

NHẬN DIỆN KHUÔN MẶT ĐỘ CHÍNH XÁC CAO VỚI INSIGHT FACE VÀ MẠNG NƠ-RON SÂU

Học Phần: Trí Tuệ Nhân Tạo

Giảng Viên: Lê Trung Hiếu

1st Đinh Ngọc Chính
Khoa công nghệ thông tin
Đại học Đại Nam
Hà Nội, Việt Nam
MSV: 1771020103

2nd Phạm Văn Trà
Khoa công nghệ thông tin
Đại học Đại Nam
Hà Nội, Việt Nam
MSV: 1771020676

3rd Trần Dương Anh
Khoa công nghệ thông tin
Đại học Đại Nam
Hà Nội, Việt Nam
MSV: 1771020052

4th Trương Hữu Vinh
Khoa công nghệ thông tin
Đại học Đại Nam
Hà Nội, Việt Nam
MSV: 1771020759

Tóm tắt nội dung—Bài báo cáo trình bày hệ thống nhận dạng khuôn mặt kết hợp InsightFace và mạng nơ-ron sâu, hỗ trợ điểm danh tự động trong lớp học với độ chính xác cao và tốc độ xử lý nhanh. Hệ thống hoạt động hiệu quả ngay cả trong điều kiện ánh sáng và góc chụp khác nhau, đảm bảo khả năng tổng quát tốt trên nhiều bộ dữ liệu. Kết quả thực nghiệm cho thấy phương pháp này không chỉ nâng cao độ chính xác mà còn có tiềm năng ứng dụng rộng rãi trong quản lý lớp học thông minh và các lĩnh vực liên quan.

Keywords—InsightFace, nhận dạng khuôn mặt, mạng nơ-ron sâu, thời gian thực, hệ thống an ninh

I. GIỚI THIỆU ĐỀ TÀI

A. Nhận diện khuôn mặt trong thời đại số

Nhận diện khuôn mặt là một trong những ứng dụng quan trọng của trí tuệ nhân tạo và thị giác máy tính, đóng vai trò then chốt trong nhiều lĩnh vực như an ninh, giám sát, kiểm soát ra vào, xác thực danh tính, thanh toán điện tử, chăm sóc khách hàng và thậm chí trong y tế. Công nghệ này cho phép máy tính phân tích và nhận diện khuôn mặt của một hay nhiều người trong ảnh hoặc video, từ đó xác định danh tính hoặc phân loại đối tượng dựa trên đặc điểm sinh trắc học.

Trước đây các phương pháp nhận diện khuôn mặt theo cách truyền thống đã phụ thuộc vào các kỹ thuật xử lý ảnh như Haar Cascade và Local Binary Patterns đã được áp dụng rộng rãi. Tuy vậy chúng vẫn tồn tại một số hạn chế, đặc biệt là với điều kiện ánh sáng không ổn định, sự thay đổi góc mặt hay các biến đổi tự nhiên của khuôn mặt theo thời gian như lão hóa hoặc có những biểu cảm khác nhau. Những hạn chế này đang ảnh hưởng đến sự chính xác và ổn định của hệ thống nhận diện khuôn mặt, gây ra khó khăn và không đáp ứng tốt cho nhu cầu thực tế.

Sự phát triển vượt bậc của các mô hình học sâu đặc biệt là mạng nơ-ron tích chập đã tạo nên những bước tiến mang tính cách mạng trong lĩnh vực nhận diện khuôn mặt. Các mô hình hiện đại có khả năng tự động trích xuất và học hỏi các đặc trưng khuôn mặt từ lượng dữ liệu lớn từ đó nâng cao đáng kể độ chính xác ngay cả trong những điều kiện môi trường không thuận lợi. Bên cạnh đó, các kỹ thuật nhúng khuôn mặt như FaceNet, DeepFace và InsightFace đã mở ra những khả năng mới trong việc so sánh và phân loại khuôn mặt. Những phương pháp này

không chỉ cải thiện hiệu suất mà còn tăng cường tính ổn định và độ tin cậy của hệ thống nhận diện khuôn mặt, mở đường cho nhiều ứng dụng thực tiễn trong các lĩnh vực như an ninh, xác thực danh tính và tương tác người-máy.

Sự tiến bộ vượt bậc trong lĩnh vực nhận diện khuôn mặt ngày nay là nhờ vào sự kết hợp giữa trí tuệ nhân tạo và các kỹ thuật học sâu. Giờ đây, các hệ thống không chỉ đạt độ chính xác cao đáng kinh ngạc mà còn có khả năng hiểu và nhận diện khuôn mặt một cách tổng quát, bất kể sự khác biệt trong các tập dữ liệu. Chúng ta có thể thấy chúng trong các hệ thống điểm danh tự động tại trường học, giúp đơn giản hóa việc quản lý danh tính trong các giao dịch tài chính hoặc tăng cường an ninh bằng cách giám sát thông minh trong các thành phố hiện đại. Nhận diện khuôn mặt không chỉ là một công nghệ tiên tiến, mà còn là một phần không thể thiếu trong cuộc sống số hóa của chúng ta, giúp nâng cao hiệu quả và an toàn trong nhiều khía cạnh của đời sống.

B. Bối cảnh thực tế và động lực nghiên cứu đề tài

Nhận diện khuôn mặt là một trong những bài toán cốt lõi của thị giác máy tính và trí tuệ nhân tạo, đóng vai trò quan trọng trong nhiều lĩnh vực khác nhau như an ninh, kiểm soát truy cập, thanh toán điện tử, nhận dạng danh tính, và cá nhân hóa trải nghiệm người dùng. Với sự phát triển mạnh mẽ của công nghệ học sâu (Deep Learning), các phương pháp truyền thống dựa trên trích xuất đặc trưng thủ công đang dần được thay thế bởi các mô hình học sâu có khả năng học biểu diễn đặc trưng khuôn mặt một cách tự động và chính xác hơn.

Trong những năm gần đây, sự ra đời của các mạng nơ-ron tích chập đã giúp cải thiện đáng kể hiệu suất của các hệ thống nhận diện khuôn mặt. Đặc biệt là những mô hình tiên tiến như ArcFace, CosFace, SphereFace đã chứng minh hiệu quả vượt trội nhờ khả năng tối ưu hóa khoảng cách giữa các đặc trưng khuôn mặt trên không gian nhúng. Trong số đó, InsightFace là một framework hiện đại được xây dựng dựa trên ArcFace, kết hợp nhiều kỹ thuật tiên tiến nhằm cải thiện độ chính xác và khả năng phân biệt khuôn mặt trên tập dữ liệu lớn.

Tuy nhiên, việc triển khai InsightFace trong các môi trường thực tế vẫn còn nhiều thách thức, chẳng hạn như khả năng tổng

quát hóa đối với nhiều điều kiện ánh sáng, góc nhìn khác nhau, và tốc độ xử lý thời gian thực. Do đó, việc nghiên cứu và tối ưu hóa InsightFace cho các ứng dụng thực tế không chỉ là một vấn đề khoa học cấp thiết mà còn mang lại cho chúng ta giá trị thực tiễn to lớn.

C. Mục tiêu của đề tài

Mục tiêu của nhóm em là hướng đến việc cải thiện hiệu suất của hệ thống nhận diện khuôn mặt bằng cách khai thác những yếu tố của InsightFace và mạng nơ-ron sâu. Cụ thể, đề tài sẽ hướng tới những mục tiêu sau:

- **Nâng cao độ chính xác khi nhận diện gương mặt:**
 - Tối ưu hóa việc sử dụng framework InsightFace kết hợp với kiến trúc mạng nơ-ron sâu nhằm cải thiện khả năng phân biệt các đặc trưng khuôn mặt.
- **Tăng cường khả năng nhận diện khi có các điều kiện bất lợi:**
 - Đảm bảo hệ thống có khả năng nhận diện chính xác trong các điều kiện môi trường khác nhau, bao gồm sự thay đổi về ánh sáng, góc nhìn và biểu cảm khuôn mặt khác nhau.
 - Đánh giá hiệu suất của mô hình trên các tập dữ liệu đa dạng, bao gồm cả dữ liệu tiêu chuẩn và dữ liệu thực tế để đảm bảo tính ổn định và độ tin cậy của hệ thống.
- **Đảm bảo hiệu suất thời gian thực:** Tối ưu hóa mô hình để đạt được tốc độ xử lý nhanh chóng, đáp ứng yêu cầu của các ứng dụng thời gian thực như giám sát an ninh và kiểm soát ra vào.
- **Xây dựng một ứng dụng thực tiễn:**
 - Phát triển một ứng dụng nhận diện khuôn mặt có khả năng triển khai trong các môi trường thực tế, với các ứng dụng tiềm năng trong lĩnh vực an ninh, xác thực danh tính và kiểm soát truy cập.
 - Đảm bảo tính dễ dàng tích hợp và mở rộng của hệ thống.

D. Phạm vi và đối tượng nghiên cứu

Báo cáo này nhóm chúng em tập trung vào việc nghiên cứu và phát triển một hệ thống nhận diện khuôn mặt có độ chính xác cao, sử dụng framework InsightFace kết hợp với mạng nơ-ron sâu. Phạm vi và đối tượng nghiên cứu được xác định như sau:

- **Thu thập và xử lý dữ liệu:** Nhóm em thu thập và xử lý dữ liệu hình ảnh của các thành viên trong lớp qua các điều kiện ánh sáng và góc chụp khác nhau.
- **Xây dựng và tối ưu hóa mô hình:**
 - Nhóm em sẽ tập trung vào việc xây dựng, huấn luyện và tối ưu hóa mô hình nhận diện khuôn mặt sử dụng framework InsightFace và mạng nơ-ron sâu.
 - Nghiên cứu và áp dụng các phương pháp tối ưu hóa hàm mất mát, kiến trúc mạng và các kỹ thuật huấn luyện để nâng cao độ chính xác và hiệu năng của mô hình.
- **Đánh giá hiệu năng:**

- Phân tích và so sánh kết quả nhận diện khuôn mặt, đánh giá hiệu năng của mô hình dựa trên những điều kiện khác nhau.
- Sử dụng các chỉ số đánh giá hiệu năng phù hợp để đo lường và so sánh hiệu quả của mô hình trong các điều kiện thực nghiệm khác nhau.

E. Những khó khăn của đề tài

Nhận diện khuôn mặt là một trong những bài toán mặc dù đã được nghiên cứu từ lâu nhưng bài toán này vẫn còn nhiều thách thức cần giải quyết để đạt được độ chính xác cao trong các điều kiện thực tế. Hiện nay, các phương pháp hiện đại như InsightFace và mạng nơ-ron sâu đã giúp cải thiện đáng kể độ chính xác, tuy nhiên vẫn còn nhiều yếu tố ảnh hưởng đến hiệu suất của hệ thống.

Các thách thức của bài toán nhận diện khuôn mặt có thể kể đến như sau:

- **Tư thế và góc chụp:** Ảnh khuôn mặt có thể thay đổi tùy vào góc chụp, ánh sáng và các tư thế khác nhau như chụp thẳng, chụp chéo, chụp từ trên xuống hoặc từ dưới lên có thể làm thay đổi cấu trúc khuôn mặt và gây khó khăn cho chương trình khi nhận diện.
- **Sự xuất hiện hoặc thiếu hụt một số đặc điểm trên khuôn mặt:** Những yếu tố như râu, kính mắt hoặc trang sức có thể làm thay đổi đáng kể hình dạng khuôn mặt. Điều này khiến hệ thống nhận diện gặp khó khăn trong việc xác định các yếu tố đặc trưng của khuôn mặt.
- **Biểu cảm khuôn mặt:** Cùng một khuôn mặt đó nhưng có thể có những biến đổi lớn khi họ biểu lộ cảm xúc như vui vẻ, tức giận, buồn bã hay ngạc nhiên,... Điều này làm thay đổi đáng kể các đặc trưng của khuôn mặt, gây ảnh hưởng đến độ chính xác của mô hình.
- **Điều kiện ánh sáng và chất lượng hình ảnh:** Ảnh chụp trong điều kiện ánh sáng yếu hoặc bị nhiễu do chất lượng camera kém cũng có thể làm giảm hiệu suất khi nhận diện.
- **Sự thay đổi theo thời gian:** Gương mặt bị thay đổi do lão hóa hay người đó mới đi phẫu thuật chỉnh sửa một bộ phận nào đó trên gương mặt thì sẽ làm cho hệ thống khó nhận diện chính xác người đó là ai khi so sánh với tấm ảnh đã chụp cách đó nhiều năm.
- **Hệ thống quy mô lớn:** Trong thực tế thì hệ thống nhận diện khuôn mặt cần phải xử lý một lượng dữ liệu khổng lồ, ví dụ như cơ sở dữ liệu khuôn mặt của cảnh sát có thể chứa hàng triệu đến hàng tỷ hình ảnh. Việc tối ưu hóa hiệu suất và tốc độ nhận diện trên các tập dữ liệu lớn là một thách thức quan trọng.

F. Cấu trúc bài báo cáo

Báo cáo này được tổ chức thành các phần chính như sau:

- **Phần 1: Giới thiệu** – Trình bày bối cảnh nghiên cứu, động lực thực hiện đề tài, mục tiêu nghiên cứu, phạm vi và cấu trúc bài báo cáo.
- **Phần 2: Cơ sở lý thuyết** – Cung cấp nền tảng lý thuyết về nhận diện khuôn mặt, mạng nơ-ron sâu, và mô hình InsightFace.

- **Phần 3: Phương pháp nghiên cứu** – Mô tả chi tiết về kiến trúc mô hình, quy trình huấn luyện và các chiến lược tối ưu hóa được áp dụng.
- **Phần 4: Thực nghiệm và đánh giá** – Trình bày kết quả thí nghiệm, phân tích hiệu suất của mô hình và so sánh với các phương pháp nhận diện khác.
- **Phần 5: Kết luận và hướng phát triển** – Tổng kết các đóng góp chính của nghiên cứu và đề xuất hướng phát triển trong tương lai.

II. CƠ SỞ LÝ THUYẾT

A. Mạng Nơ-ron Sâu trong nhận diện khuôn mặt

1) *Tổng quan về mạng nơ-ron sâu:* Mạng Nơ-ron Sâu (Deep Neural Networks - DNNs) là một loại mạng nơ-ron nhân tạo với nhiều lớp giữa lớp đầu vào và lớp đầu ra. Chúng bao gồm các thành phần như nơ-ron, khớp nối, trọng số, độ chệch và các hàm số, hoạt động mô phỏng chức năng của não người. DNNs có khả năng học các mối quan hệ phi tuyến phức tạp và đã được áp dụng thành công trong nhiều lĩnh vực như nhận dạng hình ảnh và xử lý ngôn ngữ tự nhiên.[1]

Trong bài toán nhận diện khuôn mặt, mạng nơ-ron sâu có thể tự động trích xuất các đặc trưng quan trọng từ hình ảnh đầu vào, giúp cải thiện độ chính xác so với các phương pháp truyền thống. Nhờ vào khả năng học từ dữ liệu lớn, DNNs có thể nhận diện khuôn mặt ngay cả trong điều kiện ánh sáng kém, các góc chụp khác nhau hoặc khuôn mặt có những biểu cảm đa dạng.

2) *Mô hình ArcFace và vai trò trong nhận diện khuôn mặt:* ArcFace là một phương pháp tối ưu hóa trong nhận dạng khuôn mặt, dựa trên hàm mất mát biên góc cộng (Additive Angular Margin Loss). Phương pháp này giúp tăng cường khả năng phân biệt của mô hình bằng cách thêm một biên góc vào logit mục tiêu. Điều này giúp thu hẹp khoảng cách giữa các đặc trưng trong cùng một lớp (intra-class compactness) và mở rộng khoảng cách giữa các lớp khác nhau (inter-class discrepancy), từ đó cải thiện độ chính xác trong nhận dạng khuôn mặt. [2]

Công thức của ArcFace được định nghĩa như sau:

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_i+m))}}{e^{s(\cos(\theta_i+m))} + \sum_{j \neq y_i} e^{s \cos(\theta_j)}} \quad [2]$$

Trong đó:

- θ_i là góc giữa vector đặc trưng của mẫu và vector trọng số của lớp tương ứng
- m là hệ số biên
- s là hệ số co giãn để điều chỉnh khoảng cách giữa các lớp

ArcFace giúp mô hình tạo ra các embeddings có khoảng cách lớn hơn giữa những người khác nhau giúp tăng độ chính xác của hệ thống nhận diện khuôn mặt.

3) *Kiến trúc ResNet/MobileFaceNet trong InsightFace:* InsightFace là một hệ thống phân tích khuôn mặt tiên tiến, sử dụng các mô hình học sâu như ResNet và MobileFaceNet để trích xuất đặc trưng khuôn mặt. [2]

ResNet (Residual Network) là một kiến trúc mạng nơ-ron sâu phổ biến, được đề xuất để giải quyết vấn đề suy giảm gradient khi độ sâu của mạng tăng lên. ResNet được thiết kế để giải quyết vấn đề suy giảm gradient khi độ sâu của mạng

tăng lên, bằng cách sử dụng các kết nối tắt (skip connections) giúp duy trì luồng thông tin và cải thiện khả năng huấn luyện mạng sâu. [3] Công thức của một lớp Residual Block được biểu diễn như sau:

$$Y = X + F(X, W) \quad [3]$$

Trong đó:

X là đầu vào

$F(X, W)$ là phép biến đổi phi tuyến với trọng số W giúp cải thiện khả năng học các đặc trưng phức tạp.

MobileFaceNet là một kiến trúc mạng nơ-ron tích chập nhẹ, được thiết kế tối ưu cho các thiết bị di động. Mạng này sử dụng các khối tích chập phân tách theo chiều sâu (Depthwise Separable Convolutions) nhằm giảm số lượng tham số và chi phí tính toán, đồng thời vẫn đảm bảo hiệu suất cao trong các tác vụ xử lý ảnh. [4]

Trong InsightFace thì việc lựa chọn giữa ResNet hay MobileFaceNet phải tùy thuộc vào yêu cầu của từng ứng dụng cụ thể, ResNet thường được sử dụng khi chương trình yêu cầu độ chính xác cao nhất, trong khi đó thì MobileFaceNet được ưu tiên khi cần tối ưu hóa hiệu suất trên các thiết bị di động. Cả hai kiến trúc này đều được tích hợp chặt chẽ với các thành phần khác của InsightFace, chẳng hạn như hàm mất mát ArcFace để tạo ra một hệ thống nhận diện khuôn mặt hiệu quả.

4) *Kỹ thuật Embedding Learning trong nhận diện khuôn mặt:* Học biểu diễn (Embedding Learning) là một phương pháp quan trọng trong lĩnh vực nhận diện khuôn mặt, giúp ánh xạ hình ảnh khuôn mặt vào một không gian nhúng (embedding space), trong đó khoảng cách giữa các vector đặc trưng phản ánh mức độ tương đồng giữa các khuôn mặt. Phương pháp này cho phép chuyển đổi hình ảnh khuôn mặt thành các vector có tính phân biệt cao, giúp hệ thống nhận diện có thể so sánh và phân biệt các khuôn mặt một cách hiệu quả. Cụ thể, khuôn mặt của cùng một người sẽ có khoảng cách gần nhau trong không gian nhúng, trong khi khuôn mặt của những người khác nhau sẽ có khoảng cách xa hơn. [5]

Trong hệ thống nhận diện khuôn mặt, Embedding Learning đóng vai trò quan trọng trong việc tạo ra các vector đặc trưng có khả năng phân biệt cao. Một phương pháp phổ biến để tối ưu hóa quá trình này là sử dụng các hàm mất mát dựa trên lề góc, giúp tăng cường độ phân biệt giữa các vector nhúng. Kỹ thuật này đảm bảo rằng khoảng cách giữa các vector thuộc về các danh tính khác nhau được mở rộng một cách hiệu quả, từ đó nâng cao độ chính xác của hệ thống nhận diện khuôn mặt ngay cả trong các điều kiện phức tạp. [2]

Quá trình Embedding Learning trong nhận diện khuôn mặt bao gồm ba bước chính: huấn luyện mô hình, tạo vector đặc trưng và so sánh vector đặc trưng. Trước tiên, mô hình mạng nơ-ron sâu được huấn luyện trên một tập dữ liệu lớn nhằm học cách ánh xạ hình ảnh khuôn mặt vào một không gian vector có tính phân biệt cao. Sau khi được huấn luyện, mô hình có thể chuyển đổi bất kỳ khuôn mặt nào thành một vector đặc trưng trong không gian nhúng. Để xác định mức độ giống nhau giữa hai khuôn mặt, ta so sánh khoảng cách giữa các vector đặc trưng của chúng, với khuôn mặt của cùng một người có khoảng cách gần nhau hơn so với các khuôn mặt khác nhau. [5]

B. InsightFace và phương pháp nhận diện khuôn mặt

1) *Giới thiệu về InsightFace*: InsightFace là một thư viện mã nguồn mở chuyên về nhận diện khuôn mặt, cung cấp các giải pháp tiên tiến cho các tác vụ như phát hiện, căn chỉnh và nhận diện khuôn mặt. Hệ thống này được xây dựng dựa trên các mô hình học sâu hiệu quả, trong đó ArcFace đóng vai trò quan trọng như một phương pháp hiện đại giúp cải thiện độ chính xác của nhận diện khuôn mặt.[6]

Một trong những điểm nổi bật của InsightFace là việc sử dụng hàm mất mát ArcFace, giúp tăng cường khoảng cách giữa các vector đặc trưng của các cá nhân khác nhau, từ đó nâng cao độ chính xác của hệ thống nhận diện khuôn mặt. Hàm mất mát này tạo ra một biên độ góc bổ sung trong không gian nhúng, giúp cải thiện khả năng phân biệt giữa các danh tính khác nhau. Bên cạnh đó, InsightFace còn áp dụng các kiến trúc mạng nơ-ron sâu như ResNet và MobileFaceNet, giúp tối ưu hóa quá trình trích xuất đặc trưng khuôn mặt, đảm bảo hiệu suất cao ngay cả trên các thiết bị có tài nguyên hạn chế. [2]

2) *Cách InsightFace trích xuất đặc trưng khuôn mặt*: Hệ thống InsightFace thực hiện trích xuất đặc trưng khuôn mặt thông qua các bước sau:

• Tiền xử lý hình ảnh khuôn mặt

- Ảnh khuôn mặt được phát hiện và căn chỉnh sao cho vị trí mắt, mũi và miệng nằm ở các vị trí cố định.
- Các phép biến đổi như chuẩn hóa màu sắc, xoay và cắt ảnh được áp dụng để đảm bảo tính đồng nhất trong tập dữ liệu đầu vào.

• Mạng nơ-ron sâu trích xuất đặc trưng

- Hình ảnh đã được căn chỉnh sẽ được đưa vào một mô hình mạng nơ-ron sâu.
- Mạng nơ-ron sẽ chuyển đổi ảnh đầu vào thành một vector đặc trưng có kích thước cố định (ví dụ như 512 chiều).
- Các đặc trưng được học từ dữ liệu giúp hệ thống nhận diện các điểm khác biệt và tương đồng giữa các khuôn mặt.

• Chuẩn hóa và ánh xạ đặc trưng

- Sau khi trích xuất, vector đặc trưng sẽ được chuẩn hóa để có độ dài cố định.
- Các vector này được ánh xạ vào một không gian Euclidean sao cho các khuôn mặt của cùng một người nằm gần nhau, còn khuôn mặt của những người khác nằm xa nhau.

3) *Phương pháp so sánh Embeddings để nhận diện*: Sau khi trích xuất được vector đặc trưng của khuôn mặt, hệ thống InsightFace thực hiện quá trình nhận diện bằng cách so sánh các embeddings trong không gian đặc trưng. Quá trình này bao gồm:

• Tính toán khoảng cách giữa các embeddings

- Hệ thống sử dụng *khoảng cách cosine* để đo độ tương đồng giữa hai vector đặc trưng.
- Công thức tính khoảng cách cosine giữa hai vector A và B như sau:

$$d_{\cos}(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad [5]$$

Trong đó:

- * $A \cdot B$ là tích vô hướng (dot product) của hai vector A và B . Tích vô hướng được tính bằng tổng của tích các phần tử tương ứng của hai vector.
- * $\|A\|$ và $\|B\|$ là độ dài (norm) của vector A và B tương ứng. Độ dài của một vector được tính bằng căn bậc hai của tổng bình phương các phần tử của vector đó.
- * $d_{\cos}(A, B)$ là giá trị Cosine Similarity giữa hai vector A và B . Giá trị này nằm trong khoảng $[-1, 1]$. Giá trị càng gần 1 thì hai khuôn mặt càng giống nhau, giá trị càng gần -1 thì hai khuôn mặt càng khác nhau, và giá trị gần 0 thì hai vector không có sự tương đồng.

Giá trị này càng gần 1 thì hai khuôn mặt càng giống nhau.

C. Xử lý hình ảnh trong nhận diện khuôn mặt

Xử lý hình ảnh là một bước quan trọng trong nhận diện khuôn mặt, giúp cải thiện chất lượng đầu vào và tăng độ chính xác của hệ thống. Quá trình xử lý ảnh thường bao gồm các giai đoạn như tiền xử lý, phát hiện khuôn mặt và điều chỉnh ảnh để giảm ảnh hưởng của các yếu tố môi trường như góc chụp và ánh sáng.

1) *Tiền xử lý ảnh với OpenCV*: OpenCV (viết tắt của Open Source Computer Vision Library) là một thư viện mã nguồn mở chuyên dùng trong xử lý ảnh và thị giác máy tính. Công nghệ cung cấp các công cụ và thư viện để phân tích và xử lý ảnh, video từ việc xác định các đối tượng trong ảnh đến việc nhận diện khuôn mặt hoặc theo dõi chuyển động khác. [7] Các bước chính của tiền xử lý ảnh bao gồm:

– Bước 1: Đọc ảnh

- * Sử dụng hàm `cv2.imread()` để đọc ảnh từ file.
- * OpenCV hỗ trợ nhiều định dạng ảnh phổ biến như JPEG, PNG, BMP.

– Bước 2: Phát hiện khuôn mặt

- * Sử dụng các thuật toán phát hiện khuôn mặt như Haar cascades hoặc DNN (Deep Neural Network).
- * Các mô hình đã được huấn luyện sẵn có thể được sử dụng để phát hiện khuôn mặt một cách nhanh chóng và hiệu quả.

– Bước 3: Cắt và căn chỉnh khuôn mặt

- * Cắt phần khuôn mặt từ ảnh gốc dựa trên vị trí đã phát hiện.
- * Căn chỉnh khuôn mặt để đảm bảo các đặc điểm như mắt, mũi, miệng nằm ở vị trí cố định.
- * Việc căn chỉnh giúp giảm thiểu sự ảnh hưởng của góc chụp và tư thế khuôn mặt.

– Bước 4: Chuyển đổi ảnh xám

- * Chuyển đổi ảnh màu sang ảnh xám bằng hàm `cv2.cvtColor()`.
- * Ảnh xám chỉ chứa thông tin về độ sáng, giúp giảm tải cho mô hình nhận diện.

– Bước 5: Chuẩn hóa ảnh

- * Chuẩn hóa độ sáng và độ tương phản của ảnh để đảm bảo tính đồng nhất.
- * Các phương pháp chuẩn hóa phổ biến bao gồm histogram equalization và adaptive histogram equalization.

– Bước 6: Thay đổi kích thước ảnh

- * Thay đổi kích thước ảnh về kích thước cố định bằng hàm `cv2.resize()`.
- * Giúp phù hợp với yêu cầu của mô hình nhận diện.

– Bước 7: Làm mờ ảnh

- * Áp dụng các bộ lọc làm mờ (blur) để giảm nhiễu và làm mịn ảnh.
- * Các bộ lọc làm mờ phổ biến bao gồm Gaussian blur và median blur.

2) **Phát hiện và cắt khuôn mặt từ ảnh:** Phát hiện và cắt khuôn mặt từ ảnh là một bước quan trọng trong quá trình xử lý ảnh cho nhận diện khuôn mặt. Mục đích của bước này là xác định chính xác vị trí khuôn mặt, loại bỏ các phần không liên quan, giúp tối ưu quá trình nhận diện ở các bước xử lý tiếp theo.

Các phương pháp phát hiện khuôn mặt phổ biến:

- **Haar cascades:** Haar Cascades là một phương pháp phát hiện khuôn mặt cổ điển, dựa trên việc sử dụng các đặc trưng Haar để nhận diện các đặc điểm của khuôn mặt. Phương pháp này có ưu điểm là tốc độ nhanh, dễ triển khai, nhưng có thể gặp khó khăn khi xử lý các điều kiện ánh sáng phức tạp hoặc góc chụp không thuận lợi. [8]
- **DNN (Deep Neural Network):** Các mô hình Deep Neural Network (DNN) đóng vai trò quan trọng trong việc phát hiện khuôn mặt với độ chính xác cao. Nhờ khả năng học sâu, các mô hình này có thể xử lý tốt các biến thể về ánh sáng, góc chụp và biểu cảm khuôn mặt. So với các phương pháp truyền thống như Haar Cascade, mô hình DNN có hiệu suất tốt hơn trong môi trường phức tạp. Tuy nhiên, một nhược điểm lớn của DNN là yêu cầu tài nguyên tính toán cao hơn, đòi hỏi phần cứng mạnh mẽ hơn để có thể hoạt động hiệu quả. [9]

Các bước thực hiện phát hiện và cắt khuôn mặt:

– Bước 1: Đọc ảnh

- * Sử dụng thư viện như OpenCV để đọc ảnh từ file.

– Bước 2: Phát hiện khuôn mặt

- * Sử dụng một trong các phương pháp trên (Haar cascades hoặc DNN) để xác định vị trí khuôn mặt trong ảnh.
- * Kết quả trả về là tọa độ của các hình chữ nhật bao quanh khuôn mặt.

– Bước 3: Cắt khuôn mặt

- * Sử dụng tọa độ hình chữ nhật để cắt phần khuôn mặt ra khỏi ảnh gốc.
- * Phần khuôn mặt đã cắt có thể được lưu lại hoặc sử dụng trực tiếp cho các bước xử lý tiếp theo.

Xử lý ảnh là một bước quan trọng trong nhận diện khuôn mặt, giúp cải thiện chất lượng đầu vào và nâng cao độ chính xác của hệ thống. Các kỹ thuật như tiền xử lý với OpenCV, phát hiện khuôn mặt bằng MTCNN và căn chỉnh khuôn mặt giúp hệ thống nhận diện ổn định ngay cả khi gặp các yếu tố gây nhiễu như góc chụp và ánh sáng.

D. Hệ thống nhận diện khuôn mặt thời gian thực

Nhận diện khuôn mặt thời gian thực là một ứng dụng quan trọng của công nghệ nhận diện khuôn mặt, được sử dụng trong giám sát an ninh, kiểm soát ra vào và các hệ thống tương tác thông minh. Hệ thống này yêu cầu tốc độ xử lý nhanh, độ chính xác cao và khả năng mở rộng để xử lý nhiều người dùng cùng lúc.

1) **Cấu trúc hệ thống nhận diện thời gian thực:** Một hệ thống nhận diện khuôn mặt thời gian thực bao gồm các thành phần chính sau: [10]

- **Nguồn dữ liệu đầu vào:** Camera giám sát, webcam hoặc luồng video trực tiếp từ thiết bị di động.
- **Tiền xử lý dữ liệu:**
 - * Chuyển đổi ảnh sang ảnh xám để giảm kích thước dữ liệu.
 - * Cân bằng độ sáng để cải thiện chất lượng nhận diện.
 - * Phát hiện và cắt khuôn mặt bằng MTCNN, Haar Cascade hoặc Dlib.
 - * Căn chỉnh khuôn mặt để đảm bảo đầu vào nhất quán.
- **Trích xuất đặc trưng khuôn mặt:** Sử dụng các mô hình học sâu như ResNet, MobileFaceNet hoặc InsightFace để tạo Embeddings.
- **So sánh và nhận diện:** Sử dụng phương pháp đo khoảng cách như:
 - * **Cosine Similarity:** Đo độ tương đồng giữa hai vector đặc trưng.
 - * **Euclidean Distance:** Đánh giá mức độ khác biệt giữa hai khuôn mặt.
- **Quản lý dữ liệu và kết quả:** Lưu trữ Embeddings vào cơ sở dữ liệu và hiển thị kết quả nhận diện.

2) **Xử lý đa luồng (Multithreading) trong nhận diện khuôn mặt:** Nhận diện khuôn mặt thời gian thực yêu cầu tốc độ xử lý cao, do đó sử dụng đa luồng giúp cải thiện hiệu suất bằng cách thực hiện các tác vụ song song:

- **Luồng đọc dữ liệu (Data Acquisition Thread):** Thu nhận hình ảnh liên tục từ camera.
- **Luồng tiền xử lý (Preprocessing Thread):** Cắt, chuyển đổi ảnh xám, căn chỉnh khuôn mặt.
- **Luồng trích xuất đặc trưng (Feature Extraction Thread):** Chạy mô hình học sâu để tạo Embeddings.
- **Luồng nhận diện và so sánh (Recognition Thread):** So sánh Embeddings với cơ sở dữ liệu.
- **Luồng quản lý dữ liệu (Data Management Thread):** Lưu trữ và cập nhật Embeddings.

Sử dụng đa luồng giúp giảm độ trễ và tối ưu tài nguyên hệ thống, đảm bảo quá trình nhận diện diễn ra mượt mà.

3) *Lưu trữ và quản lý dữ liệu Embeddings*: Dữ liệu Embeddings trong nhận diện khuôn mặt là các vector đặc trưng có số chiều cố định, được trích xuất từ mô hình học sâu nhằm biểu diễn khuôn mặt trong không gian số học. Theo Schroff, Kalenichenko và Philbin (2015), FaceNet trực tiếp huấn luyện đầu ra của mô hình thành một embedding nhỏ gọn bằng cách sử dụng hàm mất mát bộ ba, thay vì dựa vào một lớp trung gian. Quá trình này giúp tối ưu hóa khoảng cách giữa các vector, đảm bảo rằng các khuôn mặt của cùng một người sẽ gần nhau, trong khi khuôn mặt của những người khác sẽ xa nhau trong không gian đặc trưng. Nhờ đó, hệ thống có thể so sánh và phân biệt các khuôn mặt một cách hiệu quả mà không cần sử dụng hình ảnh gốc.[11]

Phương pháp lưu trữ dữ liệu

- **Lưu trữ trong bộ nhớ (In-memory Storage)**: Sử dụng Redis hoặc Numpy để lưu trữ tạm thời các Embeddings, giúp truy xuất nhanh chóng.
- **Lưu trữ trong cơ sở dữ liệu (Database Storage)**: Sử dụng các hệ quản trị cơ sở dữ liệu như SQLite, PostgreSQL hoặc MongoDB để lưu trữ lâu dài, đảm bảo tính bền vững và khả năng truy vấn phức tạp.

Tìm kiếm và so sánh Embeddings

- **Tìm kiếm gần đúng (Approximate Nearest Neighbors - ANN)**: Triển khai các thư viện như FAISS hoặc Annoy để tăng tốc độ truy vấn trong không gian vector đặc trưng lớn.
- **Chỉ mục KD-Tree hoặc Ball-Tree**: Xây dựng chỉ mục bằng KD-Tree hoặc Ball-Tree để tối ưu hóa việc tìm kiếm các Embeddings tương tự trong không gian vector.

Cập nhật và bảo trì dữ liệu

- Cập nhật các Embeddings mới khi có người dùng mới gia nhập hệ thống.
- Thực hiện xóa dữ liệu cũ hoặc không còn sử dụng để tối ưu hóa bộ nhớ và tăng hiệu suất truy vấn, đảm bảo hệ thống luôn hoạt động hiệu quả.

Hệ thống nhận diện khuôn mặt thời gian thực yêu cầu một kiến trúc tối ưu để đảm bảo tốc độ và độ chính xác cao. Việc sử dụng đa luồng giúp cải thiện hiệu suất, trong khi quản lý dữ liệu Embeddings đúng cách giúp hệ thống nhận diện nhanh chóng và chính xác. Những yếu tố này giúp hệ thống hoạt động ổn định và có thể triển khai vào các ứng dụng thực tế như kiểm soát an ninh, chăm công và dịch vụ khách hàng thông minh.

III. PHƯƠNG PHÁP NGHIÊN CỨU VÀ KỸ THUẬT SỬ DỤNG

A. Mô hình nghiên cứu

Đề tài này nhóm em sẽ nghiên cứu cách sử dụng học sâu (Deep Learning) và các mô hình mạng nơ-ron tích chập (CNN) để phát hiện, trích xuất đặc trưng và nhận diện khuôn mặt. Cụ thể:

- **Phát hiện khuôn mặt**: Sử dụng mô hình **InsightFace** với **RetinaFace** để xác định chính xác vị trí khuôn mặt trong ảnh.
- **Trích xuất đặc trưng**: Áp dụng **ArcFace** để tạo vector đặc trưng (embedding) đại diện cho mỗi khuôn mặt.
- **So sánh và nhận diện**: Sử dụng phương pháp **Cosine Similarity** hoặc **Euclidean Distance** để tính độ tương đồng giữa các khuôn mặt.

B. Phương pháp triển khai

Quy trình thực hiện gồm các bước:

1) Thu thập dữ liệu

- Dữ liệu gồm nhiều hình ảnh khuôn mặt các thành viên trong lớp với góc độ, điều kiện ánh sáng và biểu cảm khác nhau để đảm bảo tính đa dạng.
- Mỗi ảnh được lưu kèm thông tin định danh của từng thành viên như mã sinh viên, họ tên, lớp. Để đảm bảo chất lượng và tính khả dụng của dữ liệu, quá trình thu thập cần xem xét các yếu tố quan trọng sau:
- **Độ phân giải ảnh**: Hình ảnh cần có độ phân giải đủ cao để nhận dạng các đặc trưng khuôn mặt một cách rõ ràng.
- **Điều kiện ánh sáng**: Thu thập dữ liệu trong nhiều điều kiện ánh sáng khác nhau (tự nhiên, nhân tạo, ánh sáng yếu) để tăng tính đa dạng.
- **Góc chụp**: Ảnh chụp từ nhiều góc độ khác nhau (chính diện, nghiêng trái/phải, trên/dưới) giúp mô hình học được các biến đổi tự nhiên của khuôn mặt.
- **Biểu cảm khuôn mặt**: Thu thập dữ liệu với nhiều trạng thái cảm xúc khác nhau (vui, buồn, bình thường, cười) để hệ thống có thể nhận diện chính xác ngay cả khi biểu cảm thay đổi.
- **Yếu tố môi trường**: Bổ sung dữ liệu với các trường hợp có nhiễu như nền phức tạp, người đeo kính, đeo khẩu trang hoặc che một phần khuôn mặt để tăng khả năng tổng quát hóa của mô hình.

2) Tiền xử lý dữ liệu

- Chuẩn hóa ảnh về cùng kích thước để thống nhất dữ liệu đầu vào.
- Tăng cường dữ liệu bằng cách xoay, lật, điều chỉnh độ sáng nhằm giúp mô hình học tốt hơn.
- Mỗi ảnh được gán nhãn bằng thông tin của các thành viên theo định dạng MSV - Họ tên - Lớp. Quá trình này được thực hiện thủ công.



- 3) **Huấn luyện mô hình** Quá trình huấn luyện bao gồm phát hiện khuôn mặt, trích xuất đặc trưng và lưu trữ dữ liệu.

Bước 1: Phát hiện và cắt khuôn mặt bằng RetinaFace

- Sử dụng mô hình RetinaFace để phát hiện vị trí khuôn mặt trong ảnh.
- Khuôn mặt được cắt ra dựa trên bounding box được mô hình cung cấp.
- Nếu ảnh có nhiều khuôn mặt, chỉ lấy khuôn mặt lớn nhất hoặc lựa chọn theo một quy tắc xác định.

Bước 2: Trích xuất đặc trưng bằng ArcFace

- Sau khi phát hiện khuôn mặt, sử dụng mô hình ArcFace để trích xuất embedding (vector đặc trưng).
- Embedding này là một biểu diễn số hóa của khuôn mặt, giúp so sánh giữa các khuôn mặt hiệu quả.

Bước 3: Lưu trữ dữ liệu huấn luyện

- Các embedding thu được từ ArcFace được lưu vào cơ sở dữ liệu cùng với thông tin người dùng.
- Dữ liệu này sẽ được sử dụng trong quá trình nhận diện.

- 4) **Nhận diện khuôn mặt** Quá trình nhận diện khuôn mặt dựa trên việc so sánh embedding của khuôn mặt mới với dữ liệu đã được lưu.

Bước 1: Trích xuất embedding khuôn mặt mới

- Khi một người xuất hiện trước camera, hệ thống chụp ảnh và phát hiện khuôn mặt bằng RetinaFace.
- Sử dụng ArcFace để trích xuất embedding của khuôn mặt đó.

Bước 2: So sánh với cơ sở dữ liệu

- So sánh embedding mới với các embedding trong cơ sở dữ liệu để tìm độ tương đồng.

Bước 3: Xác nhận danh tính

- Nếu độ tương đồng lớn hơn một ngưỡng xác định trước, hệ thống xác nhận danh tính của người đó.
- Nếu không có khuôn mặt nào khớp, hệ thống trả về kết quả "Không xác định".

C. Kỹ thuật sử dụng

1) **Phát hiện khuôn mặt:** Hệ thống sử dụng RetinaFace – một mô hình dựa trên CNN giúp phát hiện khuôn mặt với độ chính xác cao ngay cả khi có góc quay hoặc điều kiện ánh sáng khác nhau. RetinaFace còn có thể xác định các điểm đặc trưng trên khuôn mặt như mắt, mũi, miệng, giúp cải thiện quá trình trích xuất đặc trưng.

2) **Trích xuất đặc trưng (Face Embedding):** Sau khi phát hiện khuôn mặt, hệ thống sử dụng ArcFace để tạo ra vector đặc trưng cho mỗi khuôn mặt. Các vector này sẽ được lưu trữ vào cơ sở dữ liệu (`face_db.pkl`) để phục vụ cho quá trình nhận diện.

3) **So sánh và nhận diện khuôn mặt:** Hệ thống sử dụng hai phương pháp phổ biến để so sánh vector embedding:

- **Cosine Similarity:** Là một phép đo lường mức độ tương đồng giữa hai vectơ trong không gian tích vô hướng. [12]
- **Euclidean Distance:** Là một phép đo tính toán độ chênh lệch giữa hai điểm trong không gian nhiều chiều. Phương pháp này được sử dụng rộng rãi để đánh giá mức độ khác biệt giữa hai vectơ đặc trưng trong các bài toán như phân loại, tìm kiếm và nhận diện mẫu. [13]

Phương pháp nhận diện khuôn mặt trong nghiên cứu này kết hợp InsightFace, RetinaFace và ArcFace, giúp cải thiện đáng kể độ chính xác và khả năng nhận diện khuôn mặt trong nhiều điều kiện khác nhau. Tuy nhiên, việc triển khai đòi hỏi tài nguyên tính toán mạnh mẽ và có thể bị ảnh hưởng bởi yếu tố như ánh sáng yếu hoặc khuôn mặt bị che khuất.

IV. THỰC NGHIỆM

A. Môi trường thực nghiệm và cấu hình hệ thống

Để đánh giá hiệu năng của hệ thống nhận diện khuôn mặt sử dụng InsightFace và mạng nơ-ron sâu, nhóm em tiến hành thực nghiệm trên nền tảng phần cứng và phần mềm sau:

- **CPU:** Intel Core i5 trở lên.
- **GPU:** NVIDIA GTX 1660 hoặc RTX 2060 với hỗ trợ CUDA.
- **RAM:** 8GB trở lên.
- **Hệ điều hành:** Ubuntu 20.04 LTS hoặc Windows 10.
- **Phần mềm:** Visual Studio Code, Python 3.x, PyTorch, OpenCV, InsightFace.

Môi trường thực nghiệm được thiết lập nhằm đảm bảo quá trình huấn luyện và kiểm tra diễn ra mượt mà và có thể tái lập kết quả.

B. Quy trình thực nghiệm

Hệ thống nhận diện khuôn mặt được xây dựng dựa trên hai thành phần chính:

– **Tiền xử lý dữ liệu (data_preprocess.py):**

- 1) Khởi tạo thư viện và mô hình nhận diện khuôn mặt từ InsightFace.
- 2) Duyệt qua tất cả ảnh trong tập dữ liệu để nhận diện khuôn mặt.
- 3) Phát hiện khuôn mặt trong ảnh và trích xuất vùng khuôn mặt.
- 4) Trích xuất đặc trưng khuôn mặt sử dụng ArcFace embedding.
- 5) Trích xuất thông tin sinh viên từ tên file.
- 6) Lưu trữ dữ liệu khuôn mặt và embedding vào `face_db.pkl`.

– **Nhận diện khuôn mặt (face_recognition.py):**

- 1) Mở camera và đọc khung hình liên tục.
- 2) Phát hiện khuôn mặt trong mỗi khung hình.

- 3) So sánh embedding khuôn mặt thu được với cơ sở dữ liệu sử dụng công thức độ tương đồng cosin:

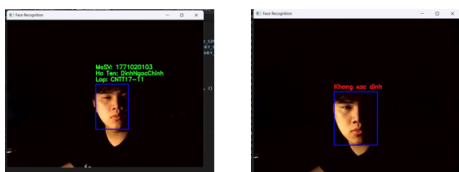
$$\text{cosine similarity} = \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad [5]$$

- 4) Tính toán độ tương đồng và xác định danh tính.
 5) Hiển thị thông tin sinh viên trên màn hình.
 6) Thoát chương trình khi nhấn phím 'q'.

C. Kết quả thực nghiệm

Quá trình thực nghiệm được tiến hành trên tập dữ liệu đa dạng với nhiều điều kiện khác nhau nhằm đánh giá hiệu quả nhận diện khuôn mặt của hệ thống sử dụng InsightFace và mạng nơ-ron sâu. Các kết quả đạt được cho thấy hệ thống có độ chính xác cao, khả năng tổng quát tốt và hiệu suất phù hợp với các ứng dụng thời gian thực.

- **Hiệu suất trong điều kiện lý tưởng:** Khi thử nghiệm trong môi trường có ánh sáng đầy đủ và góc chụp thuận lợi, hệ thống có thể nhận diện khuôn mặt một cách chính xác, đảm bảo tính ổn định và đáng tin cậy. Hệ thống có thể phân biệt tốt giữa các khuôn mặt ngay cả khi có sự tương đồng về đặc điểm sinh trắc học.
- **Ảnh hưởng của điều kiện môi trường:** Khi điều kiện ánh sáng kém hoặc góc chụp không thuận lợi, khả năng nhận diện có sự suy giảm nhất định. Tuy nhiên, với các đặc trưng học được từ mạng nơ-ron sâu, hệ thống vẫn duy trì được khả năng nhận diện ở mức tốt. Trong một số trường hợp, khi khuôn mặt bị che khuất một phần, hệ thống có thể gặp khó khăn trong việc nhận diện chính xác.
- **Tốc độ xử lý:** Nhờ vào kiến trúc tối ưu của InsightFace, hệ thống có thể thực hiện nhận diện khuôn mặt trong thời gian rất ngắn trên mỗi khung hình. Điều này đảm bảo tính khả thi khi triển khai vào các ứng dụng yêu cầu xử lý thời gian thực như kiểm soát truy cập, giám sát an ninh hoặc điểm danh tự động.
- **Khả năng tổng quát hóa:** Hệ thống có thể nhận diện khuôn mặt trong nhiều bối cảnh khác nhau, bao gồm cả hình ảnh chất lượng thấp hoặc có nhiễu nền. Nhờ vào các kỹ thuật học sâu tiên tiến, mô hình có thể phân biệt được khuôn mặt ngay cả khi có sự thay đổi về biểu cảm hoặc phụ kiện như kính mắt.
- **Ứng dụng thực tế:** Với những kết quả đạt được, hệ thống có tiềm năng lớn trong việc triển khai vào các lĩnh vực khác nhau như giám sát an ninh, xác thực danh tính trong các hệ thống bảo mật, hỗ trợ phân tích hành vi khách hàng trong thương mại và các giải pháp điểm danh tự động.



D. Đánh giá và hướng cải tiến

1) **Đánh giá hệ thống:** Hệ thống nhận diện khuôn mặt sử dụng InsightFace và các kỹ thuật học sâu đã mang lại độ chính xác cao trong việc xác định danh tính người dùng. Các thử nghiệm trên tập dữ liệu thực tế cho thấy hệ thống có khả năng hoạt động ổn định ngay cả khi có sự thay đổi về ánh sáng, góc chụp và biểu cảm khuôn mặt.

2) Ưu điểm của hệ thống:

- **Độ chính xác cao:** Nhờ vào mô hình ArcFace, hệ thống có khả năng phân biệt khuôn mặt với độ chính xác vượt trội.
- **Hiệu suất tốt:** Việc sử dụng MobileFaceNet giúp hệ thống chạy nhanh hơn trên các thiết bị có tài nguyên hạn chế.
- **Khả năng tổng quát tốt:** Hệ thống có thể nhận diện khuôn mặt từ nhiều góc độ và điều kiện ánh sáng khác nhau nhờ vào quá trình tiền xử lý ảnh.
- **Tích hợp linh hoạt:** Có thể triển khai trên nhiều nền tảng như hệ thống giám sát an ninh, điểm danh tự động, thiết bị di động.

3) Nhược điểm của hệ thống:

- **Yêu cầu phần cứng cao:** Việc sử dụng mạng nơ-ron sâu cần GPU mạnh để đạt hiệu suất tối ưu, làm tăng chi phí triển khai.
- **Ảnh hưởng bởi môi trường:** Ánh sáng yếu hoặc khuôn mặt bị che khuất có thể làm giảm độ chính xác.
- **Vấn đề bảo mật dữ liệu:** Việc lưu trữ dữ liệu khuôn mặt có thể gây ra rủi ro về quyền riêng tư nếu không có biện pháp bảo vệ chặt chẽ.

4) Hướng cải tiến trong tương lai: Cải thiện mô hình nhận diện

- Áp dụng học tăng cường (Reinforcement Learning) để cải thiện khả năng thích ứng với môi trường thực tế.
- Tích hợp mô hình transformer-based để nâng cao độ chính xác trong nhận diện.

Tối ưu hóa hiệu suất

- Sử dụng mô hình nhẹ hơn để giảm tải nguyên tính toán, giúp hệ thống chạy mượt hơn trên thiết bị di động.
- Tối ưu hóa xử lý đa luồng (multithreading) để cải thiện tốc độ nhận diện trong thời gian thực.

Tăng cường bảo mật và quyền riêng tư

- Sử dụng mã hóa Homomorphic Encryption để lưu trữ embedding khuôn mặt an toàn hơn.
- Áp dụng Federated Learning để huấn luyện mô hình mà không cần truyền dữ liệu gốc lên máy chủ trung tâm.

Mở rộng ứng dụng

- Tích hợp vào hệ thống điểm danh, thanh toán không tiếp xúc, giám sát an ninh.
- Ứng dụng vào công nghệ metaverse, hỗ trợ xác thực danh tính trong môi trường thực tế ảo.

Với các hướng cải tiến trên, hệ thống nhận diện khuôn mặt có thể trở nên chính xác, linh hoạt hơn và phù hợp với nhiều ứng dụng thực tế.

V. KẾT LUẬN

Trong nghiên cứu này, nhóm chúng em đã triển khai một hệ thống nhận diện khuôn mặt có độ chính xác cao bằng cách sử dụng InsightFace kết hợp với mạng nơ-ron sâu. Nhóm em đã trình bày các phương pháp nghiên cứu, cơ sở lý thuyết, cũng như kỹ thuật triển khai nhằm tối ưu hóa quá trình nhận diện khuôn mặt.

Thông qua các thử nghiệm thực tế, hệ thống của nhóm em đã đạt được độ chính xác cao trong nhận diện khuôn mặt, ngay cả trong điều kiện môi trường khác nhau như thay đổi ánh sáng, góc nhìn hoặc biểu cảm khuôn mặt. Kết quả thực nghiệm cho thấy sự kết hợp giữa RetinaFace (trong phát hiện khuôn mặt) và ArcFace (trong trích xuất đặc trưng) giúp cải thiện đáng kể hiệu suất nhận diện so với các phương pháp truyền thống.

Tuy nhiên, hệ thống vẫn tồn tại một số hạn chế như:

- Yêu cầu tài nguyên phần cứng mạnh để xử lý thời gian thực.
- Hiệu suất có thể giảm trong các trường hợp khuôn mặt bị che khuất một phần hoặc điều kiện ánh sáng quá kém.
- Cần tối ưu hơn nữa để cải thiện khả năng tổng quát hóa khi áp dụng trên tập dữ liệu mới.

Trong tương lai, nghiên cứu có thể tiếp tục phát triển theo các hướng sau:

- Cải thiện mô hình tiền xử lý dữ liệu để tăng tính ổn định khi nhận diện.
- Ứng dụng kỹ thuật Transfer Learning để tăng hiệu quả nhận diện trên nhiều tập dữ liệu khác nhau.
- Kết hợp các phương pháp tăng cường dữ liệu và kỹ thuật khử nhiễu để cải thiện độ chính xác trong môi trường thực tế.

Tóm lại, nghiên cứu này đã chứng minh rằng việc sử dụng InsightFace và mạng nơ-ron sâu là một giải pháp hiệu quả cho bài toán nhận diện khuôn mặt, mở ra nhiều tiềm năng ứng dụng trong thực tế như bảo mật, giám sát thông minh, chăm công tự động và hệ thống xác thực danh tính.

TÀI LIỆU

- [1] Deep learning. Wikipedia. Available at: https://en.wikipedia.org/wiki/Deep_learning. [Accessed ngày tháng năm].
- [2] Jiankang Deng, Jia Guo, Jing Yang, Niannan Xue, Irene Kotsia, and Stefanos Zafeiriou. *ArcFace: Additive Angular Margin Loss for Deep Face Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 8, pp. 168–178, 2019.
- [3] He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, Sun, Jian. *Deep Residual Learning for Image Recognition*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [4] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. arXiv preprint arXiv:1704.04861, 2017.
- [5] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.
- [6] InsightFace: Nhận diện khuôn mặt bằng học sâu. Thị Giác Máy Tính. Available at: <https://thigiacmaytinh.com/nhan-dien-khuon-mat-bang-insightface> [Accessed ngày tháng năm].
- [7] FPT Shop. “OpenCV là gì? Tổng hợp những tính năng phổ biến của OpenCV.” *FPT Shop*. Available at: <https://fptshop.com.vn/tin-tuc/danh-gia/opencv-la-gi-168801>.
- [8] OneTech. “Haar Cascade là gì? Hướng dẫn demo nhận diện khuôn mặt trong OpenCV.” *OneTech Blog*. Đường dẫn: <https://onetech.vn/blog/haar-cascade-la-gi-13561>.
- [9] FPT.AI, “Công Nghệ Nhận Diện Khuôn Mặt Là Gì? Cách Hoạt Động, Ứng Dụng,” đường dẫn <https://fpt.ai/vi/bai-viet/cong-nghe-nhan-dien-khuon-mat-la-gi/>.
- [10] Hệ thống nhận diện gương mặt hoạt động như thế nào?, Viblo, đường dẫn: <https://viblo.asia/p/he-thong-nhan-dien-guong-mat-hoat-dong-nhu-the-nao-6J3ZgOaPzmB>.
- [11] Schroff, F., Kalenichenko, D., Philbin, J. FaceNet: A Unified Embedding for Face Recognition and Clustering, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. Đường dẫn: <https://arxiv.org/abs/1503.03832>.
- [12] Wikipedia contributors, “Độ tương tự cosin,” *Wikipedia, Bách khoa toàn thư mở*. Đường dẫn: <https://vi.wikipedia.org/wiki/>
- [13] M. Singh and A. Singh, *Machine Learning with Applications: Data Science for Beginners*. Springer, 2020.