BỘ GIÁO DỤC VÀ ĐÀO TẠO



**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HỒ CHÍ MINH**

**ĐỒ ÁN CƠ SỞ**

**TÌM HIỂU TỔNG QUAN VỀ CÁC KỸ THUẬT NHẬN DẠNG KHUÔN MẶT VÀ ỨNG DỤNG THỰC NGHIỆM**

Ngành : **CÔNG NGHỆ THÔNG TIN**

Chuyên ngành: **KHOA HỌC DỮ LIỆU**

Giảng viên hướng dẫn : ThS. Lê Nhật Tùng

Sinh viên thực hiện :

+ Nguyễn Đức Trường MSSV: 2286400868 Lớp: 22DKHA1

+ Trần Tuấn Đạt MSSV: 2286400007 Lớp: 22DKHA1

+ Lê Trương Duy Khôi MSSV: 2286400012 Lớp: 22DKHA1

TP. Hồ Chí Minh, 2025

BỘ GIÁO DỤC VÀ ĐÀO TẠO



**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HỒ CHÍ MINH**

**ĐỒ ÁN CƠ SỞ**

**TÌM HIỂU TỔNG QUAN VỀ CÁC KỸ THUẬT NHẬN DẠNG KHUÔN MẶT VÀ ỨNG DỤNG THỰC NGHIỆM**

Ngành : **CÔNG NGHỆ THÔNG TIN**

Chuyên ngành: **KHOA HỌC DỮ LIỆU**

Giảng viên hướng dẫn : ThS. Lê Nhật Tùng

Sinh viên thực hiện :

+ Nguyễn Đức Trường MSSV: 2286400868 Lớp: 22DKHA1

+ Trần Tuấn Đạt MSSV: 2286400007 Lớp: 22DKHA1

+ Lê Trương Duy Khôi MSSV: 2286400012 Lớp: 22DKHA1

TP. Hồ Chí Minh, 2025

**NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN**

…………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………….…………………………………..……………………………………………….………………………………………………………………...…………………….…………………………………………………………………………………….…………………………………………………………………………………….………………………………..…………………………………………………….……………………………………………………………………………………………………

|  |
| --- |
| TPHCM, Ngày … tháng … năm 2024 |
| **Giáo viên hướng dẫn** |
| (Ký tên, đóng dấu) |

## LỜI CAM ĐOAN

Chúng tôi, gồm ba thành viên là Trần Tuấn Đạt, Lê Trương Duy Khôi, Nguyễn Đức Trường.

Xin cam đoan rằng mọi thông tin được trình bày trong báo cáo này đều chính xác và đầy đủ nhất theo sự hiểu biết của chúng tôi. Toàn bộ nội dung của báo cáo đều được trình bày dựa trên quan điểm, kiến thức cá nhân và tích lũy của từng thành viên trong nhóm và được chọn lọc từ nhiều nguồn tài liệu có đính kèm chi tiết và hợp lệ. Chúng tôi cũng cam đoan rằng bài báo cáo này không sao chép từ bất kỳ nguồn nào mà không được chỉ rõ trong văn bản hoặc được đánh dấu trong phần tài liệu tham khảo. Chúng tôi hoàn toàn chịu trách nhiệm về nội dung của báo cáo và sẵn sàng bổ sung hoặc sửa lỗi nếu có yêu cầu, nhằm đảm bảo tính trung thực và trách nhiệm đối với báo cáo này.

Xin chân thành cảm ơn!

MỤC LỤC

[LỜI CAM ĐOAN ii](#_Toc196469182)

[DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT iv](#_Toc196469183)

[DANH MỤC CÁC BẢNG, HÌNH VẼ, ĐỒ THỊ v](#_Toc196469184)

[LỜI MỞ ĐẦU 1](#_Toc196469185)

[CHƯƠNG 1 TỔNG QUAN 2](#_Toc196469186)

[1.1 Giới thiệu đề tài 2](#_Toc196469187)

[1.2 Nhiệm vụ của đề tài 2](#_Toc196469188)

[1.2.1 Tính cấp thiết của đề tài 3](#_Toc196469189)

[1.2.2 Ý nghĩa khoa học và thực tiễn của đề tài 4](#_Toc196469190)

[1.3 Mục tiêu nghiên cứu 5](#_Toc196469191)

[1.3.1 Mục tiêu tổng quát 5](#_Toc196469192)

[1.3.2 Mục tiêu cụ thể 5](#_Toc196469193)

[1.4 Đối tượng và phạm vi nghiên cứu 5](#_Toc196469194)

[1.4.1 Đối tượng nghiên cứu 5](#_Toc196469195)

[1.4.2 Phạm vi nghiên cứu 5](#_Toc196469196)

[1.5 Phương pháp nghiên cứu 6](#_Toc196469197)

[1.5.1 Phương pháp nghiên cứu sơ bộ 6](#_Toc196469198)

[1.5.2 Phương pháp nghiên cứu tài liệu 6](#_Toc196469199)

[1.5.3 Phương pháp nghiên cứu thống kê 6](#_Toc196469200)

[1.5.4 Phương pháp thực nghiệm 7](#_Toc196469201)

[1.5.5 Phương pháp đánh giá 7](#_Toc196469202)

[1.6 Những đóng góp nghiên cứu của đề tài 7](#_Toc196469203)

[1.6.1 Đóng góp về mặt lý thuyết 7](#_Toc196469204)

[1.6.2 Đóng góp trong thực tiễn 8](#_Toc196469205)

[CHƯƠNG 2 CƠ SỞ LÝ THUYẾT 9](#_Toc196469206)

[2.1 Giới thiệu tổng quan về thị giác máy tính (computer vision) 9](#_Toc196469207)

[2.1.1 Định nghĩa và vai trò của nhận dạng khuôn mặt 9](#_Toc196469208)

[2.1.2 Ưu điểm và hạn chế 9](#_Toc196469209)

[2.2. Deep Learning 10](#_Toc196469210)

[2.3. OpenCV (Open Source Computer Vision Library) 11](#_Toc196469211)

[2.4 Các phương pháp, kỹ thuật tiền xử lý ảnh. 12](#_Toc196469212)

[**2.4.1 Chuyển sang ảnh xám (Grayscale Conversion)** 12](#_Toc196469213)

[**2.4.2** **Cân bằng Histogram (Histogram Equalization)** 13](#_Toc196469214)

[**2.4.3** **Làm mịn ảnh (Image Smoothing-blurring)** 14](#_Toc196469215)

[**2.4.4** **Biến đổi hình học (Affine / Perspective Transformations)** 15](#_Toc196469216)

[2.4.4.1.1Biến đổi hình học (Affine Transformation) 15](#_Toc196469217)

[2.4.4.2 Biến đổi phối cảnh (Perspective Transformation) 15](#_Toc196469218)

[**2.4.5 Data Augmentation (Phóng đại dữ liệu)** 16](#_Toc196469219)

[**2.4.6 GAN-based Enhancement (Generative Adversarial Network)** 17](#_Toc196469220)

[2.5 Phát hiện đặc trưng và mô tả (Feature Detection & Description) 18](#_Toc196469221)

[**2.5.1 Edge Detection**: 18](#_Toc196469222)

[**2.5.2 Blob Detection:** 19](#_Toc196469223)

[**2.5.4 Learned Features** 19](#_Toc196469224)

[2.5.4.1 ResNet 20](#_Toc196469225)

[2.5.4.2 EfficientNet: 20](#_Toc196469226)

[**2.5.5 Vision Transformers (ViT)** 21](#_Toc196469227)

[2.6 Phát hiện đối tượng (Object Detection) 22](#_Toc196469228)

[**2.6.1** **R-CNN** 22](#_Toc196469229)

[**2.6.2 YOLO (You Only Look Once)** 22](#_Toc196469230)

[2.7 Nhận dạng và phân loại đối tượng (Object Classification & Recognition) 23](#_Toc196469231)

[**2.7.1 HOG Kết Hợp SVM** 23](#_Toc196469232)

[*2.7.1.1* *HOG - Histogram of Oriented Gradients* 23](#_Toc196469233)

[2.7.1.2 SVM - Support Vector Machine 24](#_Toc196469234)

[**2.7.2 Template Matching** 25](#_Toc196469235)

[**2.7.3 Transformer-based: Vision Transformer (ViT), Swin Transformer** 25](#_Toc196469236)

[2.7.3.1Swin Transformer 25](#_Toc196469237)

## DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

## DANH MỤC CÁC HÌNH ẢNH

## DANH MỤC CÁC BẢNG, HÌNH VẼ, ĐỒ THỊ

## LỜI MỞ ĐẦU

Trong cuộc sống hiện đại, công nghệ ngày càng phát triển và được ứng dụng rộng rãi vào nhiều lĩnh vực khác nhau và nó đóng vai trò quan trọng trong hầu hết các lĩnh vực của đời sống, từ giáo dục, y tế cho đến giao thông và sản xuất. Bên cạnh những công nghệ quen thuộc như trí tuệ nhân tạo hay học máy Một trong những lĩnh vực đang được quan tâm nhiều hiện nay là thị giác máy tính, nơi mà máy móc có thể nhìn và hiểu hình ảnh giống như con người.

Và với mong muốn tìm hiểu sâu hơn về lĩnh vực này, nhóm chúng em đã chọn đề tài **“**Tìm Hiểu Tổng Quan Về Các Kỹ Thuật Nhận Dạng Khuôn Mặt Và Ứng Dụng Thực Tiễn**”**. Qua đề tài, nhóm không chỉ muốn nắm rõ cách các kỹ thuật nhận dạng hoạt động, mà còn hiểu được quá trình phát triển và ứng dụng của chúng trong thực tế.

# CHƯƠNG 1 TỔNG QUAN

## **Giới thiệu đề tài**

Trong thế giới số hóa hiện nay, thị trường trí tuệ nhân tạo toàn cầu đang có nhiều bước tiến mạnh mẽ và với tốc độ đổi mới chưa từng có. Đặc biệt là lĩnh vực nhận dạng khuôn mặt, nó được áp dụng rộng rãi và đóng vai trò trong nhiều ứng dụng thực tiễn. Nhu cầu sử dụng các hệ thống nhận dạng khuôn mặt đang không ngừng tăng cao, khi mà việc xác thực nhanh, chính xác và không tiếp xúc trở nên quan trọng hơn bao giờ hết. Xuất phát từ nhu cầu thực tiễn và tiềm năng ứng dụng to lớn, thì nhóm đã nghiên cứu và lựa chọn đề tài **“**Tìm Hiểu Tổng Quan Về Các Kỹ Thuật Nhận Dạng Khuôn Mặt Và Ứng Dụng Thực Tiễn**”** nhằm tổng hợp, phân tích tất cả các kỹ thuật trong lĩnh vực nhận dạng khuôn mặt và triển khai thử nghiệm trên một số bài toán thực tế.

## **Nhiệm vụ của đề tài**

Nhiệm vụ của đề tài " Tìm Hiểu Tổng Quan Về Các Kỹ Thuật Nhận Dạng Khuôn Mặt Và Ứng Dụng Thực Tiễn " là tiến hành khảo sát, phân loại và phân tích các phương pháp nhận dạng khuôn mặt trong lĩnh vực thị giác máy tính. Đề tài tập trung làm rõ nguyên lý hoạt động, quy trình thực hiện, cũng như ưu nhược điểm của từng kỹ thuật. Từ đó, phân tích khả năng áp dụng của từng loại kỹ thuật vào từng bài toán, từng giai đoạn khác nhau của một dự án. Điều này giúp các nhà phát triển hiểu rõ và hoạch định trước quy trình phát triển của dự án. Và với các nhà đầu tư và doanh nghiệp, các dự án nhận dạng khuôn mặt có thể giúp họ dễ dàng quản lý nhân sự và học viên, nâng cao mức độ an toàn trong việc kiểm soát truy cập, xác minh danh tính trong giao dịch.

* + 1. Tính cấp thiết của đề tài

Trong bối cảnh hiện nay các nhà phát triển có thể học hỏi và khai thác cũng như mở rộng công nghệ một cách dễ tiếp cận hơn lúc trước, cùng với sự ra đời và cải tiến của nhiều công trình nghiên cứu mang tính lịch sử và có nhiều ông lớn trong lĩnh vực công nghệ tham gia sẽ tạo ra một môi trường cạnh tranh khốc liệt. Điều này đặt ra nhu cầu cấp bách cho các nhà nghiên cứu và doanh nghiệp công nghệ phải liên tục cập nhật, cải tiến và tối ưu hóa các thuật toán nhận dạng khuôn mặt nhằm nâng cao độ chính xác, tốc độ xử lý và khả năng ứng dụng thực tế.

Việc nghiên cứu tổng quan các kỹ thuật nhận dạng khuôn mặt là một bước quan trọng trong việc nâng cao hiệu quả của các hệ thống xác thực sinh trắc học. Việc này giúp các tổ chức và doanh nghiệp hiểu rõ đặc điểm, ưu, nhược điểm của từng phương pháp, từ đó lựa chọn giải pháp phù hợp với nhu cầu thực tế, qua đó giúp định hướng phát triển lâu dài với định hướng nhận dạng chính xác, thông minh và an toàn. Cụ thể hơn, việc tìm hiểu bao quát về các kỹ thuật nhận dạng có thể giúp các nhà phát triển và doanh nghiệp:

Phù hợp với bài toán thực tế: Mỗi kỹ thuật nhận dạng có ưu nhược điểm riêng. Việc hiểu rõ các phương pháp giúp lựa chọn công nghệ phù hợp với yêu cầu về độ chính xác, tốc độ và môi trường triển khai.

Tiết kiệm chi phí và tài nguyên: Việc lựa chọn đúng kỹ thuật giúp doanh nghiệp tối ưu hóa chi phí triển khai. Các mô hình hiệu quả sẽ giảm thiểu tài nguyên tính toán, đồng thời giảm thời gian xử lý, giúp doanh nghiệp tiết kiệm chi phí vận hành trong dài hạn.

Tăng khả năng mở rộng và linh hoạt trong ứng dụng: Các kỹ thuật nhận dạng khuôn mặt tiên tiến, khi được tích hợp vào các hệ thống, có thể mở rộng và áp dụng linh hoạt trong nhiều tình huống khác nhau.

Từ những lý do trên, có thể thấy rằng việc tìm hiểu tổng quan kỹ thuật nhận dạng không chỉ giúp doanh nghiệp nâng cao hiệu quả vận hành và bảo mật, mà còn thúc đẩy triển khai các giải pháp số thông minh, đáp ứng tốt hơn nhu cầu thị trường hiện đại.

* + 1. Ý nghĩa khoa học và thực tiễn của đề tài

Ý nghĩa khoa học: Đề tài đóng góp vào lĩnh vực thị giác máy tính và trí tuệ nhân tạo thông qua việc tổng hợp, phân tích và so sánh các kỹ thuật nhận dạng khuôn mặt hiện đại như Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG) và các mô hình deep learning tiên tiến khác. Việc khảo sát này không chỉ củng cố kiến thức nền tảng mà còn làm sáng tỏ phạm vi áp dụng của từng kỹ thuật trong các bài toán cụ thể như xác thực danh tính, điểm danh tự động, hay kiểm soát truy cập. Từ đó, cung cấp cơ sở khoa học vững chắc để các nhà phát triển và chuyên gia lựa chọn kỹ thuật phù hợp với mục tiêu và điều kiện triển khai thực tế.

Ý nghĩa thực tiễn: Đề tài hỗ trợ các doanh nghiệp trong việc lựa chọn và ứng dụng các kỹ thuật nhận dạng khuôn mặt phù hợp với nhu cầu thực tiễn. Việc áp dụng phù hợp không chỉ giúp nâng cao độ chính xác và tốc độ xử lý mà còn giảm thiểu chi phí vận hành, đồng thời tăng cường trải nghiệm người dùng và khả năng cạnh tranh trên thị trường. Ngoài ra, đề tài còn hỗ trợ các nhà phát triển và chuyên gia kỹ thuật trong việc nắm bắt xu hướng công nghệ mới, rút ngắn thời gian triển khai giải pháp thực tế, đồng thời cung cấp nền tảng dữ liệu và kiến thức có thể tái sử dụng hoặc mở rộng trong các hệ thống lớn hơn.

## **1.3 Mục tiêu nghiên cứu**

### 1.3.1 Mục tiêu tổng quát

Đề tài hướng đến việc nghiên cứu, tổng hợp và phân tích các kỹ thuật hiện đại trong lĩnh vực nhận dạng khuôn mặt, từ đó đánh giá khả năng ứng dụng của từng kỹ thuật trong các tình huống thực tế. Mục tiêu chính là đề xuất hướng triển khai phù hợp cho các doanh nghiệp có nhu cầu ứng dụng nhận diện khuôn mặt nhằm tối ưu hóa hoạt động vận hành, nâng cao hiệu quả quản lý và tăng cường trải nghiệm người dùng.

### 1.3.2 Mục tiêu cụ thể

Với bài nghiên cứu này, bước đầu tiên sẽ là thu thập và xử lý một tập dữ liệu hình ảnh khuôn mặt có gắn nhãn. Sau đó, nghiên cứu, so sánh và triển khai các kỹ thuật nhận dạng khuôn mặt như LBP, HOG và các phương pháp dựa trên kiến trúc CNN kết hợp để tìm ra phương pháp hiệu quả nhất. Quá trình nghiên cứu sẽ bao gồm các bước như phát hiện khuôn mặt, trích xuất đặc trưng, ánh xạ vector và so sánh độ tương đồng giữa các khuôn mặt. Mô hình sẽ được đánh giá dựa trên các tiêu chí như độ chính xác, tốc độ xử lý và khả năng hoạt động ổn định. Cuối cùng, đề tài sẽ đề xuất các giải pháp cụ thể và nền tảng cho các tổ chức và doanh nghiệp trong việc triển khai công nghệ nhận dạng khuôn mặt nhằm tối ưu hóa quy trình vận hành và nâng cao trải nghiệm người dùng.

**1.4 Đối tượng và phạm vi nghiên cứu**

### 1.4.1 Đối tượng nghiên cứu

Đối tượng nghiên cứu của đề tài là các kỹ thuật và phương pháp nhận dạng khuôn mặt, được sử dụng trong hệ thống xác thực người dùng. Qua các bước thu thập và xử lý dữ liệu ảnh tiếp đó thực hiện áp dụng tất cả các phương pháp trên một tập dữ liệu cụ thể. Qua đó đưa ra phương pháp tối ưu và đề xuất cải thiện và nâng cao chất lượng dự án.

1.4.2 Phạm vi nghiên cứu

Đề tài tập trung vào thu thập, xử lý dữ liệu ảnh khuôn mặt của người dùng từ tất cả nguồn dữ liệu có sẵn và đảm bảo tuân thủ các quy định về quyền riêng tư và bảo mật dữ liệu người dùng. Chúng tôi sẽ áp dụng các phương pháp phân tích dữ liệu và nhận dạng khuôn mặt để phân tích và so sánh hiệu quả của từng phương pháp, kĩ thuật cụ thể. Sau đó nghiên cứu đưa ra quy trình cải thiện và nâng cao chất lượng hệ thống, để đảm bảo rằng khi đưa vào thực tế thì dự án có thể mang đến độ chính xác, tốc độ và an toàn nhất. Điều này giúp mang lại giá trị lý thuyết và thực tiễn cho các doanh nghiệp trên thị trường trí tuệ nhân tạo

**1.5 Phương pháp nghiên cứu**

1.5.1 Phương pháp nghiên cứu sơ bộ

Trước khi tiến hành thu thập và xử lý dữ liệu, chúng tôi sẽ thực hiện một nghiên cứu sơ bộ để hiểu rõ hơn về nghiên cứu và các yếu tố quan trọng liên quan. Nghiên cứu này bao gồm việc tìm hiểu về các kỹ thuật nhận dạng khuôn mặt, ưu nhược điểm của từng phương pháp, và các yếu tố ảnh hưởng đến hiệu suất và độ chính xác của mô hình. Thông qua nghiên cứu sơ bộ, chúng tôi sẽ xác định các vấn đề cụ thể cần giải quyết và đề xuất các phương pháp nghiên cứu phù hợp.

1.5.2 Phương pháp nghiên cứu tài liệu

Chúng tôi sẽ tiến hành nghiên cứu tài liệu để thu thập thông tin về các phương pháp và công cụ nhận dạng khuôn mặt trong lĩnh vực thị giác máy tính và học máy. Qua việc đánh giá các nghiên cứu trước đây và các công trình khoa học liên quan, chúng tôi sẽ xác định các kỹ thuật nhận dạng khuôn mặt phù hợp nhất cho nghiên cứu của mình và áp dụng chúng vào việc triển khai các hệ thống xác thực người dùng.

1.5.3 Phương pháp nghiên cứu thống kê

Trong quá trình phân tích dữ liệu, chúng tôi sẽ sử dụng các phương pháp thống kê như phân tích phương sai hay kiểm tra độ tương quan hoặc sử dụng các phương pháp học máy (machine learning) để mô tả và phân tích các đặc trưng quan trọng trong nhận dạng khuôn mặt. Thông qua việc áp dụng các phương pháp thống kê này, chúng tôi sẽ đánh giá mối quan hệ giữa các đặc trưng và xác định các yếu tố ảnh hưởng đến độ chính xác và hiệu quả của hệ thống nhận dạng khuôn mặt.

1.5.4 Phương pháp thực nghiệm

Chúng tôi sẽ tiến hành thực nghiệm trên tập dữ liệu ảnh được thu thập từ các nguồn chính thống. Quá trình này bao gồm việc tiền xử lý dữ liệu hình ảnh, áp dụng các phương pháp nhận dạng khuôn mặt như LBP, HOG, hoặc các mô hình học sâu (CNN) để phân loại và xác minh khuôn mặt. Đồng thời, chúng tôi sẽ đánh giá hiệu quả của các kỹ thuật nhận dạng. Thông qua việc thực nghiệm này, chúng tôi sẽ kiểm tra và đảm bảo tính khả thi và hiệu quả của phương pháp nghiên cứu.

1.5.5 Phương pháp đánh giá

Cuối cùng, chúng tôi sẽ thực hiện phương pháp đánh giá để đo lường hiệu quả của các phương pháp nhận dạng khuôn mặt. Quá trình này bao gồm việc so sánh các chỉ số, độ chính xác, tốc độ xử lý, cũng như khả năng nhận dạng trong các điều kiện thực tế. Chúng tôi sẽ tiến hành đánh giá hiệu quả của từng kỹ thuật nhận dạng khuôn mặt trên từng bộ dữ liệu thực nghiệm, từ đó rút ra những nhận định về hiệu suất và khả năng áp dụng của mỗi phương pháp.

**1.6 Những đóng góp nghiên cứu của đề tài**

1.6.1 Đóng góp về mặt lý thuyết

Đề tài đóng góp vào việc áp dụng và phát triển các kỹ thuật nhận dạng khuôn mặt trong các hệ thống xác thực người dùng. Việc nghiên cứu và áp dụng các mô hình và thuật toán nhận dạng khuôn mặt, như các phương pháp học sâu (Deep Learning) và các mô hình phân loại, vào bài toán nhận diện các đặc trưng cùa khuôn mặc giúp mở rộng ứng dụng trong môi trường thực tế. Đề tài cũng cung cấp các giải pháp tối ưu hóa và cải thiện độ chính xác, từ đó tạo ra nhiều bước tiến vượt bậc cho các nghiên cứu sau này trong lĩnh vực nhận dạng khuôn mặt và bảo mật.

1.6.2 Đóng góp trong thực tiễn

Đề tài cung cấp các giải pháp và chiến lược cụ thể cho các doanh nghiệp trong việc áp dụng hệ thống xác thực người dùng bằng nhận dạng khuôn mặt. Việc cải thiện độ chính xác và hiệu quả của hệ thống xác thực giúp doanh nghiệp nâng cao trải nghiệm người dùng và tối ưu hóa quy trình xác thực. Đồng thời, các kết quả nghiên cứu có thể giúp các doanh nghiệp xây dựng các hệ thống bảo mật mạnh mẽ hơn, giảm thiểu rủi ro gian lận, và tăng cường sự tin cậy từ phía khách hàng. Điều này có thể dẫn đến việc nâng cao chất lượng dịch vụ và tăng trưởng trong môi trường thương mại điện tử ngày càng phát triển.

# **CHƯƠNG 2 CƠ SỞ LÝ THUYẾT**

## **2.1** **Tổng quan về nhận dạng khuôn mặt**

### 2.1.1 Giới thiệu về nhận dạng khuôn mặt (Face Recogniton)

Nhận diện khuôn mặt thực chất là quá trình xác định danh tính dựa vào đặc điểm sinh trắc học của khuôn mặt thông qua hình ảnh hoặc video. Công nghệ này kết hợp các lĩnh vực như thị giác máy tính và học sâu để phân tích và đánh giá đặc trưng khuôn mặt.

* Phát hiện khuôn mặt: Xác định vị trí có khuôn mặt của người trong ảnh hay video.
* Căn chỉnh khuôn mặt: Đưa khuôn mặt về chính giữa, xác định các điểm trên khuôn mặt như: mắt, mũi, miệng, sau đó xoay hay cắt ảnh sao cho các điểm trở nên cân đối.
* Trích xuất các đặc trưng: Hình ảnh sau căn chỉnh được chuyển thành chuỗi vector số học chứa các đặc trưng nhận diện, nói đơn giản là chuyển thông tin khuôn mặt thành dữ liệu máy tính có thể xử lý.
* So sánh với các dữ liệu đã lưu: Vector đặc trưng này sẽ được đối chiếu với các vector đã lưu trong hệ thống, thuật toán tự động nhận diện sự tương đồng để xác định danh tính nếu có khớp.
* Kết quả output: Với trường hợp xác thực thành công, hệ thống sẽ trả về định danh người dùng tương ứng. Nếu không có sự trùng khớp, hệ thống sẽ từ chối truy cập.

### 2.1.2 Ưu điểm và hạn chế

Ưu điểm:

* Tự động hóa cao: Giúp con người hay các doanh nghiệp có thể giảm bớt đi khối lượng công việc, thay thế trong các việc lặp đi lặp lại.
* Tốc độ xử lý nhanh: Có thể phân tích các video và hình ảnh trong thời gian ngắn để đưa ra các kết quả chính xác.
* Độ chính xác cao: Khi được huấn luyện tốt thì mô hình có thể đưa ra được kết quả chính xác cao, ổn định.
* Ứng dụng rộng rãi: Có thể áp dụng trong nhiều lĩnh vực như: nhận diện, y tế, an ninh, công nghiệp, ô tô tự lái, nông nghiệp, …

Hạn chế:

* Phụ thuộc dữ liệu: Cần lượng dữ liệu lớn và chất lượng để huấn luyện cho mô hình.
* Yêu cầu phần cứng mạnh: Các mô hình Deep Learning cần GPU để hoạt động hiệu quả.
* Ứng dụng nhận diện khuôn mặt có thể ảnh đến quyền riêng tư cá nhân và cũng có thể dùng trong việc lừa đảo.

## **2.2 Thư viện OpenCV (Open Source Computer Vision Library)**

OpenCV là một thư viện mã nguồn mở cho thị giác máy tính, học máy và xử lý hình ảnh. Đóng một vai trò quan trọng trong hoạt động thời gian thực. Bằng cách đó, người ta có thể xử lý hình ảnh và video để xác định các đối tượng, khuôn mặt hoặc thậm chí là nhận diện.

Khi nó được tích hợp với các thư viện khác nhau, chẳng hạn như NumPy, Python có khả năng xử lý cấu trúc mảng opencv để phân tích. Để xác định một mẫu hình ảnh và các tính năng khác nhau của nó, sử dụng không gian vector và thực hiện các phép toán trên các tính năng này.

OpenCV cho phép thực hiện các thao tác khác nhau trong hình ảnh như:

* Đọc hình ảnh : OpenCV giúp đọc hình ảnh từ tệp hoặc trực tiếp từ máy ảnh để có thể truy cập và thực hiện các bước xử lý thêm.
* Tăng cường hình ảnh: Có thể nâng cao hình ảnh bằng cách điều chỉnh độ sáng, độ sắc nét hoặc độ co lại của hình ảnh..
* Phát hiện đối tượng: Đối tượng cũng có thể được phát hiện bằng cách sử dụng OpenCV,. Điều này cũng có thể nhận dạng khuôn mặt, hình dạng hoặc thậm chí là đối tượng.
* Lọc hình ảnh: Thay đổi hình ảnh bằng cách áp dụng các bộ lọc khác nhau như làm mờ hoặc làm sắc nét.
* Nhận dạng đối tượng
* Học máy và phân cụm (ml, flann)
* Tăng tốc CUDA (gpu)
  1. **Các phương pháp, kỹ thuật tiền xử lý ảnh.**
     1. Làm mịn ảnh (Image Smoothing-blurring)

Trong quá trình tiền xử lý và đầu vào nhận diện, ảnh thường chứa nhiều nhiễu hoặc chi tiết nhỏ không cần thiết gây khó khăn cho việc trích xuất đặc trưng. Để giải quyết vấn đề này, kỹ thuật làm mịn ảnh hay còn gọi là làm mờ được áp dụng giúp giảm mức độ biến thiên giữa các pixel lân cận, giúp ảnh trở nên mượt hơn và loại bỏ nhiễu cục bộ.

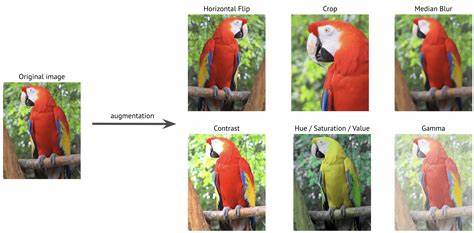
* Làm mờ trung bình: Mỗi pixel được thay bằng trung bình cộng của các pixel xung quanh, sử dụng khi nhiễu không đồng nhất và ít tập trung.
* Làm mờ Gaussian: Sử dụng phân phối chuẩn để tính trung bình có trọng số, sử dụng khi cần làm mờ mịn và giảm nhiễu hiệu quả mà vẫn giữ biên rõ ràng
* Làm mờ trung vị: Mỗi pixel được thay bằng giá trị trung vị của vùng xung quanh, sử dụng khi ảnh bị nhiễu muối tiêu (một số pixel trong ảnh bị thay đổi đột ngột thành màu đen hoàn toàn (0) hoặc trắng hoàn toàn (255))

### 2.3.2 Phóng đại dữ liệu (Data Augmentation)

Trong nhận dạng khuôn mặt, dữ liệu đầu vào thường bị giới hạn về số lượng hoặc thiếu tính đa dạng. Khiến mô hình dễ bị overfitting. Kỹ thuật Data được sử dụng nhằm khắc phục vấn đề này bằng cách tạo ra nhiều phiên bản biến đổi khác nhau của ảnh gốc nhưng vẫn giữ nguyên nhãn hoặc nội dung chính, từ đó giúp mô hình học được tốt hơn và tổng quát hơn.

Các kỹ thuật phóng đại dữ liệu phổ biến gồm:

* Lật ảnh : Lật ngang thường được sử dụng trong nhận diện khuôn mặt vì khuôn mặt thường đối xứng.
* Xoay ảnh: Xoay ảnh với một góc nhất định như ±15° để mô phỏng các tư thế đầu khác nhau.
* Phóng to/thu nhỏ: Làm thay đổi tỷ lệ kích thước đối tượng trong ảnh.
* Dịch chuyển :Di chuyển ảnh theo trục X hoặc Y.
* Thay đổi độ sáng, độ tương phản: Mô phỏng điều kiện ánh sáng khác nhau.
* Thêm nhiễu : Làm cho dữ liệu giống với môi trường thực hơn. Giúp mô hình thích ứng tốt với ảnh thực tế bị nhiễu vd: camera chất lượng thấp.
* Cắt ảnh ngẫu nhiên :cắt các vùng nhỏ ngẫu nhiên từ ảnh gốc để tạo thêm các biến thể.



<https://th.bing.com/th/id/OIP.Za3VLUEHu7JhiLRWO3Lv_AHaDp?cb=iwc2&rs=1&pid=ImgDetMain>

## **Phát hiện đặc trưng và mô tả (Feature Detection & Description)**

Phát hiện đặc trưng là bước quan trọng nhằm tìm ra các điểm đặc trưng như khóe mắt, sống mũi, miệng, cằm... giúp mô hình có thể phân biệt giữa các khuôn mặt khác nhau. Từ đó phục vụ cho các tác vụ so khớp, xác thực hoặc phân loại khuôn mặt.

### 2.4.1 Phát hiện khuôn mặt bằng MTCNN

MTCNN (Multi-task Cascaded Convolutional Networks) là một phương pháp phát hiện khuôn mặt hiện đại, tích hợp hai tác vụ chính là phát hiện khung khuôn mặt và dự đoán điểm đặc trưng trên khuôn mặt như: mắt, mũi, miệng,...

Mô hình này hoạt động theo cấu trúc gồm ba mạng nơ-ron tích chồng:

* P-Net (Proposal Network): thực hiện quét sơ bộ ảnh để đề xuất các vùng có thể chứa khuôn mặt.
* R-Net (Refine Network): lọc lại các vùng được P-Net đề xuất và loại bỏ các hộp dự đoán sai.
* O-Net (Output Network): cung cấp vị trí chính xác hơn của khuôn mặt và 5 điểm đặc trưng (2 mắt, mũi, 2 khóe miệng) để phục vụ căn chỉnh khuôn mặt.

#### Trích xuất đặc trưng khuôn mặt bằng FaceNet

FaceNet là một mô hình học sâu nhằm ánh xạ mỗi ảnh khuôn mặt vào một không gian đặc trưng có kích thước cố định (128 chiều hoặc 512 chiều), gọi là face embedding. Khoảng cách giữa hai vector embedding phản ánh mức độ giống nhau giữa hai khuôn mặt.

Đặc điểm của embedding từ FaceNet:

* Hai khuôn mặt giống nhau sẽ có embedding gần nhau (khoảng cách nhỏ).
* Hai khuôn mặt khác nhau sẽ có khoảng cách embedding lớn hơn.

Để xác định xem 2 khuôn mặt có giống nhau không thì ta cần so sánh độ gần nhau giữa hai vector embedding.

Ta sử dụng khoảng cách Euclidean là độ dài đoạn thẳng giữa hai vector trong không gian nhiều chiều.

Công thức khoảng cách Euclidean:

A black square with numbers and a square in it

AI-generated content may be incorrect.

Nếu hai khoảng cách hai vector gần nhau trong không gian, thì hai khuôn mặt tương ứng giống nhau

* 1. **Mô Hình**

2.5.1 Logistic Regression

Hồi quy Logistic Regression là một thuật toán máy học có giám sát được sử dụng cho các nhiệm vụ phân loại trong đó là dự đoán xác suất một cá thể thuộc về một lớp nhất định hay không.

Mô hình hồi quy logistic chuyển đổi đầu ra giá trị liên tục của hàm hồi quy tuyến tính thành đầu ra giá trị phân loại bằng cách sử dụng hàm sigmoid, ánh xạ bất kỳ tập hợp đầu vào các biến độc lập có giá trị thực nào thành giá trị từ 0 đến 1. Chức năng này được gọi là hàm logistic.

Các tính năng đầu vào độc lập là biến X và biến phụ thuộc Y chỉ có tính nhị phân tức là 0 hoặc 1 sau đó áp dụng hàm đa tuyến tính cho các biến đầu vào X.

Ở đây là quan sát thứ i của X, là trọng số hoặc hệ số và b là thuật ngữ thiên vị còn được gọi là chặn. Điều này có thể được biểu diễn dưới dạng dot của trọng lượng và thiên vị.

Chức năng của hàm sigmoid có đầu vào là z và đầu ra là y dự đoán với xác suất từ 0 đến 1.

* có xu hướng về 1 nếu
* có xu hướng về 0 nếu
* luôn có giới hạn giữa 0 và 1

Đối với y = 1 thì xác suất dự đoán sẽ được tính như sau:

Đối với y = 0 thì xác suất dự đoán sẽ được tính như sau:

* + 1. SVM

Support Vector Machine (SVM) là một thuật toán học máy có giám sát tương đối đơn giản được sử dụng để phân loại hoặc hồi quy. Về cơ bản, SVM tìm thấy một siêu mặt phẳng tạo ra ranh giới giữa các loại dữ liệu. Trong không gian 2 chiều, siêu mặt phẳng này không là gì ngoài một đường thẳng. Trong SVM, từng mục dữ liệu trong tập dữ liệu trong không gian N chiều, trong đó N là số lượng tính năng/thuộc tính trong dữ liệu. Sau đó tìm siêu mặt phẳng tối ưu để tách dữ liệu.

Hàm kernel trong SVM kernel cho bạn biết rằng với hai điểm dữ liệu trong không gian tính năng ban đầu, sự tương đồng giữa các điểm trong không gian tính năng mới được chuyển đổi là gì. Có nhiều chức năng hạt nhân khác nhau có sẵn, nhưng có hai chức năng rất phổ biến:

* Radial Basis Function Kernel (RBF): Sự tương đồng giữa hai điểm trong không gian tính năng được chuyển đổi là một hàm phân rã theo cấp số nhân của khoảng cách giữa các vectơ và không gian đầu vào ban đầu.
* Polynomial Kernel: Polynomial Kernel nhận một tham số bổ sung, 'degree' kiểm soát độ phức tạp của mô hình và chi phí tính toán của quá trình biến đổi.

Ưu điểm:

* Chúng hoạt động rất tốt trên một loạt các bộ dữ liệu.
* Chúng rất linh hoạt: các hàm hạt nhân khác nhau có thể được chỉ định hoặc các hạt nhân tùy chỉnh cũng có thể được xác định cho các kiểu dữ liệu cụ thể.
* Chúng hoạt động tốt cho cả dữ liệu chiều cao và thấp.

Nhược điểm :

* Hiệu quả (thời gian chạy và sử dụng bộ nhớ) giảm khi kích thước của tập huấn luyện tăng lên.
* Cần chuẩn hóa cẩn thận dữ liệu đầu vào và điều chỉnh tham số.
* Không cung cấp một công cụ ước tính xác suất trực tiếp.
* Khó giải thích tại sao một dự đoán được đưa ra.
  + 1. Random Forest

Random Forest là một thuật toán học máy sử dụng nhiều cây quyết định để đưa ra dự đoán tốt hơn. Mỗi cây xem xét các phần ngẫu nhiên khác nhau của dữ liệu và kết quả của chúng được kết hợp bằng cách bỏ phiếu để phân loại hoặc tính trung bình cho hồi quy. Điều này giúp cải thiện độ chính xác và giảm lỗi.

A diagram of different trees

AI-generated content may be incorrect.

Quy trình hoạt động

* Tạo nhiều cây quyết định: Thuật toán tạo ra nhiều cây quyết định, mỗi cây sử dụng một phần dữ liệu ngẫu nhiên. Vì vậy, mỗi cây đều khác nhau một chút.
* Chọn các tính năng ngẫu nhiên: Khi xây dựng mỗi cây, nó không xem tất cả các tính năng (cột) cùng một lúc. Nó chọn một vài ngẫu nhiên để quyết định cách phân tách dữ liệu. Điều này giúp các cây luôn khác biệt với nhau.
* Mỗi cây đưa ra một dự đoán: Mỗi cây đưa ra câu trả lời hoặc dự đoán riêng dựa trên những gì nó học được từ phần dữ liệu của nó.
* Kết hợp các dự đoán:

1. Để phân loại, câu trả lời cuối cùng là câu trả lời mà hầu hết các cây đều đồng ý, tức là bỏ phiếu đa số.
2. Đối với hồi quy, câu trả lời cuối cùng là trung bình của tất cả các dự đoán cây.

* Tại sao nó hoạt động tốt: Sử dụng dữ liệu và tính năng ngẫu nhiên cho mỗi cây giúp tránh quá khớp và làm cho dự đoán tổng thể chính xác và đáng tin cậy hơn.

Các tính năng của Random Forest:

* Xử lý dữ liệu bị thiếu: Nó có thể hoạt động ngay cả khi một số dữ liệu bị thiếu, vì vậy bạn không phải lúc nào cũng cần phải tự điền vào khoảng trống.
* Hiển thị tầm quan trọng của tính năng: Nó cho biết tính năng (cột) nào hữu ích nhất để đưa ra dự đoán, giúp hiểu dữ liệu của mình tốt hơn.
* Hoạt động tốt với dữ liệu lớn và phức tạp: Nó có thể xử lý các bộ dữ liệu lớn với nhiều tính năng mà không làm chậm hoặc giảm độ chính xác.
* Được sử dụng cho các nhiệm vụ khác nhau: có thể sử dụng nó cho cả phân loại như dự đoán loại hoặc nhãn và hồi quy như dự đoán số hoặc số tiền.

Ưu điểm:

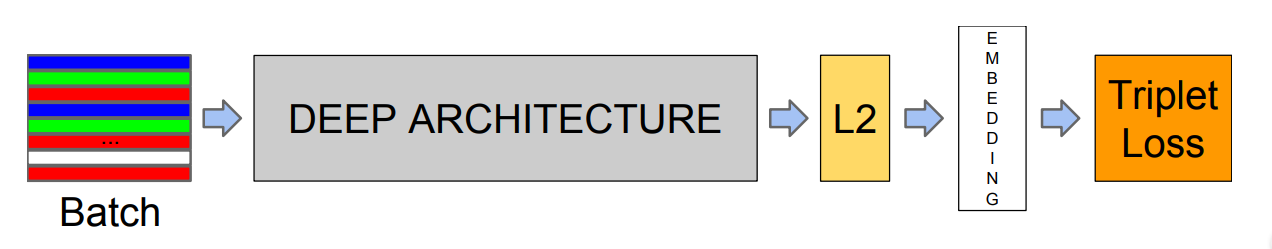
* Random Forest cung cấp các dự đoán rất chính xác ngay cả với bộ dữ liệu lớn.
* Random Forest có thể xử lý tốt dữ liệu bị thiếu mà không ảnh hưởng đến độ chính xác.
* Nó không yêu cầu chuẩn hóa hoặc tiêu chuẩn hóa trên tập dữ liệu.
* Khi chúng ta kết hợp nhiều cây quyết định, nó làm giảm nguy cơ quá phù hợp với mô hình.

Hạn chế :

* Nó có thể tốn kém về mặt tính toán, đặc biệt là với một số lượng lớn cây.
* Khó giải thích mô hình hơn so với các mô hình đơn giản hơn như cây quyết định.
  + 1. FaceNet

FaceNet là tên của hệ thống nhận dạng khuôn mặt được các nhà nghiên cứu của Google đề xuất vào năm 2015 trong bài báo có tựa đề FaceNet: A Unified Embedding for Face Recognition and Clustering.

Họ đề xuất một cách tiếp cận trong đó nó tạo ra bản đồ khuôn mặt chất lượng cao từ các hình ảnh bằng cách sử dụng các kiến trúc học sâu như ZF-Net và Inception Network. Sau đó, nó sử dụng một phương pháp gọi là triplet lossnhư một hàm mất mát để đào tạo mô hình này.



FaceNet sử dụng ZF-Net hoặc Inception Network làm kiến trúc cơ bản của nó. Nó cũng thêm một số tích chập 1 \* 1 để giảm số lượng tham số. Các mô hình deep learning này xuất ra một nhúng hình ảnh f(x) với L2 chuẩn hóa được thực hiện trên nó. Các nhúng này sau đó được chuyển vào hàm mất mát để tính toán tổn thất.

Mục tiêu của hàm mất này là làm cho khoảng cách bình phương giữa hai hình ảnh nhúng (độc lập với điều kiện hình ảnh và tư thế) của cùng một nhận dạng nhỏ, trong khi khoảng cách bình phương giữa hai hình ảnh có danh tính khác nhau là lớn. Do đó, một hàm mất mát mới được gọi là ***triplet loss*** được sử dụng. Ý tưởng sử dụng tổn thất bộ ba trong kiến trúc là nó giúp mô hình thực thi một biên độ giữa các khuôn mặt của các danh tính khác nhau.

Triplet loss: Việc nhúng một hình ảnh được biểu thị bằng f (x) chẳng hạn như x ∈ R. Việc nhúng này ở dạng vectơ có kích thước 128 và nó được chuẩn hóa sao cho

Công thức Triplet loss được tính như sau:

2.5.4 MobileNetV2

## *2.5.4.1 Giới thiệu về mô hình MobileNet*

MobileNet là một dòng mạng neural sâu, nó không phải là một mô hình nhận diện khuôn mặt chuyên biệt như FaceNet, nhưng nó thường được dùng làm kiến trúc nền (backbone) [ct] trong các hệ thống xác thực khuôn mặt để tối ưu hóa cho các thiết bị có tài nguyên hạn chế. Mục tiêu chính của MobileNet là duy trì hiệu suất tốt trong khi giảm thiểu số lượng tham số và chi phí tính toán.

*2.5.4.2 Nền tảng toán học*

MobileNet thực hiện các phép tính theo kiểu đơn giản, chia nhỏ công việc thay vì dùng một phép tích chập lớn như các mạng CNN truyền thống.

Bước 1: Mở rộng chiều dữ liệu (Expansion Layer) [ct]: ta dùng một phép tính convolution [ct] 1x1 để nhân bản thông tin sau đó dữ liệu sẽ trở nên dày hơn, giúp các bước sau xử lý tốt hơn.

*z = ReLU6(W1​∗x)* (2.)

Trong đó:  
- x: đầu vào  
- W1​: bộ lọc 1×1 dùng để tăng số kênh  
- z: đầu ra sau bước mở rộng

Và được giới hạn bởi hàm kích hoạt ReLU [ct]

*ReLU6(x)=min(max(0,x),6)*

Diễn giải:

* Nếu x ≤ 0 : đầu ra là 0
* Nếu 0 < x <6 : đầu ra là chính x
* Nếu x ≥ 6 : đầu ra bị giới hạn ở mức 6

Dùng ReLU6 vì giá trị 6 vừa đủ lớn để giữ tính phi tuyến và vừa đủ nhỏ để dễ lượng tử hóa xuống 8 – bit, đây là giá trị tối ưu để mô hình chạy nhẹ và nhanh trên thiết bị di dộng.

Bước 2: Tích chập Depthwise [ct], thay vì ta dùng một phép tích chập toàn bộ thì Depthwise thực hiện tích chập độc lập mỗi kênh để giảm mạnh số phép tính và tham số.

*di​=Ki​∗ zi​*

Trong đó:  
- Ki ​: bộ lọc áp dụng riêng cho kênh i  
- zi: kênh thứ i của đầu ra từ bước 1  
- di​: kết quả tích chập depthwise kênh i

Bước 3: Projection Layer [ct]: Sau khi các phép tích chập hoàn thành công việc sẽ được kết hợp các kênh với convolution 1x1 để đưa về số kênh ban đầu.

*y=W2​∗ d (2. )*

Trong đó:  
- W2​: bộ lọc 1×1 giảm số kênh  
- y: đầu ra cuối cùng của block

Bước 4: Residual Connection [ct]: Nếu đầu vào và đầu ra có cùng kích thước thì ta cộng lại đầu vào với kết quả đầu ra, giúp giữ lại thông tin gốc, tránh mất mát, và giúp máy học nhanh hơn.

*x* + *y (2. )*

*2.5.4.3 Diễn giải thuật toán*

Giai đoạn 1:

* Ảnh được đưa vào MobileNetV2, qua chuỗi các khối Inverted Residual [ct] và Linear Bottleneck [ct].
* Các lớp depthwise và convolution giúp giảm tính toán nhưng vẫn giữ được đặc trưng quan trọng.

Giai đoạn 2:

* Với nhận diện khuôn mặt, đầu ra là vector đặc trưng (embedding).
* Vector này có thể:
  + Dùng để huấn luyện bộ phân loại.
  + So sánh với vector từ các ảnh khác (ví dụ: nhận diện người lạ hay người quen).

*2.5.4.4 Phân tích độ phức tạp*

MobileNetV2 sử dụng một kỹ thuật tích chập đặc biệt giúp chia nhỏ quá trình xử lý thành hai bước đơn giản hơn thay vì làm tất cả cùng lúc. Đồng thời, còn sử dụng các khối xử lý giúp giữ lại thông tin quan trọng mà vẫn tiết kiệm được công sức tính toán. Nhờ đó, tổng số phép tính của mạng giảm nhiều lần so với các mô hình học sâu thông thường.

*2.5.4.5 Ưu Nhược điểm*

Ưu điểm:

* Nhẹ, nhanh, dễ triển khai trên thiết bị có tài nguyên hạn chế.
* Hiệu quả tốt với tốc độ xử lý nhanh.
* Phù hợp làm backbone cho nhiều ứng dụng như nhận diện khuôn mặt, phân loại ảnh, v.v.

Nhược điểm:

* Độ chính xác thấp hơn so với các mô hình lớn hơn như ResNet50 hay EfficientNet.
* Có thể bỏ sót các chi tiết nhỏ trong ảnh vì tối ưu chủ yếu cho tốc độ.
  1. **Nhận dạng đối tượng (Object Recognition)**

Nhận dạng đối tượng là quá trình xác định sự hiện diện và loại của một đối tượng cụ thể trong ảnh, điều này đồng nghĩa với việc không chỉ phát hiện có khuôn mặt, mà còn biết đó là ai trong danh sách đã được ghi nhớ.

* Thu nhận hình ảnh: Camera liên tục ghi lại các khung hình video làm dữ liệu đầu vào.
* Xử lý hình ảnh: Khung hình được xử lý giảm nhiễu, điều chỉnh màu và chuẩn hóa định dạng.
* Phát hiện khuôn mặt: Hệ thống xác định vị trí khuôn mặt và tách riêng vùng chứa khuôn mặt.
* Trích xuất đặc trưng: Mỗi khuôn mặt được chuyển thành một vector đặc trưng duy nhất.
* Nhận diện danh tính: Vector được đưa vào mô hình phân loại để so sánh và xác định danh tính.
* Hiển thị kết quả: Hệ thống đánh giá độ tin cậy và hiển thị tên cùng tỉ lệ chính xác qua độ tin cậy trên video.

CHƯƠNG 3 PHƯƠNG PHÁP THỰC NGHIỆM

**3.1 Định nghĩa mục tiêu của bài toán**

Ở bước này, mục tiêu của chúng tôi là xác định rõ phạm vi bài toán để thiết kế giải pháp hợp lý. Bài toán ở đây là phân loại, cụ thể là nhận diện danh tính người dùng thông qua ảnh khuôn mặt. Dữ liệu đầu vào gồm các ảnh mặt người, được phân chia theo từng thư mục, mỗi thư mục đại diện cho một cá nhân nhất định. Các ảnh này sẽ được trích xuất đặc trưng, sau đó đưa vào mô hình dự đoán để nhận về đầu ra là tên người tương ứng cùng với độ chính xác. Ngoài ra, hệ thống phải có khả năng phát hiện khuôn mặt chưa từng xuất hiện trong tập huấn luyện và trả về kết quả “Unknown”, đảm bảo phù hợp với yêu cầu về độ tin cậy và bảo mật thực tế.

**3.2 Thu thập dữ liệu**

Trong các hệ thống nhận diện khuôn mặt, chất lượng dữ liệu đầu vào quyết định hiệu quả chung của mô hình. Vì vậy, quá trình thu thập dữ liệu sẽ được tiến hành thủ công, chúng tôi phải thực sự đảm bảo chất lượng và tính chính xác của dữ liệu.

3.2.1 Phương pháp thu thập dữ liệu

Nhóm chúng tôi tiến hành thu thập dữ liệu ảnh khuôn mặt bằng cách chụp ảnh trực tiếp từng cá nhân trong điều kiện môi trường thực tế, sử dụng camera có độ phân giải cao để đảm bảo chất lượng ảnh rõ nét. Đối với mỗi đối tượng, số lượng ảnh thu thập dao động từ tối thiểu 5 đến gần 30 ảnh, tùy thuộc vào thời gian cũng như mức độ hợp tác của người tham gia.

3.2.2 Đảm bào chất lượng của dữ liệu

Để đảm bảo hệ thống nhận diện hoạt động hiệu quả trong điều kiện thực tế, mỗi cá nhân cần được chụp ảnh ở nhiều góc mặt khác nhau, bao gồm chính diện, nghiêng trái, nghiêng phải. Ngoài ra, các biểu cảm cũng phải đa dạng từ bình thường, cười nhẹ cho đến nghiêm túc. Các điều kiện ánh sáng cũng được kiểm tra, ví dụ như ánh sáng tự nhiên và ánh sáng phòng, nhằm tối ưu hóa khả năng nhận diện.

Về mặt kỹ thuật, phông nền của ảnh cần giữ đơn giản, hạn chế tối đa các yếu tố gây nhiễu như nhiều người xuất hiện trong khung hình hoặc vật thể che khuất khuôn mặt. Điều này giúp quá trình xử lý dữ liệu ở các bước sau diễn ra thuận lợi và nhất quán hơn.

3.2.3 Lưu trữ dữ liệu

Sau khi chụp, ảnh được lưu trữ theo cấu trúc thư mục, trong thư mục dataset mỗi thư mục con tương ứng với một người, được đặt tên theo tên của người đó để thuận tiện trong quá trình đánh nhãn và huấn luyện. Trong các thư mục con sẽ có tất cả các ảnh khuôn mặt đã được chụp của người đó.

**3.3 Tiền xử lý dữ liệu**

3.3.1 Làm sạch dữ liệu ảnh đầu vào

Đầu tiên, dữ liệu ảnh được xử lý thông qua một bước kiểm tra nhằm loại bỏ các tệp không hợp lệ. Trước hết, hệ thống sẽ duyệt qua từng thư mục con, mỗi thư mục tương ứng với một người trong bộ dữ liệu. Trong mỗi thư mục con, tất cả các tệp sẽ được rà soát lần lượt để kiểm tra xem chúng có phải là ảnh hợp lệ hay không. Cụ thể, nếu một tệp dữ liệu không thuộc định dạng ảnh chuẩn như .jpg, .jpeg hoặc .png, hoặc là ảnh bị lỗi chẳng hạn không thể mở được do hỏng file hoặc định dạng không đúng thì hệ thống sẽ phát hiện và tự động xóa tệp đó khỏi bộ dữ liệu.

A screenshot of a computer screen

AI-generated content may be incorrect.

Về cơ bản, bước này đảm bảo rằng chỉ những ảnh có định dạng và chất lượng đạt chuẩn mới được giữ lại để phục vụ cho các giai đoạn xử lý tiếp theo.

3.3.2 Mô tả dữ liệu

Dữ liệu được thu thập ban đầu bao gồm tổng cộng 754 ảnh khuôn mặt của các đối tượng người dùng khác nhau. Mỗi ảnh được chụp trực tiếp bằng camera với độ phân giải ổn định, đảm bảo rõ nét khuôn mặt, ánh sáng đồng đều và ít nhiễu.

Các ảnh được thu thập đều ở định dạng .jpg, .jpeg hoặc .png, với độ phân giải ổn định. Kích thước phổ biến nhất của các ảnh trong bộ dữ liệu là (640, 480).

3.3.3 Phân tích sự mất cân bằng giữa các lớp

Sau khi đã nắm sơ bộ cấu trúc và mô tả của dữ liệu, chúng tôi tiến hành thống kê số lượng ảnh tương ứng với từng lớp (mỗi lớp đại diện cho một người). Mỗi lớp được đặt tên theo tên người, và số lượng ảnh phản ánh số lần gương mặt của người đó được ghi nhận trong các quá trình tiếp theo.

Để trực quan hóa sự phân bố, chúng tôi sử dụng biểu đồ cột để thể hiện số lượng ảnh của từng lớp, từ đó đánh giá mức độ cân bằng của dữ liệu huấn luyện. Bên cạnh đó, biểu đồ này cũng được sử dụng nhằm chỉ rõ hơn mức trung vị, tứ phân vị, các điểm ngoại lai và độ lệch giữa các lớp.

A graph with blue lines

AI-generated content may be incorrect.

Qua biểu đồ, ta thấy sự mất cân bằng rõ rệt ví dụ như có nhiều lớp có số ảnh rất cao (trên 35 ảnh) và cũng có vài lớp có số lượng ảnh rất ít ( dưới 10 ảnh). Nếu đưa dữ liệu này vào thực thi mô hình thì có thể gây ra sự thiên lệch mô hình (model bias [ct])về các lớp có nhiều ảnh hơn, qua đó làm giảm hiệu quả nhận diện của các lớp có ít dữ liệu

3.3.4 Tăng cường dữ liệu

Sau khi phân tích số lượng ảnh trong từng lớp, nhận thấy dữ liệu phân bố không đồng đều, nhiều lớp có số lượng ảnh thấp dưới ngưỡng cần thiết để mô hình có thể học tốt. Do đó, chúng tôi tiến hành bước tăng cường dữ liệu (data augmentation [ct]) nhằm mở rộng số lượng mẫu huấn luyện cho những lớp có ít ảnh.

*3.3.4.1 Xác định ngưỡng số lượng ảnh tối thiểu*

- Vì số lượng ảnh của các lớp trải dài từ khoảng 10 – 35 ảnh nên chúng tôi quyết định ngưỡng tối thiểu của mỗi lớp cần có ít nhất 30 ảnh để đảm bảo đủ số lượng để có thể thực hiện trích xuất thông tin đặc trưng cho quá trình huấn luyện mô hình.

- Các lớp có ít hơn 30 ảnh sẽ được tăng cường để đạt đến mức tối thiểu này.

*3.3.4.2 Áp dụng tăng cường dữ liệu*

Nếu một lớp chứa ít hơn 30 ảnh, hệ thống sẽ tự động xác định số lượng ảnh cần bổ sung để đảm bảo mỗi lớp đạt ngưỡng tối thiểu là 30 ảnh. Cụ thể, số lượng ảnh cần tăng cường được tính bằng cách lấy 30 trừ đi số ảnh hiện có trong thư mục.

Để tạo ra các ảnh bổ sung, chương trình sẽ chọn ngẫu nhiên một ảnh từ tập ảnh sẵn có trong lớp đó rồi áp dụng các phép biến đổi đã được định nghĩa trước như xoay ảnh một góc nhỏ, lật gương, tăng độ sáng, hoặc điều chỉnh độ bão hòa màu. Mỗi lần lặp, thuật toán sẽ ngẫu nhiên lựa chọn một kỹ thuật tăng cường từ danh sách, nhằm đảm bảo tính đa dạng cho các ảnh mới tạo ra và tránh sự lặp lại.

* Xoay 10; 15 độ : Xoay ảnh theo chiều kim đồng hồ một góc 10,15 độ
* Xoay -10; -15 độ : Xoay ảnh ngược chiều kim đồng hồ một góc 10,15 độ
* Lật ngang : Phản chiếu ảnh theo trục dọc
* Tăng độ sáng : Làm sáng ảnh để mô phỏng điều kiện ánh sáng khác nhau
* Tăng độ bão hòa màu : Làm tăng độ rực màu để mô phỏng biến thiên môi trường

A collage of a person

AI-generated content may be incorrect.

Quy trình này được thực hiện đến khi tổng số ảnh trong lớp đạt mức tối thiểu yêu cầu. Các ảnh mới sẽ được đặt tên với tiền tố “aug\_” kèm theo số thứ tự, giúp phân biệt với ảnh gốc.

3.3.5 Phân bố dữ liệu sau khi tăng cường

Biểu đồ cột dưới đây minh họa số lượng ảnh trong từng lớp sau quá trình tăng cường.

A green and white graph

AI-generated content may be incorrect.

Từ biểu đồ có thể thấy rằng phần lớn các lớp đều đã đạt được ngưỡng tối thiểu là 30 ảnh như đã đặt ra trong tiêu chí tăng cường điều này là yếu tố quan trọng giúp cải thiện tính cân bằng của tập dữ liệu huấn luyện, từ đó hạn chế hiện tượng thiên lệch lớp.

**3.4 Trích xuất đặc trưng**

Sau khi đã có tập ảnh khuôn mặt đã được xử lý, tổ chức theo cấu trúc định sẵn, chúng tôi tiến hành xây dựng đoạn mã để trích xuất vector đặc trưng từ mỗi ảnh bằng cách sử dụng mô hình FaceNet.

Ở giai đoạn này, các ảnh khuôn mặt gốc vốn chứa nhiều thông tin thừa và nhiễu sẽ được chuyển đổi thành các vector đặc trưng có kích thước cố định, đảm bảo tính phân biệt tốt giữa các cá nhân.

3.4.1 Thiết lập môi trường

Ở phần đầu chương trình, hệ thống xác định thiết bị xử lý cho phép tự động ưu tiên GPU nếu máy tính có hỗ trợ CUDA [ct], giúp tăng tốc xử lý đáng kể khi làm việc với ảnh số lượng lớn. Nếu không có GPU, chương trình vẫn hoạt động bình thường trên CPU.

3.4.2 MTCNN ( **(Multi-task Cascaded Convolutional Networks)**

Với mỗi ảnh trong dataset, chương trình lần lượt thực hiện mở ảnh, chuyển đổi sang định dạng RGB và đưa vào mô hình MTCNN.

Mạng MTCNN sẽ giúp ta phát hiện các khuôn mặt có trong ảnh hoặc frame trong video webcam với 3 lớp mạng khác biệt, tượng trưng cho 3 stage [ct] chính là P-Net, R-Net và O-Net

*3.4.2.1 P – Net* (Proposal Network).

Trước hết, một bức ảnh thường sẽ có nhiều hơn một người. MTCNN có một giải pháp, bằng cách sử dụng phép Resize [ct] ảnh, để tạo một loạt các bản sao từ ảnh gốc với kích cỡ khác nhau, từ to đến nhỏ, tạo thành một Image Pyramid [ct].

Với mỗi một phiên bản copy-resize của ảnh gốc, ta sử dụng kernel [ct] 12x12 pixel và stride = 2 để đi qua toàn bộ bức ảnh, dò tìm khuôn mặt. Sau đó, ta sẽ đưa những kernels được cắt ra từ trên và truyền qua mạng P-Net . Kết quả cho ra các bounding boxes nằm trong mỗi kernel, mỗi bounding boxes sẽ chứa tọa độ 4 góc để xác định vị trí trong kernel chứa nó

A diagram of a block diagram

AI-generated content may be incorrect.

Để loại trừ bớt các bounding boxes và các kernels, ta sử dụng 2 phương pháp chính là lập mức Threshold confident [ct] nhằm xóa đi các box có mức confident thấp và sử dụng NMS [ct] (Non-Maximum Suppression) để xóa các box có tỷ lệ trùng nhau.

A screenshot of a computer

AI-generated content may be incorrect.

Sau khi đã xóa bớt các box không hợp lý, ta sẽ chuyển các tọa độ của các box về với tọa độ gốc của bức ảnh thật. Kết quả của quá trình trên sẽ là những tọa độ của box tương ứng ở trên ảnh kích thước ban đầu.

*3.4.2.2 R – Net (Refine Network)*

R-Netthực hiện các bước như P-Net. Tuy nhiên, mạng còn sử dụng một phương pháp tên là padding [ct], thực hiện việc chèn thêm các zero-pixels vào các phần thiếu của bounding box nếu bounding box bị vượt quá biên của ảnh. Tất cả các bounding box lúc này sẽ được resize về kích thước 24x24, được coi như 1 kernel và feed vào mạng R. Kết quả sau cũng là những tọa độ mới của các box còn lại.

A diagram of a network

AI-generated content may be incorrect.

*3.4.2.2 O – Net (Output Network)*

A diagram of a graph

AI-generated content may be incorrect.

Mạng này cũng thực hiện tương tự như việc trong mạng R. Tuy nhiên, kết quả đầu ra của mạng lúc này không còn chỉ là các tọa độ của các box nữa, mà trả về 3 giá trị bao gồm: 4 tọa độ của bounding , tọa độ 5 điểm landmark trên mặt, bao gồm 2 mắt, 1 mũi, 2 bên cánh môi và điểm confident của mỗi box, Tất cả sẽ được lưu vào thành 1 dictionary với 3 biến đầu ra kể trên.

3.4.3 Trích xuất vector đặc trưng bằng FaceNet

Khi đã có ảnh khuôn mặt chuẩn hóa từ MTCNN, chương trình tiếp tục đưa ảnh này vào mô hình Inception để trích xuất vector đặc trưng trước khi đưa vào mô hình.

*3.4.3.1 Các khái niệm cơ bản*

* Embedding Vector:  
  Là vector đặc trưng có kích thước cố định được học để biểu diễn khuôn mặt trong không gian đặc trưng sao cho các khuôn mặt giống nhau nằm gần nhau và khác nhau nằm xa nhau.
* Inception V1:  
  Là kiến trúc CNN với các khối Inception, khối này cho phép mạng được học theo cấu trúc song song, nghĩa là với 1 đầu vào có thể được đưa vào nhiều các lớp Convolution [ct] khác nhau để đưa ra các kết quả khác nhau, sau đó sẽ được gọp vào thành một output.
* Triplet Loss:

Thay vì sử dụng các hàm mất mát truyền thống, khi mà ta chỉ so sánh giá trị đầu ra của mạng với sự thật thực tế của dữ liệu, Triplet Loss đưa ra một công thức mới bao gồm 3 giá trị đầu vào gồm: Anchor, Positive, Negative.

Mạng được huấn luyện sao cho embedding của anchor gần positive và xa negative nhất có thể.

*3.4.3.2 Huấn luyên*

- Xây dựng mạng trích xuất đặc trưng (Feature Extractor):  
Mạng Inception V1 được sử dụng để chuyển ảnh khuôn mặt đầu vào thành một vector embedding cố định có kích thước 128 chiều

**-** Chuẩn hóa vector embedding:  
Áp dụng chuẩn hóa L2 [ct] cho các vector embedding để đưa tất cả embedding về cùng một không gian đơn vị. Việc này giúp việc so sánh khoảng cách Euclidean [ct] giữa các vector trở nên nhất quán và có ý nghĩa hơn.

- Tối ưu với Triplet Loss:  
Mạng được huấn luyện với hàm mất mát Triplet Loss, trong đó mỗi lần cập nhật trọng số sẽ dựa trên một bộ ba ảnh gồm:

* Anchor: ảnh gốc cần xác định danh tính,
* Positive: ảnh khác nhưng cùng người với anchor,
* Negative: ảnh của người khác.

Điều kiện mất mát đặt ra là khoảng cách giữa Anchor và Positive phải nhỏ hơn khoảng cách giữa Anchor và Negative ít nhất một ngưỡng.

Để thúc đẩy quá trình học hiệu quả hơn,chiến lược chọn bộ ba (Triplet Selection) được sử dụng, trong đó ưu tiên chọn các negative khó**,** tức là các mẫu khác người nhưng có embedding gần với anchor nhất. Điều này buộc mạng phải học kỹ hơn để phân biệt các khuôn mặt có đặc điểm tương đồng.

- Kiểm tra và triển khai:

Sau khi quá trình huấn luyện hoàn tất, mạng có thể được sử dụng để thực hiện các tác vụ như xác thực khuôn mặt hoặc nhận diện người.  
Khi có một ảnh khuôn mặt mới cần kiểm tra, ta đưa ảnh này qua mạng trích xuất đặc trưng đã được huấn luyện để thu được một vector embedding 128 chiều.  
Sử dụng khoảng cách L2 (Euclidean) để so sánh embedding của ảnh mới với các embedding đã lưu trong hệ thống.

* Nếu khoảng cách nhỏ hơn một ngưỡng định trước, thì hai ảnh được coi là cùng người.
* Nếu khoảng cách lớn hơn ngưỡng, thì được coi là người khác.

**3.5 Phân vùng tập dữ liệu.**

Sau khi hoàn tất quá trình trích xuất vector đặc trưng từ ảnh khuôn mặt, dữ liệu được chia thành hai tập là tập huấn luyện và tập kiểm tra. Đảm bảo quá trình huấn luyện được thực hiện trên một phần dữ liệu, trong khi hiệu quả của mô hình được đánh giá trên phần dữ liệu còn lại chưa từng được sử dụng trong huấn luyện.

- Đầu tiên, chương trình sẽ duyệt qua toàn bộ các tệp đặc trưng đã được tạo trong bước trích xuất embedding. Mỗi tệp chứa danh sách các vector đặc trưng và nhãn tương ứng của một người.

- Trước khi chia tập, toàn bộ vector đặc trưng trong mỗi tệp được xáo trộn để đảm bảo tính ngẫu nhiên và giảm thiểu sai lệch do thứ tự dữ liệu.

- 70% số mẫu được sử dụng cho tập huấn luyện, 30% số mẫu còn lại được sử dụng cho tập kiểm tra, trong trường hợp số lượng mẫu quá nhỏ chương trình sẽ đảm bảo có ít nhất một mẫu trong tập kiểm tra, giúp tránh tình huống không có dữ liệu test sẽ gây lỗi khi đánh giá mô hình.

Sau khi chia xong, toàn bộ dữ liệu huấn luyện và kiểm tra của tất cả đối tượng được gộp lại thành hai tập tổng hợp. Các tập này được chuẩn hóa và lưu trữ nhằm phục vụ cho các bước huấn luyện mô hình phân loại ở giai đoạn tiếp theo.

**3.6 Huấn luyện mô hình.**

**………………**

**3.7 Đánh giá hiệu suất**

**………….**

3.8. Hyper-parameter Tuning

Việc triển khai ba mô hình Machine Learning và thu được best\_score\_ là 0.89 cho thấy kết quả rất tốt. Tuy nhiên, đối với bài toán nhận diện khuôn mặt thì việc triển khai các mô hình máy học không phải là ý hay và còn phụ thuộc vào nhiều yếu tố. Nên chúng em quyết định triển khai FaceNet và MobileNet để thực hiện bài toán nhận diện khuôn mặt.

3.9. Build Pipeline

Đối với mô hình FaceNet chúng em thực hiện dự đoán xác suất bằng hàm kích hoạt relu và softmax với 3 lớp và mô hình được tối ưu bằng adam, hàm mất mát categorical\_crossentropy. Sau đó lưu mô hình và triển khai nhận diện bằng mô hình đã được tối ưu bằng cách mở webcam để nhận diện trực tiếp và kết quả thu được khá tốt. Trong 30 lần nhận diện thì 27 lần nhận diện chính xác đối tượng. Điều này cho thấy mô hình FaceNet rất phù hợp với bài toán nhận diện.

Mô hình MobileNet giúp giảm thiểu số lượng tham số và thời gian tính toán. Mặc dù triển khai mô hình này với các tham số tối ưu nhất vẫn không đưa ra được kết quả như mong muốn. Kết quả trong 36 lần nhận diện chỉ thu về 8 lần nhận diện chính xác. Nên việc áp dụng MobileNet vào bài toán chỉ mang mục đích so sánh và thử nghiệm mô hình.

A screenshot of a graph

AI-generated content may be incorrect.

Để đánh giá hiệu quả các mô hình nhận diện khuôn mặt, chúng tôi sử dụng chỉ số **accuracy (độ chính xác)** trên cả **tập huấn luyện** và **tập kiểm tra**.

* **Logistic Regression** có độ chính xác thấp nhất (train 77%, test 71%), cho thấy mô hình này không học được tốt đặc trưng khuôn mặt.
* **SVM** và **Random Forest** đạt accuracy cao trên tập huấn luyện (lần lượt là 93% và 100%), nhưng độ chính xác giảm đáng kể khi áp dụng vào tập kiểm tra (82% và 84%), phản ánh hiện tượng **overfitting** – đặc biệt rõ rệt ở Random Forest.
* **MobileNet** – một mô hình học sâu nhẹ – đạt độ chính xác tuyệt đối (train 100%, test 100%). Tuy nhiên, do mô hình tương đối đơn giản và kích thước nhỏ, việc đạt accuracy 100% có thể là do phù hợp hoàn hảo với tập dữ liệu hiện tại, và có thể không phản ánh khả năng tổng quát tốt trong trường hợp dữ liệu phức tạp hơn.
* **FaceNet** đạt độ chính xác rất cao và ổn định (train 94%, test 93%). Mặc dù không đạt 100% như MobileNet trên tập này, nhưng đây là một mô hình học sâu mạnh mẽ được thiết kế chuyên biệt cho việc **học biểu diễn khuôn mặt (face embeddings)**. Do đó, FaceNet cho thấy **khả năng tổng quát hóa tốt và đáng tin cậy hơn**, đặc biệt khi áp dụng cho các tình huống thực tế đa dạng.

Với các kết quả thu được, nhóm chúng tôi quyết định chọn **FaceNet là mô hình chính thức** để thực hiện bài toán nhận diện khuôn mặt. Đây là mô hình cân bằng tốt giữa **độ chính xác**, **tính ổn định**, và **khả năng mở rộng** khi áp dụng vào các hệ thống thực tế hoặc tập dữ liệu lớn hơn.

Các mô hình học máy như SVM, Random Forest, và Logistic Regression vẫn được sử dụng trong quá trình nghiên cứu, với vai trò là **đối chứng và so sánh hiệu năng**, nhờ thời gian huấn luyện nhanh và dễ triển khai. Tuy nhiên, những mô hình này không có khả năng tự trích xuất đặc trưng và thường đòi hỏi ảnh phải được xử lý trước thông qua các kỹ thuật đặc trưng hóa (như HOG, PCA hoặc embedding).