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](#_Toc195956866)

[DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT iv](#_Toc195956867)

[DANH MỤC CÁC BẢNG, HÌNH VẼ, ĐỒ THỊ iv](#_Toc195956868)

[LỜI MỞ ĐẦU 1](#_Toc195956869)

[CHƯƠNG 1 TỔNG QUAN 2](#_Toc195956870)

[1.1 Giới thiệu đề tài 2](#_Toc195956871)

[1.2 Nhiệm vụ của đề tài 2](#_Toc195956872)

[1.2.1 Tính cấp thiết của đề tài 3](#_Toc195956873)

[1.2.2 Ý nghĩa khoa học và thực tiễn của đề tài 4](#_Toc195956874)

[1.3 Mục tiêu nghiên cứu 5](#_Toc195956875)

[1.3.1 Mục tiêu tổng quát 5](#_Toc195956876)

[1.3.2 Mục tiêu cụ thể 5](#_Toc195956877)

[1.4 Đối tượng và phạm vi nghiên cứu 5](#_Toc195956878)

[1.4.1 Đối tượng nghiên cứu 5](#_Toc195956879)

[1.4.2 Phạm vi nghiên cứu 5](#_Toc195956880)

[1.5 Phương pháp nghiên cứu 6](#_Toc195956881)

[1.5.1 Phương pháp nghiên cứu sơ bộ 6](#_Toc195956882)

[1.5.2 Phương pháp nghiên cứu tài liệu 6](#_Toc195956883)

[1.5.3 Phương pháp nghiên cứu thống kê 6](#_Toc195956884)

[1.5.4 Phương pháp thực nghiệm 7](#_Toc195956885)

[1.5.5 Phương pháp đánh giá 7](#_Toc195956886)

[1.6 Những đóng góp nghiên cứu của đề tài 7](#_Toc195956887)

[1.6.1 Đóng góp về mặt lý thuyết 7](#_Toc195956888)

[1.6.2 Đóng góp trong thực tiễn 8](#_Toc195956889)

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

## 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 Nhậns 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. Chính vì vậy, nhận dạng khuôn mặt đã trở thành một điều gì đó khá quen thuộc với chúng ta trong cuộc sống hàng ngày đơn giản là việc quét xác thực gương mặt để mở khóa điện thoại, máy tính cũng như các thiết bị công nghệ khác. Nhưng không chỉ vậy, 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 hơn về đặc điểm, ưu nhược điểm của từng phương pháp, từ đó lựa chọn giải pháp phù hợp với nhu cầu thực tế giúp tiết kiệm thời gian và chi phí trong quá trình triển khai, 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.

Cải thiện trải nghiệm người dùng: tạo ra một quy trình nhận diện nhanh chóng và chính xác, giảm thời gian chờ đợi của người dùng. Trải nghiệm người dùng được cải thiện sẽ gia tăng sự hài lòng và lòng trung thành của khách hàng.

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**

**1 Các phương pháp, kỹ thuật tiền xử lý ảnh.**

* 1. **Grayscale Conversion.**

Một bức ảnh ban đầu thường có rất nhiều màu, những màu đó đều là kết quả của sự pha trộn giữa 3 màu là Red, Green và Blue, hay được gọi là nhóm màu RGB. Nhưng đối với những bài toán xử lý ảnh nói chung và nhận dạng khuôn mặt nói riêng thì việc sử dụng những bức ảnh RGB này khiến việc xử lý và nhận dạng trở nên phức tạp, mất thời gian và độ chính xác cũng không được tối ưu. Nhiều tác vụ trong lĩnh vực thị giác máy tính đặc biệt là nhận dạng khuôn mặt chỉ cần thông tin về cường độ ánh sáng chứ không cần màu sắc.

Nên việc chuyển ảnh sang ảnh xám là một bước vô cùng quan trọng và cần thiết khi giảm từ 3 kênh màu xuống còn 1 kênh duy nhất giúp đơn giản hóa dữ liệu đầu vào và giảm chi phí tính toán

**Quy trình thực hiện:**

* Bước 1: Đọc ảnh đầu vào: Sử dụng các thư viện như OpenCV, Pillow để thực hiện đọc ảnh từ file vào chương trình để xử lý.

+ OpenCV : *import cv2 – cv2.imread*

+ Pillow : *from PIL import Image – Image.open*

* Bước 2: Thực hiện chuyển đổi: Mỗi pixel màu được chuyển thành giá trị xám bằng công thức có trọng số: Gray= 0.299 x R +0.587 x G +0.114 xR.

+ OpenCV : *cv2.cvtColor(img, cv2.COLOR\_BGR2GRAY)*

+ Pillow : *img\_gray = img.convert('L')*

* 1. **Cân bằng Histogram (Histogram Equalization)**

Đôi khi một bức ảnh có tổng thể hoặc vài điểm ảnh quá tối hoặc quá sáng sẽ làm cho quá trình xử lý và nhận dạng gặp khó khăn, Điều này khiến các đặc điểm quan trọng như mắt, mũi, miệng bị mờ nhạt hoặc mất chi tiết, đặc biệt trong điều kiện ánh sáng yếu hoặc quá gắt.

Kỹ thuật cân bằng histogram giúp cải thiện điều này bằng cách trảiđều phân bố mức xám của ảnh trong khoảng [0,255].

Vùng [0,255] là phạm vi phổ biến của pixel trong ảnh số, hoặc có thể nói mỗi pixel của 1 ảnh số được lưu trữ bằng 8 bit tức là có thể biểu diễn được 2⁸ = 256 mức độ sáng khác nhau.

Các mức độ này được đánh số từ 0 đến 255, trong đó:

* 0 đại diện cho đen hoàn toàn (không có ánh sáng).
* 255 đại diện cho trắng hoàn toàn (sáng nhất).
* Các giá trị ở giữa biểu diễn các mức xám trung gian.

**Quy trình thực hiện:**

* Bước 1: Chuyển ảnh sang grayscale
* Bước 2: Tính histogram – đếm số pixel ở mỗi mức sáng (0–255).
* Bước 3: Tính hàm phân phối tích lũy CDF
  + CDF (Cumulative Distribution Function) được dùng để tính tần suất tích lũy của giá trị mức xám trong ảnh.
* Bước 4: Chuẩn hóa CDF về khoảng [0, 255].
* Bước 5: Dùng CDF để ánh xạ lại các giá trị pixel cũ → tạo ảnh mới có độ tương phản tốt hơn.

**Thực hiện chuyển đổi**:

+ OpenCV : i*mg\_eq = cv2.equalizeHist(img)*

+ Pillow : *img\_eq = ImageOps.equalize(img)*

* 1. **Làm mịn ảnh (Image Smoothing-blurring)**

Trong quá trình tiền xử lý ảnh, ảnh đầu vào 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 (blurring**)** 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 mịn ảnh đặc biệt quan trọng trong các bài toán như nhận diện khuôn mặt, phát hiện cạnh, phân đoạn ảnh,… vì nó giúp cải thiện độ chính xác bằng cách loại bỏ những chi tiết không quan trọng hoặc gây nhiễu.

Các kỹ thuật làm mờ phổ biến:

* 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 (Gaussian distribution) 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))

Với OpenCV :

Làm mờ trung bình: *img\_blur = cv2.blur(img, (5, 5))*

Làm mờ Gaussian : *img\_gaussian = cv2.GaussianBlur(img, (5, 5), 0*

Làm mờ trung vị *: img\_median = cv2.medianBlur(img, 5)*

Với Pillow :

Làm mờ Gaussian*: img\_blurimg.filter(ImageFilter.GaussianBlur(radius=2))*

Làm mờ trung bình*: img\_box = img.filter(ImageFilter.BoxBlur(2))*

* 1. **Biến đổi hình học (Affine / Perspective Transformations)**

Thực tế, hình ảnh không phải lúc nào cũng hoàn hảo để chúng ta sử dụng, đôi lúc sẽ gặp vài trường hợp bị lệch, nghiêng, xoay, co giãn hoặc chụp từ các góc nhìn khác nhau khiến cho đối tượng cần nhận dạng không nằm đúng vị trí hoặc tỷ lệ. Do đó nếu không căn chỉnh lại, các thuật toán phía sau có thể nhận sai, dẫn đến giảm độ chính xác của hệ thống.

*1.4.1Biến đổi hình học (Affine Transformation):*

Đây là một biến đổi hình học tuyến tính, bao gồm các thao tác như: tịnh tiến (di chuyển), xoay, phóng to, thu nhỏ, và làm nghiêng hình ảnh. Biến đổi này giữ nguyên các đường thẳng và tỷ lệ song song, nhưng không giữ nguyên khoảng cách hay góc. Được sử dụng khi ảnh chỉ bị nghiêng, lệch góc nhẹ, hoặc căn chỉnh đối tượng cơ bản.

Quy trình thực hiện:

* Bước 1: Chọn 3 điểm ảnh (input) có vị trí biết trước trong ảnh gốc.
* Bước 2: Chọn 3 điểm tương ứng (output) trong ảnh đích nơi muốn đưa đối tượng về
* Bước 3 :Tính ma trận biến đổi Affine (2x3) từ 3 cặp điểm.

*cv2.getAffineTransform(pts1, pts2)*

* Bước 4: Áp dụng biến đổi để tạo ảnh mới.

*cv2.warpAffine(img, M, (width, height)*

*1.4.2 Biến đổi phối cảnh (Perspective Transformation):*

Đây là một loại biến đổi mô phỏng góc nhìn không gian 3 chiều. Biến đổi phối cảnh không giữ nguyên song song, dùng để tạo hiệu ứng như nhìn ảnh từ một góc nghiêng, làm cho các đường thẳng hội tụ về một điểm xa giống như cách mắt người nhìn cảnh vật ngoài đời thật.Được sử dụng khi ảnh bị chụp lệch góc nghiêng mạnh.

Quy trình thực hiện:

* Bước 1: Chọn 4 điểm input tạo thành 1 tứ giác (vật thể bị méo hoặc lệch góc).
* Bước 2: Xác định 4 điểm đích tương ứng với hình chữ nhật chuẩn.
* Bước 3: Tính ma trận perspective (3x3).

*cv2.getPerspectiveTransform(pts1, pts2)*

* Bước 4: Áp dụng phép biến đổi toàn cục.

*cv2.warpPerspective(img, M, (w, h))*

**1.5 Data Augmentation (Phóng đại dữ liệu)**

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.

*cv2.flip*

* Xoay ảnh: Xoay ảnh với một góc nhất định như ±15° để mô phỏng các tư thế đầu khác nhau.

*cv2.getRotationMatrix2D*

* Phóng to/thu nhỏ: Làm thay đổi tỷ lệ kích thước đối tượng trong ảnh.

*cv2.resize*

* Dịch chuyển :Di chuyển ảnh theo trục X hoặc Y.

*cv2.warpAffine*

* Thay đổi độ sáng, độ tương phản: Mô phỏng điều kiện ánh sáng khác nhau.

*img\_new = alpha \* img + beta (alpha: độ tương phản, beta: độ sáng)*

* 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.

*np.random.normal(0, sigma, img.shape)*

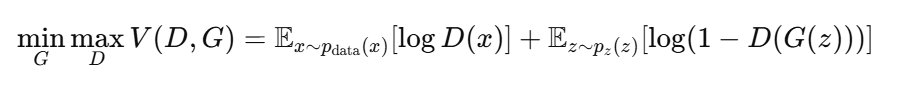
* Cắt ảnh ngẫu nhiên :cắt các vùng nhỏ ngẫu nhiên từ ảnh gốc để tạo thêm các biến thể.

*img[y:y+h, x:x+w]*

* 1. **GAN-based Enhancement (Generative Adversarial Network)**

GAN (Generative Adversarial Network) là một mô hình học sâu bao gồm hai mạng nơ-ron hoạt động đối kháng: Generator (Sinh ảnh) và Discriminator (Phân biệt ảnh). Generator cố gắng tạo ra ảnh giống thật, còn Discriminator cố gắng phân biệt ảnh thật và ảnh giả. Khi đó Generator bị ép phải sáng tạo ra dữ liệu ngày càng giống thật, bởi vì nếu nó tạo ra ảnh quá giả thì sẽ bị Discriminator phát hiện. Và khi Discriminator không còn phân biệt được ảnh thật ảnh giả nữa, nghĩa là Generatoe đã học được phân phối dữ liệu thật.

GAN được huấn luyện dựa trên hàm mất mát sau:



Trong đó: D(x) là xác xuất ảnh thật.

G(z) là ảnh giả do Generator tạo ra từ nhiễu z

Quy trình hoạt động:

* **Generator (1)** tạo ra dữ liệu giả từ một vector nhiễu z ~ p(z).
* **Discriminator (1)** nhận cả ảnh thật và giả, và cố gắng phân biệt chúng.
* **Generator (2)** được cập nhật để tạo ảnh giả tốt hơn (làm Discriminator nhầm lẫn).
* **Discriminator (2)** được cập nhật để nhận diện ảnh giả tốt hơn.
* Quá trình lặp lại cho đến khi ảnh giả gần giống với ảnh thật.

Các ứng dụng thực tế của GAN-based Enhancement:

* Sinh ảnh khuôn mặt: StyleGAN, Face GAN.
* Tăng độ phân giải ảnh : SRGAN, Real-ESRGAN.
* Khôi phục ảnh mờ/nhiễu: DeblurGAN, Denoising GAN.

**2. 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.

* 1. **Edge Detection**:

Phát hiện các đường biên như viền mặt, viền mắt, mũi, miệng – rất hữu ích trong việc định vị hình dạng khuôn mặt.

* **Laplacian:** Laplacian tìm cạnh bằng cách đo sự thay đổi độ sáng theo mọi hướng. Phát hiện tốt nhưng nhạy với nhiễu, nên dùng để bổ sung chi tiết trong ảnh.

*laplacian = cv2.Laplacian(image, cv2.CV\_64F)*

* **Sobel:** Sobel phát hiện cạnh theo chiều ngang và dọc bằng cách tính độ biến thiên độ sáng. Nhanh, dễ dùng nhưng dễ bị nhiễu và biên không quá sắc nét. Thích hợp cho xử lý ảnh cơ bản hơn.

*sobelx = cv2.Sobel(image, cv2.CV\_64F, 1, 0, ksize=3) # trục X*

*sobely = cv2.Sobel(image, cv2.CV\_64F, 0, 1, ksize=3) # trục Y*

* **Canny**: Canny là phương pháp tối ưu với nhiều bước lọc nhiễu và phát hiện biên rõ nét. Kết quả đẹp, ít nhiễu, rất phù hợp

*canny = cv2.Canny(image, threshold1=100, threshold2=200)*

* 1. **Corner Detection**

Là một kỹ thuật xác định các điểm góc vì chúng thường xuất hiện tại các giao điểm giữa các đường viền trong hình ảnh như góc của mắt, mũi, miệng,….

* **Harris Corner**: Phương pháp này phát hiện góc bằng cách tính toán ma trận tự tương quan của ảnh tại mỗi điểm, từ đó xác định nơi có sự thay đổi nhiều về độ sáng ở hai hướng. Các điểm góc sẽ có giá trị ma trận tự tương quan cao.

*harris\_corners = cv2.cornerHarris(gray, , , )*

* **Shi-Tomasi** : Shi-Tomasi cải tiến hơn Harris, thay vì sử dụng ma trận tự tương quan, Shi-Tomasi tìm kiếm các điểm có đặc tính góc rõ ràng bằng cách tính toán các giá trị riêng của ma trận tự tương quan.

*s\_tomasi = cv2.goodFeaturesToTrack(gray, , , )*

* 1. **Blob Detection:**

Chủ yếu tập trung vào việc phát hiện các vùng sáng hoặc tối trong ảnh mà có độ thay đổi đồng nhất về màu sắc hoặc độ sáng.

* LoG (Laplacian of Gaussian): như tên của nó, Phương pháp này áp dụng bộ lọc Gaussian để làm mịn ảnh, sau đó tính toán Laplacian (đạo hàm bậc hai) của ảnh đã được làm mịn. Giúp phát hiện các vùng có sự thay đổi mạnh về cường độ sáng.

*gaussian= cv2.GaussianBlur(image, (5, 5), 0)*

*laplacian = cv2.Laplacian(blurred\_image, cv2.CV\_64F)*

* DoG (Difference of Gaussian): là một phương pháp phát hiện blob tương tự như LoG, nhưng thay vì sử dụng Laplacian, DoG tính hiệu giữa hai bộ lọc Gaussian. Tương tự như đạo hàm bậc hai mà không cần tính toán trực tiếp Laplacian, giúp tiết kiệm tài nguyên tính toán.

*gaussian1 = cv2.GaussianBlur(image, (5, 5), 1.0)*

*gaussian2= cv2.GaussianBlur(image, (5, 5), 2.0)*

*DoG = gaussian1 – gaussian2*

* 1. **Learned Features**

Learned Features là các đặc trưng được học trực tiếp từ dữ liệu ảnh thông qua quá trình huấn luyện mạng sâu thay vì định nghĩa thủ công.

Đối với các phương pháp này, các đặc trưng không còn được trích xuất bằng tay như các phương pháp (Sobel, Harris, LoG…) mà được học tự động từ dữ liệu thông qua mạng nơ-ron tích chập (CNN). Giúp mô hình nhận diện khuôn mặt đạt độ chính xác cao hơn trong môi trường phức tạp.

Mạng CNN gồm nhiều tầng tích chập,mỗi tầng sẽ học một tập hợp bộ lọc khác nhau.

* Các tầng đầu học các đặc trưng đơn giản như cạnh, góc, kết cấu khuôn mặt (giống như Sobel, Harris,... nhưng tối ưu hơn).
* Các tầng giữa và tầng sâu học các đặc trưng phức tạp hơn như mắt, mũi, miệng, và cấu trúc tổng thể của khuôn mặt.
* Kết quả cuối cùng là một vector đặc trưng đại diện duy nhất cho khuôn mặt đầu vào.

**2.4.1 ResNet**

Mạng sâu sử dụng cơ chế residual connection để tránh mất thông tin khi mạng quá sâu

Giống như khi giải một bài toán đơn giản, ban đầu chỉ cần thực hiện vài bước là giải được. Nhưng nếu bài toán đó có quá nhiều bước rối rắm, phức tạp thì đôi khi làm taquên mất điều quan trọng ban đầu của bài toán. Trong mạng nơ-roncũng vậy. Khi mạng quá nhiều lớp, máy tính cũng dễ bị quên những gì quan trọng ban đầu trong ảnh.

Qua đó cơ chế residual connection của ResNet xuất hiện, nó có chức năng giúp giữ nguyên đặc trưng gốc của khuôn mặt ban đầu trong khi vẫn học được dữ liệu mới, Khi mạng học cái mới thì nó không quên thông tin cũ vì đường tắt (residual connection) đã nối thông tin cũ vào kết quả mới.

Quy trình hoạt động:

|  |
| --- |
| * + Trích xuất đặc trưng cơ bản |

|  |
| --- |
| * + Dùng khối residual để học cái mới mà không quên cái cũ |

|  |  |
| --- | --- |
|  | * + Lặp lại nhiều lần để học đặc trưng ngày càng chi tiết |

|  |  |
| --- | --- |
|  | Rút gọn thông tin thành vector |

|  |
| --- |
| * + Đưa ra dự đoán cuối cùng |

**2.4.2 EfficientNet:**

Khi muốn nâng cấp một mạng CNN, thường có ba cách:

|  |  |
| --- | --- |
| * + Tăng chiều sâu : | Thêm nhiều lớp để học đặc trưng sâu hơn |

|  |  |
| --- | --- |
| * + Tăng chiều rộng : | Thêm nhiều nơ-ron/lọc ở mỗi lớp |

|  |  |
| --- | --- |
| * + Tăng độ phân giải ảnh : | Dùng ảnh to hơn để có chi tiết tốt hơn |

Nhưng nếu chỉ tăng một thứ trong ba chỉ số trên mô hình có thể rất chậm hoặc không hiệu quả. Cho nên EfficientNet chọn cách tăng cả 3 thứ một cách cân bằng gọi là compound scaling giúp tối ưu cả độ sâu, chiều rộng và độ phân giải hình ảnh.

**2.5 Vision Transformers (ViT).**

Vision Transformer (ViT) là một mô hình không dùng mạng nơ-ron tích chập, mà sử dụng Transformer là một kiến trúc rất thành công trong xử lý ngôn ngữ tự nhiên. Và bây giờ, ViT mang sức mạnh của Transformer vào thị giác máy tính, đặc biệt là nhận dạng khuôn mặt.

Transformer là một mô hình trí tuệ nhân tạo được thiết kế để xử lý dữ liệu dạng chuỗi bằng cách sử dụng cơ chế self-attention giúp mô hình chú ý đúng chỗ và tìm mối liên hệ giữa các phần trong hình ảnh, từ đó hiểu nhanh và nhớ lâu hơn

Quy trình hoạt động:

* Bước 1: Chia ảnh thành các patches

Thay vì đưa cả ảnh vào mạng như CNN, ViT cắt ảnh thành các miếng nhỏ.

* Bước 2: Biến các patch thành vector

Mỗi patch sẽ được chuyển thành một vector số học

* Bước 3: Thêm vị trí

Vị trí các patch không còn như trong ảnh gốc, ViT sẽ thêm thông tin vị trí vào mỗi patch, để mô hình biết miếng nào nằm ở đâu.

* Bước 4: Đưa vào Transformer

Tất cả các patch này sẽ được đưa vào Transformer học các mối quan hệ giữa các phần trong ảnh.

* Bước 5: Dự đoán

ViT dùng một patch đặc biệt gọi là **[CLS]** để tóm tắt toàn bộ ảnh sau đó được dùng để phân loại.

**3. Nhận dạng và phân loại đối tượng (Object Classification & Recognition)**

Phân loại đối tượng là quá trình gán một nhãn cho toàn bộ hình ảnh hoặc một vùng ảnh chứa đối tượng, nhằm xác định đối tượng đó thuộc loại nào, có phải khuôn mặt hay không.

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.

* 1. **HOG Kết Hợp SVM .**

Đây là một quy trình gồm hai bước kết hợp giữa kỹ thuật xử lý ảnh (HOG) và thuật toán học máy (SVM) để nhận dạng và phân loại đối tượng.

* + 1. **HOG - Histogram of Oriented Gradients**

Kỹ thuật này thực hiện trích xuất đặc trưng hình học từ ảnh, cụ thể là các cạnh và đường viền của vật thể trong ảnh, rất phù hợp để mô tả hình dạng khuôn mặt, mắt, miệng. Và kỹ thuật này chỉ tập trung vào cấu trúc hình học, không bị ảnh hưởng nhiều bởi ánh sáng hay màu sắc.

Quy trình hoạt động:

* + Bước 1:Ảnh được chia thành các ô nhỏ
  + Bước 2: Trong mỗi ô, dùng các toán tử để tính
    - gradient (Gx: thay đổi sáng tối theo chiều ngang)
    - gradient (Gy: thay đổi sáng tối theo chiều dọc)
    - độ lớn của các cạnh : (magnitude): sqrt(Gx² + Gy²)
    - hướng (orientation): arctan(Gy / Gx)
* Bước 3 : Tạo histogram cho các hướng cạnh trong mỗi vùng, mô tả cách các cạnh phân bố.
  + - * Chia hướng cạnh thành các bins, ví dụ: 0–20°, 20–40°
      * Tìm bin phù hợp với hướng của nó.
* Bước 4: Chuẩn hóa và ghép đặc trưng
* Thực hiện gọp các ô lại thành một khối (2x2), sau đó sẽ được nối lại thành 1 vector dài.
* Chuẩn hóa vector bằng cách dùng L1-norm (theo đường ngang, dọc) hoặc L2-norm (theo đường chéo)
* Sau đó, duyệt qua ảnh bằng cách trượt các khối, sẽ sinh ra nhiều vector đặc trưng. Cuối dùng là nối tất cả lại sẽ tạo ra vector HOG, biểu diễn toàn bộ ảnh
  + 1. **SVM - Support Vector Machine**

SVM là một thuật toán học máy có khả năng phân loại hiệu quả với các bài toán phân loại nhị phân như có hay không có khuôn mặt, mô hình hoạt động bằng cách tìm một mặt phẳng tối ưu, sao cho khoảng cách giữa nó và các điểm dữ liệu gần nhất của mỗi lớp là lớn nhất giúp việc phân loại trở nên rõ ràng và chính xác

Khi kết hợp với HOG, SVM giúp phân biệt các đặc trưng hình học đã trích xuất, sau đó đưa ra quyết định cuối cùng về nhãn của ảnh hoặc vùng ảnh.

Quy trình hoạt động:

* Bước 1: Nhận đầu vào là vector đặc trưng( có thể là vector HOG đã trích xuất từ ảnh.
* Bước 2: Tìm siêu mặt phẳng phân chia hai lớp dữ liệu (có khuôn mặt và không khuôn mặt).
* Bước 3: Tối đa hóa khoảng cách (margin) giữa siêu mặt phẳng và các điểm gần nhất thuộc hai lớp từ đó tạo ra một vùng an toàn giữa hai loại đối tượng.
* Bước 4: Với ảnh mới, kiểm tra xem nó nằm về phía nào của siêu mặt phẳng từ đó dự đoán nhãn cho hình ảnh.
  1. **Template Matching**

Template Matching là một kỹ thuật dùng để tìm kiếm và định vị một phần cụ thể trong ảnh, bằng cách so sánh với một mẫu ảnh đã biết.

**Quy trình hoạt động:**

* **Bước 1:** Chọn một mẫu ảnh như mắt, mũi miệng,…
* **Bước 2:** Di chuyển mẫu này khắp ảnh lớn (quét từ trái sang phải, từ trên xuống dưới).
* **Bước 3:** Ở mỗi vị trí, tính toán độ giống nhau giữa mẫu và phần ảnh tại đó. Thường dùng các công thức như:
  + *Normalized Cross-Correlation (cv2.TM\_CCOEFF\_NORMED)*
  + *Sum of Squared Differences (cv2.TM\_SQDIFF)*
* **Bước 4:** Chọn vị trí có điểm số giống nhau cao nhất đó là nơi khớp nhất với mẫu.

Kỹ thuật này hiệu quả trong các trường hợp đơn giản khi đối tượng không xoay, không thay đổi kích thước, bị che khuất, biến dạng hoặc ở góc nhìn khác.

* 1. **Transformer-based: Vision Transformer (ViT), Swin Transformer**
     1. **Swin Transformer**

Khác với ViT (xử lý toàn ảnh), Swin Transformer chia ảnh đầu vào thành các cửa sổ, mỗi cửa sổ được xem là một vùng cục bộ, và self-attention chỉ áp dụng trong cửa sổ đó, giúp giảm chi phí tính toán.

Swin Transformer học đặc trưng theo tầng giống như CNN

* Lớp đầu : học đặc trưng cục bộ như đường viền, cạnh, hình dáng.
* Lớp sau : học đặc trưng trừu tượng hơn như hình dạng, khuôn mặt, đối tượng.

Nhưng nếu chỉ dùng cửa sổ cố định, mô hình sẽ không học được mối quan hệ giữa các vùng cửa sổ khác nhau. Vì vậy, Swin Transformer dùng kỹ thuật:

Shifted Window (dịch chuyển cửa sổ):

* Ở lớp tiếp theo, các cửa sổ được dịch chuyển một ít so với lớp trước.
* Nhờ vậy, vùng tiếp giáp giữa các cửa sổ cũ được xử lý trong cùng một cửa sổ mới.
* Kết quả: mô hình học được mối liên kết giữa các vùng ảnh khác nhau, tương tự như khi ta quét mắt qua ảnh.

**4. Phát hiện đối tượng (Object Detection)**

Mục tiêu: Xác định vị trí và nhãn của đối tượng trong ảnh.

**Phương pháp cổ điển:**

* Sliding Window + HOG + SVM (VD: phát hiện người)
* Viola–Jones (phát hiện khuôn mặt)

**Phương pháp hiện đại:**

* CNN-based: R-CNN, Fast R-CNN, Faster R-CNN
* One-stage detectors: YOLO (You Only Look Once), SSD, RetinaNet
* Transformer-based: DETR, Deformable DETR

**6. Dự đoán điểm đặc biệt trên cơ thể/người/vật (Keypoint Detection & Pose Estimation)**

Mục tiêu: Tìm các điểm đặc biệt trên vật thể (như khớp người).

**Phương pháp cổ điển:**

* HOG + SVM
* Optical Flow tracking

**Phương pháp hiện đại:**

* OpenPose, AlphaPose
* HRNet, MediaPipe Pose
* MMPose (OpenMMLab)

**8. Thị giác máy tính trong video ( Chỉ phù hợp khi sử dụng dữ liệu video)**

Mục tiêu: Phát hiện hành vi, theo dõi đối tượng trong chuỗi khung hình.

**Phương pháp cổ điển:**

* Background Subtraction
* Optical Flow
* Kalman Filter, Mean-Shift / CamShift tracking

**Phương pháp hiện đại:**

* Deep SORT (tracking)
* Action Recognition: I3D, C3D, SlowFast
* Temporal Transformer, TimeSformer

**CÁC MÔ HÌNH**

**4.1. Histogram of Oriented Gradients (HOG) + Sliding Window + SVM**

**Định nghĩa:**  
HOG là kỹ thuật trích xuất đặc trưng bằng cách tính toán hướng gradient của ảnh và gom nhóm chúng vào các histogram. Khi kết hợp với cửa sổ trượt (sliding window) và bộ phân loại SVM, đây là một phương pháp cổ điển để phát hiện đối tượng trong ảnh.

**Quy trình thực hiện:**

1. Tiền xử lý ảnh (chuyển ảnh về grayscale, chuẩn hóa).
2. Tính toán hướng gradient tại mỗi điểm ảnh.
3. Chia ảnh thành các ô ,ính histogram hướng gradient cho từng ô.
4. Gom các ô thành khối ,chuẩn hóa đặc trưng trong mỗi block.
5. Kết hợp đặc trưng từ tất cả các block để tạo vector HOG tổng thể.
6. Áp dụng cửa sổ trượt trên ảnh đầu vào với các tỉ lệ khác nhau.
7. Tính HOG đặc trưng tại mỗi cửa sổ, đưa vào SVM để phân loại.
8. Áp dụng phương pháp non-maximum suppression để loại bỏ các phát hiện trùng lặp.

**4.2. R-CNN (Regions with CNN features)**

**Định nghĩa:**  
R-CNN là phương pháp kết hợp giữa trích xuất vùng đề xuất (region proposal) và học sâu để phát hiện đối tượng. Nó sử dụng CNN để trích xuất đặc trưng từ các vùng đề xuất và sau đó phân loại chúng.

**Quy trình thực hiện:**

1. Áp dụng thuật toán như Selective Search để tạo các vùng đề xuất (region proposals).
2. Resize từng vùng đề xuất về cùng kích thước cố định.
3. Dùng CNN (như AlexNet) để trích xuất đặc trưng từ mỗi vùng đề xuất.
4. Đưa đặc trưng vào một bộ phân loại SVM để xác định loại đối tượng.
5. Dự đoán bounding box chính xác hơn bằng cách sử dụng hồi quy (bounding box regression).
6. Áp dụng non-maximum suppression để loại bỏ phát hiện trùng lặp.

**4.3. Fast R-CNN**

**Định nghĩa:**  
Fast R-CNN cải tiến từ R-CNN bằng cách tính đặc trưng toàn cục cho toàn bộ ảnh, sau đó sử dụng Region of Interest (RoI) Pooling để trích xuất đặc trưng cho từng vùng đề xuất từ đặc trưng toàn ảnh.

**Quy trình thực hiện:**

1. Đưa toàn bộ ảnh vào mạng CNN để tạo bản đồ đặc trưng (feature map).
2. Sử dụng thuật toán như Selective Search để tạo vùng đề xuất.
3. Dùng RoI Pooling để trích xuất đặc trưng từ các vùng trên feature map.
4. Đưa đặc trưng qua mạng fully connected.
5. Phân loại đối tượng (softmax) và dự đoán bounding box (bounding box regression).
6. Áp dụng non-maximum suppression.

**4.4. Faster R-CNN**

**Định nghĩa:**  
Faster R-CNN thay thế bước tạo vùng đề xuất bằng một mạng học sâu gọi là Region Proposal Network (RPN), giúp tăng tốc quá trình phát hiện đối tượng.

**Quy trình thực hiện:**

1. Đưa ảnh vào CNN để tạo feature map.
2. RPN quét các anchor boxes trên feature map để dự đoán vùng có thể chứa đối tượng.
3. Lọc và chọn ra các vùng đề xuất tốt nhất.
4. Sử dụng RoI Pooling để lấy đặc trưng từ các vùng đề xuất.
5. Phân loại đối tượng và dự đoán bounding box.
6. Áp dụng non-maximum suppression.

**4.5. YOLO (You Only Look Once)**

**Định nghĩa:**  
YOLO là một kiến trúc phát hiện đối tượng một bước (one-stage), chia ảnh thành các lưới (grid) và dự đoán đồng thời bounding boxes và class cho mỗi ô.

**Quy trình thực hiện:**

1. Chia ảnh thành lưới S×SS \times SS×S, mỗi ô chịu trách nhiệm phát hiện đối tượng nếu trung tâm đối tượng nằm trong ô đó.
2. Sử dụng CNN để dự đoán:
   * Bounding box (tọa độ + độ tin cậy).
   * Xác suất các lớp.
3. Tính toán hàm mất mát kết hợp giữa tọa độ, độ tin cậy và phân loại.
4. Sử dụng threshold và non-maximum suppression để lọc ra các bounding box cuối cùng.

**4.6. SSD (Single Shot MultiBox Detector)**

**Định nghĩa:**  
SSD cũng là một mô hình one-stage như YOLO, nhưng sử dụng nhiều tỉ lệ và kích thước hộp (default boxes) tại nhiều lớp đặc trưng để phát hiện đối tượng ở nhiều mức độ khác nhau.

**Quy trình thực hiện:**

1. Trích xuất feature maps ở nhiều tầng CNN.
2. Ở mỗi tầng, gán các default boxes với các tỷ lệ và tỉ lệ khung hình khác nhau.
3. Dự đoán class và tọa độ cho từng default box.
4. Sử dụng threshold và non-maximum suppression để lấy các kết quả cuối cùng.

**4.7. RetinaNet**

**Định nghĩa:**  
RetinaNet là mô hình one-stage cải tiến với focal loss để giải quyết vấn đề mất cân bằng giữa mẫu dương và âm khi huấn luyện mô hình phát hiện đối tượng.

**Quy trình thực hiện:**

1. Dùng backbone CNN (như ResNet + FPN) để tạo multi-scale feature maps.
2. Gán anchor boxes tại các điểm khác nhau trong feature maps.
3. Dự đoán class và bounding box cho mỗi anchor.
4. Tính focal loss để giảm trọng số các mẫu dễ phân loại.
5. Áp dụng non-maximum suppression.