Quy trình eKYC điển hình.....	11
Hình 1. 4: Nhận diện khuôn mặt trong eKYC	12
Hình 2. 1: Các phương pháp kim tự tháp ảnh và kim tự tháp đặc trưng	25
Hình 2. 2: Mô tả về đầu vào và đầu ra của mô hình RetinaFace	27
Hình 2. 3: Kiến trúc mô hình RetinaFace	28
Hình 2. 4: Năm điểm mốc trong khuôn mặt được mô hình RetinaFace dự đoán.....	29
Hình 2. 5: Dữ liệu đầu vào và đầu ra của mô hình ArcFace.....	30
Hình 2. 6: Kiến trúc mô hình ArcFace.....	32
Hình 2. 7: ArcFace không chỉ là một mô hình phân biệt mà còn là một mô hình sinh	32
Hình 2. 8: Hàm mất mát ArcFace	34
Hình 2. 9: Độ tương đồng cosine	36
Hình 2. 10: Graph Attention Network	37
Hình 2. 11: Kiến trúc mô hình phát hiện khuôn mặt giả mạo	40
Hình 3. 1: Kiến trúc tổng thể đề xuất.....	45
Hình 3. 2: Tập dữ liệu thử nghiệm.....	48

MỞ ĐẦU

Hiện nay, ngành ngân hàng đang chuyển mình mạnh mẽ sang hướng số hoá với sự gia tăng của các dịch vụ ngân hàng điện tử (e-banking) và ngân hàng số (digital banking). Trong bối cảnh này, eKYC (Electronic Know Your Customer) trở thành một giải pháp không thể thiếu để tối ưu hoá quy trình xác thực khách hàng, giảm bớt sự phụ thuộc vào giao dịch trực tiếp và giấy tờ truyền thống.

Công nghệ nhận diện khuôn mặt nổi bật với khả năng nhanh chóng, chính xác và tiện lợi trong việc xác minh danh tính. So với các phương pháp truyền thống như nhập mã OTP hay đối chiếu thủ công, nhận diện khuôn mặt mang lại mức độ bảo mật cao hơn, đồng thời cải thiện trải nghiệm người dùng bằng cách đơn giản hoá quy trình đăng ký và đăng nhập. Việc sử dụng nhận diện khuôn mặt trong eKYC giúp phát hiện và ngăn chặn các hành vi gian lận như sử dụng giấy tờ giả hoặc mạo danh. Ngoài ra, công nghệ này còn hỗ trợ kiểm tra chéo thông tin khách hàng với cơ sở dữ liệu một cách tự động và chính xác, giảm thiểu sai sót trong khâu xác thực.

Nhiều tổ chức tài chính lớn trên thế giới đã áp dụng thành công công nghệ nhận diện khuôn mặt trong eKYC, mở ra cơ hội áp dụng rộng rãi tại Việt Nam để nâng cao năng lực cạnh tranh.

Từ những lý do trên, học viên đã lựa chọn đề tài “**Nghiên cứu ứng dụng nhận diện khuôn mặt trong xác thực eKYC tại ngân hàng VPBank**” để thực hiện đề án tốt nghiệp thạc sĩ. Hy vọng rằng nghiên cứu này sẽ mang lại những hiểu biết sâu hơn và khám phá mới trong lĩnh vực này, tạo ra những đóng góp ý nghĩa cho cộng đồng nghiên cứu và mang lại giá trị thực tế cho người dùng.

Mục tiêu của đề tài “Nghiên cứu ứng dụng nhận diện khuôn mặt trong xác thực eKYC tại ngân hàng VPBank” nhằm nghiên cứu và ứng dụng một giải pháp nhận diện khuôn mặt hiệu quả, đáp ứng các yêu cầu về độ chính xác, bảo mật và tính tiện lợi trong quá trình xác thực danh tính khách hàng của ngân hàng.

Đối tượng nghiên cứu của đề án là các vấn đề liên quan đến các mô hình nhận diện khuôn mặt sử dụng học sâu đáp ứng các yêu cầu về độ chính xác, bảo mật và tính tiện lợi trong quá trình xác thực danh tính khách hàng.

Phạm vi nghiên cứu của đề án tập trung vào ứng dụng nhận diện khuôn mặt trong hệ thống eKYC tại các ngân hàng thương mại ở Việt Nam, nơi các dịch vụ chuyển đổi số và áp dụng công nghệ tiên tiến đang ngày càng phát triển.

Nội dung đề án được trình bày thành ba chương theo cấu trúc sau:

Chương 1: Tổng quan về nhận diện khuôn mặt và xác thực eKYC

Trình bày tổng quan về bài toán nhận diện khuôn mặt, các phương pháp, ứng dụng của nhận diện khuôn mặt, quy trình eKYC và về các vấn đề bảo mật và pháp lý liên quan.

Chương 2: Nghiên cứu mô hình nhận diện khuôn mặt

Trình bày về các khái niệm liên quan đến các mô hình nhận diện khuôn mặt sử dụng học sâu. Sau đó sẽ trình bày về các tiêu chí đánh giá mô hình nhận diện khuôn mặt. Cuối cùng sẽ lựa chọn mô hình nhận diện khuôn mặt cho xác thực eKYC của ngân hàng.

Chương 3: Giải pháp ứng dụng nhận diện khuôn mặt trong xác thực eKYC tại ngân hàng VPBank

Chương này sẽ giới thiệu chung về eKYC tại ngân hàng VPBank. Sau đó, sẽ trình bày giải pháp triển khai nhận diện khuôn mặt trong xác thực. Cuối cùng sẽ trình bày về quá trình thực nghiệm và đánh giá kết quả.

CHƯƠNG 1. TỔNG QUAN VỀ NHẬN DIỆN KHUÔN MẶT VÀ XÁC THỰC eKYC

Chương này sẽ trình bày tổng quan về bài toán nhận diện khuôn mặt, các phương pháp, ứng dụng của nhận diện khuôn mặt, quy trình eKYC. Cuối cùng sẽ trình bày về các vấn đề bảo mật và pháp lý liên quan.

1.1. Giới thiệu bài toán nhận diện khuôn mặt

Nhận diện khuôn mặt (Face Recognition) là một kỹ thuật sinh trắc học quan trọng dùng để xác thực danh tính, được ứng dụng rộng rãi trong nhiều lĩnh vực như quân sự, tài chính, an ninh công cộng và đời sống thường nhật [1]. Bài toán này nhằm xác định hoặc xác thực danh tính của một người dựa trên ảnh khuôn mặt của họ như minh họa trong hình 1.1. So với các phương thức sinh trắc học khác như vân tay hay mống mắt, nhận diện khuôn mặt có ưu điểm là không xâm lấn, có thể thực hiện từ khoảng cách xa và khó bị giả mạo hơn bằng cách thông thường (mặc dù vẫn có rủi ro bị đánh lừa bằng ảnh hoặc mặt nạ, ...). Tuy nhiên, nhận diện khuôn mặt cũng đối mặt với nhiều thách thức do sự đa dạng lớn về điều kiện ánh sáng, góc nhìn, biểu cảm, độ tuổi, và sự xuất hiện của các vật cản (như kính, khẩu trang). Những yếu tố này làm cho việc nhận dạng khuôn mặt (xác định danh tính trong một tập nhiều người) và xác thực khuôn mặt (so khớp 1-1 để kiểm tra đúng người) trở nên khó khăn hơn so với nhận dạng các đối tượng cố định.



Hình 1. 1: Ví dụ minh họa nhận diện khuôn mặt

Trong thập kỷ qua, lĩnh vực nhận diện khuôn mặt đã có những bước đột phá mạnh mẽ nhờ sự phát triển của học sâu. Hàng năm có hàng chục công trình nghiên cứu về nhận diện khuôn mặt được công bố tại các hội nghị và tạp chí hàng đầu (CVPR, ICCV, PAMI, ...), và nhiều kỹ thuật đã nhanh chóng được ứng dụng trong thực tiễn [2]. Đặc biệt, từ năm 2014, sự ra đời của các mô hình học sâu như DeepFace và DeepID đã mở ra kỷ nguyên mới, giúp nâng cao đáng kể độ chính xác của nhận diện khuôn mặt [2]. Độ chính xác trên các bộ dữ liệu chuẩn đã tăng lên mức ngang bằng hoặc vượt qua con người. Chẳng hạn, mô hình GaussianFace (2014) đạt tỷ lệ xác minh 98,52% trên bộ LFW, lần đầu tiên vượt qua mức 97,53% của con người trên cùng bộ dữ liệu. Những hệ thống hiện đại còn có thể đạt độ chính xác trên 99,8% ở LFW nhờ các kỹ thuật học sâu tiên tiến. Nhận diện khuôn mặt đã và đang đóng vai trò quan trọng trong đời sống, xuất hiện trong các ứng dụng quen thuộc như mở khóa thiết bị bằng khuôn mặt và thanh toán di động.

1.2. Tổng quan các phương pháp nhận diện khuôn mặt

Nhận diện khuôn mặt đã trải qua nhiều giai đoạn phát triển, từ các phương pháp truyền thống dựa trên đặc trưng thủ công đến các kỹ thuật hiện đại sử dụng học sâu. Dưới đây là tổng quan về các phương pháp chính:

1.2.1. Phương pháp cổ điển

Trước khi học sâu trở nên phổ biến, các hệ thống nhận diện khuôn mặt chủ yếu dựa vào trích xuất đặc trưng thủ công và các mô hình học máy truyền thống. Phân tích thành phần chính (PCA) là một kỹ thuật nền tảng đã được Turk và Pentland giới thiệu trong phương pháp Eigenface (1991) – phương pháp đầu tiên giải quyết nhận diện khuôn mặt trong thị giác máy tính [2]. Ý tưởng chính của Eigenface là chuyển ảnh khuôn mặt xám thành vector một chiều và sử dụng PCA để giảm số chiều, tìm ra không gian đặc trưng có phương sai lớn nhất mà vẫn giữ được thông tin phân biệt giữa các khuôn mặt. Phương pháp Fisherface (1997) sau đó cải tiến bằng cách sử dụng Phân tích biệt tuyến tính (LDA), tối ưu hóa tỉ số giữa phương sai liên-lớp và nội-lớp, nhằm tạo không gian đặc trưng phân biệt tốt hơn giữa các người khác nhau

[2]. Bên cạnh PCA/LDA, nhiều đặc trưng thủ công khác đã được áp dụng: ví dụ, Local Binary Patterns (LBP) trích xuất mô hình vân da mặt dựa trên histogram các mẫu kết cấu cục bộ và đã cho kết quả tốt trong nhận diện khuôn mặt. Ahonen và các cộng sự (2004) cho thấy việc dùng LBP trên từng vùng của ảnh mặt và kết hợp lại có thể cải thiện độ chính xác đáng kể. Ngoài LBP, các đặc trưng SIFT, SURF, HOG... vốn thành công trong nhận dạng vật thể cũng được điều chỉnh để nhận dạng khuôn mặt. Sau khi trích xuất đặc trưng, các thuật toán học máy như SVM (máy vector hỗ trợ) thường được sử dụng để phân loại hoặc xác thực khuôn mặt. Chẳng hạn, nhiều nghiên cứu kết hợp SVM với đặc trưng PCA, LDA hoặc ICA (Phân tích thành phần độc lập) để tăng độ chính xác nhận diện. Nhìn chung, các phương pháp cổ điển đã đạt được thành công nhất định trong môi trường có kiểm soát (ví dụ ảnh chân dung cố định) nhưng thường suy giảm hiệu suất khi áp dụng trên ảnh trong điều kiện tự nhiên (ngoài trời, độ phân giải thấp, biến dạng mạnh).

1.2.2. Phương pháp dựa trên học sâu

Từ năm 2014, sự bùng nổ của mạng nơ-ron tích chập (CNN) [8] đã cách mạng hóa lĩnh vực nhận diện khuôn mặt. Các mô hình học sâu có khả năng tự học đặc trưng phân biệt từ dữ liệu ảnh lớn, thay vì dựa vào thiết kế thủ công. Những công trình tiên phong như DeepFace (2014) [4] và DeepID (2014) [5] lần đầu tiên áp dụng CNN cho nhận dạng khuôn mặt, đạt kết quả vượt trội so với phương pháp cũ [3]. Tiếp đó, FaceNet (2015) [6] giới thiệu cách học biểu diễn nhúng 128 chiều cho mỗi khuôn mặt sao cho khoảng cách giữa hai vector đặc trưng phản ánh mức độ giống nhau của hai mặt – cách tiếp cận này đạt 99,63% độ chính xác trên LFW, gần như tuyệt đối. Trong 5 năm trở lại đây, hàng loạt cải tiến đã được đề xuất để hoàn thiện hệ thống nhận diện khuôn mặt học sâu:

- **Mạng sâu và kiến trúc hiệu quả hơn:** Sử dụng các mạng CNN sâu như ResNet, SENet, hoặc gần đây là các mô hình Transformer, giúp cải thiện hiệu suất nhận dạng. Việc tiền xử lý ảnh đầu vào cũng được chú trọng (ví dụ: căn

chỉnh khuôn mặt theo 3D như DeepFace đề xuất để giảm ảnh hưởng của pose lệch) [2].

- **Hàm mất mát đặc biệt:** Thay vì hàm Softmax thông thường, các nghiên cứu đưa ra các hàm mất mát nhúng khoảng cách với biên nhằm tăng độ phân biệt giữa các lớp nhận dạng. Tiêu biểu là ArcFace (2019) [7] với hàm mất mát cosin cộng biên góc. ArcFace không chỉ có diễn giải hình học rõ ràng mà còn giúp tối đa hóa khoảng cách giữa các lớp khác nhau trên không gian đặc trưng. Nhờ đó, mô hình ArcFace đạt hiệu năng hàng đầu: trên tập LFW, ArcFace (ResNet-50) đạt độ chính xác tới 99,81%.
- **Dữ liệu huấn luyện lớn và đa dạng:** Hiệu quả của mô hình nhận diện khuôn mặt tỷ lệ thuận với quy mô và độ đa dạng của dữ liệu huấn luyện. Những bộ dữ liệu khổng lồ như MSCeleb-1M, VGGFace2, CASIA-WebFace, v.v., chứa hàng triệu ảnh của hàng ngàn người, đã được sử dụng để huấn luyện các mô hình sâu. Nghiên cứu cũng chỉ ra tầm quan trọng của việc tăng cường dữ liệu (data augmentation) và tạo dữ liệu tổng hợp để mô phỏng các biến dạng khuôn mặt, nhằm cải thiện tính tổng quát của mô hình [2].
- **Nhận diện khuôn mặt trong các điều kiện đặc thù:** Gần đây, nhiều hướng nghiên cứu tập trung vào những bài toán con như nhận diện khuôn mặt có đeo khẩu trang, nhận diện khuôn mặt trong video, nhận diện thời gian thực trên thiết bị di động, v.v. Chẳng hạn, việc nhận diện khi đeo khẩu trang trở nên cấp thiết trong đại dịch COVID-19; các giải pháp kết hợp học sâu với mô hình 3D hoặc học chuyển tiếp (transfer learning) đã được phát triển để giải quyết vấn đề suy giảm hiệu suất khi khuôn mặt bị che một phần. Ngoài ra, kỹ thuật đối mặt (face frontalization) và cân bằng miền (domain adaptation) cũng được áp dụng để tăng cường khả năng nhận diện trong các điều kiện ánh sáng, góc chụp khác nhau hoặc giữa các miền dữ liệu khác nhau.

Phương pháp học sâu hiện đại đã vượt trội so với phương pháp cổ điển về độ chính xác và độ bền vững. Hệ thống nhận diện khuôn mặt ngày nay thường bao gồm

cả một pipeline hoàn chỉnh: từ phát hiện khuôn mặt, căn chỉnh, trích xuất đặc trưng bởi mạng CNN, đến so sánh đặc trưng để nhận dạng/ xác thực. Trong đó, mạng CNN đóng vai trò cốt lõi tạo ra biểu diễn đặc trưng (face embedding) cho mỗi khuôn mặt – các vector này có tính phân tách cao giữa các danh tính khác nhau và gần nhau cho cùng một người. Kết quả là, nhận diện khuôn mặt tự động đã đạt độ chính xác rất cao ngay cả trong môi trường tự nhiên, mở đường cho nhiều ứng dụng thực tiễn quan trọng như trình bày ở phần tiếp theo.

1.3. Ứng dụng của nhận diện khuôn mặt

Nhận diện khuôn mặt là một trong những công nghệ sinh trắc học được ứng dụng rộng rãi trong nhiều lĩnh vực của đời sống hiện đại. Nhờ sự phát triển mạnh mẽ của trí tuệ nhân tạo và thị giác máy tính, công nghệ này ngày càng được tích hợp vào các hệ thống thực tế nhằm nâng cao hiệu quả, tính bảo mật và trải nghiệm người dùng. Dưới đây là một số ứng dụng tiêu biểu:

- **An ninh công cộng và giám sát:** Nhận diện khuôn mặt hỗ trợ lực lượng an ninh xác định tội phạm hoặc đối tượng tình nghi trong đám đông (ví dụ qua camera giao thông, camera an ninh tại nơi công cộng). Nhiều chính phủ đã triển khai hệ thống nhận dạng khuôn mặt để giám sát an ninh; thống kê cho thấy có tới 7/10 chính phủ trên thế giới sử dụng công nghệ này một cách rộng rãi. Chẳng hạn, Trung Quốc xây dựng mạng lưới camera thông minh tích hợp AI để nhận dạng khuôn mặt người dân tại các thành phố lớn, hỗ trợ việc tìm kiếm tội phạm hoặc người mất tích. Tuy nhiên, việc ứng dụng trong giám sát công cộng cũng dấy lên lo ngại về quyền riêng tư.
- **Kiểm soát biên giới và hàng không:** Tại các cửa khẩu và sân bay, nhận diện khuôn mặt được dùng để tự động kiểm tra hộ chiếu và nhận diện hành khách, giúp tăng tốc độ làm thủ tục. Nhiều sân bay hiện đại cho phép hành khách làm thủ tục lên máy bay bằng “hộ chiếu khuôn mặt” – hệ thống sẽ đối chiếu khuôn mặt hành khách với ảnh lưu trong cơ sở dữ liệu hộ chiếu hoặc visa. Việc này không chỉ thuận tiện mà còn nâng cao an ninh, vì giảm thiểu nguy cơ sử dụng giấy tờ giả.

- **Mở khóa thiết bị và ứng dụng di động:** Một trong những ứng dụng phổ biến nhất của nhận diện khuôn mặt là mở khóa điện thoại thông minh, máy tính bảng, laptop... Thống kê năm 2024 cho thấy khoảng 68% người dùng sử dụng nhận dạng khuôn mặt để mở khóa thiết bị cá nhân của họ. Tiêu biểu, công nghệ Face ID trên iPhone (Apple) sử dụng camera hồng ngoại và thuật toán nhận dạng khuôn mặt 3D để xác thực người dùng một cách an toàn, với xác suất nhận nhầm dưới 1/1.000.000 (theo Apple). Nhiều ứng dụng di động (ví dụ ứng dụng ngân hàng, ví điện tử) cũng tích hợp đăng nhập bằng khuôn mặt để tăng tính tiện lợi và bảo mật, thay thế cho mật khẩu hoặc vân tay.
- **Thanh toán và dịch vụ tài chính:** Nhận diện khuôn mặt đang được ứng dụng trong thanh toán điện tử và xác thực khách hàng tài chính. Tại một số nước, khách hàng có thể thanh toán tại cửa hàng chỉ bằng cách quét khuôn mặt (ví dụ: hệ thống "Smile to Pay" của Alipay tại Trung Quốc). Các ngân hàng triển khai giải pháp eKYC cho phép mở tài khoản từ xa bằng xác thực khuôn mặt thay vì gặp trực tiếp. Việc này giúp chống gian lận trong tài chính – khuôn mặt của người mở tài khoản được đối chiếu với giấy tờ tùy thân để đảm bảo họ đúng là người thật, tránh sử dụng danh tính giả.
- **Mạng xã hội và quản lý hình ảnh:** Các nền tảng mạng xã hội như Facebook từng sử dụng thuật toán nhận diện khuôn mặt để tự động gợi ý gắn thẻ (tag) bạn bè trong ảnh tải lên. Tính năng này giúp người dùng dễ dàng nhận ra ai có mặt trong ảnh và quản lý album ảnh cá nhân. Các dịch vụ lưu trữ ảnh (Google Photos, Apple Photos) cũng dùng nhận diện khuôn mặt để nhóm ảnh theo từng người, giúp việc tra cứu ảnh theo khuôn mặt trở nên thuận tiện. Nhưng do lo ngại về quyền riêng tư, Facebook đã phải tắt tính năng nhận diện khuôn mặt gợi ý tag và xóa dữ liệu khuôn mặt của hơn 1 tỷ người dùng vào năm 2021.
- **Kiểm soát ra vào và chấm công:** Nhiều cơ quan, doanh nghiệp triển khai hệ thống kiểm soát cửa ra vào sử dụng nhận diện khuôn mặt thay cho thẻ từ hoặc vân tay. Nhân viên chỉ cần đứng trước camera để hệ thống nhận diện và mở khóa cửa nếu hợp lệ. Tương tự, máy chấm công bằng khuôn mặt giúp ghi nhận

giờ vào/ra của nhân viên tự động, tránh việc gian lận (như nhờ chấm công hộ). Việc này đặc biệt hữu ích trong bối cảnh đề cao vệ sinh khi tiếp xúc (như sau đại dịch), do người dùng không cần chạm tay vào thiết bị.



Hình 1. 2: Minh họa một số ứng dụng của nhận diện khuôn mặt

Như vậy từ an ninh, giao thông, tài chính đến đời sống hằng ngày, công nghệ nhận diện khuôn mặt đã chứng tỏ tính hữu ích và ngày càng phổ biến. Tất nhiên, đi kèm với sự thuận tiện là những yêu cầu cao hơn về độ chính xác và an toàn của hệ thống, đặc biệt trong các ứng dụng nhạy cảm như tài chính (eKYC) hay giám sát công cộng. Phần tiếp theo sẽ tìm hiểu sâu hơn về quy trình eKYC – một ứng dụng quan trọng của nhận diện khuôn mặt trong xác thực danh tính khách hàng

1.4. Quy trình eKYC và ứng dụng nhận diện khuôn mặt

eKYC (Electronic Know Your Customer) [30] là quy trình định danh khách hàng điện tử, cho phép xác thực danh tính người dùng từ xa thông qua các phương tiện kỹ thuật số. Đây là bước phát triển từ quy trình KYC truyền thống (khách hàng phải xuất trình giấy tờ tùy thân trực tiếp) sang mô hình trực tuyến hoàn toàn. eKYC ngày càng được áp dụng phổ biến trong lĩnh vực tài chính – ngân hàng, viễn thông, bảo hiểm, y tế..., nhờ tính tiện lợi và khả năng mở rộng phạm vi phục vụ khách

hàng. Chẳng hạn, các ngân hàng sử dụng eKYC để mở tài khoản trực tuyến cho khách hàng mới mà không cần họ đến quầy giao dịch; các nhà mạng viễn thông định danh thuê bao khi bán SIM điện tử; bệnh viện xác thực danh tính bệnh nhân trong khám chữa bệnh từ xa. Lợi ích của eKYC bao gồm: rút ngắn thời gian xử lý (từ vài ngày xuống còn vài phút), giảm chi phí nhân lực, mở rộng được dịch vụ ra phạm vi địa lý xa, và tăng độ chính xác nhờ ứng dụng các công nghệ tự động như nhận dạng ký tự (OCR) và nhận diện khuôn mặt.

Quy trình eKYC điển hình thường bao gồm các bước chính sau đây [30]:

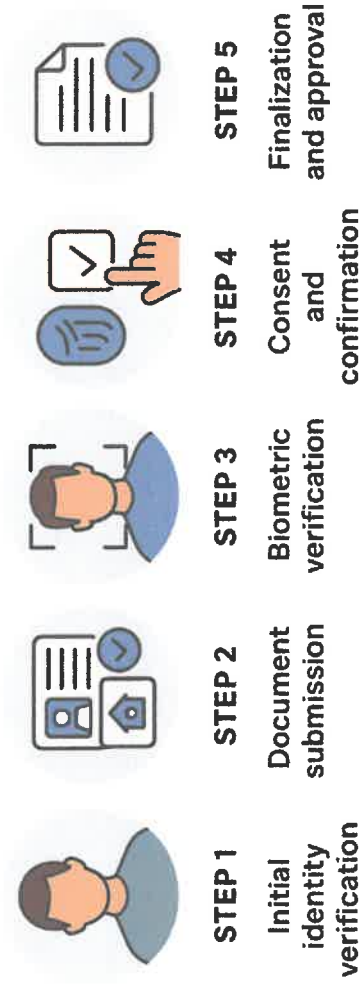
- **Thu thập thông tin khách hàng:** Khách hàng bắt đầu bằng việc cung cấp các thông tin nhận dạng cá nhân cơ bản (họ tên, ngày sinh, số CMND/CCCD/Hộ chiếu, địa chỉ, ...). Thông tin này có thể được điền vào biểu mẫu trực tuyến hoặc trích xuất từ ảnh chụp giấy tờ tùy thân bằng kỹ thuật OCR. Hệ thống sẽ kiểm tra chéo thông tin với cơ sở dữ liệu thích hợp (ví dụ: đối chiếu số ID với cơ quan nhà nước) để đảm bảo tính hợp lệ.
- **Gửi và xác thực giấy tờ tùy thân:** Khách hàng chụp ảnh (hoặc quay video) các giấy tờ tùy thân được yêu cầu, phổ biến là chứng minh nhân dân/căn cước công dân hoặc hộ chiếu, và tải lên hệ thống. Hệ thống eKYC sẽ tự động kiểm tra tính xác thực của giấy tờ này – ví dụ: kiểm tra các dấu hiệu chống giả, đọc thông tin trên thẻ và so sánh với thông tin khách hàng đã khai. Công nghệ thị giác máy tính được áp dụng để phát hiện giấy tờ giả mạo hoặc chỉnh sửa.

- **Xác thực sinh trắc học (nhận diện khuôn mặt):** Đây là bước quan trọng nhằm đảm bảo người đang thực hiện eKYC chính là chủ nhân của giấy tờ tùy thân đã cung cấp. Hệ thống sẽ yêu cầu khách hàng cung cấp dữ liệu sinh trắc học, thường là ảnh selfie khuôn mặt hoặc video ngắn. Sau đó, thuật toán nhận diện khuôn mặt sẽ đối chiếu khuôn mặt của người dùng với ảnh trên giấy tờ tùy thân đã gửi để kiểm tra xem có trùng khớp hay không. Đồng thời, để ngăn ngừa gian lận, hệ thống eKYC thường tích hợp thêm cơ chế phát hiện giả mạo khuôn mặt (liveness detection) nhằm đảm

bảo người đang xác thực là người thật đang trực tiếp trước camera, chứ không phải ảnh chụp hoặc video phát lại. Ví dụ, ứng dụng có thể yêu cầu người dùng nhìn theo hướng dẫn, chớp mắt hoặc đọc một dãy số ngẫu nhiên để xác minh tính sống động. Đây chính là ứng dụng trực tiếp của công nghệ nhận diện khuôn mặt trong eKYC dùng để xác thực danh tính sinh trắc học từ xa một cách nhanh chóng và chính xác.

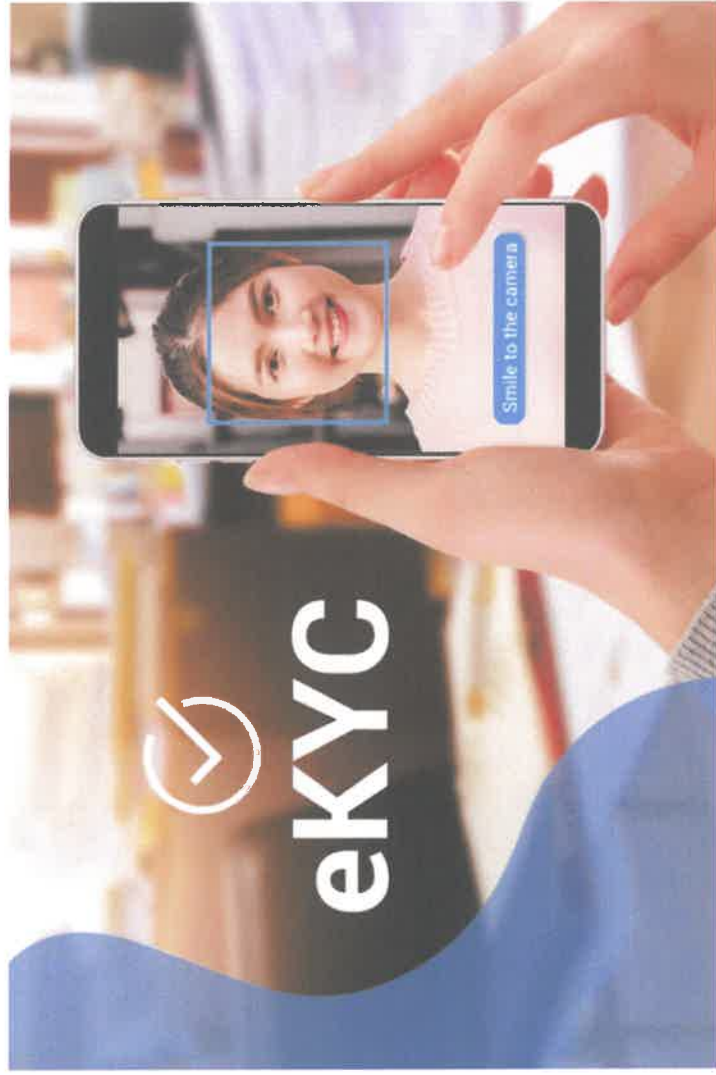
- **Xác nhận và đồng ý (Consent):** Sau khi các bước xác minh hoàn tất, hệ thống sẽ yêu cầu khách hàng xác nhận lại toàn bộ thông tin đã cung cấp là đúng và đồng ý với các điều khoản về eKYC. Bước này nhằm đảm bảo khách hàng tự nguyện tham gia quy trình định danh điện tử và hiểu rõ quyền lợi, trách nhiệm cũng như việc dữ liệu của họ sẽ được sử dụng ra sao.
- **Phê duyệt và lưu trữ:** Toàn bộ hồ sơ eKYC của khách hàng sẽ được gửi tới bộ phận có thẩm quyền (ví dụ: hệ thống tự động hoặc nhân viên ngân hàng) để phê duyệt lần cuối. Nếu đạt yêu cầu, khách hàng sẽ được xác thực thành công và có thể truy cập dịch vụ (mở tài khoản, kích hoạt thẻ SIM, ...). Dữ liệu định danh điện tử của khách hàng sau đó được lưu trữ an toàn phục vụ cho kiểm tra sau này và tuân thủ quy định (ví dụ: nhật ký giao dịch, báo cáo cơ quan quản lý nếu cần).

eKYC PROCESS



Hình 1. 3: Quy trình eKYC điển hình

Trong quy trình trên, nhận diện khuôn mặt giữ vai trò then chốt ở bước xác thực sinh trắc học (bước 3). Công nghệ này giúp kết nối dữ liệu số (ảnh trên giấy tờ) với con người thực một cách tự động. Nhờ nhận diện khuôn mặt, eKYC có thể loại bỏ sự can thiệp của con người trong khâu xác minh, vừa tăng tốc độ xử lý vừa giảm sai sót chủ quan. Theo phân loại, phương pháp eKYC sử dụng sinh trắc học khuôn mặt như trên được gọi là eKYC xác thực sinh trắc [30]. Đây là phương thức đang được nhiều ngân hàng, công ty tài chính và viễn thông áp dụng do tính an toàn cao: khuôn mặt mỗi người là duy nhất, khó làm giả hơn thông tin giấy tờ, và việc đối chiếu khuôn mặt bằng AI cho độ chính xác rất cao (thường trên 99% đối với ảnh rõ nét). Một số hình thức eKYC khác cũng tồn tại như: eKYC qua video call (nhân viên gọi video trực tiếp để đối chiếu khách hàng), eKYC dựa trên danh tính số (xác minh qua tài khoản định danh có sẵn, chẳng hạn dùng thông tin từ nhà mạng hoặc tài khoản chính phủ điện tử), nhưng nhìn chung eKYC sinh trắc học bằng khuôn mặt đang là xu hướng chủ đạo nhờ sự phát triển của công nghệ AI.



Hình 1. 4: Nhận diện khuôn mặt trong eKYC

Tại Việt Nam, nhiều ngân hàng (VPBank, TPBank, ...) đã triển khai mở tài khoản trực tuyến bằng eKYC. Khách hàng chỉ cần dùng ứng dụng ngân hàng chụp ảnh CMND/CCCD và chụp một video ngân khuôn mặt theo hướng dẫn; hệ thống sẽ tự động nhận dạng và đối chiếu khuôn mặt với ảnh trên CMND để xác thực. Toàn bộ quy trình diễn ra trong vài phút, sau đó tài khoản ngân hàng được mở mà không cần khách hàng đến quầy trực tiếp. Đây chính là sự kết hợp hiệu quả giữa thị giác máy tính (nhận dạng ký tự, nhận diện khuôn mặt) và công nghệ số để đơn giản hóa thủ tục “Know Your Customer”.

Tuy eKYC mang lại nhiều lợi ích, việc triển khai cũng đòi hỏi chú ý đến các yếu tố bảo mật và pháp lý. Hệ thống phải đảm bảo an toàn dữ liệu cá nhân của khách hàng, chống lại các cuộc tấn công mạo danh tinh vi (như sử dụng ảnh giả, video deepfake), đồng thời tuân thủ các quy định pháp luật về bảo vệ dữ liệu và quyền riêng tư. Phần tiếp theo sẽ trình bày cụ thể hơn về những vấn đề bảo mật và pháp lý trong nhận diện khuôn mặt và eKYC.

1.5. Các vấn đề bảo mật và pháp lý liên quan

Mặc dù mang lại nhiều tiện ích, việc ứng dụng rộng rãi nhận diện khuôn mặt và eKYC cũng đặt ra không ít vấn đề về bảo mật thông tin và khuôn khổ pháp lý. Dưới đây là các khía cạnh cần quan tâm:

1.5.1. Vấn đề bảo mật và kỹ thuật

Về bảo mật kỹ thuật thường xảy ra các vấn đề sau:

Nguy cơ tấn công mạo danh (spoofing): Hệ thống nhận diện khuôn mặt có thể bị đánh lừa bởi các hình thức mạo danh như sử dụng ảnh chụp, video ghi sẵn hoặc thậm chí mặt nạ 3D của người khác. Đây là mối đe dọa nghiêm trọng đối với các ứng dụng bảo mật cao như mở khóa thiết bị hay eKYC, nơi kẻ gian có thể cố gắng giả mạo danh tính người dùng hợp pháp. Trong thực tế, các hệ thống nhận dạng khuôn mặt thương mại luôn tích hợp bước chống giả mạo (face anti-spoofing) trước khi nhận dạng nhằm đảm bảo khuôn mặt đầu vào là thật [2]. Các kỹ thuật chống spoofing bao gồm: yêu cầu người dùng thực hiện hành động ngẫu nhiên (như chớp mắt, quay đầu), sử dụng camera hồng ngoại đo độ sống của da, hoặc phân tích texture ảnh để

phát hiện điểm bất thường (ví dụ: độ phẳng của ảnh in). Nhiều nghiên cứu đã phát triển thuật toán chuyên biệt cho bài toán phát hiện khuôn mặt giả mạo. Chẳng hạn, Boulkenafet và cộng sự (2015) dùng phân tích màu sắc và kết cấu trên ảnh RGB để phân biệt mặt thật với ảnh in màu [2]; Patel và cộng sự (2016) đề xuất giải pháp bảo mật mở khóa khuôn mặt trên điện thoại bằng cách kết hợp camera thường và các cảm biến khác để xác minh đối tượng không phải là ảnh hoặc video giả. Ngoài ra, sự xuất hiện của deepfake (video/hình ảnh giả do AI tạo nên) trong những năm gần đây càng làm tăng nguy cơ qua mặt hệ thống nhận diện. Tội phạm có thể dùng deepfake để tạo video khuôn mặt chuyển động giống hệt nạn nhân nhằm đánh lừa eKYC. Để đối phó, các công ty đang phát triển công cụ phát hiện deepfake tự động dựa trên dấu hiệu nhiễu, artefact mà video deepfake để lại (ví dụ: bất thường ở vùng nháy mắt, vùng chuyển tiếp khuôn mặt với phông nền, v.v.). Cuộc chiến giữa kỹ thuật nhận diện và kỹ thuật giả mạo bằng AI (AI vs AI) vẫn đang tiếp diễn, đòi hỏi các hệ thống eKYC phải liên tục cập nhật các biện pháp bảo mật mới nhất.

Rò rỉ và lạm dụng dữ liệu khuôn mặt: Dữ liệu khuôn mặt (bao gồm ảnh, video, và đặc trưng khuôn mặt đã trích xuất) là dữ liệu sinh trắc học nhạy cảm. Nếu hệ thống không bảo vệ chặt chẽ, dữ liệu này có thể bị rò rỉ do tấn công mạng hoặc do nội gián, dẫn đến hệ lụy nghiêm trọng. Khác với mật khẩu có thể đổi, khi dữ liệu khuôn mặt của một người bị lộ, họ không thể thay đổi “mật khẩu sinh trắc” là khuôn mặt mình. Kẻ xấu có thể sử dụng dữ liệu khuôn mặt bị đánh cắp để mạo danh, vượt qua các hệ thống bảo mật khác hoặc xâm phạm đời tư của nạn nhân. Một ví dụ là công ty Clearview AI (Mỹ) từng thu thập hàng tỷ bức ảnh khuôn mặt từ internet và cung cấp dịch vụ nhận diện cho cơ quan pháp luật, nhưng cũng vấp phải nhiều chỉ trích và các vụ kiện do việc thu thập dữ liệu không xin phép người dùng. Do đó, các tổ chức triển khai nhận diện khuôn mặt và eKYC phải đầu tư mạnh vào an ninh mạng: mã hóa dữ liệu khuôn mặt trong cơ sở dữ liệu, kiểm soát chặt chẽ truy cập, thường xuyên kiểm toán hệ thống, và có phương án ứng phó nếu xảy ra sự cố rò rỉ (thông báo cho người dùng, vô hiệu hóa dữ liệu cũ,).

Độ chính xác và sai sót thuật toán: Mặc dù độ chính xác trung bình của nhận diện khuôn mặt đã rất cao, nhưng vẫn có trường hợp sai sót, đặc biệt là với dữ liệu phức tạp. Lỗi nhận dạng có thể chia làm hai loại: nhận nhầm người này thành người khác (false positive) hoặc không nhận ra đúng người (false negative). Trong bối cảnh an ninh, nhận nhầm có thể dẫn đến bất oan (nếu hệ thống nhận diện sai một người vô tội thành tội phạm truy nã), còn bỏ sót có thể bỏ lọt nguy cơ. Các nghiên cứu trước đây cũng chỉ ra thuật toán nhận diện khuôn mặt có thể thiên lệch (bias) về giới tính, chủng tộc nếu dữ liệu huấn luyện không đa dạng, dẫn đến độ chính xác thấp hơn với một số nhóm người nhất định. Đây là vấn đề kỹ thuật lẫn đạo đức mà người phát triển cần lưu ý khắc phục, bằng cách cải thiện tập dữ liệu huấn luyện cũng như thuật toán để đảm bảo công bằng và chính xác trên mọi đối tượng.

1.5.2. Vấn đề pháp lý và quyền riêng tư

a. *Nước ngoài*

Trên thế giới, vấn đề pháp lý và quyền riêng tư liên quan đến công nghệ nhận diện khuôn mặt và xác thực eKYC được quản lý nghiêm ngặt và đang trong quá trình hoàn thiện khung pháp lý. Một trong những văn bản quan trọng nhất là Quy định chung về bảo vệ dữ liệu cá nhân (General Data Protection Regulation - GDPR) của Liên minh châu Âu, có hiệu lực từ năm 2018. Theo GDPR, dữ liệu sinh trắc học, trong đó bao gồm dữ liệu nhận diện khuôn mặt, được coi là một dạng dữ liệu cá nhân đặc biệt nhạy cảm và yêu cầu sự bảo vệ nghiêm ngặt hơn so với các dữ liệu thông thường [9]. GDPR yêu cầu các tổ chức phải nhận được sự đồng ý rõ ràng (explicit consent) từ cá nhân trước khi thu thập và xử lý dữ liệu sinh trắc học của họ. Đồng thời, các tổ chức phải đảm bảo tính minh bạch trong việc xử lý dữ liệu, thông báo rõ ràng mục đích thu thập dữ liệu, thời gian lưu trữ, và biện pháp bảo vệ dữ liệu được áp dụng [9].

Ngoài ra, một số quốc gia và khu vực khác cũng có các quy định riêng biệt về việc xử lý dữ liệu sinh trắc học và nhận diện khuôn mặt. Tại Hoa Kỳ, luật pháp liên bang chưa có khung thống nhất, nhưng một số tiểu bang như Illinois đã ban hành Đạo

luật bảo vệ thông tin sinh trắc học (Biometric Information Privacy Act - BIPA). Luật này quy định rõ các tổ chức phải thông báo và nhận được sự đồng ý bằng văn bản từ người dùng trước khi thu thập dữ liệu sinh trắc như khuôn mặt, dấu vân tay. Vi phạm BIPA có thể dẫn đến những hình phạt nặng nề; ví dụ điển hình là Facebook đã phải trả khoản tiền phạt lên tới 650 triệu USD vào năm 2021 để dàn xếp vụ kiện tập thể liên quan đến việc sử dụng công nghệ nhận diện khuôn mặt mà không được phép [10].

Cùng với đó, một số quốc gia châu Âu như Luxembourg, Bỉ đã cấm sử dụng công nghệ nhận diện khuôn mặt tại các khu vực công cộng, nhằm bảo vệ quyền riêng tư cá nhân trước các nguy cơ giám sát hàng loạt và lạm dụng dữ liệu [11]. Hiện nay, Liên minh châu Âu đang tiếp tục xây dựng Đạo luật AI, nhằm giới hạn và quy định chặt chẽ hơn nữa việc ứng dụng nhận diện khuôn mặt theo thời gian thực trong giám sát công cộng. Qua đó, cho thấy rõ sự thận trọng và quan tâm cao của cộng đồng quốc tế đối với việc sử dụng công nghệ nhận diện khuôn mặt, nhằm cân bằng giữa lợi ích công nghệ và bảo vệ quyền riêng tư cá nhân.

b. Tại Việt Nam

Tại Việt Nam, vấn đề bảo vệ dữ liệu cá nhân và quyền riêng tư liên quan đến nhận diện khuôn mặt và eKYC cũng đã bắt đầu được chú trọng hơn. Ngày 17/4/2023, Chính phủ Việt Nam ban hành Nghị định số 13/2023/NĐ-CP về bảo vệ dữ liệu cá nhân, chính thức có hiệu lực từ tháng 7 năm 2023. Nghị định này được xem là một bước tiến quan trọng trong việc xây dựng khung pháp lý cho bảo vệ dữ liệu cá nhân tại Việt Nam, bao gồm cả dữ liệu sinh trắc học [12]. Theo nghị định này, dữ liệu sinh trắc học (như dữ liệu khuôn mặt) được coi là dữ liệu nhạy cảm, và việc thu thập, xử lý dữ liệu này phải có sự đồng ý rõ ràng và được thông báo đầy đủ cho cá nhân có liên quan. Các tổ chức, doanh nghiệp phải thông báo rõ mục đích xử lý, biện pháp bảo vệ dữ liệu và thời gian lưu trữ. Đồng thời, nghị định yêu cầu phải có các biện pháp bảo vệ thích hợp, chống lại nguy cơ rò rỉ, mất mát hoặc sử dụng sai mục đích.

Bên cạnh Nghị định 13/2023/NĐ-CP, Việt Nam cũng đang từng bước hoàn thiện thêm các văn bản hướng dẫn thực hiện, đặc biệt liên quan đến ứng dụng công nghệ số như eKYC. Ngân hàng Nhà nước Việt Nam đã ban hành Thông tư số 16/2020/TT-NHNN cho phép các ngân hàng thương mại thực hiện định danh khách hàng điện tử (eKYC) từ xa, trong đó nhận diện khuôn mặt được coi là giải pháp chính để xác thực danh tính khách hàng [13]. Tuy nhiên, trong quá trình áp dụng thực tiễn, nhiều tổ chức vẫn gặp khó khăn do thiếu các hướng dẫn chi tiết hơn về tiêu chuẩn kỹ thuật, phương pháp chống giả mạo khuôn mặt, và vấn đề bảo mật an toàn thông tin.

Ngoài các quy định pháp lý, vấn đề quyền riêng tư cũng bắt đầu được cộng đồng và các nhà làm luật tại Việt Nam quan tâm hơn. Một số ý kiến cho rằng Việt Nam cần sớm xây dựng các tiêu chuẩn kỹ thuật cụ thể và tăng cường giám sát hoạt động thu thập dữ liệu cá nhân, đặc biệt trong bối cảnh công nghệ AI và nhận diện khuôn mặt ngày càng phổ biến [14]. Điều này nhằm hạn chế tối đa các nguy cơ tiềm ẩn từ việc lạm dụng dữ liệu cá nhân hoặc giám sát quá mức, đồng thời đảm bảo việc ứng dụng công nghệ tuân thủ chặt chẽ quyền riêng tư của người dân.

Việt Nam đang trong quá trình xây dựng và hoàn thiện khung pháp lý để quản lý công nghệ nhận diện khuôn mặt và các ứng dụng liên quan như eKYC, với mục tiêu đảm bảo cân bằng giữa phát triển công nghệ và bảo vệ quyền riêng tư của cá nhân

1.6. Kết luận chương

Nội dung chương 1 đã giới thiệu về bài toán nhận diện khuôn mặt, sau đó trình bày về các phương pháp nhận diện khuôn mặt và ứng dụng của nó. Bên cạnh đó, chương 1 cũng trình bày về quy trình eKYC, các vấn đề bảo mật và pháp lý liên quan. Ở chương tiếp theo, đề án sẽ trình bày về các mô hình nhận diện khuôn mặt để ứng dụng trong xác thực eKYC.

CHƯƠNG 2. NGHIÊN CỨU MÔ HÌNH NHẬN DIỆN KHUÔN MẶT

Chương này sẽ trình bày về các khái niệm liên quan đến các mô hình nhận diện khuôn mặt sử dụng học sâu. Sau đó sẽ trình bày về các tiêu chí đánh giá mô hình nhận diện khuôn mặt. Cuối cùng sẽ lựa chọn mô hình nhận diện khuôn mặt cho xác thực eKYC của ngân hàng.

2.1. Khảo sát một số mô hình nhận diện khuôn mặt sử dụng học sâu

Nhận diện khuôn mặt đã trải qua nhiều giai đoạn phát triển, đặc biệt với sự bùng nổ của học sâu kể từ năm 2014. Các mô hình học sâu được giới thiệu trong giai đoạn này đã mang lại những bước tiến lớn, đưa hiệu suất nhận diện khuôn mặt tiệm cận hoặc vượt ngưỡng con người. Nội dung dưới đây trình bày các mô hình nhận diện khuôn mặt dựa trên học sâu tiêu biểu.

2.1.1. DeepFace (2014)

Mô hình DeepFace, được đề xuất bởi Taigman và cộng sự vào năm 2014, là một trong những mô hình đầu tiên áp dụng học sâu để giải quyết bài toán nhận diện khuôn mặt với hiệu suất cao [4]. Mô hình này kết hợp hai bước quan trọng: căn chỉnh khuôn mặt ba chiều và trích xuất đặc trưng bằng mạng nơ-ron tích chập (CNN) nhiều lớp. Cụ thể, DeepFace thực hiện chuẩn hóa hình dạng khuôn mặt bằng cách căn chỉnh các điểm đặc trưng lên một mô hình 3D trung bình trước khi đưa ảnh vào mạng CNN có khoảng 120 triệu tham số để học biểu diễn đặc trưng.

Bằng việc huấn luyện trên tập dữ liệu lớn hơn 4 triệu ảnh, DeepFace đã đạt độ chính xác 97,35% trên tập Labeled Faces in the Wild (LFW), lần đầu tiên thu hẹp đáng kể khoảng cách với khả năng nhận diện khuôn mặt của con người [4]. Công trình này đã chứng minh tính khả thi của việc học trực tiếp các đặc trưng khuôn mặt thông qua các mạng học sâu mà không cần thiết kế thủ công.

2.1.2. DeepID (2014)

Cùng thời điểm, nhóm nghiên cứu của Sun và cộng sự đã giới thiệu mô hình DeepID, với cách tiếp cận khác biệt dựa trên phân loại danh tính [5]. Thay vì đơn

thuần học đặc trưng phân biệt cho từng cặp ảnh, DeepID huấn luyện mạng CNN để phân loại khuôn mặt thành 10.000 danh tính khác nhau, từ đó tạo ra các đặc trưng giàu tính phân biệt.

DeepID sử dụng nhiều mạng CNN nhỏ học các vùng cục bộ trên khuôn mặt, sau đó hợp nhất các đặc trưng học được. Các đặc trưng trung gian (DeepID features) sau khi huấn luyện được sử dụng cho bài toán xác thực khuôn mặt mới, không thuộc tập huấn luyện. Phương pháp này giúp DeepID đạt độ chính xác 97,45% trên LFW, tương đương với DeepFace nhưng với quy trình huấn luyện phân loại danh tính quy mô lớn [5]. Các phiên bản cải tiến như DeepID2 và DeepID3 tiếp tục cải thiện hiệu năng nhờ kết hợp nhiều mạng và sử dụng tập dữ liệu mở rộng.

2.1.3. FaceNet (2015)

Năm 2015, Schroff và nhóm nghiên cứu đã giới thiệu FaceNet, mang đến một cách tiếp cận mới khi học trực tiếp ánh xạ khuôn mặt vào một không gian những Euclid [6]. Không giống DeepFace hay DeepID, FaceNet không tối ưu hóa bài toán phân loại danh tính, mà học một hàm nhúng sao cho khoảng cách giữa các vector đặc trưng phản ánh chính xác mức độ giống nhau về danh tính.

Mô hình FaceNet sử dụng mạng CNN sâu dựa trên kiến trúc Inception và được huấn luyện bằng hàm mất mát triplet loss. Loss này được xây dựng từ các bộ ba ảnh: một anchor, một positive (cùng người) và một negative (khác người), nhằm tối ưu hóa khoảng cách trong không gian những theo hướng làm giảm khoảng cách anchor-positive và tăng khoảng cách anchor-negative [6].

Với chiến lược này và tập dữ liệu huấn luyện cực lớn (~200 triệu ảnh từ ~8 triệu người), FaceNet đạt độ chính xác 99,63% trên LFW và 95,12% trên tập Youtube Faces, thiết lập kỷ lục mới tại thời điểm đó [6]. Công trình này đã đặt nền tảng cho các phương pháp nhận diện khuôn mặt dựa trên embedding vector thay vì phân lớp trực tiếp.

2.1.4. ArcFace (2019)

Trong những năm tiếp theo, hướng nghiên cứu tập trung vào cải tiến hàm mất mát để tăng tính phân biệt giữa các vector đặc trưng. Năm 2019, Deng và cộng sự đã đề xuất ArcFace, sử dụng hàm mất mát Additive Angular Margin Loss nhằm tối ưu hóa trực tiếp khoảng cách góc giữa các lớp khuôn mặt trên hypersphere [7].

ArcFace cải tiến trên nền tảng của SphereFace và CosFace, đơn giản hóa việc thêm khoảng cách margin bằng cách cộng một hằng số vào góc phân loại. Nhờ cách tiếp cận hình học rõ ràng và đơn giản trong huấn luyện, ArcFace đã đạt 99,83% accuracy trên LFW và dẫn đầu trên nhiều benchmark nhận diện khuôn mặt quy mô lớn như MegaFace [7]. Từ đó, ArcFace trở thành lựa chọn mặc định cho nhiều hệ thống nhận diện khuôn mặt hiện đại.

2.1.5. Các mô hình gần đây

Tiếp nối ArcFace, nhiều công trình trong những năm gần đây tập trung vào việc cải thiện kỹ thuật huấn luyện và khai thác thêm thông tin từ dữ liệu.

CurricularFace [19] áp dụng kỹ thuật học có giáo trình, điều chỉnh trọng số mẫu trong hàm mất mát theo độ khó của mẫu qua từng giai đoạn huấn luyện, nhằm nâng cao hiệu quả học tập và khả năng phân biệt với dữ liệu nhiễu.

MagFace [20] giới thiệu một cơ chế trong đó độ lớn (magnitude) của embedding vector phản ánh chất lượng khuôn mặt đầu vào. Điều này không những giúp cải thiện độ chính xác nhận diện mà còn cho phép đánh giá nhanh độ tin cậy của mẫu đầu vào. Ngoài ra, các nghiên cứu mới đang khám phá các kiến trúc transformer thay thế CNN truyền thống, cũng như khai thác học tự giám sát (self-supervised learning) để giảm phụ thuộc vào dữ liệu gán nhãn, hứa hẹn các hướng phát triển tiếp theo cho nhận diện khuôn mặt.

2.2. Các tiêu chí đánh giá mô hình nhận diện khuôn mặt

Trong các nghiên cứu về nhận diện khuôn mặt, việc đánh giá hiệu năng mô hình cần dựa trên những tiêu chí định lượng chuẩn hóa nhằm đảm bảo khả năng so sánh, phân tích và tối ưu hệ thống một cách khách quan. Các chỉ số phổ biến bao gồm

độ chính xác (Accuracy), tỷ lệ chấp nhận sai (False Acceptance Rate - FAR), tỷ lệ từ chối sai (False Rejection Rate - FRR), đường cong ROC (Receiver Operating Characteristic Curve) và điểm lỗi cân bằng (Equal Error Rate - EER). Dưới đây là chi tiết từng tiêu chí cùng công thức liên quan.

2.2.1. Độ chính xác (Accuracy)

Độ chính xác (Accuracy) là thước đo tổng quát đánh giá tỷ lệ các dự đoán đúng (bao gồm cả dương tính và âm tính) so với tổng số dự đoán. Accuracy thể hiện khả năng tổng thể của hệ thống trong việc phân loại chính xác các đối tượng.

Công thức xác định độ chính xác:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

Trong đó:

- TP (True Positive): Số lượng mẫu thực sự dương tính được phân loại đúng.
- TN (True Negative): Số lượng mẫu thực sự âm tính được phân loại đúng.
- FP (False Positive): Số lượng mẫu thực sự âm tính nhưng bị phân loại nhầm thành dương tính.
- FN (False Negative): Số lượng mẫu thực sự dương tính nhưng bị phân loại nhầm thành âm tính.

Ý nghĩa: Accuracy phản ánh xác suất để một mẫu bất kỳ được hệ thống dự đoán đúng nhãn thực tế. Giá trị Accuracy dao động từ 0 đến 1 (hoặc biểu diễn dưới dạng phần trăm từ 0% đến 100%). Accuracy cao cho thấy mô hình có khả năng tổng quát tốt đối với tập dữ liệu thử nghiệm. Tuy nhiên, trong các bài toán mà dữ liệu mất cân bằng (số mẫu dương tính và âm tính chênh lệch nhiều), Accuracy đơn thuần có thể không phản ánh đầy đủ hiệu năng thực sự của hệ thống.

2.2.2. Độ đo F1 (F1 Score)

Độ đo F1 (F1 Score) là một chỉ số tổng hợp giữa độ chính xác theo từng lớp (Precision) và khả năng bao phủ (Recall) của mô hình, nhằm cân bằng giữa hai loại

lỗi khác nhau. F1 Score đặc biệt hữu ích trong các bài toán phân loại nhị phân, nơi cần đánh giá sự cân đối giữa việc nhận dạng đúng mẫu thuộc lớp mục tiêu và việc không nhận nhầm mẫu ngoài lớp mục tiêu.

- Độ chính xác theo lớp (Precision):

$$Precision = \frac{TP}{TP + FP}$$

- Khả năng bao phủ (Recall):

$$Recall = \frac{TP}{TP + FN}$$

Sau đó, F1 Score được định nghĩa là trung bình điều hòa (harmonic mean) của Precision và Recall:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Ý nghĩa: F1 Score đo lường mức độ cân bằng giữa Precision và Recall. Một mô hình có F1 Score cao chứng tỏ rằng nó không chỉ nhận diện đúng số lượng lớn các mẫu thuộc lớp mục tiêu (Precision cao), mà còn bao phủ được phần lớn số mẫu cần nhận diện (Recall cao).

- Nếu Precision cao nhưng Recall thấp, hoặc ngược lại, F1 Score sẽ giảm, phản ánh sự mất cân đối trong hiệu năng mô hình
- Nếu cả Precision và Recall đều cao, F1 Score cũng sẽ cao, thể hiện hệ thống có độ chính xác và độ bao phủ tốt đồng thời.

Giá trị F1 Score nằm trong khoảng $[0,1]$, với 1 là mức hoàn hảo (Precision = Recall = 1) và 0 là mức kém nhất.

2.2.3. Tỷ lệ chấp nhận sai (*False Acceptance Rate - FAR*)

Tỷ lệ chấp nhận sai (FAR) đo lường khả năng hệ thống phân loại nhầm mẫu âm tính thành dương tính. FAR thường được sử dụng trong các hệ thống xác thực, nơi yêu cầu hạn chế việc nhận nhầm đối tượng không hợp lệ.

Công thức tính FAR:

$$FAR = \frac{FP}{FP + TN}$$

Ý nghĩa: FAR đại diện cho xác suất mà hệ thống sai sót kiểu I (Type I Error), tức là chấp nhận nhầm một mẫu không hợp lệ. Giá trị FAR thấp chứng tỏ hệ thống có khả năng phân biệt tốt giữa các đối tượng khác biệt và giảm thiểu nguy cơ sai sót nghiêm trọng trong ứng dụng thực tế. Khi thiết kế hệ thống nhận diện khuôn mặt có yêu cầu an ninh cao, FAR là chỉ số cần được tối ưu hóa ở mức thấp nhất.

2.2.4. Tỷ lệ từ chối sai (*False Rejection Rate - FRR*)

Tỷ lệ từ chối sai (FRR) đo lường khả năng hệ thống phân loại nhầm mẫu dương tính thành âm tính. Đây là loại lỗi ngược lại với FAR và thường liên quan đến mức độ khắt khe của hệ thống trong việc chấp nhận các đối tượng hợp lệ.

Công thức tính FAR:

$$FRR = \frac{FN}{FN + TP}$$

Ý nghĩa: FRR phản ánh xác suất xảy ra sai sót kiểu II (Type II Error), tức là từ chối nhầm một mẫu hợp lệ. FRR thấp cho thấy hệ thống có khả năng chấp nhận đúng đối tượng thực sự. Tuy nhiên, giảm FRR một cách cực đoan có thể làm tăng FAR, do đó trong thiết kế và tối ưu hóa, cần đạt được sự cân bằng hợp lý giữa FAR và FRR tùy theo yêu cầu ứng dụng. Các hệ thống chú trọng trải nghiệm người dùng thường yêu cầu FRR thấp để hạn chế gây phiền hà cho đối tượng hợp lệ.

2.2.5. Đường cong ROC (*Receiver Operating Characteristic Curve*)

Đường cong ROC mô tả mối quan hệ giữa True Accept Rate (TAR) và False Acceptance Rate (FAR) khi thay đổi ngưỡng quyết định của hệ thống.

Công thức tính True Accept Rate (TAR):

$$TAR = 1 - FRR = \frac{TP}{TP + FN}$$

Để xây dựng đường cong ROC, ta thay đổi liên tục ngưỡng phân loại và tính toán các giá trị FAR và TAR tương ứng. Các điểm (FAR, TAR) thu được sẽ tạo thành đường cong ROC.

Ý nghĩa: đường cong ROC thể hiện khả năng phân biệt của mô hình giữa hai lớp (dương tính và âm tính) trên toàn bộ dải ngưỡng. Một hệ thống lý tưởng có đường

cong ROC tiến sát góc trên trái của đồ thị, tương ứng với TAR cao và FAR thấp. Giá trị Area Under Curve (AUC) cũng thường được tính toán để tóm tắt chất lượng phân loại: AUC càng gần 1, hiệu suất mô hình càng tốt.

2.2.6. Điểm lỗi cân bằng (*Equal Error Rate - EER*)

Điểm lỗi cân bằng (EER) là giá trị tại đó tỷ lệ chấp nhận sai (FAR) bằng với tỷ lệ từ chối sai (FRR). EER được xác định bằng cách tìm ngưỡng phân loại sao cho:

$$FAR = FRR$$

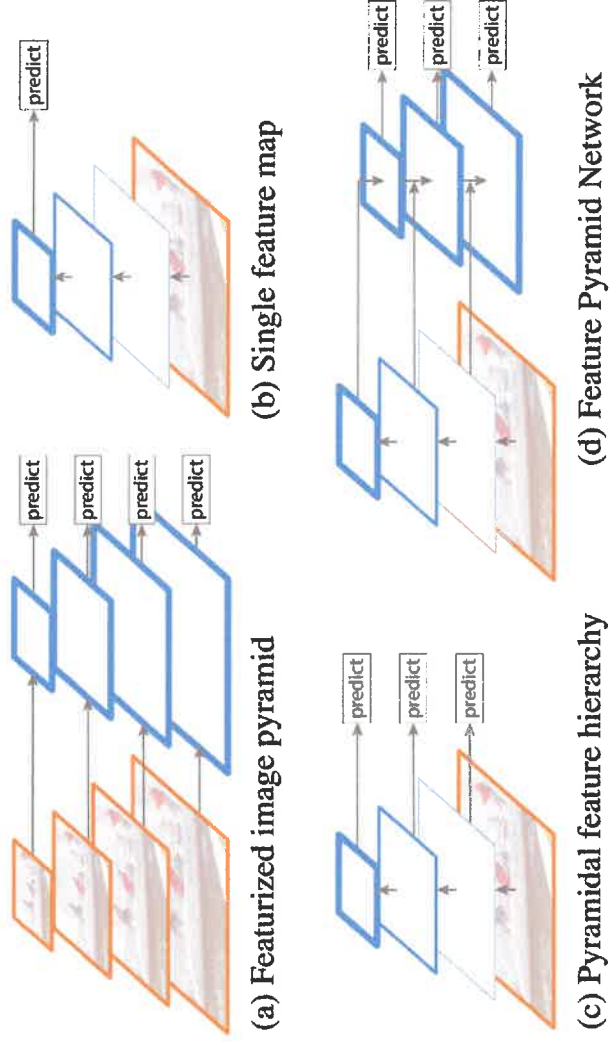
Ý nghĩa: EER cung cấp một chỉ số đơn giản nhưng hiệu quả để so sánh các hệ thống nhận diện khuôn mặt. Giá trị EER thấp cho thấy mô hình có khả năng cân bằng tốt giữa việc chấp nhận đúng và từ chối đúng, đồng thời giảm thiểu cả hai loại lỗi. Trong các báo cáo đánh giá hệ thống sinh trắc học, EER thường được sử dụng như một thước đo chính thức để so sánh hiệu năng giữa các phương pháp.

2.3. Một số yếu tố trong bài toán phát hiện khuôn mặt

2.3.1. Kim tự tháp ảnh so với kim tự tháp đặc trưng

Các bài toán phát hiện khuôn mặt truyền thống được giải quyết bằng mô hình cửa sổ trượt kết hợp với bộ phân loại tuyến tính trên toàn ảnh. Các phương pháp trích xuất đặc trưng được minh họa như trong Hình 2.1. Tiêu biểu là phương pháp Viola-Jones, trong đó mô hình sử dụng chuỗi phân tầng (cascade) để loại bỏ các vùng không chứa khuôn mặt từ kim tự tháp ảnh (Hình 2.1b), tức ảnh được nén đa tỉ lệ. Mô hình này đạt hiệu suất thời gian thực và từng là chuẩn mực trong suốt một thời gian dài.

Các phương pháp kim tự tháp ảnh và kim tự tháp đặc trưng [24]: (a) Kim tự tháp đặc trưng hình ảnh; (b) Bản đồ đặc trưng đơn; (c) Phân cấp tính năng hình kim tự tháp; (d) Mạng kim tự tháp đặc trưng. Tuy nhiên, với sự phát triển của mạng neuron tích chập (CNN), cách tiếp cận hiện đại chuyển sang kim tự tháp đặc trưng (feature pyramid network). Thay vì xử lý ảnh ở nhiều tỉ lệ, mô hình trích xuất đặc trưng từ các tầng khác nhau của mạng CNN, kết hợp với cơ chế anchor trượt (sliding-anchor) để bao phủ đa dạng kích thước và vị trí khuôn mặt. Cách tiếp cận này mang lại hiệu quả cao hơn cả về độ chính xác lẫn tốc độ xử lý.



Hình 2.1: Các phương pháp kim tự tháp ảnh và kim tự tháp đặc trưng

2.3.2. Mô hình hai giai đoạn và mô hình một giai đoạn

Các phương pháp phát hiện khuôn mặt hiện nay kế thừa nhiều thành tựu từ các mô hình phát hiện vật thể tổng quát và có thể chia thành hai nhóm:

- **Phương pháp hai giai đoạn:** điển hình như Faster R-CNN
- **Phương pháp một giai đoạn:** ví dụ như các mô hình SSD hay RetinaNet thực hiện phân loại và hồi quy trực tiếp trên lưới đặc trưng, tiết kiệm thời gian xử lý đáng kể

Phương pháp hai giai đoạn sử dụng cơ chế “đề xuất và tinh chỉnh” nên đạt độ chính xác định vị cao. Ngược lại, phương pháp một giai đoạn thực hiện lấy mẫu dày đặc các vị trí và tỉ lệ khuôn mặt, dẫn đến tình trạng mất cân bằng nghiêm trọng giữa mẫu dương và âm trong quá trình huấn luyện. Để xử lý vấn đề này, các phương pháp lấy mẫu chọn lọc và tái trọng số đã được áp dụng rộng rãi. So với phương pháp hai giai đoạn, phương pháp một giai đoạn hiệu quả hơn và có tỉ lệ phát hiện lại (recall) cao hơn, nhưng đi kèm là tỉ lệ dương tính giả cao hơn và độ chính xác định vị thấp hơn.

2.3.3. Mô hình hóa ngữ cảnh

Với các khuôn mặt nhỏ, bị che khuất hoặc chất lượng kém, việc mô hình hóa ngữ cảnh xung quanh khuôn mặt trở nên quan trọng. Các mô hình như SSH [21] và PyramidBox [22] mở rộng vùng cảm nhận (receptive field) thông qua các module ngữ cảnh đặt lên kim tự tháp đặc trưng. Ngoài ra, để xử lý biến dạng phi tuyến (non-rigid), Deformable Convolutional Networks (DCN) [23] sử dụng các tầng biến dạng giúp mô hình linh hoạt thích ứng với biến đổi hình học phức tạp.

Minh chứng cho hiệu quả của mô hình hóa ngữ cảnh là kết quả của đội chiến thắng WIDER Face Challenge 2018 [25] cho thấy sự kết hợp giữa mở rộng ngữ cảnh cứng (bằng receptive field) và ngữ cảnh mềm (bằng biến dạng) có thể nâng cao rõ rệt hiệu suất phát hiện.

2.3.4. Học đa tác vụ

Nhằm tăng độ chính xác và độ tin cậy trong phát hiện khuôn mặt, nhiều mô hình hiện đại tích hợp thêm nhánh dự đoán landmark hoặc các thông tin hình học khác, biến bài toán trở thành học đa tác vụ (multi-task learning). Việc căn chỉnh khuôn mặt chính xác (alignment) không chỉ cải thiện phát hiện mà còn nâng cao chất lượng đặc trưng cho các bài toán kế tiếp như nhận dạng.

Mask R-CNN [26] là một ví dụ điển hình, nơi việc thêm nhánh mask segmentation song song với phát hiện làm tăng đáng kể độ chính xác tổng thể. DensePose [27] kế thừa kiến trúc này để thu được các nhãn chi tiết và toạ độ cấu trúc con người từ ảnh.

2.4. Lựa chọn mô hình nhận diện khuôn mặt cho xác thực eKYC của ngân hàng

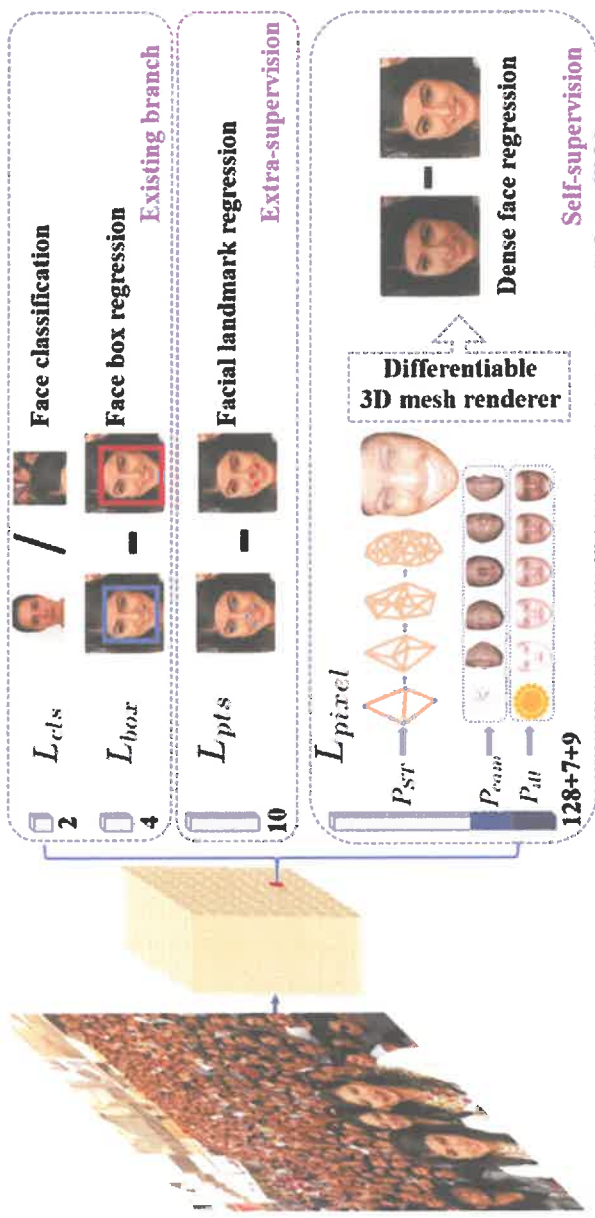
Phương pháp đề án đề xuất, tận dụng các kỹ thuật học sâu tiên tiến dùng để nhận diện khuôn mặt cho xác thực eKYC của ngân hàng bao gồm 5 nội dung chính:

(1) Phát hiện khuôn mặt, (2) Trích xuất đặc trưng, (3) So sánh độ tương đồng của các vector đặc trưng, (4) Mạng chú ý đồ thị (Graph Attention Network - GAT), (5) Phát hiện khuôn mặt giả mạo

2.4.1. Dữ liệu đầu vào và đầu ra của mô hình

RetinaFace nhận vào một ảnh RGB với kích thước $H \times W \times 3$, đã được chuẩn hóa (mean-std của ImageNet) để đảm bảo ổn định khi huấn luyện. Mạng sinh ra tại mỗi vị trí của các bản đồ đặc trưng một tập anchors với nhiều kích thước (ví dụ 16^2 , 32^2 , 64^2) và tỉ lệ khung hình cố định, giúp phủ mọi kích thước khuôn mặt. Đầu ra cuối cùng sau một lần forward với mỗi anchor positive gồm:

- (1) Face score (Điểm số khuôn mặt) – xác suất tồn tại khuôn mặt ở trong ảnh đầu vào
- (2) Bounding box (Hộp bao) – bốn giá trị offset Δx , Δy , Δw , Δh so với anchor
- (3) Landmark gồm 5 điểm mốc của một khuôn mặt chuẩn hóa theo anchor
- (4) Dense 3D mesh (khi dùng head pixel-wise) – tham số shape, camera và illumination để tái tạo lưới mặt 3D

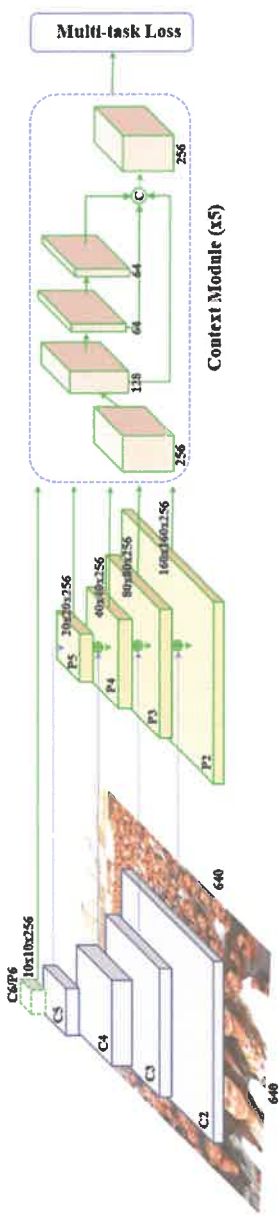


Hình 2. 2: Mô tả về đầu vào và đầu ra của mô hình RetinaFace

2.4.2. Kiến trúc mô hình RetinaFace

Kiến trúc mô hình RetinaFace được mô tả như trong Hình 2.3, bao gồm ba thành phần chính: Khối Backbone sử dụng các mô hình CNN để trích xuất các đặc trưng của hình ảnh đầu vào; Khối mạng đặc trưng kim tự tháp (Feature Pyramid Network) để trích xuất đặc trưng ở các kích thước ảnh khác nhau; cuối cùng là Khối

Multi-task Head để trả về đầu ra bao gồm face score, bounding box, landmark, 3D mesh. Phần backbone của RetinaFace thường sử dụng Resnet50 [16] hoặc Resnet101 [16] đã được pre-trained trên ImageNet [17], đảm bảo trích xuất những đặc trưng hình ảnh giàu ý nghĩa. Ba cấp feature map C3, C4, C5 được xuất ra từ các block conv3_x, conv4_x và conv5_x. Tiếp đó, mạng đặc trưng kim tự tháp (FPN) kết hợp cơ chế top-down và lateral connections để tạo ra ba bản đồ đa tỉ lệ P3, P4, P5, giữ lại cả thông tin về chi tiết nhỏ lẫn ngữ cảnh toàn cục. Trên mỗi Pn, các nhánh convolution 3×3 , 5×5 và 7×7 chạy song song, mở rộng receptive field mà không làm giảm độ phân giải, cải thiện khả năng phát hiện các khuôn mặt nhỏ và bị che khuất.



Hình 2. 3: Kiến trúc mô hình RetinaFace

Với mỗi anchor trên P3–P5, head classification sẽ tính xác suất “face” so với “background” thông qua một lớp convolution 3×3 nối tiếp bởi convolution 1×1 và hàm sigmoid. RetinaFace sử dụng Focal Loss [18] ($\gamma=2$, $\alpha=0.25$) để tự động giảm trọng số của các negative anchor để phân biệt, từ đó dồn lực học vào những trường hợp khó và hiếm, giúp tăng độ nhạy trong phát hiện khuôn mặt dù trong điều kiện ánh sáng hoặc góc nhìn phức tạp.

Song song với đó, head regression cho bounding box dự đoán bốn giá trị offset (Δx , Δy , Δw , Δh) so với mỗi anchor gốc. Để tối ưu, mạng dùng Smooth L1 Loss, một biến thể của Huber Loss, để loại bỏ các dữ liệu outlier, giữ cho gradient ổn định trong quá trình lan truyền ngược. Kết quả offset sau đó được giải mã thành tọa độ bounding box cuối cùng trong bước inference.



Hình 2. 4: Năm điểm mốc trong khuôn mặt được mô hình RetinaFace dự đoán

Để tăng độ chính xác trong việc phát hiện khuôn mặt, RetinaFace còn trang bị một head chuyên dự đoán 5 điểm mốc quan trọng (hai mắt, mũi, hai góc môi) như trong Hình 2.4. Head này hồi quy trực tiếp 10 giá trị (x, y) chuẩn hóa theo kích thước anchor và tối ưu bằng L1 Loss. Nhờ supervision bổ sung này, mô hình học được cấu trúc hình học của khuôn mặt, hỗ trợ rất tốt cho các bài toán face alignment hoặc normalization trước khi nhận dạng.

2.4.3. Hàm mất mát của mô hình RetinaFace

Để đồng thời tối ưu cả bốn nhiệm vụ (phân loại, hồi quy bounding box, hồi quy landmark và dense regression), RetinaFace định nghĩa một hàm mất mát tổng hợp như sau:

$$L = L_{cls}(p_i, p_i^*) + \lambda_1 p_i^* L_{box}(t_i, t_i^*) + \lambda_2 p_i^* L_{pts}(l_i, l_i^*) + \lambda_3 p_i^* L_{pixel}$$

Trong đó bao gồm:

- Hàm mất mát phân loại mặt $L_{cls}(p_i, p_i^*)$ với p_i là xác suất dự đoán anchor i chứa mặt, còn p_i^* là nhãn thực (bằng 1 nếu anchor là positive, 0 nếu negative).
- Hàm mất mát hồi quy hộp giới hạn $L_{box}(t_i, t_i^*)$ với $t_i = \{t_x, t_y, t_w, t_h\}$ và $t_i^* = \{t_x^*, t_y^*, t_w^*, t_h^*\}$ lần lượt là toạ độ hộp dự đoán và hộp ground-truth của

anchor positive. Chúng ta chuẩn hoá mục tiêu hồi quy (tọa độ tâm, chiều rộng, chiều cao)

- Hàm mất mát hồi quy landmark $L_{pts}(l_i, l_i^*)$ với $l_i = \{l_{x_1}, l_{y_1}, \dots, l_{x_5}, l_{y_5}\}_i$ và $l_i^* = \{l_{x_1}^*, l_{y_1}^*, \dots, l_{x_5}^*, l_{y_5}^*\}_i$, là tọa độ năm điểm landmark dự đoán và ground-truth tương ứng với anchor positive. Tương tự box regression, các tọa độ landmark cũng được chuẩn hoá dựa trên tâm anchor trước khi tính L1 Loss.
- Hàm mất mát dense regression $L_{pixel} = \frac{1}{W * H} \sum_i^W \sum_j^H \|\mathcal{R}(\mathcal{D}_{PT'})$

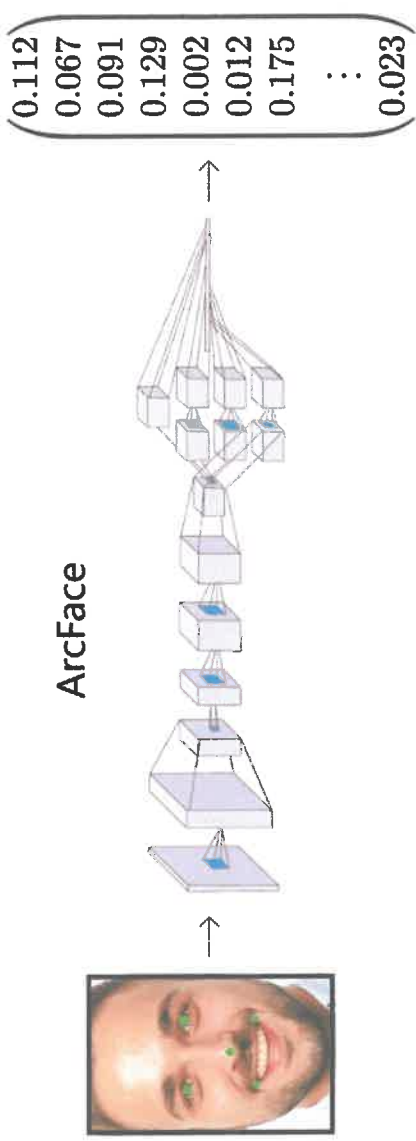
$P_{cam}, P_{ill}\}_{i,j} - l_{i,j}^*$, trong đó W và H lần lượt là chiều rộng và chiều cao của $l_{i,j}^*$

Các hệ số cân bằng $\lambda_1, \lambda_2, \lambda_3$ được cài đặt lần lượt là 0.25, 0.1 và 0.01, nhằm tăng độ ảnh hưởng của tín hiệu giám sát bounding box và landmark so với tín hiệu self-supervised. Khi chạy inference, RetinaFace chỉ cần một lần forward để trả về kết quả lớp, hộp giới hạn, 5 landmark và lưới 3D của khuôn mặt.

2.4.4. Trích xuất đặc trưng sử dụng ArcFace

2.4.4.1. Dữ liệu đầu vào và đầu ra của mô hình

Sau khi phát hiện khuôn mặt trong ảnh đầu vào, các điểm mốc (facial landmarks) được xác định bằng mô hình RetinaFace. Dựa trên năm điểm mốc chính – gồm hai mắt, chóp mũi và hai khóe miệng – ảnh khuôn mặt được căn chỉnh hình học nhằm giảm thiểu sai khác do tư thế chụp (pose variation) và biểu cảm khuôn mặt. Quá trình căn chỉnh này bao gồm các phép biến đổi hình học như xoay, thay đổi tỉ lệ và tịnh tiến để đưa khuôn mặt về một dạng chuẩn hóa thống nhất.



Hình 2. 5: Dữ liệu đầu vào và đầu ra của mô hình ArcFace

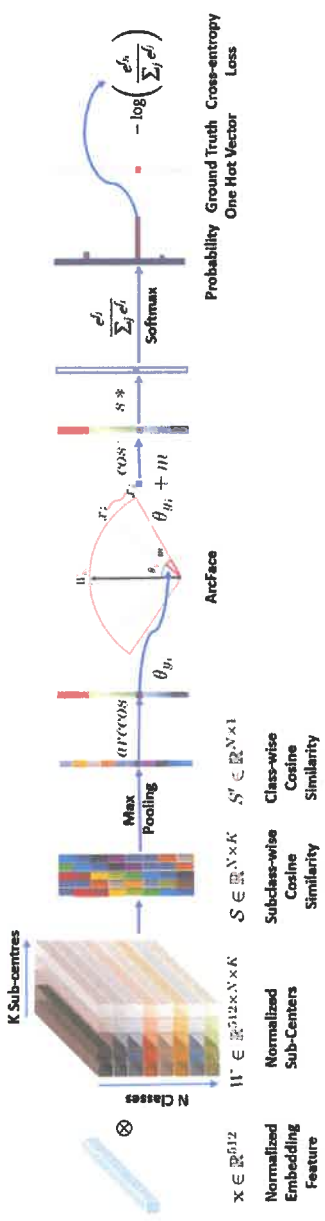
Sau bước căn chỉnh, vùng ảnh chứa khuôn mặt được cắt (cropping) và resize về kích thước chuẩn 112×112 pixel, nhằm phù hợp với yêu cầu đầu vào của kiến trúc mạng nơ-ron sâu trong ArcFace. Bức ảnh sau xử lý được chuẩn hóa giá trị màu (theo thống kê mean-std của ImageNet hoặc chuẩn tương đương) để ổn định quá trình huấn luyện.

Tại giai đoạn suy luận (inference), ảnh đầu vào được đưa qua phần trích xuất đặc trưng (feature extractor) – tức toàn bộ mạng ArcFace ngoại trừ lớp phân loại cuối cùng. Đầu ra là một vector những có kích thước cố định 512 chiều, đại diện cho biểu diễn đặc trưng của khuôn mặt đầu vào. Vector này chứa thông tin nhận dạng cốt lõi, đã được tối ưu trong quá trình huấn luyện để có khả năng phân biệt mạnh giữa các danh tính khác nhau.

2.4.4.2. Kiến trúc mô hình và phương pháp huấn luyện

Kiến trúc ArcFace được xây dựng dựa trên mạng nơ-ron tích chập sâu (DCNN) với backbone phổ biến là ResNet50 hoặc ResNet100, được mô tả chi tiết như trong Hình 2.6. Các biến thể này được điều chỉnh bằng cách loại bỏ tầng bottleneck để đảm bảo thông tin không bị nén quá mức trong quá trình lan truyền. Ảnh đầu vào sau khi qua các tầng tích chập và pooling sẽ tạo ra một tensor đặc trưng trung gian, sau đó được xử lý bởi tầng Global Average Pooling để tạo thành vector có kích thước 512 chiều.

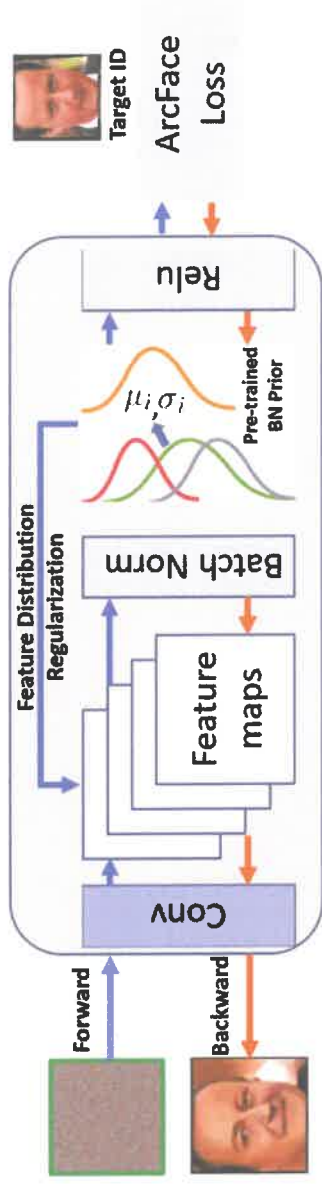
Phía sau phần backbone là một chuỗi tầng xử lý embedding. Chuỗi này thường bao gồm: một tầng Batch Normalization để chuẩn hóa phân phối đặc trưng, một tầng Dropout (nếu sử dụng) nhằm giảm overfitting, một tầng Fully Connected để cố định đầu ra là vector 512 chiều, và một tầng BatchNorm cuối để ổn định embedding. Vector này được chuẩn hóa L2, đảm bảo tất cả điểm nằm trên mặt cầu đơn vị trong không gian 512 chiều. Nhờ đó, mô hình học trực tiếp mối quan hệ góc (cosine similarity) giữa các đặc trưng khuôn mặt.



Hình 2. 6: Kiến trúc mô hình ArcFace

Mô hình được huấn luyện với hàm mất mát *Additive Angular Margin Loss* (sẽ được trình bày chi tiết trong mục 2.4.2.3), được thiết kế để nhằm tăng cường độ chặt nội lớp và độ tách biệt liên lớp. Thay vì sử dụng softmax truyền thống dựa trên tích vô hướng, ArcFace chuyển không gian tối ưu sang không gian góc bằng cách chuẩn hóa vector đặc trưng và vector trọng số lớp, từ đó biến tích vô hướng thành $\cos(\theta)$, trong đó θ là góc giữa hai vector.

Điểm then chốt của ArcFace là cộng thêm một biên góc cố định m vào θ của lớp đúng, làm tăng khoảng cách trắc địa giữa các lớp trong không gian cầu. Sau đó, toàn bộ logit được nhân với một hệ số scale s nhằm khuếch đại gradient. Nhờ cơ chế này, ArcFace tạo ra các vùng quyết định hình nón sắc nét hơn quanh tâm lớp, từ đó giảm đáng kể lỗi nhận dạng giữa các danh tính gần nhau.



Hình 2. 7: ArcFace không chỉ là một mô hình phân biệt mà còn là một mô hình sinh

Mô hình ArcFace được huấn luyện trên các tập dữ liệu lớn như MS1MV2, VGGFace2 [29] và IBUG-500K, với ảnh đã được căn chỉnh và gán nhãn. Mỗi batch huấn luyện có thể chứa 256–512 ảnh. Mô hình được tối ưu bằng SGD (có sử dụng momentum) hoặc Adam, với learning rate khởi điểm khoảng 0.1 và giảm dần theo

lịch trình xác định trước (step decay hoặc theo validation score). Số epoch thường từ 12–20 tùy theo kích thước và chất lượng của tập dữ liệu.

Khi huấn luyện hoàn tất, lớp softmax phía sau bị loại bỏ, và mô hình lúc này hoạt động như một bộ sinh vector embedding. Việc so sánh giữa hai khuôn mặt được thực hiện bằng cách đo cosine giữa hai vector đặc trưng (được mô tả chi tiết trong mục 3). Nếu cosine gần 1 (tức góc nhỏ), hai ảnh được xác định là cùng danh tính. Điều này giúp ArcFace dễ dàng tích hợp vào các hệ thống xác thực, nhận diện hoặc truy xuất khuôn mặt với độ chính xác cao.

ArcFace không chỉ là một mô hình phân biệt mà còn là một mô hình sinh. Mô hình không chỉ phân tách các đặc trưng khuôn mặt một cách hiệu quả mà còn tạo ra các biểu diễn có tính liên tục trong không gian nhúng. Điều này cho phép mô hình không chỉ nhận diện mà còn mô phỏng sự thay đổi của các đặc điểm khuôn mặt. Nhờ đó, ArcFace có thể được ứng dụng trong cả các nhiệm vụ phân loại và sinh tổng hợp khuôn mặt.

2.4.4.3. Hàm mất mát

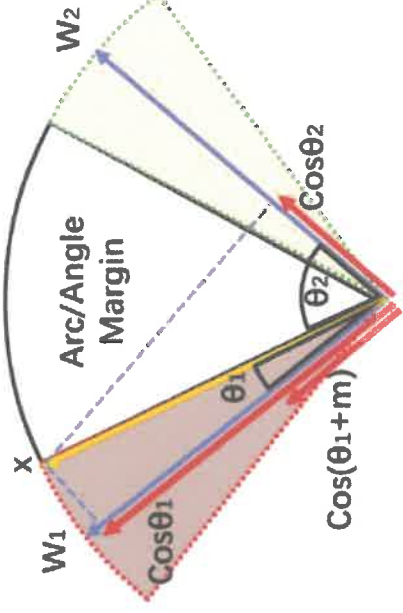
Trong các hệ thống nhận dạng khuôn mặt hiện đại, việc tối ưu hóa khoảng cách giữa các đặc trưng khuôn mặt (embeddings) đóng vai trò then chốt để đảm bảo hiệu quả phân biệt giữa các danh tính khác nhau. Các hàm mất mát truyền thống như Softmax mặc dù mang lại khả năng phân loại nhất định, nhưng lại không trực tiếp tối ưu hóa sự chặt chẽ nội lớp và sự tách biệt liên lớp trong không gian đặc trưng. Nhằm giải quyết hạn chế này, ArcFace đã đề xuất một phương pháp tiếp cận mới thông qua hàm mất mát Additive Angular Margin Loss, nhằm tối ưu hóa trực tiếp khoảng cách góc (angular distance) giữa các đặc trưng và trọng số lớp.

Trong ArcFace, mỗi mẫu đầu vào x_i sau khi qua backbone mạng học sâu sẽ được chuyển thành vector nhúng 512 chiều và chuẩn hóa L2 để nằm trên mặt cầu đơn vị. Tương tự, trọng số lớp W_j cũng được chuẩn hóa. Khi đó, tích vô hướng giữa vector nhúng và vector trọng số tương ứng trở thành giá trị cosine của góc giữa hai vector: $W_j^T x_i = \cos(\theta_j)$. Đối với nhãn đúng y_i , thay vì sử dụng trực tiếp $\cos(\theta_{y_i})$ như trong

Softmax, ArcFace bổ sung thêm một biên góc cộng thêm m vào góc θ_{y_i} tạo thành $\cos(\theta_{y_i} + m)$. Công thức hàm mất mát được định nghĩa như sau:

$$L = -\log \left(\frac{e^{s \cdot \cos(\theta_{y_i} + m)}}{e^{s \cdot \cos(\theta_{y_i} + m)} + \sum_{j \neq y_i} e^{s \cdot \cos(\theta_j)}} \right)$$

Trong đó, s là hệ số scale để khuếch đại chênh lệch logit (thường chọn là 64), và m là biên góc cộng thêm (thường lấy 0.5 radians). Cách thiết kế này mang ý nghĩa hình học rất rõ ràng: việc cộng thêm m trực tiếp vào góc tạo ra một vùng cấm (margin gap) rõ ràng trong không gian góc, tương đương với việc tối ưu hóa một khoảng cách trắc địa (geodesic distance margin) có định trên mặt cầu đơn vị. Kết quả là các đặc trưng của cùng một danh tính sẽ được ép lại gần nhau hơn, trong khi đặc trưng của các danh tính khác nhau sẽ buộc phải nằm cách nhau ít nhất một khoảng cố định về góc. Minh họa trực quan về sự khác biệt này được thể hiện trong hình dưới đây:

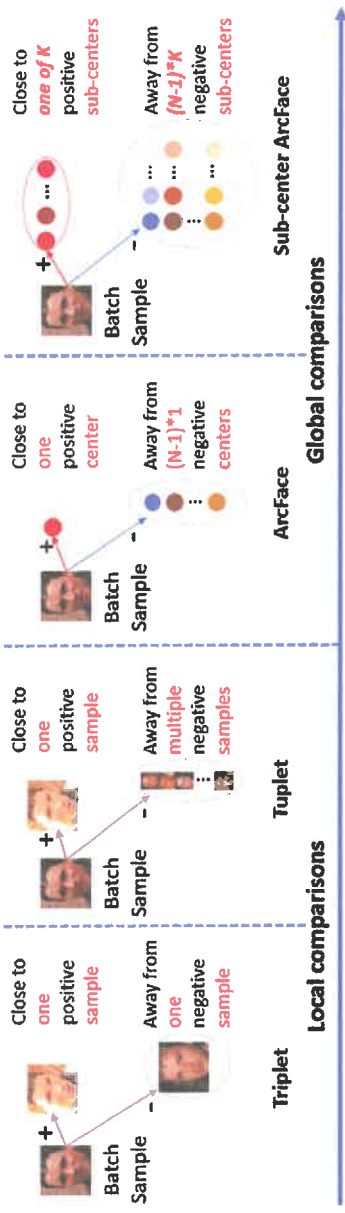


Hình 2. 8: Hàm mất mát ArcFace

Trong hình, không gian đặc trưng với Softmax truyền thống cho phép các đặc trưng của các lớp khác nhau khá gần nhau, dẫn đến ranh giới quyết định mờ nhạt. Với ArcFace, các đặc trưng bị buộc phải cách nhau một biên góc rõ ràng, đảm bảo mỗi danh tính chiếm một vùng quyết định hình nón nhỏ hơn và không chồng lấn lên nhau, từ đó tăng cường mạnh mẽ khả năng phân biệt.

Sơ với các phương pháp trước đó như SphereFace và CosFace, cách tiếp cận của ArcFace thể hiện sự ưu việt về sự đơn giản, hiệu quả và tính ổn định trong quá trình huấn luyện. SphereFace sử dụng một biên nhân trên góc, yêu cầu kỹ thuật huấn

luyện phức tạp (như annealing); CosFace áp dụng biên cosine cộng tính, tuy đơn giản hơn SphereFace, nhưng vẫn không tối ưu hóa khoảng cách cung tròn trực tiếp như ArcFace. Việc ArcFace duy trì được biên góc tuyến tính và cố định không chỉ cải thiện độ chính xác mà còn giúp việc hội tụ nhanh hơn, dễ dàng hơn trong thực tế triển khai.



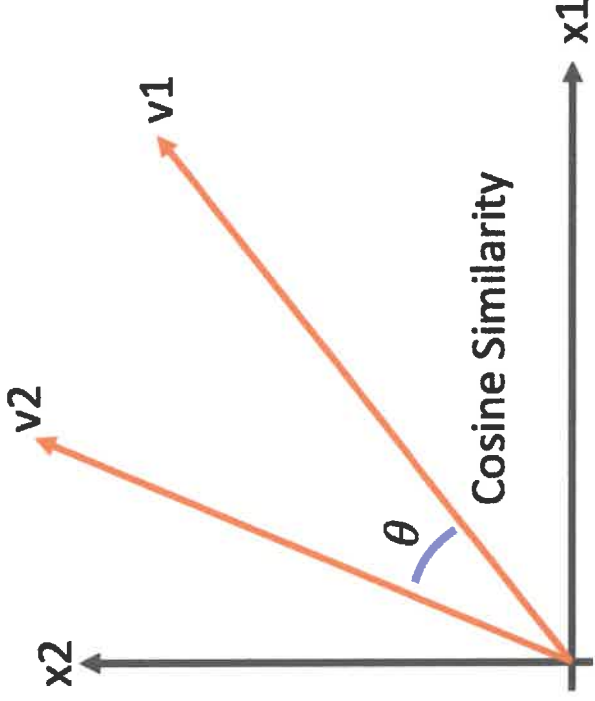
So sánh về đặc điểm các hàm mất mát

Hơn nữa, ArcFace không cần phải kết hợp thêm các hàm mất mát phụ như Triplet loss hay Center loss để đạt được khả năng phân biệt mạnh. Các thí nghiệm thực nghiệm trong bài báo cho thấy ArcFace đạt kết quả vượt trội trên các tập dữ liệu nhận dạng khuôn mặt tiêu chuẩn như LFW, CFP-FP và AgeDB-30, chứng minh rằng hàm mất mát Additive Angular Margin Loss có thể tự mình đạt được sự cân bằng tối ưu giữa tính đơn giản và hiệu quả, mà không cần phức tạp hóa pipeline huấn luyện.

Tóm lại, Additive Angular Margin Loss là một bước tiến mang tính đột phá trong bài toán nhận dạng khuôn mặt, với đóng góp chính là việc tối ưu hóa trực tiếp khoảng cách trắc địa, mang lại cả tính chính xác hình học và hiệu quả huấn luyện thực tế, đồng thời đơn giản hóa quá trình triển khai và ứng dụng thực tế.

2.4.5. So sánh độ tương đồng của các vector đặc trưng

Độ tương đồng cosine (Cosine similarity) là một phép đo phổ biến được sử dụng để đánh giá mức độ tương đồng giữa hai vector, đặc biệt trong các bài toán nhận dạng mẫu như nhận dạng khuôn mặt. Trong bối cảnh nhận diện khuôn mặt, độ tương đồng cosine cho biết mức độ giống nhau giữa hai khuôn mặt dựa trên biểu diễn vector đặc trưng của chúng trong không gian nhiều chiều.



Hình 2. 9: Độ tương đồng cosine

Điểm số độ tương đồng được tính bằng cách lấy cosine của góc giữa hai vector, với kết quả được chuẩn hóa trong khoảng từ -1 đến 1. Một giá trị cosine bằng 1 cho thấy hai khuôn mặt là hoàn toàn giống nhau về mặt đặc trưng nhận dạng, trong khi giá trị -1 cho thấy sự khác biệt hoàn toàn. Công thức tính độ tương đồng cosine được định nghĩa như sau:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Trong đó:

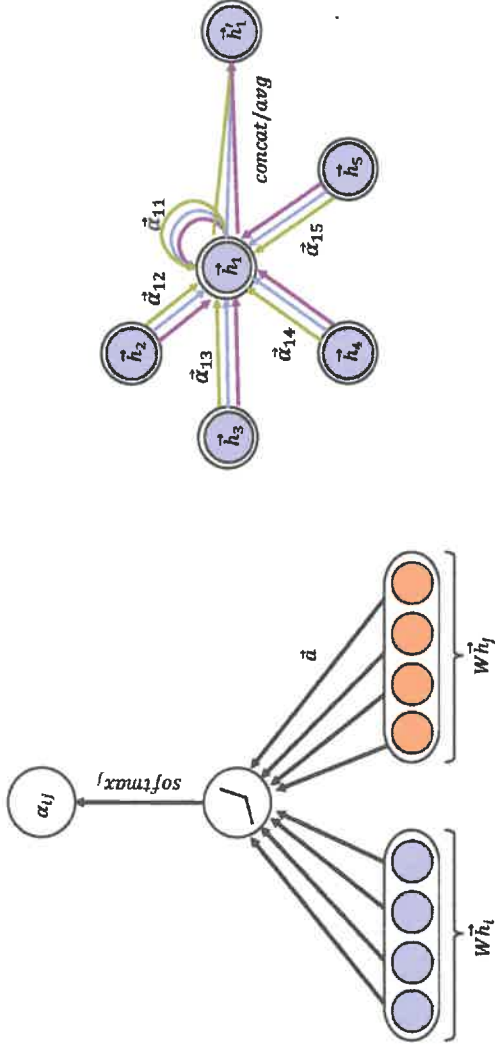
- $\mathbf{A}$ và $\mathbf{B}$ là hai vector đặc trưng khuôn mặt
- $\mathbf{A} \cdot \mathbf{B}$ là tích vô hướng giữa hai vector
- $\|\mathbf{A}\| \|\mathbf{B}\|$ là độ dài (norm) của từng vector
- n là số chiều của vector.

Trong phương pháp của chúng tôi, độ tương đồng cosine được sử dụng để so sánh các vector đặc trưng được sinh ra bởi mô hình nhận dạng khuôn mặt (embedding

model). Mỗi khuôn mặt trong cơ sở dữ liệu được ánh xạ thành một vector đặc trưng có ý nghĩa về mặt nhận dạng. Khi một khuôn mặt mới cần nhận dạng, chúng tôi tính toán độ tương đồng cosine giữa vector đặc trưng của khuôn mặt truy vấn q và các vector đặc trưng x_i trong tập dữ liệu đã lưu. Nếu giá trị $\cos(\theta)$ giữa q và một x_i vượt qua một ngưỡng xác định trước, hai khuôn mặt được xem là cùng danh tính.

2.4.6. Mạng chú ý đồ thị (GAT)

Trong bài toán nhận dạng khuôn mặt, việc sử dụng một ảnh duy nhất có thể không đủ thông tin để mô hình phân biệt chính xác danh tính của một người, đặc biệt trong các điều kiện ánh sáng, tư thế hoặc biểu cảm thay đổi. Do đó, một cách tiếp cận hiệu quả là thu thập nhiều ảnh khuôn mặt từ các góc nhìn khác nhau (chẳng hạn năm góc: chính diện, nghiêng trái, nghiêng phải, cúi, ngẩng), mỗi ảnh được mã hóa thành một vector embedding. Tuy nhiên, khi số lượng người trong hệ thống trở nên lớn, việc so sánh từng vector với toàn bộ cơ sở dữ liệu trở nên tốn kém và làm giảm tốc độ suy luận. Để giải quyết vấn đề này, chúng tôi đề xuất sử dụng Graph Attention Network (GAT) [28] để tổng hợp các vector embedding của cùng một người thành một biểu diễn duy nhất, giàu ngữ cảnh và hiệu quả hơn về mặt tính toán.



Hình 2. 10: Graph Attention Network

Lớp chú ý (attention layer) trong GAT được thiết kế để xử lý dữ liệu có cấu trúc đồ thị, nơi mỗi đỉnh (node) không tồn tại độc lập mà có thể học đặc trưng mới bằng cách "tổng hợp có chọn lọc" thông tin từ các đỉnh lân cận. Điểm nổi bật của

GAT so với các kiến trúc mạng đồ thị truyền thống (như GCN – Graph Convolutional Network) là khả năng học trọng số liên kết động (adaptive weights) giữa các nút bằng cơ chế attention thay vì dùng trọng số tĩnh hoặc được định nghĩa trước.

Giả sử có một đồ thị $G=(V, E)$ gồm N đỉnh, với mỗi đỉnh $i \in V$ có đặc trưng đầu vào $\vec{h}_i \in R^F$, tức là vector đặc trưng có F chiều. Tập lân cận của đỉnh i được ký hiệu là N_i , bao gồm các đỉnh có liên kết trực tiếp đến i , kể cả chính nó. Các bước biến đổi trong GAT được thực hiện lần lượt như sau:

- *Bước 1: Biến đổi tuyến tính đặc trưng*

Mỗi vector $\vec{h}_i$ trước tiên được biến đổi tuyến tính qua một ma trận trọng số học được $W \in R^{F' \times F}$, để ánh xạ đặc trưng sang không gian mới có kích thước F' . Ta có:

$$\vec{h}'_i = W\vec{h}_i$$

Việc này tương tự như một lớp fully-connected áp dụng cho từng node, nhằm chuyển đặc trưng sang không gian embedding phù hợp hơn với tác vụ.

- *Bước 2: Tính điểm chú ý cặp nút (attention coefficient)*

Đối với mỗi cặp nút lân cận (i, j) , điểm chú ý e_{ij} được tính bằng cách ghép nối hai vector $\vec{h}'_i || \vec{h}'_j$, sau đó nhân với vector trọng số học được $\vec{a} \in R^{2F'}$, rồi đưa qua hàm kích hoạt phi tuyến LeakyReLU:

$$e_{ij} = \text{LeakyReLU}(\vec{a}^T [\vec{h}'_i || \vec{h}'_j])$$

Điểm số này phản ánh mức độ “chú ý” mà đỉnh i nên dành cho đỉnh j . Việc dùng LeakyReLU thay vì ReLU giúp tránh trường hợp gradient bị triệt tiêu với đầu vào âm.

- *Bước 3: Chuẩn hóa attention bằng Softmax*

Để đảm bảo tính cộng dồn và dễ giải thích, các điểm chú ý e_{ij} được chuẩn hóa qua hàm softmax trong phạm vi các nút lân cận của i :

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})}$$

Kết quả là một phân phối xác suất trên các nút kề, thể hiện mức độ đóng góp tương đối của từng nút đối với nút trung tâm.

- *Bước 4: Tổng hợp đặc trưng có trọng số*

Sau khi có các hệ số attention, đặc trưng đầu ra của nút i được tính bằng cách tổng hợp có trọng số các đặc trưng của các nút lân cận:

$$\vec{h}_i^{\text{out}} = \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \vec{h}_j' \right)$$

Hàm σ là hàm kích hoạt phi tuyến như ELU hoặc ReLU. Quá trình này tương đương với một phép "trộn thông tin" giữa các node gần nhau, nhưng có trọng số linh hoạt được học trực tiếp từ dữ liệu.

- *Bước 5: Multi-head Attention*

Để tăng độ bền vững và khả năng biểu diễn, GAT thường sử dụng nhiều "đầu chú ý" (multi-head attention). Có hai cách tổng hợp kết quả từ các head: Ghép nối (concatenation): dùng cho các tầng giữa (intermediate layers):

$$\vec{h}_i^{\text{out}} = \parallel_{k=1}^K \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(k)} W^{(k)} \vec{h}_j \right)$$

Trung bình (average): dùng cho tầng cuối cùng (output layer) để đảm bảo đầu ra nhất quán về chiều.

$$\vec{h}_i^{\text{out}} = \frac{1}{K} \sum_{k=1}^K \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(k)} W^{(k)} \vec{h}_j \right)$$

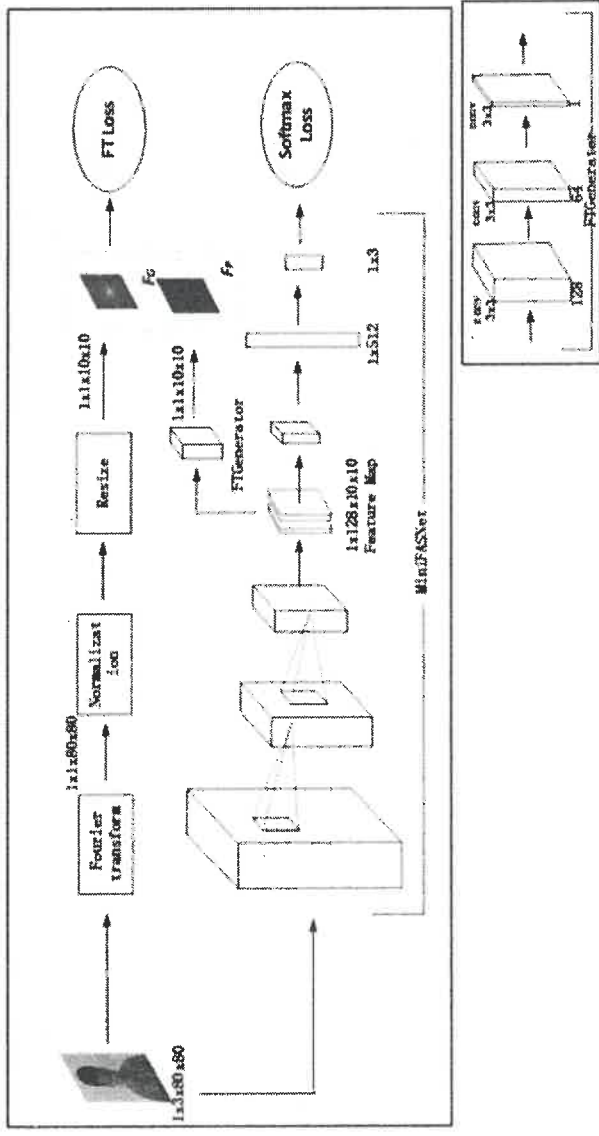
Trong đó K là số lượng head, và mỗi head có bộ tham số riêng. Lớp chú ý trong GAT không chỉ cho phép mô hình xác định mối quan hệ ngữ nghĩa giữa các đặc trưng đầu vào mà còn học cách đánh giá mức độ tin cậy và đóng góp thông tin của từng ảnh (góc mặt) khi tổng hợp vào vector đại diện cuối cùng. Điều này đặc biệt quan trọng trong nhận dạng khuôn mặt đa góc, nơi

mỗi ảnh có thể có chất lượng khác nhau hoặc bị ảnh hưởng bởi các yếu tố như bóng tối, che khuất hoặc lệch góc.

2.4.7. Phát hiện khuôn mặt giả mạo

Trong các hệ thống nhận dạng khuôn mặt thực tế, một trong những thách thức lớn nhất là khả năng bị qua mặt (spoofed) bởi các ảnh chụp in ra, video phát lại hoặc các thiết bị trình chiếu. Để chống lại các tấn công như vậy, bài toán Face Anti-Spoofing được đặt ra nhằm phân biệt khuôn mặt thật (real face) và khuôn mặt giả mạo (fake face). Trong kiến trúc được trình bày dưới đây, mô hình được thiết kế đặc biệt để học ra các đặc trưng nhạy cảm với tín hiệu tần số cao xuất hiện khi ảnh giả được thu lại qua các thiết bị như màn hình hoặc giấy in. Kiến trúc mô hình phát hiện khuôn mặt giả mạo được đề xuất như Hình 2.11. Trong đó dữ liệu đầu vào là ảnh khuôn mặt kích thước $1 \times 80 \times 80$ (grayscale). Quá trình xử lý chia thành hai nhánh song song:

- Nhánh Fourier (FT Path): khai thác tín hiệu miền tần số (frequency domain)
- Nhánh học sâu (MiniFASNet Path): học đặc trưng không gian truyền thống từ ảnh gốc



Hình 2. 11: Kiến trúc mô hình phát hiện khuôn mặt giả mạo

Hai nhánh này được huấn luyện đồng thời để cùng đóng góp vào quá trình phân loại ảnh giả/thật thông qua hai hàm mất mát khác nhau: FT Loss và Softmax Loss.

Ở nhánh Fourier, ảnh đầu vào được chuyển sang miền tần số bằng phép biến đổi Fourier 2D. Đầu ra là tensor $1 \times 80 \times 80$ biểu diễn phổ tần số của ảnh. Tiếp theo, phổ này được chuẩn hóa nhằm làm nổi bật các thành phần biên độ cao – vốn thường xuất hiện nhiều hơn ở ảnh giả do hiện tượng quét lại (re-capture) qua thiết bị.

Sau đó, tensor được resize về kích thước nhỏ hơn $1 \times 10 \times 10$ nhằm giảm độ phức tạp và đưa vào một nhánh con có tên là FTGenerator. Nhánh này là một mạng CNN đơn giản gồm ba lớp tích chập (3×3) với số kênh lần lượt là $128 \rightarrow 64 \rightarrow 2$, như mô tả ở hình góc dưới bên phải. FTGenerator có nhiệm vụ sinh ra hai bản đồ đặc trưng:

- F_g : biểu diễn đặc trưng từ phổ gốc của ảnh.
- F_p : biểu diễn đặc trưng từ ảnh giả.

Sự khác biệt giữa hai phổ này được học thông qua hàm FT Loss, giúp mô hình tập trung vào các nhiễu Fourier – vốn là tín hiệu phân biệt ảnh giả và thật trong môi trường thực tế.

Song song với nhánh Fourier, ảnh đầu vào cũng được đưa vào mạng học sâu MiniFASNet, một kiến trúc nhẹ phù hợp với các hệ thống thời gian thực. Mạng gồm nhiều tầng convolution nối tiếp, sinh ra bản đồ đặc trưng có kích thước $1 \times 128 \times 10 \times 10$. Đây là biểu diễn đặc trưng không gian giàu thông tin hình học của khuôn mặt.

Bản đồ đặc trưng sau đó được đưa qua một tầng FTGenerator chung với nhánh Fourier (chia sẻ trọng số), rồi qua một chuỗi xử lý để tạo ra vector nhúng 1×512 , và cuối cùng là lớp phân loại đầu ra 1×3 – tương ứng với ba lớp: real face, print attack, replay attack. Lớp này được huấn luyện với Softmax Loss, đóng vai trò là bộ phân loại chính.

Hai nhánh được thiết kế không tách biệt hoàn toàn, mà cùng chia sẻ bộ sinh đặc trưng FTGenerator. Trong quá trình huấn luyện, FT Loss đóng vai trò là supervision bổ sung, giúp mô hình học ra được các nhiễu Fourier tiềm ẩn trong ảnh

giả mà nhánh chính có thể bỏ sót. Khi inference, mô hình có thể chỉ cần sử dụng đầu ra từ nhánh chính để phân loại, nhờ đã học được đặc trưng nhiều trong giai đoạn huấn luyện.

Mô hình đề xuất là một giải pháp hiệu quả cho bài toán chống giả mạo khuôn mặt, nhờ khai thác đồng thời thông tin không gian và tần số. Việc tích hợp đặc trưng miền Fourier cho phép mô hình nhận diện được các đặc điểm không thể nhìn thấy bằng mắt thường, chẳng hạn nhiều sóng hoặc các pattern đặc trưng của ảnh in lại. Hơn nữa, với thiết kế nhẹ và có thể triển khai thời gian thực, mô hình này phù hợp với các ứng dụng xác thực khuôn mặt trong môi trường mở, nơi các nguy cơ tấn công bằng ảnh hoặc video là phổ biến.

2.5. Kết luận chương

Nội dung chương 2 đã trình bày về các khái niệm liên quan đến các mô hình nhận diện khuôn mặt sử dụng học sâu. Bên cạnh đó, chương 2 cũng trình bày về các tiêu chí đánh giá mô hình nhận diện khuôn mặt, và sau đó lựa chọn mô hình nhận diện khuôn mặt cho xác thực eKYC của ngân hàng. Ở chương tiếp theo, đề án sẽ trình bày về giải pháp triển khai nhận diện khuôn mặt trong xác thực eKYC cùng với quá trình thực nghiệm và đánh giá kết quả.

CHƯƠNG 3: GIẢI PHÁP ỨNG DỤNG NHẬN DIỆN KHUÔN MẶT TRONG XÁC THỰC eKYC TẠI NGÂN HÀNG VPBANK

Chương này sẽ giới thiệu chung về eKYC tại ngân hàng VPBank. Sau đó, sẽ trình bày giải pháp triển khai nhận diện khuôn mặt trong xác thực. Cuối cùng sẽ trình bày về quá trình thực nghiệm và đánh giá kết quả.

3.1. Giới thiệu chung về eKYC tại ngân hàng VPBank

Ngân hàng Thương mại Cổ phần Việt Nam Thịnh Vượng (VPBank) là đơn vị tiên phong trong việc triển khai giải pháp định danh điện tử (eKYC) tại Việt Nam. Công nghệ eKYC cho phép VPBank vượt qua mọi rào cản về địa lý và thời gian, hỗ trợ khách hàng thực hiện quy trình định danh hoàn toàn trực tuyến, không cần đến ngân hàng. Với eKYC, khách hàng có thể mở tài khoản thanh toán trực tuyến ngay lập tức mà không phải chờ đợi, đồng thời được trải nghiệm các dịch vụ ngân hàng số một cách nhanh chóng và thuận tiện.

Thông qua eKYC, khách hàng VPBank có thể thực hiện chuyển khoản, thanh toán hóa đơn, liên kết ví điện tử, nạp tiền cho các dịch vụ... với hạn mức nạp tiền tối đa 10.000.000 đồng mỗi ngày và tổng hạn mức nạp tiền tối đa 300.000.000 đồng. Chỉ mất ít phút đăng ký, xác thực thông qua eKYC người dùng đã có ngay tài khoản thanh toán và Internet Banking mà không cần đến ngân hàng.

Trong hệ thống xác thực eKYC của VPBank, công nghệ nhận diện khuôn mặt đóng vai trò cốt lõi và không thể thiếu. Đây là một trong những thành phần chính giúp ngân hàng thực hiện quy trình định danh khách hàng một cách chính xác, nhanh chóng và bảo mật cao. Cụ thể, ứng dụng nhận diện khuôn mặt được tích hợp trực tiếp vào quy trình đăng ký tài khoản trực tuyến trên Internet Banking hoặc ứng dụng VPBank NEO.

Khi khách hàng thực hiện mở tài khoản, hệ thống sẽ yêu cầu khách hàng chụp ảnh chân dung hoặc quay video khuôn mặt. Công nghệ nhận diện khuôn mặt sẽ tiến hành đối chiếu và xác thực khuôn mặt của khách hàng với ảnh trên giấy tờ tùy thân (như CMND/CCCD) đã được chụp hoặc quét trước đó. Quá trình này không chỉ giúp

kiểm tra tính hợp lệ của giấy tờ mà còn đảm bảo người thực hiện giao dịch là chính chủ, ngăn chặn các hành vi giả mạo.

Ứng dụng eKYC của VPBank mang đến nhiều lợi ích nổi bật, cụ thể như sau [31]:

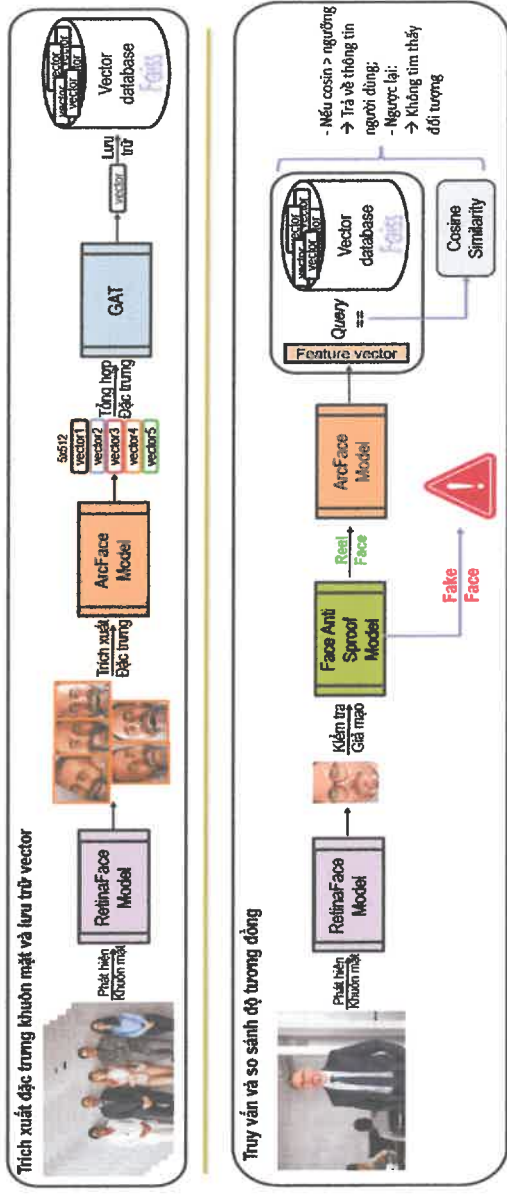
- Khách hàng có thể đăng ký tài khoản thông qua Internet Banking hoặc ứng dụng VPBank NEO trên điện thoại di động chỉ trong vài phút, mà không cần đến trực tiếp chi nhánh ngân hàng.
- Khách hàng được tự do lựa chọn số tài khoản theo sở thích, bao gồm cả các số tài khoản đẹp.
- Miễn phí toàn bộ phí chuyển khoản nội bộ, chuyển khoản liên ngân hàng, thanh toán hóa đơn và chi tiêu cá nhân.
- Rút tiền miễn phí tại hơn 17.000 cây ATM liên kết trên toàn quốc.
- Không yêu cầu số dư tối thiểu khi mở tài khoản eKYC, trong khi tài khoản truyền thống yêu cầu số dư tối thiểu 50.000 đồng.
- Hỗ trợ thanh toán nhanh chóng, an toàn thông qua mã QR.
- Cung cấp dịch vụ SMS Banking miễn phí, giúp khách hàng theo dõi biến động số dư và thay đổi thông tin tài khoản một cách kịp thời.
- Việc mở tài khoản trực tuyến qua eKYC tại VPBank được miễn phí trong năm đầu tiên. Từ năm thứ hai, ngân hàng áp dụng mức phí quản lý tài khoản 11.000 đồng/tháng nếu số dư bình quân tháng dưới 2.000.000 đồng.

3.2. Giải pháp triển khai nhận diện khuôn mặt trong xác thực eKYC

Kiến trúc hệ thống nhận diện khuôn mặt đề xuất bao gồm hai giai đoạn chính:

- **Giai đoạn 1: Trích xuất đặc trưng khuôn mặt và xây dựng cơ sở dữ liệu.**
Ảnh đầu vào được đi qua mô hình RetinaFace để phát hiện và căn chỉnh khuôn mặt, sau đó dùng ArcFace để trích xuất vector đặc trưng 512 chiều.
- **Giai đoạn 2: Truy vấn và so sánh độ tương đồng.** Thực hiện nhận diện bằng cách xử lý ảnh truy vấn qua quy trình tương tự, tìm kiếm vector tương ứng trong cơ sở dữ liệu bằng thư viện FAISS, và so sánh độ tương đồng cosine với ngưỡng để xác định danh tính. Kiến trúc này giải quyết bài toán nhận diện 1:N (một-đối-nhiều) trong bối cảnh các thách thức của nhận diện khuôn mặt trong

môi trường không kiểm soát (ánh sáng, tư thế, che khuất, v.v.), nơi học sâu đã cho thấy ưu thế so với các phương pháp truyền thống. Dưới đây, đề án sẽ trình bày chi tiết về 2 giai đoạn:



Hình 3. 1: Kiến trúc tổng thể đề xuất

❖ Giai đoạn 1:

Ở giai đoạn đầu, hệ thống thu thập nhiều ảnh khuôn mặt của từng người từ nhiều góc nhìn khác nhau, nhằm tăng độ bao phủ về hình thái và giảm rủi ro do tư thế hoặc ánh sáng. Mỗi ảnh được xử lý qua mô hình RetinaFace, một mạng phát hiện khuôn mặt một giai đoạn có khả năng dự đoán chính xác bounding box, 5 điểm landmark (mắt, mũi, miệng). Landmark được sử dụng để căn chỉnh hình học ảnh khuôn mặt (chuẩn hóa xoay, tỷ lệ và tịnh tiến), giúp tăng tính nhất quán giữa các mẫu.

Ảnh căn chỉnh sau đó được đưa qua mô hình ArcFace để trích xuất vector đặc trưng có kích thước 512 chiều. ArcFace được huấn luyện với hàm mất mát Biên Góc Cộng (Additive Angular Margin Loss), giúp tăng độ chặt của vector trong cùng danh tính và đẩy xa giữa các danh tính khác nhau, thông qua việc thêm một biên cố định vào góc giữa vector đặc trưng và vector lớp. Cách tiếp cận này giúp các vector phân bố trên mặt cầu đơn vị và so sánh dễ dàng bằng cosine similarity.

Thay vì lưu toàn bộ 5 vector embedding từ 5 góc mặt riêng biệt, hệ thống sử dụng một mô hình Graph Attention Network (GAT) để học cách tổng hợp chúng thành một vector đặc trưng duy nhất, giàu thông tin ngữ cảnh. GAT học trọng số chú

ý giữa các góc mặt, giảm thiểu ảnh hưởng của ảnh nhiễu hoặc thiếu sáng, đồng thời duy trì tính đại diện toàn diện cho một người. Vector đầu ra cuối cùng của GAT được lưu trữ trong cơ sở dữ liệu vector để phục vụ cho giai đoạn truy vấn sau này.

Để chuẩn bị cho bước nhận diện trong giai đoạn tiếp theo, các vector đặc trưng đầu ra từ GAT được lưu trữ trong một hệ thống cơ sở dữ liệu vector chuyên biệt. Để phục vụ cho việc tìm kiếm hiệu quả trong tập vector có kích thước lớn, hệ thống sử dụng FAISS (Facebook AI Similarity Search) – một thư viện mã nguồn mở được phát triển bởi Facebook AI Research. FAISS được thiết kế đặc biệt để hỗ trợ tìm kiếm gần đúng (approximate nearest neighbor – ANN) trên không gian vector nhiều chiều với tốc độ cao và hiệu quả bộ nhớ vượt trội. FAISS hỗ trợ nhiều thuật toán chỉ mục tối ưu, bao gồm:

- IVF (Inverted File Index): phân vùng không gian embedding và chỉ tìm kiếm trong các vùng liên quan.
- HNSW (Hierarchical Navigable Small World): một cấu trúc đồ thị điều hướng, cho phép truy vấn với chi phí logarit trong khi vẫn giữ độ chính xác cao.
- PQ (Product Quantization): nén vector để tiết kiệm bộ nhớ và tăng tốc độ truy vấn.

Nờ khả năng mở rộng linh hoạt và khả năng xử lý hàng triệu vector trong thời gian thực, FAISS là thành phần quan trọng trong các hệ thống nhận diện quy mô lớn, đóng vai trò như cầu nối giữa vector đặc trưng và thông tin định danh người dùng. Các vector sau khi được GAT tổng hợp và chuẩn hóa sẽ được gán nhãn định danh và lưu vào hệ thống FAISS để phục vụ bước so khớp trong giai đoạn 2.

❖ *Giai đoạn 2:*

Khi người dùng truy cập hệ thống với một ảnh đầu vào (query), ảnh này sẽ đi qua pipeline tương tự như giai đoạn 1. Trước hết, RetinaFace phát hiện và căn chỉnh khuôn mặt, sau đó ảnh được kiểm tra qua mô hình Face Anti-Spoofing, nhằm loại bỏ các trường hợp giả mạo bằng hình ảnh hoặc video trình chiếu. Mô hình chống giả sử dụng đặc trưng hình học và miền tần số (Fourier) để phân biệt ảnh thật và ảnh bị làm

giả. Nếu mẫu đầu vào bị phát hiện là Fake, hệ thống từ chối truy vấn. Nếu ảnh là Real, ảnh tiếp tục được đưa qua ArcFace để sinh ra vector đặc trưng 512 chiều đại diện cho khuôn mặt truy vấn.

Sau đó, FAISS (Facebook AI Similarity Search) [32] được sử dụng để tìm kiếm hiệu quả các vector gần nhất trong cơ sở dữ liệu. FAISS là thư viện tối ưu cho tìm kiếm tương đồng vector quy mô lớn, thường sử dụng các thuật toán Tìm kiếm Hàng xóm Gần nhất Xấp xỉ (ANN) như IVF (Inverted File) hoặc HNSW (Hierarchical Navigable Small World) để đánh đổi một phần độ chính xác lấy tốc độ và hiệu quả bộ nhớ. IVF phân vùng dữ liệu và tìm kiếm trong các phân vùng liên quan, trong khi HNSW sử dụng cấu trúc đồ thị phân cấp. Lựa chọn chỉ mục FAISS và các tham số cần được tinh chỉnh dựa trên dữ liệu và yêu cầu ứng dụng.

Độ tương đồng Cosine được dùng để so sánh vector truy vấn với các vector ứng viên từ FAISS. Phép đo này tính cosine góc giữa hai vector, bỏ qua độ lớn, phù hợp với các embedding đã chuẩn hóa của ArcFace vốn tập trung vào hướng vector trên siêu cầu. Giá trị tương đồng (từ -1 đến 1) sau đó được so sánh với một ngưỡng xác định trước. Việc chọn ngưỡng là sự đánh đổi quan trọng giữa Tỷ lệ Chấp nhận Sai (FAR - False Acceptance Rate) và Tỷ lệ Từ chối Sai (FRR - False Rejection Rate). Ngưỡng cao giảm FAR (tăng bảo mật) nhưng tăng FRR (giảm tiện lợi), và ngược lại. Ngưỡng tối ưu phụ thuộc vào yêu cầu ứng dụng (ví dụ: nhận diện 1:N thường cần FAR thấp hơn) và phải được xác định thực nghiệm qua phân tích đường cong ROC/DET trên tập dữ liệu xác thực.

3.3. Thử nghiệm và đánh giá

3.3.1. Xây dựng cơ sở dữ liệu

Để đánh giá khả năng ứng dụng thực tế của hệ thống nhận diện khuôn mặt trong bài toán xác thực khách hàng điện tử (eKYC) của ngân hàng VPBank, đề án tiến hành xây dựng một tập dữ liệu khuôn mặt đa dạng, phản ánh chân thực các điều kiện mà người dùng có thể gặp phải khi đăng ký tài khoản hoặc xác thực từ xa thông qua điện thoại di động.

Về phần cứng, hệ thống thử nghiệm được triển khai trên bộ vi xử lý Intel Core i7, bộ nhớ RAM 32GB. Bên cạnh đó, card đồ họa NVIDIA RTX 3080 (VRAM 10GB GDDR6X) được sử dụng để tăng tốc các tác vụ huấn luyện mạng nơ-ron sâu và suy luận theo thời gian thực.

Về phần mềm, sử dụng môi trường Python 3.10 để chạy, được tích hợp với các thư viện học sâu như PyTorch 2.0.1 để xây dựng và huấn luyện mô hình ArcFace, thư viện OpenCV 4.7.0 cho các hoạt động xử lý ảnh. Thư viện FAISS 1.7.4 được sử dụng nhằm tối ưu hoá tìm kiếm vector trong không gian vector, nhằm cải thiện tốc độ xác thực khuôn mặt. Ngoài ra, các thư viện NumPy, Scikit-learn và Matplotlib cũng được sử dụng cho các hoạt động phân tích dữ liệu và đánh giá kết quả.

Hệ thống được triển khai dưới dạng notebook trên nền tảng Visual Studio Code, cho phép dễ dàng theo dõi toàn bộ pipeline xử lý, tổ chức thí nghiệm lặp lại và đánh giá tính nhất quán của kết quả. Thời gian suy luận trung bình được đo đặc bằng thư viện *time*, giúp giảm thiểu ảnh hưởng của nhiều yếu tố ngẫu nhiên và tăng tính khách quan cho đánh giá hiệu suất.

Tổng thể, môi trường thử nghiệm được thiết kế đảm bảo tính khả thi triển khai thực tế, đồng thời cho phép đánh giá chi tiết và đáng tin cậy đối với các chỉ số như độ chính xác, tốc độ xử lý và mức ổn định của hệ thống trong điều kiện khai thác thực tế. Các công cụ mã nguồn mở được ứng dụng trong thiết lập cũng giúp tiết kiệm chi phí và tăng tính linh hoạt trong việc chuyển giao hệ thống cho các đơn vị sử dụng cuối như khối công nghệ thông tin của ngân hàng.

3.3.3. Kết quả thử nghiệm và đánh giá

Kết quả thử nghiệm được trình bày trong bảng 3.1.

Để đánh giá, học viên sử dụng các chỉ số chuẩn trong nhận diện: accuracy, precision, recall và F1-score. Kết quả được so sánh với 4 phương pháp khác:

- CNN
- MTCNN, Facenet sử dụng VGG16
- ArcFace + Voting (so sánh riêng từng góc, lấy kết quả đa số)

- FaceNet (baseline model không dùng margin loss)

Kết quả cho thấy phương pháp ArcFace kết hợp GAT đạt độ chính xác rất cao với 98.32%, vượt qua tất cả các phương pháp đối chứng khác. Precision đạt 98.6%, cho thấy hệ thống rất ít xảy ra nhận nhầm giữa các người dùng. Recall cũng đạt tới 98.1%, phản ánh khả năng phát hiện đúng gần như toàn bộ người thật trong cơ sở dữ liệu. F1-score cân bằng giữa precision và recall đạt mức 98.35%, chứng minh hệ thống có hiệu suất tổng thể ổn định và mạnh mẽ.

Phương pháp	Accuracy (%)	Precision (%)	Recal (%)	F1-score (%)	Thời gian suy luận (ms)
CNN	88.52	89.40	87.60	88.49	23.4
MTCNN	92.21	93.00	91.00	91.99	35.2
ArcFace + Voting	94.12	94.30	93.70	94.00	97.5
FaceNet	95.47	95.80	94.90	95.34	26.4
Phương pháp đề xuất	98.32	98.60	98.10	98.35	21.3

Bảng 3. 1: Kết quả thực nghiệm

Để so sánh, chúng tôi cũng thử nghiệm với các phương pháp phổ biến như: trung bình vector (average pooling), chọn giá trị lớn nhất (max pooling), bỏ phiếu đa số (voting) và sử dụng mô hình FaceNet làm baseline. Bảng 1 trình bày kết quả so sánh giữa phương pháp đề xuất và một số phương pháp đối chứng phổ biến, bao gồm CNN truyền thống, MTCNN, ArcFace kết hợp voting, và FaceNet. Phương pháp của chúng tôi cho kết quả vượt trội về cả bốn chỉ số đánh giá chính: accuracy đạt 98.32%, precision đạt 98.60%, recall đạt 98.10%, và F1-score đạt 98.35%. Không chỉ dừng lại ở hiệu suất nhận diện cao, hệ thống còn thể hiện ưu thế về tốc độ khi thời gian suy luận trung bình mỗi truy vấn chỉ là 21.3 ms – thấp nhất trong tất cả các phương pháp được so sánh.

Thời gian ở Bảng 3.1 chỉ đo bước tìm kiếm cosine (IndexFlatIP trên vector đã L2-normalize) không gồm bước nhận diện. Vì IndexFlatIP tính inner-product với tất cả N vector nên latency tăng tuyến tính $O(N)$. So với các mô hình còn lại, hiệu quả tổng thể của phương pháp đề xuất là rõ ràng. Cụ thể, FaceNet đạt độ chính xác khá cao (95.47%) nhưng vẫn kém hơn so với hệ thống đề xuất, trong khi MTCNN và CNN truyền thống bị giới hạn về cả hiệu suất lẫn tốc độ. Phương pháp voting sử dụng ArcFace cho kết quả tương đối (accuracy 94.12%), tuy nhiên thời gian suy luận lên tới 97.5 ms cho mỗi truy vấn khiến nó khó triển khai trong các hệ thống yêu cầu phản hồi nhanh như eKYC.

Việc sử dụng GAT không chỉ giúp tăng độ chính xác mà còn rút ngắn đáng kể thời gian xử lý mà không ảnh hưởng đến độ chính xác quá trình tìm kiếm. Nhờ khả năng học trọng số giữa các góc mặt, GAT cho phép hệ thống tổng hợp thông tin từ nhiều quan sát khác nhau một cách hiệu quả, thay vì phải lưu trữ và so sánh từng vector riêng biệt như các phương pháp truyền thống. Điều này đặc biệt hữu ích trong môi trường sản phẩm thực tế, nơi chi phí lưu trữ và tốc độ phản hồi là những yếu tố quan trọng.

Một số thử nghiệm với tham số của mô hình GAT cũng được tiến hành. Cấu hình sử dụng 4 attention heads với kích thước mỗi head là 128 cho thấy sự cân bằng tốt giữa độ chính xác và chi phí tính toán. Embedding đầu ra được giữ ở mức 512 chiều nhằm đảm bảo tương thích với ArcFace, đồng thời duy trì khả năng phân biệt tốt trong không gian đặc trưng.

Tổng thể, phương pháp đề xuất - kết hợp trích xuất đặc trưng sâu bằng ArcFace và học tổng hợp thông minh bằng Graph Attention Network – cho thấy khả năng nhận diện khuôn mặt chính xác, nhanh và ổn định trong điều kiện eKYC không kiểm soát. Với chỉ một vector lưu trữ duy nhất cho mỗi người, hệ thống vừa giảm tải bộ nhớ, vừa đảm bảo tốc độ truy vấn nhanh, giúp dễ dàng mở rộng cho hàng trăm nghìn hoặc hàng triệu người dùng.

3.4. Kết luận chương

Nội dung chương 3 đã giới thiệu chung về eKYC tại ngân hàng VPBank. Bên cạnh đó, chương 3 cũng trình bày về giải pháp triển khai nhận diện khuôn mặt trong xác thực eKYC. Và cuối cùng đã trình bày về quá trình thực nghiệm và đánh giá. Kết quả thu được phương pháp đề xuất đạt kết quả tốt và có thể ứng dụng nhận diện khuôn mặt trong xác thực eKYC tại ngân hàng VPBank.

KẾT LUẬN

Kết quả đạt được của đề án

Trong khuôn khổ bài toán xác thực danh tính điện tử (eKYC) trong ngành ngân hàng, đề án đã triển khai nghiên cứu một cách hệ thống, bao gồm các giai đoạn: phân tích lý thuyết, thiết kế mô hình kỹ thuật, thu thập và xử lý dữ liệu, huấn luyện và đánh giá hiệu suất hệ thống. Các kết quả đáng chú ý đã được xác lập như sau:

- *Nghiên cứu và hệ thống hoá kiến thức về nhận diện khuôn mặt và eKYC:*
Đề án đã phân tích toàn diện quy trình xác thực eKYC, đặc biệt tập trung vào vai trò của sinh trắc học khuôn mặt trong danh tính điện tử. Các yếu tố kỹ thuật như đặc trưng khuôn mặt (face embedding), tính kháng nhiễu, tính phân biệt và mức độ ổn định theo thời gian đã được tìm hiểu sâu.
- *Xây dựng bộ dữ liệu thực nghiệm và tiến hành đánh giá: thu thập và xử lý bộ dữ liệu khuôn mặt bao gồm nhiều mẫu khuôn mặt với sự biến đổi về ánh sáng, biểu cảm, góc quay và phong nền. Dữ liệu được gắn nhãn và phân chia để đánh giá khách quan giữa các mô hình.*
- *Đề xuất và thực thi hệ thống nhận diện khuôn mặt đồng bộ: Hệ thống gồm 3 thành phần chính: (i) RetinaFace để phát hiện và căn chỉnh khuôn mặt, (ii) ArcFace để trích xuất vector đặc trưng, (iii) mạng Graph Attention Network (GAT), và cuối cùng (iv) mô hình Anti Face để phát hiện các khuôn mặt giả mạo. Việc kết hợp bốn mô hình đã giúp tối ưu hoá về độ chính xác, tốc độ suy luận và khả năng kháng nhiễu.*
- *Đánh giá hiệu năng bằng các tiêu chí định lượng: Hệ thống được đánh giá trên các chỉ số: accuracy > 98%, precision > 97%, và average inference time ~50ms/instance. Các đánh giá này được thực hiện trong môi trường python 3.10 sử dụng GPU RTX 3080, đảm bảo khả năng tái lập và khách quan*
- *Tuân thủ yếu tố pháp lý và đạo đức AI: Toàn bộ quy trình thu thập, lưu trữ và xử lý dữ liệu khuôn mặt đã tuân thủ Nghị định 13/2023/NĐ-CP và*

các hướng dẫn đạo đức AI về bảo vệ danh tính cá nhân. Các dữ liệu nhạy cảm được mã hóa và phân quyền truy cập nghiêm ngặt.

Hướng phát triển tiếp theo

Do thời gian thực hiện đề án không nhiều, tác giả chưa có điều kiện nghiên cứu sâu rộng hơn về nhiều phương pháp mới. Trong tương lai, đề tài sẽ tiếp tục tập trung nghiên cứu các hướng phát triển sau:

- Nghiên cứu và cải tiến các mô hình nhận diện khuôn mặt để nâng cao hiệu suất nhận diện trong các điều kiện bị che khẩu trang, thiếu ánh sáng hoặc quay video ngoài trời.
- Tích hợp nhận diện khuôn mặt với phân tích hành vi, cử động khuôn mặt và giám sát video theo thời gian thực nhằm phát hiện gian lận tinh vi.
- Triển khai hệ thống trên các thiết bị biên như NVIDIA Jetson, Raspberry Pi hoặc di động để tăng tính ứng dụng linh hoạt.
- Hợp tác với các ngân hàng và tổ chức tài chính để triển khai thử nghiệm thực tế, thu thập dữ liệu và tinh chỉnh mô hình theo hướng cá nhân hoá.

Những hướng phát triển trên sẽ giúp hệ thống nhận diện khuôn mặt đạt được tính đồng bộ, độ chính xác và mức độ ứng dụng cao hơn trong thực tế, góp phần thúc đẩy chuyển đổi số và các dịch vụ tài chính thông minh.

DANH MỤC TÀI LIỆU THAM KHẢO

- [1]. M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [2]. X. Wang, J. Peng, S. Zhang, B. Chen, Y. Wang, and Y. Guo, "A Survey of Face Recognition," *arXiv preprint arXiv:2212.13038*, Dec. 2022. [Online]. Available: <https://arxiv.org/abs/2212.13038>
- [3]. Mei Wang, Weihong Deng, "Deep Face Recognition: A Survey". Available: <https://arxiv.org/abs/1804.06655>
- [4]. Y. Taigman, M. Yang, M. A. Ranzato, L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, 2014, pp. 1701–1708, doi: 10.1109/CVPR.2014.220.
- [5]. Y. Sun, X. Wang và X. Tang, "Deep Learning Face Representation from Predicting 10,000 Classes," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, 2014, pp. 1891–1898, doi: 10.1109/CVPR.2014.244.
- [6]. F. Schroff, D. Kalenichenko, J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.
- [7]. J. Deng, J. Guo, N. Xue, S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 4690–4699, doi: 10.1109/CVPR.2019.00482.
- [8]. Keiron O'Shea, Ryan Nash, "An Introduction to Convolutional Neural Networks". Available: <https://arxiv.org/abs/1511.08458>
- [9]. European Parliament, "Regulation (EU) 2016/679: General Data Protection Regulation (GDPR)," 2016. [Online]. Available: <https://gdpr-info.eu/>. [Accessed: 15-Apr-2025].

- [10]. B. Silverstein, "The dark side of facial recognition: Facebook pays \$650 million settlement for privacy violations," *Forbes*, 2021. [Online]. Available: <https://www.forbes.com/sites/bobbysilverstein/2021/03/01/the-dark-side-of-facial-recognition-facebook-pays-650-million-settlement-for-privacy-violations/>. [Accessed: 15-Apr-2025].
- [11]. J. Kaye, "Facial recognition technology banned in Luxembourg and Belgium," *The Guardian*, 2023. [Online]. Available: [Accessed: 15-Apr-2025].
- [12]. Chính phủ Việt Nam, "Nghị định 13/2023/NĐ-CP về bảo vệ dữ liệu cá nhân," 2023. [Online]. Available: [Accessed: 15-Apr-2025]. <https://vanban.chinhphu.vn/default.aspx?pageid=27160&docid=205582>
- [13]. Ngân hàng Nhà nước Việt Nam, "Thông tư 16/2020/TT-NHNN hướng dẫn về định danh khách hàng điện tử (eKYC)," 2020. [Online]. Available: <https://thuvienphapluat.vn/van-ban/Tien-te-Ngan-hang/Thong-tu-16-2020-TT-NHNN-dinh-danh-khach-hang-dien-tu-459588.aspx>. [Accessed: 15-Apr-2025].
- [14]. Bộ Thông tin và Truyền thông, "Chuyển đổi số và bảo vệ dữ liệu cá nhân tại Việt Nam," 2023. [Online]. Available: [Accessed: 15-Apr-2025]. <https://mic.gov.vn/Pages/TinTuc/161256/Bao-ve-du-lieu-ca-nhan-trong-chuyen-doi-so.html>
- [15]. Jiankeng Deng, "RetinaFace: Single-stage Dense Face Localisation in the Wild", available: <https://arxiv.org/abs/1905.00641>
- [16]. Kaiming He, "Deep Residual Learning from Image Recognition", available: <https://arxiv.org/abs/1512.03385>
- [17]. Olga Russakovsky, "ImageNet Large Scale Visual Recognition Challenge" <https://arxiv.org/abs/1409.0575>
- [18]. Tsung-Yi Lin, "Focal Loss for Dense Object Detection", available: <https://arxiv.org/abs/1708.02002>
- [19]. Y. Huang, Y. Wang, Y. Wang, and Y. Qiao, "CurricularFace: Adaptive Curriculum Learning Loss for Deep Face Recognition," *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 5901–5910, 2020.

- [20]. Q. Meng, S. Zhao, Z. Huang và F. Zhou, "MagFace: A Universal Representation for Face Recognition and Quality Assessment," *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 14225–14234, 2021.
- [21]. M. Najibi, P. Samangouei, R. Chellappa, and L. Davis, "SSH: Single Stage Headless Face Detector," *arXiv preprint arXiv:1708.03979*, 2017. [Online]. Available: <https://arxiv.org/abs/1708.03979>
- [22]. X. Tang, D. K. Du, Z. He, and J. Liu, "PyramidBox: A Context-assisted Single Shot Face Detector," *arXiv preprint arXiv:1803.07737*, 2018. [Online]. Available: <https://arxiv.org/abs/1803.07737>
- [23]. J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable Convolutional Networks," *arXiv preprint arXiv:1703.06211*, 2017. [Online]. Available: <https://arxiv.org/abs/1703.06211>
- [24]. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," *arXiv preprint arXiv:1612.03144*, 2017. [Online]. Available: <https://arxiv.org/abs/1612.03144>
- [25]. A. Barbu, N. Lay, and G. Gramajo, "Face Detection with a 3D Model," *arXiv preprint arXiv:1404.3596*, 2015. [Online]. Available: <https://arxiv.org/abs/1404.3596>
- [26]. K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *arXiv preprint arXiv:1703.06870*, 2018. [Online]. Available: <https://arxiv.org/abs/1703.06870>
- [27]. R. A. Güler, N. Neverova, and I. Kokkinos, "DensePose: Dense Human Pose Estimation In The Wild," *arXiv preprint arXiv:1802.00434*, 2018. [Online]. Available: <https://arxiv.org/abs/1802.00434>
- [28]. P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," *arXiv preprint arXiv:1710.10903*, 2018. [Online]. Available: <https://arxiv.org/abs/1710.10903>

- [29]. Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," *arXiv preprint arXiv:1710.08092*, 2018. [Online]. Available: <https://arxiv.org/abs/1710.08092>
- [30]. <https://www.socure.com/glossary/ekyc>
- [31]. <https://www.vpbank.com.vn/bi-kip-va-chia-se/retail-story-and-tips/vpb-online-category/ekyc-la-gi>
- [32]. <https://github.com/facebookresearch/faiss>

BÁO CÁO GIẢI TRÌNH
SỬA CHỮA, HOÀN THIỆN ĐỀ ÁN TỐT NGHIỆP

Họ và tên học viên: Nguyễn Thị Vân
Chuyên ngành: Hệ thống thông tin
Khóa: 2023 đợt 2

Tên đề tài: NGHIÊN CỨU ỨNG DỤNG NHẬN DIỆN KHUÔN MẶT TRONG XÁC
THỰC EKYC TẠI NGÂN HÀNG VPBANK

Người hướng dẫn khoa học: TS. Vũ Văn Thò

Ngày báo vệ: 19/07/2025

Các nội dung học viên đã sửa chữa, bổ sung trong đề án tốt nghiệp theo ý kiến đóng
góp của Hội đồng chấm đề án tốt nghiệp:

TT	Ý kiến hội đồng	Sửa chữa của học viên
1	Xem lại chỉ mục, bị sai (2.4.4.1).	Học viên đã rà soát, chỉnh sửa các lỗi về chỉ mục ở các trang 27-30.
2	Mục 2.4.5 công thức bị tràn sang trang 39.	Tiếp thu góp ý của Hội đồng, tác giả đã chỉnh sửa để công thức không bị tràn sang trang 39.
3	Bổ sung môi trường thử nghiệm, chạy như thế nào, có sử dụng GPU ko?	Tiếp thu góp ý của Hội đồng, tác giả đã bổ sung môi trường thực nghiệm, có sử dụng GPU ở mục 3.3.2 trang 48 và 49.
4	Làm rõ hơn vị trí vai trò của ứng dụng nhận diện khuôn mặt trong hệ thống xác thực của ngân hàng Vpbank.	Tiếp thu góp ý của Hội đồng, tác giả đã chỉnh sửa, làm rõ nội dung ở trang 43 và 44.
5	Làm rõ mô hình Graph Attention Network GAT có làm ảnh hưởng độ chính xác quá trình tìm kiếm không?	Tiếp thu góp ý của Hội đồng, tác giả đã làm rõ ảnh hưởng của mô hình Graph Attention Network GAT đến quá trình tìm kiếm ở trang 51.
6	Việc sử dụng công cụ tìm kiếm FAISS của Facebook chỉ phù hợp với số lượng lớn khuôn mặt, tuy nhiên nếu chỉ kiểm thử đối với 175 người thì liệu có phù hợp?	Tiếp thu góp ý của Hội đồng, tác giả đã làm rõ việc sử dụng công cụ tìm kiếm FAISS cho 175 người là phù hợp với bối cảnh xây dựng giải pháp có khả năng mở rộng như xác thực khuôn mặt

		ở trang 48 và 49.
7	Làm rõ thời gian suy luận ở bảng 3.1 trang 50, khi số lượng mặt trong kho dữ liệu tăng lên thì thời gian sẽ thay đổi theo chiều hướng nào?	Tiếp thu góp ý của Hội đồng, tác giả đã nêu cách thức tính thời gian suy luận ở bảng 3.1 ở trang 51. Và cũng làm rõ chiều hướng thay đổi của thời gian khi dữ liệu tăng lên ở trang 51.

Hà Nội, ngày tháng năm 2025

Ký xác nhận của

CHỦ TỊCH HỘI ĐỒNG
CHẤM ĐỀ ÁN

THỦ KÝ HỘI ĐỒNG
DẪN KHOA HỌC

HỌC VIÊN



PGS.TSKH. Hoàng Đăng Hải

TS. Đào Ngọc Phong

TS. Vũ Văn Thóa

Nguyễn Thị Vân

BẢN CAM ĐOAN

Tôi cam đoan đã thực hiện việc kiểm tra mức độ tương đồng nội dung đề án qua phần mềm <https://kiemtrataillieu.vn/> một cách trung thực và đạt kết quả mức độ tương đồng 07% toàn bộ nội dung đề án. Bản đề án kiểm tra qua phần mềm là bản cứng đề án đã nộp để bảo vệ trước hội đồng. Nếu sai tôi xin chịu các hình thức kỷ luật theo quy định hiện hành của Học viện.

Hà Nội, ngày 30 tháng 07 năm 2025

HỌC VIÊN CAO HỌC

(Ký và ghi rõ họ tên)



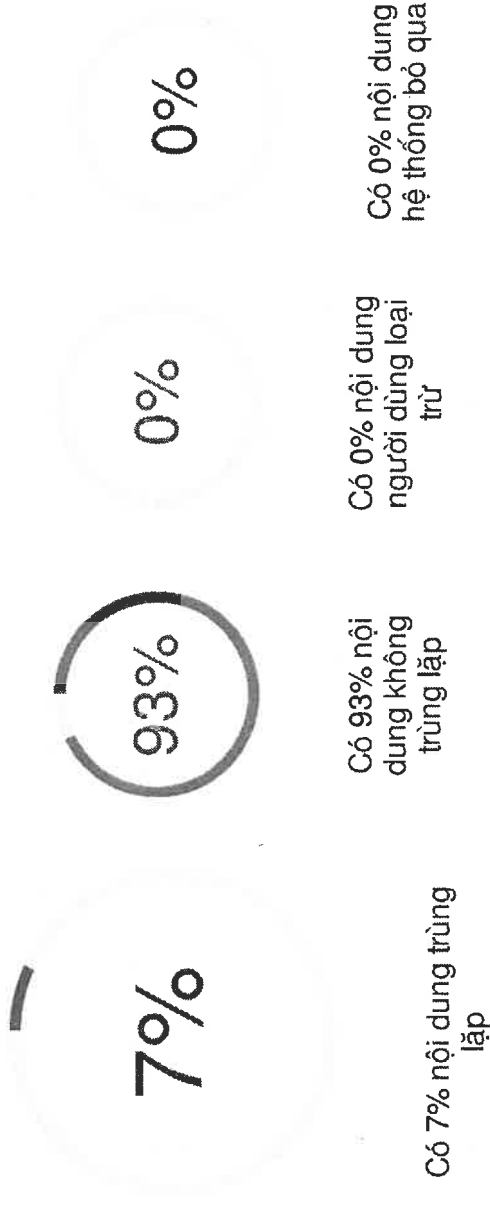
BÁO CÁO KIỂM TRA TRÙNG LẬP

Thông tin tài liệu

Tên tài liệu: Đề án_Nguyễn Thị Vân_B23CHIS065
Tác giả: Nguyễn Thị Vân
Điểm trùng lặp: 7
Thời gian tải lên: 20:12 30/07/2025
Thời gian sinh báo cáo: 20:14 30/07/2025
Các trang kiểm tra: 71/71 trang



Kết quả kiểm tra trùng lặp



Nguồn trùng lặp tiêu biểu

arxiv.org 123docz.net tailieu.vn

Học viên

Nguyễn Thị Vân

Nguyễn Thị Vân

Người hướng dẫn khoa học

TS. Vũ Văn Thò

TS. Vũ Văn Thò

BIÊN BẢN
HỌP HỘI ĐỒNG CHẤM ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

Căn cứ quyết định số Quyết định số 1098/QĐ-HV ngày 26 tháng 06 năm 2025 của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ. Hội đồng đã họp vào hồi 9...giờ...55...phút, ngày 19 tháng 07 năm 2025 tại Học viện Công nghệ Bưu chính Viễn thông để chấm đề án tốt nghiệp thạc sĩ cho:

Học viên: Nguyễn Thị Vân

Tên đề án tốt nghiệp: Nghiên cứu ứng dụng nhận diện khuôn mặt trong xác thực eKYC tại ngân hàng VPBank

Chuyên ngành: Hệ thống thông tin

Mã số: 8480104

Các thành viên của Hội đồng chấm đề án tốt nghiệp có mặt: .CS../05

TT	HỌ VÀ TÊN	TRÁCH NHIỆM TRONG HD	GHI CHÚ
1	PGS.TSKH. Hoàng Đăng Hải	Chủ tịch	
2	TS. Đào Ngọc Phong	Thư ký	
3	TS. Bùi Hải Phong	Phản biện 1	
4	PGS.TS. Trần Nguyên Ngọc	Phản biện 2	
5	PGS.TS. Hoàng Xuân Dậu	Ủy viên	

Các nội dung thực hiện:

- Chủ tịch Hội đồng điều khiển buổi họp. Công bố quyết định của Giám đốc Học viện Công nghệ Bưu chính Viễn thông về việc thành lập Hội đồng chấm đề án tốt nghiệp thạc sĩ.
- Người hướng dẫn khoa học hoặc thư ký đọc lý lịch khoa học và các điều kiện bảo vệ đề án tốt nghiệp của học viên. (có bản lý lịch khoa học và kết quả các môn học cao học của học viên kèm theo).
- Học viên trình bày tóm tắt đề án tốt nghiệp.
- Phản biện 1 đọc nhận xét (có văn bản kèm theo)
- Phản biện 2 đọc nhận xét (có văn bản kèm theo)
- Các câu hỏi của thành viên Hội đồng:
1/ Làm rõ các việc tương ứng? làm rõ...
2/ Làm rõ các việc tương ứng? làm rõ...
3/ Thuyết trình lại để được... làm rõ...
4/ Việc áp dụng công cụ tìm kiếm F8... khi chưa có 175 tương hợp...
5/ Thời gian...
7. Trả lời của học viên:

1/ Để đúng trong các giao dịch
2/ Đảm bảo tính minh bạch của GDN. Sẽ làm rõ thêm
3/ Khi cấp chúng không làm giảm đi tính chính xác

8. Thư ký đọc nhận xét về quá trình thực hiện đề án tốt nghiệp của học viên (có văn bản kèm theo).

9. Hội đồng họp riêng:

- Bầu Ban kiểm phiếu:

1. Trưởng Ban kiểm phiếu: ...Đào Ngọc Phong...
2. Ủy viên Ban kiểm phiếu: ...Bùi Hải Phong...
3. Ủy viên Ban kiểm phiếu: ...Trần Nguyễn Ngọc...

- Hội đồng chấm đề án tốt nghiệp bằng bỏ phiếu kín.
- Ban kiểm phiếu làm việc:
- Trưởng Ban kiểm phiếu báo cáo kết quả kiểm phiếu (có Biên bản họp Ban kiểm phiếu kèm theo)
- Điểm trung bình của đề án tốt nghiệp: ...8.0...

Kết luận:

1. Các nội dung cần chỉnh sửa, hoàn thiện sau bảo vệ đề án tốt nghiệp:

1. Tiếp tục đồng ý từ ý kiến của thầy đồng ý phản biện
2. Trình bày rõ ràng, sạch sẽ, không có lỗi, không có sai sót

2. Đề nghị Học viện công nhận (hoặc không) và cấp bằng (hoặc không) thạc sĩ cho học viên:
Công nhận và cấp bằng
3. Đề án tốt nghiệp có thể phát triển thành đề tài nghiên cứu cho
NCS.....Khánh.....

Buổi làm việc kết thúc vào 10h40... cùng ngày.

Chủ tịch

TS. Đào Ngọc Phong

Thư ký

PGS.TSKH. Hoàng Đăng Hải

CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập - Tự do - Hạnh phúc

NHẬN XÉT CỦA CÁN BỘ PHẢN BIỆN

ĐỀ ÁN TỐT NGHIỆP

Họ và tên cán bộ chấm phản biện: Trần Nguyễn Ngọc
Học hàm, học vị: Phó Giáo sư, Tiến sĩ
Đơn vị công tác: Học viện Kỹ thuật quân sự
Công tác chuyên môn: Quản lý, nghiên cứu và giảng dạy

NHẬN XÉT ĐỀ ÁN THẠC SĨ

Đề tài: Nghiên cứu ứng dụng nhận diện khuôn mặt trong xác thực EKYC tại ngân hàng VPBANK

Học viên thực hiện: Nguyễn Thị Vân

Đề án được kết cấu gồm 3 chương: Tổng quan về nhận diện khuôn mặt và xác thực EKYC (định danh khách hàng điện tử); Nghiên cứu mô hình nhận diện khuôn mặt; Triển khai giải pháp ứng dụng và một số thử nghiệm đánh giá Đề án tham khảo 32 nguồn tài liệu khác nhau.

Nội dung nghiên cứu là vấn đề có tính thời sự về triển khai hệ thống xác thực khách hàng điện tử sử dụng các thuật toán nhận diện khuôn mặt của trí tuệ nhân tạo.

Bố cục đề án tương đối hợp lý và tường minh, tuy nhiên nên rà soát lại cách đánh chỉ mục, ví dụ 2.4 sau đó đến ngay 2.4.1.1...Một số hình vẽ còn mờ khó quan sát như hình 2.6, 2.11. Cần rà soát bổ sung trích dẫn đến những nội dung mà học viên tham khảo như mục 2.4.5 về phát hiện khuôn mặt giả mạo, các công thức tại trang 38,39...

* Lưu ý: Bản nhận xét phải đánh máy, không viết tay.

Việc thực nghiệm còn sơ sài, mới thu thập 175 người, nên nghiên cứu bổ sung thêm dữ liệu, có thể nghiên cứu sử dụng các bộ dữ liệu mở. Hệ thống thử nghiệm cũng mới chỉ dừng ở ảnh tĩnh mà chưa có mô hình triển khai thực tế thu nhận ảnh như thế nào.

Đề nghị học viên làm rõ thêm các vấn đề sau:

- Việc sử dụng mô hình Graph Attention Network để tổng hợp các vector đặc trưng có làm giảm độ chính xác quá trình tìm kiếm không?

- Việc sử dụng cụ tìm kiếm FAISS của Facebook chỉ phù hợp với số lượng lớn khuôn mặt, tuy nhiên nếu chỉ kiểm thử đối với 175 người thì liệu có phù hợp?

- Thời gian suy luận mà tác giả đưa ra ở bảng 3.1 trang 50 được tính như thế nào, có bao gồm thời gian phát hiện khuôn mặt và so sánh với tổng cộng bao nhiêu khuôn mặt, khi số lượng khuôn mặt trong kho dữ liệu tăng lên thì thời gian sẽ thay đổi theo chiều hướng nào?

Mặc dù còn tồn tại một số điểm cần trao đổi như trên, tuy nhiên trong phạm vi nghiên cứu học viên Nguyễn Thị Vân đã hoàn thành nhiệm vụ được giao, tôi đồng ý để học viên bảo vệ trước hội đồng./.

Hà Nội, Ngày 9 tháng 7 năm 2025.

NGƯỜI CHẤM PHẢN BIỆN

Trần Nguyễn Ngọc

* Lưu ý: Bàn nhận xét phải đánh máy, không viết tay.

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập – Tự do – Hạnh phúc

BẢN NHẬN XÉT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

(Dùng cho người phân biện)

Tên đề tài đề án tốt nghiệp: Nghiên cứu ứng dụng nhận diện khuôn mặt
trên các thiết bị EKYC tài ngân hàng
Chuyên ngành: Hệ thống thông tin
Mã chuyên ngành: 8.48.01.04
Họ và tên học viên: Nguyễn Thị Vân
Họ và tên người nhận xét: Bùi Hải Phong
Học hàm, học vị: Tiến sĩ
Chuyên ngành: Khoa học máy tính
Cơ quan công tác: Trường đại học Kiến Trúc Hà Nội
Số điện thoại: 0915594033 E-mail: phong.bh@hau.edu.vn

NỘI DUNG NHẬN XÉT

I/ Cơ sở khoa học và thực tiễn, tính cấp thiết của đề tài:

Đề tài nghiên cứu các kỹ thuật nhận diện khuôn mặt,
sau đó tích hợp kết quả nhận diện vào hệ thống định
danh khách hàng. Đề tài thực hiện dựa trên khái niệm
định luật phương pháp triển khai nhận diện khuôn
mặt sử dụng các kỹ thuật học sâu, Setivaface, ArcFace
và trung tâm dựa trên dữ liệu đăng của các trang ảnh
khuôn mặt. Đề tài có tính công nghệ, khoa học và thực
tiễn cao.

II/ Nội dung của đề án tốt nghiệp, các kết quả đã đạt được:

Đề tài gồm 57 trang, 3 chương, danh mục các hình vẽ,
biểu đồ và tài liệu tham khảo. Đề tài nghiên cứu
các công nghệ nhận diện khuôn mặt, các thể thức đăng ký
tài khoản. Đề tài hình thành kiến thức nhận diện khuôn
mặt sử dụng trích chọn đặc trưng và so sánh để đăng ký

của tước tướng. Đề tài đã thử nghiệm và so sánh kết quả với các phương pháp khác để đánh giá hệ thống. Đề tài đã thu thập dữ liệu thực tế tại ngân hàng VPBank.

III/ Những vấn đề cần giải thích thêm:

Đề tài xem xét, làm rõ các vấn đề sau:

- Đề tài cần làm rõ hơn về tài, vai trò của máy dùng nhân diện khuôn mặt trong hệ thống xác thực của ngân hàng VPBank?

IV/ Kết luận:

Đồng ý cho học viên bảo vệ đề án tốt nghiệp.

Đồng ý (hoặc không đồng ý) cho phép học viên bảo vệ đề án tốt nghiệp.

Ngày.....tháng.....7.....năm 2025

NGƯỜI NHẬN XÉT


Bùi Hải Phong

BÁO CÁO GIẢI TRÌNH
SỬA CHỮA, HOÀN THIỆN ĐỀ ÁN TỐT NGHIỆP

Họ và tên học viên: Nguyễn Thị Vân
Chuyên ngành: Hệ thống thông tin
Khóa: 2023 đợt 2

Tên đề tài: NGHIÊN CỨU ỨNG DỤNG NHÂN DIỆN KHUÔN MẶT TRONG XÁC
THỰC EKYC TẠI NGÂN HÀNG VPBANK

Người hướng dẫn khoa học: TS. Vũ Văn Thỏa

Ngày báo vệ: 19/07/2025

Các nội dung học viên đã sửa chữa, bổ sung trong đề án tốt nghiệp theo ý kiến đóng
góp của Hội đồng chấm đề án tốt nghiệp:

TT	Ý kiến hội đồng	Sửa chữa của học viên
1	Xem lại chỉ mục, bị sai (2.4.4.1).	Học viên đã rà soát, chỉnh sửa các lỗi về chỉ mục ở các trang 27-30.
2	Mục 2.4.5 công thức bị tràn sang trang 39.	Tiếp thu góp ý của Hội đồng, tác giả đã chỉnh sửa để công thức không bị tràn sang trang 39.
3	Bổ sung môi trường thử nghiệm, chạy như thế nào, có sử dụng GPU ko?	Tiếp thu góp ý của Hội đồng, tác giả đã bổ sung môi trường thực nghiệm, có sử dụng GPU ở mục 3.3.2 trang 48 và 49.
4	Làm rõ hơn vị trí vai trò của ứng dụng nhận diện khuôn mặt trong hệ thống xác thực của ngân hàng Vpbank.	Tiếp thu góp ý của Hội đồng, tác giả đã chỉnh sửa, làm rõ nội dung ở trang 43 và 44.
5	Làm rõ mô hình Graph Attention Network GAT có làm ảnh hưởng độ chính xác quá trình tìm kiếm không?	Tiếp thu góp ý của Hội đồng, tác giả đã làm rõ ảnh hưởng của mô hình Graph Attention Network GAT đến quá trình tìm kiếm ở trang 51.
6	Việc sử dụng công cụ tìm kiếm FAISS của Facebook chỉ phù hợp với số lượng lớn khuôn mặt, tuy nhiên nếu chỉ kiểm thử đối với 175 người thì liệu có phù hợp?	Tiếp thu góp ý của Hội đồng, tác giả đã làm rõ việc sử dụng công cụ tìm kiếm FAISS cho 175 người là phù hợp với bối cảnh xây dựng giải pháp có khả năng mở rộng như xác thực khuôn mặt

		ở trang 48 và 49.
7	Làm rõ thời gian suy luận ở bảng 3.1 trang 50, khi số lượng mặt trong kho dữ liệu tăng lên thì thời gian sẽ thay đổi theo chiều hướng nào?	Tiếp thu góp ý của Hội đồng, tác giả đã nêu cách thức tính thời gian suy luận ở bảng 3.1 ở trang 51. Và cũng làm rõ chiều hướng thay đổi của thời gian khi dữ liệu tăng lên ở trang 51.

Hà Nội, ngày 4 tháng 8 năm 2025

Ký xác nhận của

CHỦ TỊCH HỘI ĐỒNG
CHẤM ĐỀ ÁN



PGS.TSKH. Hoàng Đăng Hải

THỦ KÝ HỘI ĐỒNG



TS. Đào Ngọc Phong

NGƯỜI HUỚNG
DẪN KHOA HỌC



TS. Vũ Văn Thỏa

HỌC VIÊN



Nguyễn Thị Vân

