TRƯỜNG ĐẠI HỌC AN GIANG

**KHOA CÔNG NGHỆ THÔNG TIN**

**KHOÁ LUẬN TỐT NGHIỆP NGÀNH KỸ THUẬT PHẦN MỀM**

**XÂY DỰNG MÔ HÌNH NHẬN DIỆN KHUÔN MẶT TRÊN RASPBERRY PI 4**

**MAI TIẾN KHÔI**

**AN GIANG, 5-2023**

TRƯỜNG ĐẠI HỌC AN GIANG

**KHOA CÔNG NGHỆ THÔNG TIN**

**KHOÁ LUẬN TỐT NGHIỆP NGHÀNH KỸ THUẬT PHẦN MỀM**

**XÂY DỰNG MÔ HÌNH NHẬN DIỆN KHUÔN MẶT TRÊN RASPBERRY PI 4**

**MAI TIẾN KHÔI**

**DPM195110**

**GIẢNG VIÊN HƯỚNG DẪN: TS.HUỲNH PHƯỚC HẢI**

**AN GIANG, 5-2023**

Khoá luận “ “

do sinh viên ……………………………………………… thực hiện dưới sự hướng dẫn của

………………………………………………………. Tác giả đã báo cáo kết quả nghiên cứu và được Hội đồng Khoa học và Đào tạo thông qua ngày ……………………….

**Phản biện 1 Phản biện 2**

*(Ký và ghi rõ chức danh, họ tên) (Ký và ghi rõ chức danh, họ tên)*

**Giảng viên hướng dẫn**

*(Ký và ghi rõ chức danh, họ tên)*

**LỜI CẢM ƠN**

Em muốn bày tỏ lòng cảm ơn sâu sắc đến những người thân, bạn bè và gia đình đã luôn ủng hộ và động viên em trong quá trình nghiên cứu và thực hiện khoá luận của em về đề tài xây dựng mô hình nhận diện khuôn mặt trên Raspberry Pi 4.

Đặc biệt, em muốn gửi lời cảm ơn đến 9 bạn lớp DH22PM đã cho phép em lấy mẫu để chạy thực nghiệm. Nhờ có sự đóng góp của những bạn đó, em đã có được dữ liệu cần thiết để hoàn thành khoá luận của mình và đạt được kết quả tốt.

Em cũng muốn cảm ơn thầy Huỳnh Phước Hải đã hướng dẫn và giúp đỡ em trong suốt quá trình nghiên cứu. Thầy đã truyền đạt cho em những kiến thức và kỹ năng cần thiết để xây dựng mô hình nhận diện khuôn mặt trên Raspberry Pi 4, cũng như giúp đỡ em vượt qua những khó khăn và thử thách trong quá trình thực hiện.

Cuối cùng, em muốn gửi lời cảm ơn đến tất cả những ai đã giúp đỡ và đóng góp vào thành công của khoá luận của em. Sự giúp đỡ của mọi người rất quan trọng đối với em và em rất biết ơn về điều đó.

**Sinh viên thực hiện**

**Mai Tiến Khôi**

**TÓM TẮT**

Trong thời đại công nghệ 4.0, việc ứng dụng trí tuệ nhân tạo vào đời sống ngày càng trở nên phổ biến và quan trọng. Nhận diện khuôn mặt là một trong những ứng dụng của trí tuệ nhân tạo được áp dụng rộng rãi trong nhiều lĩnh vực khác nhau như: bảo mật, giám sát an ninh, xác định danh tính, điểm danh...Nhận diện khuôn mặt được xem là giải pháp tốt để bảo vệ an ninh và tránh các hoạt động xâm nhập trái phép. Việc áp dụng công nghệ nhận diện khuôn mặt giúp cho các cơ quan chức năng và doanh nghiệp có thể quản lý dễ dàng và hiệu quả hơn.

Tuy nhiên, để xây dựng một mô hình nhận diện khuôn mặt cần phải sử dụng các thiết bị phần cứng đủ mạnh và phần mềm chuyên biệt. Điều này tạo ra một sự khó khăn cho các doanh nghiệp nhỏ và cá nhân muốn áp dụng công nghệ nhận diện khuôn mặt vào sản phẩm của mình. Với chi phí cao và đòi hỏi kiến thức chuyên môn, việc xây dựng mô hình nhận diện khuôn mặt trở nên phức tạp và khó khăn hơn.

Để giải quyết vấn đề này, trong đề tài em sử dụng Raspberry Pi 4, một thiết bị nhỏ gọn và có khả năng xử lý tương đối đủ để em thực hiện việc nghiên cứu cách xây dựng mô hình nhận diện khuôn mặt chạy trên thiết bị này, với việc kết hợp bởi các thư viện mà ngôn ngữ lập trình python hỗ trợ mạnh về phân tích dữ liệu cũng như các thư viện máy học để xây dựng và cho mô hình học trên tập dữ liệu khuôn mặt để thực hiện việc nhận diện. Qua quá trình nghiên cứu, em đã xây dựng thành công một mô hình nhận diện khuôn mặt cơ bản đủ để phục vụ cho việc nghiên cứu cũng như học tập trên tập dữ liệu vừa đủ, mô hình nhận diện khuôn mặt sử dụng mạng nơ-ron tích chập (CNN), một mô hình sử dụng kỹ thuật Transfer Learning và một mô hình sử dụng máy học véc-tơ (SVM) trên Raspberry Pi 4.

Kết quả của nghiên cứu này cho thấy Raspberry Pi 4 có thể được sử dụng để xây dựng mô hình nhận diện khuôn mặt phù hợp với những nhu cầu không cao về mặt giám sát hay quản lý bởi mô hình chỉ thích hợp cho một lượng dữ liệu vừa và đủ chỉ phù hợp để phục vụ việc học tập vì nghiên cứu của đề tài muốn chứng minh có thể xây dựng mô hình nhận diện khuôn mặt trên Raspberry Pi 4 nhằm giảm chi phí và đơn giản hoá quá trình phát triển sản phẩm áp dụng công nghệ nhận diện khuôn mặt cho các nhu cầu vừa đủ hoặc tương đối trong việc nhận diện khuôn mặt, việc thực hiện việc nhận diện trên lượng dữ liệu là không quá lớn. Kết quả với mong muốn đóng góp vào việc phổ biến ứng dụng trí tuệ nhân tạo trong cuộc sống của con người và nâng cao chất lượng cuộc sống của họ, đồng thời định hướng phát triển các thiết bị điện tử ngày càng thông minh và hiệu quả hơn.

MỤC LỤC

[CHƯƠNG 1: GIỚI THIỆU 1](#_Toc136162799)

[1.1 Tính cấp thiết của nghiên cứu để xây dựng mô hình 1](#_Toc136162800)

[1.2 Dữ liệu ảnh khuôn mặt 2](#_Toc136162801)

[1.3 Phát hiện người dựa trên dữ liệu ảnh khuôn mặt 3](#_Toc136162802)

[1.4 Mục tiêu, đối tượng, và phạm vi nghiên cứu 4](#_Toc136162803)

[1.5 Ý nghĩa của nghiên cứu cách xây dựng mô hình 5](#_Toc136162804)

[CHƯƠNG 2: CƠ SỞ LÝ THUYẾT 6](#_Toc136162805)

[2.1 Raspberry Pi 4 6](#_Toc136162806)

[**2.1.1 Giới thiệu về Raspberry Pi 4** 6](#_Toc136162807)

[2.2 Mạng nơ-ron tích chập sâu 7](#_Toc136162808)

[**2.2.1 Giới thiệu về Convolutional Neural Network** 7](#_Toc136162809)

[**2.2.2 Ý nghĩa các lớp cấu trúc của CNN** 9](#_Toc136162810)

[**2.2.3 Một số kỹ thuật được sử dụng để huấn luyện mô hình có hiệu quả** 11](#_Toc136162811)

[**2.3 Các mô hình học máy được ứng dụng trong đề tài để tìm ra độ chính xác cao trên tập dữ liệu** 12](#_Toc136162812)

[**2.3.1 Mô hình VGGNet** 12](#_Toc136162813)

[**2.3.2 Mô hình MobileNet** 16](#_Toc136162814)

[**2.4 Kỹ thuật Transfer Learning** 18](#_Toc136162815)

[**2.4.1 Một số vấn đề và hạn chế của các mô hình học sâu trên các tập dữ liệu thực tế** 18](#_Toc136162816)

[**2.4.2 Một số định nghĩa trong CNN** 19](#_Toc136162817)

[**2.4.3 Vai trò của Transfer learning** 20](#_Toc136162818)

[**2.4.5 Hạn chế của Transfer learning** 21](#_Toc136162819)

[2.5 Máy học véc-tơ hỗ trợ (SVM) 21](#_Toc136162820)

[**2.5.1 Giới thiệu về Support Vector Machine - SVM** 21](#_Toc136162821)

[**2.5.2 Ưu và nhược điểm của Kernel Support Vector Machine** 24](#_Toc136162822)

[**2.5.3 Một số kỹ thuật trong SVM (Support Vector Machine)** 24](#_Toc136162823)

[CHƯƠNG 3: NỘI DUNG VÀ KẾT QUẢ THỰC NGHIỆM CỦA VIỆC XÂY DỰNG MÔ HÌNH NHẬN DIỆN KHUÔN MẶT TRÊN RASPBERRY PI 4 29](#_Toc136162824)

[3.1 Xây dựng mô hình với mạng CNN 29](#_Toc136162825)

[3.2 Xây dựng mô hình sử dụng Transfer Learning 33](#_Toc136162826)

[3.3 Đánh giá mô hình mạng CNN 35](#_Toc136162827)

[3.4 Sử dụng SVM kết hợp PCA và Gridsearch cho nhận diện khuôn mặt 36](#_Toc136162828)

[3.5 Sử dụng SVM kết hợp FaceNet và MTCNN cho nhận diện khuôn mặt 39](#_Toc136162829)

[3.6 Đánh giá việc sử dụng nhận diện khuôn mặt bằng SVM và mạng CNN 44](#_Toc136162830)

[4.1 Đánh giá 44](#_Toc136162831)

[4.2 Hướng phát triển 44](#_Toc136162832)

**DANH SÁCH CÁC HÌNH VẼ**

[Hình 1: Hình ảnh mô tả ảnh khuôn mặt được phân tích bằng các mô hình máy học để nhận diện 3](#_Toc136162982)

[Hình 2:Các giai đoạn phát triển các thuật toán nhận diện khuôn mặt 4](#_Toc136162983)

[Hình 3:Bo mạch raspberry pi 4 7](#_Toc136162984)

[Hình 4:Cấu trúc cơ bản của mạng nơ-ron tích chập 8](#_Toc136162985)

[Hình 5:Kiến trúc mạng VGG 13](#_Toc136162986)

[Hình 6:Kiến trúc mô hình VGG16 14](#_Toc136162987)

[Hình 7:Kiến trúc mô hình VGG19 15](#_Toc136162988)

[Hình 8: Kiến trúc mô hình MobileNet 16](#_Toc136162989)

[Hình 9: Cấu trúc mô hình MobileNetV2 17](#_Toc136162990)

[Hình 10:Cách tạo model mới từ một pre-trained model 20](#_Toc136162991)

[Hình 11:Kết quả của việc sử dụng Transfer Learning trước và sau khi áp dụng vào mô hình 21](#_Toc136162992)

[Hình 12:Hình ảnh mô tả dữ liệu tuyến tính và dữ liệu phi tuyến tính 22](#_Toc136162993)

[Hình 13:Hình ảnh mô tả PCA giữ lại các thông tin quan trọng của ảnh 25](#_Toc136162994)

[Hình 14:Hình ảnh mô tả Gridsearch thực hiện việc tìm kiếm tham số tốt nhất cho dữ liệu 26](#_Toc136162995)

[Hình 15:Cấu trúc của MTCNN 27](#_Toc136162996)

[Hình 16:Mô tả FaceNet xử lý dữ liệu 28](#_Toc136162997)

[Hình 17:Hình ảnh mô tả việc tạo cơ sở dữ liệu để lưu thông tin của người nhận diện 30](#_Toc136162998)

[Hình 18:Hình ảnh mô tả việc thêm thông tin hoặc cập nhật thông tin trong cơ sở dữ liệu 30](#_Toc136162999)

[Hình 19: Hình ảnh mô tả việc hiển thị thông tin và cắt ảnh khuôn mặt 30](#_Toc136163000)

[Hình 20:Hình ảnh mô tả việc lấy dữ liệu và chuyển đổi ảnh xám 31](#_Toc136163001)

[Hình 21:Hình ảnh mô tả việc tạo dữ liệu cho tập huấn luyện 31](#_Toc136163002)

[Hình 22:Hình ảnh mô tả việc gán nhãn cho dữ liệu 31](#_Toc136163003)

[Hình 23:Sử dụng tăng cường dữ liệu 32](#_Toc136163004)

[Hình 24:Chia dữ liệu thành train và test, sau đó gọi datagen cho tập huấn luyện 32](#_Toc136163005)

[Hình 25:Xây dựng mô hình và tiến hành huấn luyện 32](#_Toc136163006)

[Hình 26:Tạo ra 3 kênh màu với giá trị của ảnh xám để đưa vào mô hình để huấn luyện 33](#_Toc136163007)

[Hình 27:Tinh chỉnh mô hình VGG16 33](#_Toc136163008)

[Hình 28:Tinh chỉnh mô hình VGG19 33](#_Toc136163009)

[Hình 29:Mô hình MobileNet 34](#_Toc136163010)

[Hình 30:Mô hình MobileNetV2 34](#_Toc136163011)

[Hình 31:Hình ảnh dự đoán sau khi sử dụng các model trên 34](#_Toc136163012)

[Hình 32:Dữ liệu khuôn mặt đã qua xử lý và resize về cùng kích thước 256x256 35](#_Toc136163013)

[Hình 33:Biểu đồ đánh giá độ chính xác và thời gian huấn luyện của các mô hình 35](#_Toc136163014)

[Hình 34:Thực hiện việc giảm chiều ảnh 36](#_Toc136163015)

[Hình 35:Xử lý các nhãn và đặc trưng của hình ảnh 37](#_Toc136163016)

[Hình 36:Chia dữ liệu thành 2 tập huấn luyện và kiểm tra đồng thời xử lý các thành phần của ảnh khi huấn luyện 37](#_Toc136163017)

[Hình 37:Các thành phần chính của ảnh sau khi dùng PCA để xử lý 37](#_Toc136163018)

[Hình 38:Tiến hành huấn luyện 38](#_Toc136163019)

[Hình 39:Tiến hành nhận diện bằng SVM 38](#_Toc136163020)

[Hình 40:Tiến hành load ảnh 39](#_Toc136163021)

[Hình 41:Nhận diện khuôn mặt và lưu trữ kết quả 39](#_Toc136163022)

[Hình 42:Sử dụng kết quả nhận diện và vẽ khung hình 40](#_Toc136163023)

[Hình 43:Tách và điều chỉnh kích thước khuôn mặt 40](#_Toc136163024)

[Hình 44:Thực hiện việc xử lý ảnh 41](#_Toc136163025)

[Hình 45:Gọi hàm và xử lý ảnh 41](#_Toc136163026)

[Hình 46:Hiển thị ảnh để kiểm tra 42](#_Toc136163027)

[Hình 47:Sử FaceNet để Embedding các ảnh khuôn mặt thành vector 42](#_Toc136163028)

[Hình 48:Lưu lại quá trình Embedding 42](#_Toc136163029)

[Hình 49:Tiền xử lý nhãn 'Y' 42](#_Toc136163030)

[Hình 50:Chia tập dữ liệu huấn luyện 42](#_Toc136163031)

[Hình 51:Huấn luyện mô hình 42](#_Toc136163032)

[Hình 52:Đánh giá mô hình 43](#_Toc136163033)

[Hình 53:Thực hiện việc kiểm tra và lưu lại mô hình 43](#_Toc136163034)

[Hình 54:Thực hiện việc nhận diện bằng camera 43](#_Toc136163035)

**DANH SÁCH CÁC BẢNG**

[Bảng 1:Độ chính xác và các tham số của kiến trúc mạng VGG16 và VGG 19 15](#_Toc136162913)

[Bảng 2:Độ chính xác và các tham số của kiến trúc mạng MobileNet và MobileNetV2 18](#_Toc136162914)

CHƯƠNG 1: GIỚI THIỆU

* 1. **Tính cấp thiết của nghiên cứu để xây dựng mô hình**

Khi chưa có sự phát triển mạnh mẽ của công nghệ nhận diện khuôn mặt, việc xác thực khuôn mặt đối với một cá nhân có thể trở thành một nhiệm vụ phức tạp và đòi hỏi sự chính xác cao. Trong thời kỳ đó, các phương pháp truyền thống được sử dụng để xác minh danh tính, ví dụ như xác thực bằng giấy tờ tùy thân, chữ ký hoặc mã số cá nhân.

Phương pháp xác thực bằng giấy tờ tùy thân đã được sử dụng từ lâu để xác định danh tính của một người. Tuy nhiên, phương pháp này có nhược điểm là dễ dàng bị làm giả mạo và đòi hỏi sự phụ thuộc vào việc kiểm tra hình thức giấy tờ. Đối với những ứng dụng yêu cầu mức độ xác thực cao, cần có một giải pháp tiên tiến hơn. Trong khi đó, xác thực bằng chữ ký cũng được sử dụng phổ biến. Tuy nhiên, nhược điểm của phương pháp này là chữ ký có thể được sao chép hoặc làm giả mạo dễ dàng, làm mất đi tính chính xác của quy trình xác thực. Một phương pháp khác được sử dụng là xác thực bằng mã số cá nhân. Tuy nhiên, phương pháp này dễ dàng bị xâm phạm bởi việc lừa đảo thông qua việc truy cập trái phép vào thông tin cá nhân của người khác. Trong tổng quát, các phương pháp truyền thống để xác thực khuôn mặt chưa đảm bảo tính chính xác và độ tin cậy cao. Sự ra đời của công nghệ nhận diện khuôn mặt đã mang lại một bước đột phá quan trọng trong việc giải quyết vấn đề này.

Trong những năm gần đây, công nghệ nhận diện khuôn mặt đã trải qua một sự phát triển đáng kể, mở ra nhiều tiềm năng và ứng dụng đa dạng. Nhận diện khuôn mặt là quá trình tự động xác định và xác minh danh tính của một cá nhân dựa trên các đặc điểm hình thái và cấu trúc của khuôn mặt. Điều này đã tạo ra một số tiến bộ đáng kể trong lĩnh vực an ninh, giao thông, công nghệ thông tin và nhiều lĩnh vực khác. Một trong những yếu tố chính đằng sau sự phát triển nhanh chóng của công nghệ nhận diện khuôn mặt là sự tiến bộ trong lĩnh vực học máy và trí tuệ nhân tạo. Các thuật toán nhận diện khuôn mặt ngày càng trở nên mạnh mẽ và chính xác hơn, với khả năng nhận diện và phân tích các đặc trưng của khuôn mặt, công nghệ này có thể xác định các thông tin như đặc điểm hình dạng khuôn mặt, kết cấu da, cấu trúc đường viền môi và mắt, cũng như vị trí các điểm nhận dạng quan trọng khác trên khuôn mặt con người một cách chính xác và nhanh chóng. Đặc biệt, công nghệ nhận diện khuôn mặt đã tạo ra sự đột phá vượt trội trong các lĩnh vực như an ninh và an toàn. Hệ thống nhận diện khuôn mặt đã áp dụng các mô hình được xây dựng để phục vụ cho việc nhận diện khuôn mặt được sử dụng rộng rãi trong các hệ thống kiểm soát truy cập, giúp xác định và xác minh danh tính của cá nhân trong thời gian thực. Điều này có thể áp dụng trong các cơ quan chính phủ, công ty, sân bay, trường học và các khu vực công cộng khác, giúp tăng cường đáng kể mức độ an toàn và bảo mật. Ngoài ra, công nghệ nhận diện khuôn mặt cũng đã có những tác động lớn trong lĩnh vực giải trí và trải nghiệm người dùng. Ví dụ, ứng dụng nhận diện khuôn mặt được tích hợp vào các thiết bị di động và máy tính cá nhân cho phép người dùng mở khóa điện thoại, chụp ảnh hoặc tham gia vào các trò chơi chỉ bằng cách sử dụng khuôn mặt của họ.

Trong tập dữ liệu ảnh khuôn mặt người được cắt ra với độ phân giải ảnh gốc nhằm lấy đủ thông tin của khuôn mặt trong ảnh, các ảnh được cắt ra từ đoạn video quay chân dung người sau đó dùng bộ thư viện của haarcascade để xác định khuôn mặt và cắt. Ảnh khuôn mặt được lấy ra với các kiểu khác nhau và mỗi đoạn video của một người sẽ cắt ra 200 ảnh khuôn mặt, ngoài ra trong đề tài có sử dụng thêm một tập dữ liệu từ “LFW Face Database” để tiến hành so sánh giữa tập dữ liệu thực tế và tập dữ liệu từ “LFW Face Database”. Mặc dù đã có nhiều nghiên cứu thực hiện việc xây dựng mô hình nhận diện khuôn mặt trên các thiết bị có phần cứng mạnh để giải quyết vấn đề này một cách tối ưu. Tuy nhiên với đề tài này thực hiện việc xây dựng mô hình trên thiết bị nhỏ gọn về mặt phần cứng thì vẫn chưa đủ mạnh để xử lý các tiến trình trong việc nhận diện, dù đã áp dụng các mô hình học máy nhưng kết quả vẫn không như mong đợi vì phần cứng không đủ mạnh và lượng dữ liệu chưa đủ tính tổng quát do tập dữ liệu thực tế nên việc xin phép người cho lấy mẫu cũng phụ thuộc vào vấn đề riêng tư của họ cũng giống như tập dữ liệu được lấy từ “LFW Face Database”, lượng dữ liệu chỉ đủ để phục vụ cho vấn đề nghiên cứu và học tập để chứng minh việc xây dựng mô hình nhận diện khuôn mặt vẫn có thể làm trên các thiết bị nhỏ gọn có phần cứng không quá mạnh.

Tình trạng của thế giới ngày nay đang đặt ra nhiều thách thức, việc lượng tội phạm và hành vi xâm phạm an ninh trở nên phức tạp, không những thế với sự gia tăng nhanh chóng của dân số và đô thị hoá, việc đảm bảo an toàn và hiệu quả trong quản lý dân số là một thách thức lớn. Ngoài ra việc bảo vệ dữ liệu cá nahan trở thành một ưu tiên hàng đầu nhằm tránh các vấn đề truy cập vào quyền riêng tư của mỗi người. Ở nước Việt Nam ta, đặc biệt là các sân bay, việc lưu thông giữa các nước hầu như diễn ra liên tục nên không tránh khỏi các vấn đề về tội phạm, an ninh…Vì thế việc áp dụng các mô hình nhận diện khuôn mặt trên các hệ thống kiểm tra có thế kiểm tra khuôn mặt của từng người khi vào nước ta, tránh việc tạo cơ hội cho tội phạm xâm nhập vào nước ta. Việc áp dụng công nghệ nhận diện khuôn mặt đã cho ta thấy được hiệu quả của công nghệ nhận diện ngày càng thông minh hơn.

* 1. **Dữ liệu ảnh khuôn mặt**

Dữ liệu ảnh khuôn mặt là một tập hợp các hình ảnh chụp khuôn mặt của con người. Mỗi hình ảnh trong dữ liệu đại diện cho một cá nhân cụ thể hoặc một tình huống liên quan đến khuôn mặt. Dữ liệu ảnh khuôn mặt cung cấp thông tin quan trọng về các đặc điểm hình dạng và biểu hiện trên khuôn mặt của con người. Các hình ảnh trong dữ liệu bao gồm thông tin về cấu trúc hình học của khuôn mặt như vị trí và hình dạng của mắt, mũi, miệng, và các đường viền khuôn mặt. Ngoài ra, chúng cũng bao gồm các đặc trưng như màu da, cấu trúc tóc, và các yếu tố khác có thể ảnh hưởng đến diện mạo tổng thể của khuôn mặt.

Dữ liệu ảnh khuôn mặt có thể thu thập thông qua nhiều phương pháp như máy ảnh số, máy quay video hoặc thiết bị cảm biến khác. Khi thu thập dữ liệu, các yếu tố như ánh sáng, góc chụp và độ phân giải cũng được coi là quan trọng để đảm bảo chất lượng và độ tin cậy của dữ liệu. Tuy nhiên, việc làm việc với dữ liệu ảnh khuôn mặt đòi hỏi sự cẩn thận và phức tạp. Cần phải xử lý các thách thức như ánh sáng yếu, góc chụp khác nhau, biến dạng khuôn mặt và các yếu tố nhiễu khác.

A picture containing graphical user interface

Description automatically generated

Hình 1: Hình ảnh mô tả ảnh khuôn mặt được phân tích bằng các mô hình máy học để nhận diện

* 1. **Phát hiện người dựa trên dữ liệu ảnh khuôn mặt**

Phát hiện và nhận dạng khuôn mặt đã trở thành một lĩnh vực quan trọng trong việc áp dụng công nghệ vào đời sống hàng ngày. Sử dụng dữ liệu ảnh khuôn mặt, các hệ thống phân loại và nhận dạng khuôn mặt đã đạt được những tiến bộ đáng kể trong việc xác định và phân biệt người. Một trong những phương pháp phổ biến nhất để phát hiện người dựa trên dữ liệu ảnh khuôn mặt là sử dụng kỹ thuật nhận dạng khuôn mặt dựa trên máy học. Đầu tiên, một mô hình máy học được huấn luyện với một lượng lớn dữ liệu ảnh khuôn mặt, trong đó khuôn mặt của mỗi người được xác định và gắn với một nhãn duy nhất.

Quá trình huấn luyện mô hình sẽ tìm ra các đặc trưng độc đáo trong các khuôn mặt và học cách phân loại chúng. Sau đó, mô hình có thể được sử dụng để phân loại và nhận dạng khuôn mặt mới dựa trên các đặc trưng đã học. Các thuật toán nhận dạng khuôn mặt thường sử dụng các phương pháp như học sâu (deep learning), để xác định những đặc trưng quan trọng trong ảnh khuôn mặt và phân loại chúng.

Timeline

Description automatically generated

Hình 2:Các giai đoạn phát triển các thuật toán nhận diện khuôn mặt

Tuy nhiên, việc phát hiện người dựa trên dữ liệu ảnh khuôn mặt cũng đặt ra nhiều thách thức. Đôi khi, các mô hình có thể gặp khó khăn trong việc phân biệt khuôn mặt giữa các cá nhân có ngoại hình tương đồng hoặc trong các tình huống ánh sáng yếu hoặc chất lượng ảnh kém. Để đạt được kết quả tốt nhất, cần có một tập dữ liệu lớn và đa dạng để huấn luyện mô hình, cùng với việc tiến hành điều chỉnh tham số và kiểm tra độ chính xác của mô hình.

* 1. **Mục tiêu, đối tượng, và phạm vi nghiên cứu**

Như đã đề cập ở trên, mục đích của đề tài chỉ tập trung vào việc xây dựng mô hình để nhận diện khuôn mặt trên thiết bị Raspberry Pi 4 với lượng dữ liệu ảnh khuôn mặt. Đối tượng nghiên cứu bao gồm dữ liệu ảnh khuôn mặt, phương pháp để xây dựng mô hình nhận diện và cách để áp dụng vào bài toán nhận diện khuôn mặt. Phạm vi nghiên cứu chỉ tập trung vào việc tìm hiểu cách xây dựng mô hình để nhận diện khuôn mặt trên Raspberry Pi 4.

* 1. **Ý nghĩa của nghiên cứu cách xây dựng mô hình**

Kết quả của việc nghiên cứu phương pháp cách xây dựng mô hình này để phân loại ảnh khuôn mặt và nhận diện ảnh khuôn mặt. Nghiên cứu sẽ giúp cải thiện độ chính xác khi phân loại ảnh và nhận diện ảnh khuôn mặt. Tạo ra tiền đề giúp các nhà phát triển triển khai các hệ thống ứng dụng các mô hình này vào việc nhận diện khuôn mặt trên các thiết bị nhỏ gọn có hạn chế về phần cứng nhằm phục vụ cho các nhu cầu không quá lớn trong việc kiểm soát hay an ninh, giúp giảm chi phí tối ưu hơn.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

## **2.1 Raspberry Pi 4**

### **2.1.1 Giới thiệu về Raspberry Pi 4**

Raspberry Pi 4 là một vi xử lý mạnh mẽ nhưng nhỏ gọn và giá cả phải chăng, thuộc dòng sản phẩm Raspberry Pi của Raspberry Pi Foundation. Được ra mắt vào năm 2019, Raspberry Pi 4 mang đến nhiều cải tiến đáng kể so với các phiên bản trước đó, trở thành một trong những bo mạch đơn (single-board computer) phổ biến nhất trên thị trường hiện nay. Với vi xử lý Broadcom BCM2711, Raspberry Pi 4 được trang bị một bộ xử lý 4 nhân Cortex-A72 tốc độ 1.5GHz, cùng với RAM 4GB, cho phép nó xử lý các tác vụ vừa đủ không quá nặng và có hiệu quả đạt ở mức tối ưu nhất. Ngoài ra, Raspberry Pi 4 còn có khả năng giải mã video 4K và hỗ trợ đầu ra đồ họa kép. Bo mạch này cung cấp các kết nối đa dạng như 2 cổng USB 3.0, 2 cổng USB 2.0, cổng Ethernet Gigabit, khe cắm thẻ nhớ microSD và cổng HDMI để kết nối với màn hình. Ngoài ra, Raspberry Pi 4 còn có các giao tiếp không dây như Bluetooth 5.0 và Wi-Fi chuẩn 802.11ac, giúp dễ dàng kết nối với các thiết bị ngoại vi và mạng. Một điểm đáng chú ý của Raspberry Pi 4 là khả năng mở rộng thông qua các cổng GPIO (General Purpose Input/Output). Điều này cho phép người dùng kết nối và điều khiển các linh kiện ngoại vi như cảm biến, động cơ, đèn LED và nhiều hơn nữa, mở ra nhiều ứng dụng sáng tạo trong lĩnh vực IoT (Internet of Things) và nhúng.

Vì thế việc lựa chọn thiết bị này để xây dựng mô hình nhận diện khuôn mặt là một thách thức lớn với các thiết bị phần cứng không quá mạnh để xử lý các chức năng nâng cao, đề tài chỉ nghiên cứu việc xây dựng mô hình nhận diện khuôn mặt trên thiết bị này nhằm để phục vụ mục đích nghiên cứu và học tập đồng thời cũng đóng góp một phần nhỏ vào việc nghiên cứu các mô hình phù hợp ứng dụng trên nó.

A picture containing electronics, circuit

Description automatically generated

Hình 3:Bo mạch raspberry pi 4

## **2.2 Mạng nơ-ron tích chập sâu**

### **2.2.1 Giới thiệu về Convolutional Neural Network**

Trong lĩnh vực công nghệ và trí tuệ nhân tạo, mạng nơ-ron tích chập (CNN) đã trở thành một trong những đóng góp quan trọng nhất trong việc xử lý và phân tích dữ liệu hình ảnh. Ý tưởng để sinh ra CNN xuất phát từ việc nhận ra rằng các thuật toán truyền thống không thể hiệu quả trong việc xử lý các vấn đề liên quan đến hình ảnh, do tính chất đặc biệt của dữ liệu này. CNN được lấy cảm hứng từ cách mà não bộ của con người xử lý thông tin hình ảnh. Thay vì xử lý toàn bộ hình ảnh, CNN tập trung vào các đặc trưng cục bộ và chia nhỏ hình ảnh thành các khu vực nhỏ hơn gọi là "vùng quét" (local receptive fields). Điều này cho phép CNN tập trung vào các đặc trưng quan trọng trong hình ảnh, như các cạnh, góc, hoặc các hình dạng phức tạp hơn. Một trong những ý tưởng quan trọng của CNN là sử dụng các bộ lọc tích chập (convolutional filters) để trích xuất thông tin từ các vùng quét. Những bộ lọc này hoạt động như các "mắt nhìn" đặc biệt, nhận dạng các mẫu đặc trưng trong hình ảnh. Bằng cách kết hợp các bộ lọc này, CNN có khả năng học được các đặc trưng phức tạp và sử dụng chúng để phân loại, nhận dạng và xử lý ảnh. Một lợi thế quan trọng của CNN là khả năng chia sẻ trọng số (weight sharing) và gộp thông tin (pooling), giúp giảm số lượng tham số và tính toán. Việc chia sẻ trọng số cho phép CNN học được các đặc trưng phổ biến và áp dụng chúng trên các vùng quét khác nhau của hình ảnh. Quá trình gộp thông tin giúp giảm kích thước không gian đầu ra, đồng thời giữ lại thông tin quan trọng, tăng cường khả năng tổng quát hóa của mô hình. Nhờ kiến trúc CNN, khả năng học cách trích xuất các đặc trưng quan trọng từ dữ liệu ảnh đã được nâng cao đáng kể. Điều này cho phép CNN tự động học được các đặc trưng cấp thấp như cạnh, góc, hoặc màu sắc, đến các đặc trưng cấp cao như hình dạng, đối tượng và ngữ cảnh. Điểm mạnh của CNN là khả năng tự động học và tìm ra các đặc trưng quan trọng mà không cần sự can thiệp thủ công từ con người. Với những thành tựu đáng kể của CNN, công nghệ này đã có ảnh hưởng lớn đến nhiều lĩnh vực như nhận dạng khuôn mặt, phân loại ảnh, xe tự hành và nhiều ứng dụng trong đời sống.

Cấu trúc của một CNN bao gồm một loạt các lớp khác nhau, được xếp chồng lên nhau để tạo thành một kiến trúc chồng chất. mỗi lớp có nhiệm vụ thực hiện một phần công việc đặc biệt trong việc trích xuất đặc trưng từ dữ liệu đầu vào.

****

Hình 4:Cấu trúc cơ bản của mạng nơ-ron tích chập

Lớp đầu vào (Input Layer) có nhiệm vụ nhận dữ liệu hình ảnh ban đầu và chuyển đổi chúng thành định dạng phù hợp để tiép tục xử lý. Lớp tích chập (Convolutional Layer) sẽ áp dụng các bộ lọc tích chập lên đầu vào để nhận diện các đặc trưng cục bộ. Mỗi bộ lọc sẽ tạo ra một bản đồ đặc trưng, biểu thị sự xuất hiện của các đối tượng cụ thể trong hình ảnh. Lớp gộp (Pooling Layer) sẽ giảm kích thước của bản đồ đặc trưng bằng cách thực hiện phép gộp (pooling) trên các vùng không gian. Phép gộp giúp giảm số lượng tham số và tính toán, đồng thời tạo ra tính bất biến với việc dịch chuyển nhỏ của các đối tượng trong hình ảnh. Lớp kích hoạt sẽ áp dụng một hàm kích hoạt phi tuyến tính lên các đầu ra từ lớp tích chập hoặc lớp gộp. Hàm kích hoạt giúp mô hình học các đặc trưng phi tuyến tính và tạo ra đầu ra phi tuyến tính từ các đầu vào tuyến tính. Lớp kết nối đầy đủ (Fully Connected Layer), đây là lớp cuối cùng sẽ kết nối với các đầu ra từ lớp trước và tạo ra đầu ra cuối cùng. Lớp này thường sử dụng một hoặc nhiều lớp ẩn (hidden layers) để học các biểu diễn phức tạp và thực hiện phân loại.

### **2.2.2 Ý nghĩa các lớp cấu trúc của CNN**

**Lớp tích chập (Convolution)**

Lớp tích chập (convolutional layer) đóng vai trò quan trọng trong cấu trúc Mạng nơ-ron tích chập (Convolutional Neural Network - CNN) và mang lại nhiều ý nghĩa quan trọng cho việc xử lý hình ảnh và dữ liệu không gian. Ý nghĩa chính của lớp tích chập là khả năng tìm ra các đặc trưng cục bộ và tổng hợp thông tin từ dữ liệu đầu vào. Bằng cách sử dụng các bộ lọc (filters) nhỏ có kích thước nhỏ và áp dụng phép tích chập lên các phần nhỏ của hình ảnh hoặc đầu vào, lớp tích chập giúp nhận diện các đặc trưng như cạnh, góc, đường cong, hoặc các đặc trưng phức tạp hơn như mắt, mũi, miệng trong hình ảnh. Lớp tích chập cũng đóng vai trò quan trọng trong việc giảm số lượng tham số cần học trong mạng nơ-ron. Bằng cách chia sẻ trọng số giữa các vị trí đầu vào khác nhau, lớp tích chập giúp giảm đáng kể số lượng tham số so với mạng nơ-ron thông thường, đồng thời vẫn giữ được khả năng mô hình hóa các đặc trưng không gian. Một ưu điểm khác của lớp tích chập là khả năng bảo toàn thông tin về vị trí không gian của đối tượng trong dữ liệu đầu vào. Bằng cách sử dụng các bước trượt (stride) và lớp kích thước cửa sổ (window size), lớp tích chập giúp theo dõi và bảo toàn thông tin về vị trí của đối tượng trong các tầng sau của mạng nơ-ron, tạo điều kiện thuận lợi cho việc nhận diện và phân loại đối tượng.

**Lớp phi tuyến ReLU (Rectified linear unit)**

Lớp phi tuyến ReLU (Rectified Linear Unit) đóng một vai trò quan trọng trong cấu trúc mạng nơ-ron tích chập (CNN) và mang lại nhiều ý nghĩa quan trọng. Với đặc điểm phi tuyến tính, lớp ReLU giúp mạng nơ-ron học được những đặc trưng phức tạp và đồng thời giảm thiểu sự triệt tiêu thông tin trong quá trình truyền ngược. Ý nghĩa đầu tiên của lớp phi tuyến ReLU là nó tạo điều kiện để mạng nơ-ron tích chập học được các hàm phi tuyến, đồng thời vẫn giữ được tính toán hiệu quả. Với đầu vào âm, ReLU trả về giá trị 0, điều này giúp loại bỏ sự ảnh hưởng của các đặc trưng không cần thiết hoặc nhiễu trong hình ảnh. Đồng thời, ReLU giữ nguyên giá trị đầu vào dương, tạo điều kiện cho các đặc trưng quan trọng được tăng cường và truyền ngược hiệu quả trong quá trình huấn luyện mạng.

Một ý nghĩa khác của lớp ReLU là khả năng xử lý mô hình tuyến tính và phi tuyến tính đồng thời. Khi áp dụng ReLU trên các kết quả tích chập, mạng nơ-ron có khả năng học các đặc trưng tuyến tính và phi tuyến tính một cách hiệu quả. Điều này cho phép mạng tổng hợp các đặc trưng phức tạp và non-linear từ các kết quả trước đó, làm tăng khả năng mô hình hoạt động trên dữ liệu phức tạp và đa dạng.

Cuối cùng, lớp phi tuyến ReLU cũng giúp giảm thiểu hiện tượng vanishing gradient. Trong quá trình lan truyền ngược, gradient có thể giảm dần đáng kể khi tiến về các lớp đầu của mạng nơ-ron sâu. Tuy nhiên, ReLU giữ nguyên giá trị đầu vào dương, đồng nghĩa với việc gradient không bị triệt tiêu khi lan truyền ngược qua các lớp ReLU. Điều này cải thiện khả năng học của mạng nơ-ron và giúp nâng cao hiệu suất huấn luyện.

**Lớp Pooling**

Lớp Pooling là một phần quan trọng trong cấu trúc Convolutional Neural Network (CNN) với vai trò quan trọng trong việc giảm kích thước không gian của đầu vào. Ý nghĩa của lớp Pooling không chỉ giới hạn ở việc giảm kích thước, mà còn mang lại nhiều lợi ích khác. Trong quá trình huấn luyện mạng CNN, lớp Pooling được sử dụng để giảm số lượng tham số tính toán, tăng tốc độ xử lý và tránh hiện tượng overfitting. Bằng cách áp dụng một phép tổng hợp hoặc phép lấy mẫu (sampling) trên các khu vực không gian của đầu vào, lớp Pooling giúp giảm kích thước không gian mà vẫn giữ lại thông tin quan trọng.

Ý tưởng chính của lớp Pooling là tạo ra các biểu diễn tổng quát và bất biến (invariant) đối với các biến đổi không gian như dịch chuyển hay biến dạng nhỏ. Điều này giúp mạng CNN trở nên bền vững và đáng tin cậy hơn khi xử lý ảnh, vì các đặc trưng quan trọng như cạnh, hình dạng hay mẫu cơ bản được tìm thấy ở các vùng khác nhau của ảnh, không chỉ tập trung ở một vị trí cụ thể.

Lớp Pooling có thể thực hiện các phép tổng hợp như Max Pooling hoặc Average Pooling. Trong Max Pooling, giá trị lớn nhất trong mỗi vùng pooling sẽ được chọn để đại diện cho vùng đó. Điều này giúp giữ lại thông tin quan trọng về đặc trưng cục bộ, như đường viền hay góc cạnh. Trong khi đó, Average Pooling tính trung bình các giá trị trong vùng pooling, tạo ra một biểu diễn tổng quát hơn về thông tin đặc trưng.

**Lớp kết nối đầy đủ (Fully connected layer)**

Lớp kết nối đầy đủ (fully connected layer) là một phần quan trọng trong cấu trúc của mạng nơ-ron tích chập (CNN). Ý nghĩa của lớp kết nối đầy đủ là tạo ra một liên kết giữa các đặc trưng cục bộ đã được trích xuất từ các lớp tích chập trước đó và đầu ra cuối cùng của mạng. Khi thông qua các lớp tích chập, mạng CNN có khả năng tìm ra các đặc trưng cục bộ trong hình ảnh, chẳng hạn như các cạnh, góc, hoặc (các thành phần kết cấu) textural patterns. Tuy nhiên, để thực hiện các tác vụ như phân loại hay phát hiện đối tượng, chúng ta cần một cách để kết hợp các đặc trưng này và học được mối quan hệ toàn cục giữa chúng. Đó là lúc lớp kết nối đầy đủ trở nên quan trọng.

Lớp kết nối đầy đủ hoạt động như một bộ phân loại tuyến tính, nơi mỗi nút (hoặc neuron) trong lớp này kết nối với tất cả các nút trong lớp trước đó. Việc kết nối này cho phép thông tin được truyền từ các đặc trưng cục bộ đến đầu ra cuối cùng. Nó giúp mô hình học được mối quan hệ phức tạp và trừu tượng giữa các đặc trưng, tạo ra một biểu diễn tổng thể cho hình ảnh đầu vào.

Điều quan trọng là lớp kết nối đầy đủ tạo ra khả năng học không gian đặc trưng phi tuyến tính, cho phép mô hình hiểu được sự phức tạp của dữ liệu hình ảnh. Bằng cách học các trọng số trong lớp này, mạng CNN có thể học các khái niệm trừu tượng và phức tạp như khuôn mặt, đối tượng, hoặc phong cách.

**Kết quả và đánh giá về mô hình**

Để đánh giá hiệu suất và kết quả của mô hình học máy, chúng ta cần sử dụng các độ đo phù hợp, một số độ đo được dùng.

* **Accuracy (Độ chính xác):** Đây là độ đo cơ bản nhất để đánh giá mô hình. Nó tính tỷ lệ phần trăm dự đoán chính xác trên tổng số mẫu dữ liệu.
* **Precision (Độ chính xác dương tính):** Độ đo này tập trung vào khả anngw của mô hình xác định đúng các dự đoán dương tính (positive). Nó tính tỷ lệ giữa số lượng dự đoán đúng dương tính và tổng số dự đoán dương tính.
* **Recall (Độ phục hồi):** Độ đo này đánh giá khả năng của mô hình tìm ra tất cả các trường hợp dương tính trong tập dữ liệu. Nó tính tỷ lệ giữa số lượng dự đoán đúng dương tính và tổng số thực tế dương tính.
* **F1 Score:** Độ đo kết hợp giữa Precision và Recall. Nó tính trung bình điều hoà của hai giá trị này và cung cấp một con số tổng quan về hiệu suất của mô hình.

### **2.2.3 Một số kỹ thuật được sử dụng để huấn luyện mô hình có hiệu quả**

**Dropout**

Kỹ thuật Dropout là một phương pháp hiệu quả để chống overfitting trong mạng nơ-ron. Nó hoạt động bằng cách tạm thời loại bỏ ngẫu nhiên một số lượng đơn vị đầu ra trong quá trình huấn luyện. Điều này làm giảm sự phụ thuộc quá mức vào các đặc trưng cụ thể và tạo ra một mạng nơ-ron tổng quát hơn. Dropout cũng thúc đẩy sự học tập độc lập giữa các đơn vị, dẫn đến khả năng tổng hợp thông tin tốt hơn. Bằng cách này, Dropout cải thiện tính tổng quát hóa và giúp mạng nơ-ron đạt được kết quả tốt hơn trên dữ liệu mới.

**Data augmentation**

Data augmentation là một kỹ thuật quan trọng trong việc mở rộng tập dữ liệu huấn luyện. Nó bao gồm việc tạo ra các phiên bản mới từ dữ liệu ban đầu bằng cách áp dụng các phép biến đổi nhỏ. Các biến đổi này có thể bao gồm xoay, phóng to/thu nhỏ, lật ngang/đứng và áp dụng nhiễu. Kỹ thuật này giúp cải thiện khả năng tổng quát hóa của mô hình máy học bằng cách giới thiệu sự đa dạng vào tập dữ liệu huấn luyện. Điều này không chỉ giúp giảm overfitting mà còn tăng tính linh hoạt và khả năng phát hiện đặc trưng của mô hình.

**Transfer learning**

Transfer learning là một kỹ thuật trong machine learning mà mô hình đã được huấn luyện trên một tác vụ có thể được tái sử dụng để giải quyết các tác vụ khác. Thay vì huấn luyện mô hình từ đầu, chúng ta chỉ cần điều chỉnh và tinh chỉnh lại mô hình đã tồn tại để áp dụng cho các tác vụ mới. Việc này giúp tiết kiệm thời gian và nguồn lực đào tạo. Bằng cách sử dụng kiến thức đã học từ các tác vụ trước, transfer learning giúp tăng hiệu suất và độ chính xác của mô hình trên các tác vụ mới.

**Optimizer**

Optimizer là một kỹ thuật quan trọng trong machine learning, giúp tối ưu hóa quá trình huấn luyện mô hình. Nhiệm vụ của Optimizer là điều chỉnh các tham số trong mô hình để tìm ra bộ trọng số tối ưu, dựa trên độ lỗi và hàm mất mát. Điều này được thực hiện bằng cách sử dụng các thuật toán như Gradient Descent và các biến thể của nó. Optimizer giúp tối ưu hóa quá trình huấn luyện, tăng cường hiệu suất và đảm bảo rằng mô hình hội tụ nhanh chóng và đạt được kết quả tốt nhất.

### **2.3 Các mô hình học máy được ứng dụng trong đề tài để tìm ra độ chính xác cao trên tập dữ liệu**

### **2.3.1 Mô hình VGGNet**

VGGNet, viết tắt của Visual Geometry Group Network, là một kiến trúc mạng nơ-ron tích chập (CNN) được phát triển bởi nhóm nghiên cứu Visual Geometry Group tại Đại học Oxford. Với sự đột phá của VGGNet, kiến trúc này đã chứng minh khả năng xuất sắc trong việc nhận dạng hình ảnh.

VGGNet nổi tiếng với độ sâu của mình, với các phiên bản khác nhau gồm VGG16 và VGG19, chỉ số này biểu thị số lượng lớp tích chập và lớp kết nối đầy đủ trong mạng. Với kiến trúc đồ sộ này, VGGNet đạt được hiệu suất cao và đáng tin cậy trong việc phân loại hình ảnh. Một điểm đặc biệt của VGGNet là sự đồng nhất trong cấu trúc: tất cả các lớp tích chập đều sử dụng các bộ lọc nhỏ kích thước 3x3 với bước nhảy là 1. Điều này cho phép VGGNet tập trung vào việc học các đặc trưng nhỏ và xây dựng các mô hình phân loại chính xác.

Với tầm ảnh hưởng và thành công của mình, VGGNet đã trở thành một trong những kiến trúc mạng quan trọng trong lĩnh vực thị giác máy tính. Sự đơn giản, hiệu quả và khả năng tổng quát hóa của VGGNet đã mở ra những cánh cửa mới trong việc áp dụng học sâu cho các bài toán phức tạp liên quan đến xử lý hình ảnh.

Diagram

Description automatically generated

Hình 5:Kiến trúc mạng VGG

Mạng VGG16 là một kiến trúc mạng nơ-ron sâu nổi tiếng và phổ biến trong lĩnh vực nhận dạng hình ảnh. Được đề xuất bởi nhóm nghiên cứu tại Đại học Oxford, VGG16 đạt được kết quả ấn tượng trong cuộc thi ImageNet năm 2014. Kiến trúc của VGG16 được xây dựng dựa trên việc sử dụng các lớp tích chập với kích thước nhỏ 3x3 liên tiếp nhau. Mạng này gồm 16 lớp tích chập, bao gồm các lớp Convolutional Layer và các lớp Pooling Layer. Với cấu trúc này, VGG16 có khả năng học các đặc trưng cấp thấp (như cạnh và góc) cũng như các đặc trưng cấp cao (như hình dạng và đối tượng).

Một điểm đáng chú ý của VGG16 là sự đơn giản và dễ hiểu của kiến trúc. Điều này đã giúp nó trở thành một công cụ phổ biến cho việc tiền xử lý dữ liệu hình ảnh trong các nhiệm vụ như phân loại hình ảnh, nhận dạng đối tượng và phát hiện vị trí. Với hiệu suất cao và tính ổn định, VGG16 đã trở thành một trong những mạng nơ-ron sâu được ưa chuộng trong cộng đồng nghiên cứu và ứng dụng. Kiến trúc đơn giản và khả năng học các đặc trưng phức tạp đã giúp VGG16 đóng góp quan trọng vào sự phát triển của lĩnh vực học sâu và nhận dạng hình ảnh.

Diagram

Description automatically generated

Hình 6:Kiến trúc mô hình VGG16

VGG19, viết tắt của Visual Geometry Group 19, là một kiến trúc mạng nơ-ron tích chập sâu được phát triển bởi nhóm nghiên cứu Visual Geometry Group tại Đại học Oxford. Với thiết kế mạnh mẽ và độ phức tạp cao, VGG19 là một trong những mô hình nổi tiếng và ảnh hưởng nhất trong lĩnh vực thị giác máy tính.

Kiến trúc của VGG19 bao gồm 19 lớp, trong đó có 16 lớp tích chập và 3 lớp kết nối đầy đủ. Điều đặc biệt ở VGG19 là việc sử dụng các bộ lọc nhỏ kích thước 3x3 liên tiếp nhiều lần trong quá trình tích chập, tạo ra các đặc trưng phức tạp và cung cấp khả năng học tập sâu.

Mạng VGG19 đã chứng minh sức mạnh của nó trong nhiều nhiệm vụ thị giác máy tính, bao gồm phân loại hình ảnh, nhận dạng đối tượng và trích xuất đặc trưng. Nhờ cấu trúc đơn giản và dễ hiểu, VGG19 đã trở thành một công cụ quan trọng trong việc nghiên cứu và phát triển các mô hình mạng nơ-ron tích chập. Tính đến nay, VGG19 vẫn là một trong những lựa chọn phổ biến để làm cơ sở cho việc xây dựng các mạng nơ-ron sâu mạnh mẽ. Với khả năng mô hình hóa các đặc trưng phức tạp và khả năng tổng quát hóa cao, VGG19 tiếp tục đóng vai trò quan trọng trong việc đưa thị giác máy tính đến một tầm cao mới.

A picture containing chart

Description automatically generated

Hình 7:Kiến trúc mô hình VGG19

Dưới đây đánh giá của 2 mô hình trên:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mô hình | Kích thước | Top 1 độ chính xác | Top 5 độ chính xác | Số lượng tham số | Độ sâu |
| VGG16 | 528MB | 71.3% | 90.1% | 138.4M | 16 |
| VGG19 | 549MB | 71.3% | 90.0% | 143.7M | 19 |

Bảng 1:Độ chính xác và các tham số của kiến trúc mạng VGG16 và VGG 19

### **2.3.2 Mô hình MobileNet**

**MobileNet** là một mô hình học sâu (deep learning) được phát triển nhằm đạt được cân bằng giữa hiệu suất và kích thước của mạng. Được giới thiệu bởi Google vào năm 2017, MobileNet đã trở thành một công cụ quan trọng trong lĩnh vực nhận dạng hình ảnh và xử lý ảnh trên các thiết bị di động và nhúng.

MobileNet sử dụng kiến trúc chính là mạng neural tích chập (convolutional neural network) với các đặc điểm đặc biệt để giảm kích thước của mô hình. Phương pháp này được gọi là depthwise separable convolution, tức là sự tách biệt theo chiều sâu và theo chiều không gian của ma trận đặc trưng. Thay vì áp dụng một lớp tích chập đầy đủ trực tiếp lên toàn bộ đặc trưng, MobileNet chia thành hai giai đoạn: tích chập theo chiều không gian (spatial convolution) và tích chập theo chiều sâu (depthwise convolution). Điều này giúp giảm số lượng tham số và thời gian tính toán mà vẫn giữ được hiệu suất tương đương. Một điểm đặc biệt khác của MobileNet là khả năng điều chỉnh kích thước mạng bằng cách sử dụng tham số alpha. Tham số alpha điều chỉnh số lượng kênh (channels) trong mô hình và ảnh hưởng đến cả hiệu suất và kích thước của mạng. Bằng cách điều chỉnh alpha, người dùng có thể tùy chỉnh mô hình MobileNet để đáp ứng yêu cầu về hiệu suất và kích thước trong các ứng dụng cụ thể.

MobileNet đã được sử dụng rộng rãi trong các ứng dụng di động, nhúng và thị giác máy tính. Đối với các thiết bị có tài nguyên hạn chế và yêu cầu tính toán thấp, MobileNet mang lại lợi ích về kích thước nhỏ và hiệu suất cao. Nó thích hợp cho các nhiệm vụ như nhận dạng đối tượng, phân loại hình ảnh và xử lý ảnh thời gian thực trên các nền tảng di động.

Diagram, engineering drawing

Description automatically generated

Hình 8: Kiến trúc mô hình MobileNet

**MobileNetV2** là một mô hình học sâu (deep learning) nổi tiếng và được sử dụng rộng rãi trong lĩnh vực thị giác máy tính. Được giới thiệu bởi Google vào năm 2018, MobileNetV2 là phiên bản cải tiến của MobileNetV1, với mục tiêu cung cấp một mô hình nhẹ nhàng và hiệu quả hơn cho các ứng dụng di động và nhúng. Nó được thiết kế để giảm kích thước mô hình và tăng tốc độ xử lý mà vẫn duy trì độ chính xác cao. Mô hình này sử dụng một loạt các khối học sâu gọi là Inverted Residuals và tận dụng kỹ thuật xử lý dữ liệu như bottleneck layers và linear bottlenecks. Điều này giúp mô hình trở nên hiệu quả hơn bằng cách giảm số lượng tham số và phép tính so với các mô hình truyền thống. Nó có khả năng học các đặc trưng phức tạp của hình ảnh và thực hiện các nhiệm vụ như phân loại hình ảnh, phát hiện đối tượng và tinh chỉnh (fine-tuning). Nó cho phép tùy chỉnh các đặc trưng và hiệu chỉnh mô hình theo nhu cầu cụ thể của ứng dụng. Một trong những ưu điểm của MobileNetV2 là tính di động cao. Nhờ kích thước nhỏ gọn và cấu trúc tối ưu, mô hình này rất thích hợp cho các ứng dụng nhúng và thiết bị có tài nguyên hạn chế như điện thoại di động, máy tính bảng và IoT (Internet of Things).

MobileNetV2 đã được áp dụng thành công trong nhiều lĩnh vực, bao gồm nhận dạng hình ảnh, phân loại sản phẩm, nhận diện khuôn mặt, nhận diện đối tượng và thị giác máy tính nâng cao. Với hiệu suất tốt và khả năng tích hợp dễ dàng vào các ứng dụng di động, MobileNetV2 trở thành một lựa chọn hàng đầu cho các nhà phát triển và nghiên cứu trong lĩnh vực thị giác máy tính.

Diagram

Description automatically generated

Hình 9: Cấu trúc mô hình MobileNetV2

Dưới đây đánh giá của 2 mô hình trên:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mô hình | Kích thước | Top 1 độ chính xác | Top 5 độ chính xác | Số lượng tham số | Độ sâu |
| MobileNet | 16MB | 70.4% | 89.5% | 4.3M | 55 |
| MobileNetV2 | 14MB | 71.3% | 90.1% | 3.5M | 105 |

Bảng 2:Độ chính xác và các tham số của kiến trúc mạng MobileNet và MobileNetV2

### **2.4 Kỹ thuật Transfer Learning**

### **2.4.1 Một số vấn đề và hạn chế của các mô hình học sâu trên các tập dữ liệu thực tế**

Thực tế, đối với các tập dữ liệu thực các mô hình thường dự báo không chính xác cho dù đã áp dụng những mô hình máy học có kết quả cao trên các tập dữ liệu mẫu. Các nguyên nhân thường gặp đối với các tập dữ liệu này có thể được xác đinh sau.

* **Hiện tượng overfitting:** Với tập dữ liệu nhỏ, mô hình học sâu dễ bị overfitting, tức là mô hình học cách nhận dạng và tái tạo dữ liệu huấn luyện một cách chính xác, nhưng không tổng quát hóa tốt cho các dữ liệu mới.
* **Thiếu dữ liệu đại diện:** Các mô hình học sâu yêu cầu một lượng lớn dữ liệu huấn luyện đại diện cho các biến thể của dữ liệu thực tế. Nếu tập dữ liệu không đủ đại diện, mô hình có thể không hiểu được các đặc trưng quan trọng và không đưa ra các dự đoán chính xác trên dữ liệu mới.
* **Vấn đề về hiệu suất tính toán:** Huấn luyện các mô hình học sâu trên tập dữ liệu lớn và phức tạp đòi hỏi sự tính toán mạnh mẽ và tài nguyên phần cứng cao. Quá trình huấn luyện có thể mất rất nhiều thời gian và yêu cầu cấu hình phần cứng phù hợp.
* **Vấn đề tối ưu hóa:** Tìm ra các tham số tối ưu cho mô hình học sâu là một nhiệm vụ khó khăn. Tập dữ liệu lớn và đa dạng đòi hỏi các thuật toán tối ưu hóa phải đủ mạnh mẽ để hội tụ và tránh rơi vào các điểm cực tiểu cục bộ.
* **Dữ liệu không tổng quát:** Trong tập dữ liệu thực tế, có thể tồn tại sự không tổng quát các ảnh khuôn mặt ở các điều kiện khác nhau, tập dữ liệu chỉ mang tính nghiên cứu chỉ được thực hiện tại một thời điểm nhất định, tập dữ liệu tổng quát cần có sự lấy mẫu ở các điều kiện khác nhau chẳng hạn như: ngày, đêm, thời gian, đối tượng có các trang bị trên khuôn mặt hoặc không, sự thay đổi của khuôn mặt nhất trong cách mà dữ liệu được thu thập, đánh nhãn hoặc chất lượng của dữ liệu. Điều này có thể ảnh hưởng đến hiệu suất và khả năng tổng quát hóa của mô hình.

### **2.4.2 Một số định nghĩa trong CNN**

**Pre-trained model** là một khái niệm trong machine learning và deep learning, và nó chặt chẽ liên quan đến transfer learning. Pre-trained model là một mô hình đã được huấn luyện trước độc lập trên một tác vụ tương tự hoặc liên quan. Thông thường, mô hình này được huấn luyện trên một tập dữ liệu lớn và phức tạp, chẳng hạn như ImageNet dataset trong trường hợp của các mô hình học sâu dùng cho xử lý hình ảnh.

**Transfer learning** là một kỹ thuật trong machine learning cho phép sử dụng kiến thức đã học từ một tác vụ để cải thiện hiệu suất của mô hình trên một tác vụ khác. Thay vì huấn luyện mô hình từ đầu trên tập dữ liệu mới, ta có thể sử dụng pre-trained model và tinh chỉnh (fine-tune) nó trên tập dữ liệu mới để đạt được hiệu suất tốt hơn và tiết kiệm thời gian huấn luyện.

Sự gắn kết giữa pre-trained model và transfer learning đến từ việc pre-trained model chứa kiến thức rộng và phổ quát về tác vụ mà nó đã được huấn luyện. Thông qua transfer learning, chúng ta có thể áp dụng những kiến thức này vào tác vụ mới mà chúng ta muốn giải quyết. Thay vì bắt đầu từ đầu và xây dựng một mô hình hoàn toàn mới, chúng ta có thể sử dụng pre-trained model như một "sự khởi đầu" và tiếp tục huấn luyện hoặc tinh chỉnh nó trên tập dữ liệu mới để tận dụng kiến thức đã có sẵn.

Sự kết hợp giữa pre-trained model và transfer learning mang lại nhiều lợi ích. Thứ nhất, nó giúp giảm đáng kể thời gian và công sức huấn luyện, đặc biệt là khi chúng ta chỉ có ít dữ liệu huấn luyện. Thứ hai, pre-trained model chứa những kiến thức tổng quát và trừu tượng từ tác vụ ban đầu, giúp cải thiện khả năng tổng quát hóa của mô hình trên tác vụ mới. Thứ ba, pre-trained model thường được huấn luyện trên phần lớn dữ liệu, giúp mô hình học các đặc trưng phức tạp và biểu diễn dữ liệu một cách hiệu quả.

Diagram, schematic

Description automatically generated

Hình 10:Cách tạo model mới từ một pre-trained model

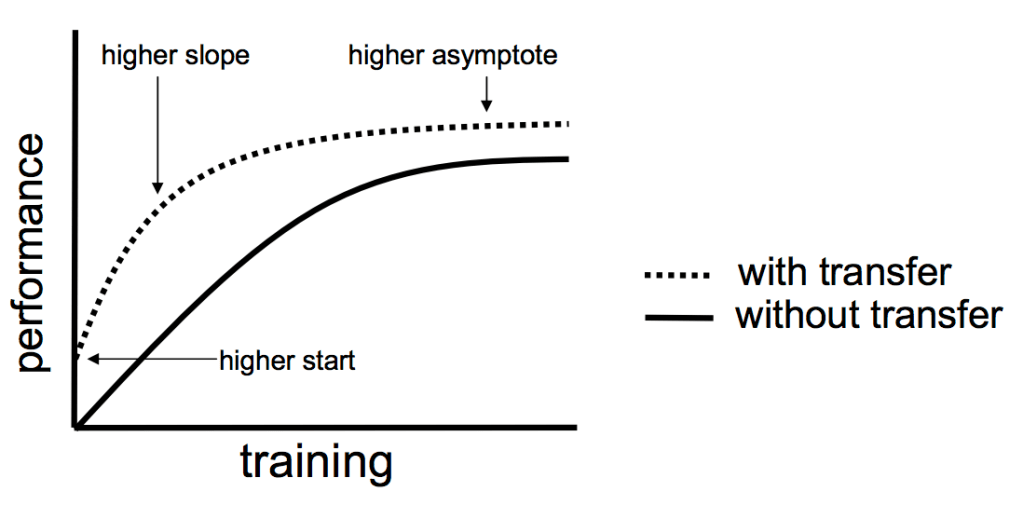
### **2.4.3 Vai trò của Transfer learning**

Transfer learning là một phương pháp trong lĩnh vực học máy, trong đó kiến thức được chuyển giao từ một tác vụ hoặc một miền dữ liệu cho tác vụ hoặc miền dữ liệu khác. Với vai trò quan trọng, transfer learning giúp chúng ta tiết kiệm thời gian và công sức trong việc huấn luyện mô hình mới từ đầu.

Với transfer learning, chúng ta có thể tận dụng kiến thức đã học được từ một mô hình có hiệu suất cao trong một nhiệm vụ tương tự để giúp cải thiện kết quả của một nhiệm vụ khác. Thay vì bắt đầu từ con số không, chúng ta sử dụng mô hình đã được huấn luyện trước đó như một bộ trích xuất đặc trưng sẵn có. Nhờ vậy, chúng ta có thể tận dụng kiến thức tổng quát được học từ các tác vụ trước đó và áp dụng chúng vào bài toán mới.

Với việc sử dụng transfer learning, chúng ta thường đạt được hai lợi ích quan trọng: tiết kiệm thời gian huấn luyện và cải thiện độ chính xác của mô hình. Thay vì phải xây dựng và huấn luyện một mô hình từ đầu, chúng ta chỉ cần điều chỉnh một số phần nhỏ của mô hình đã có sẵn để phù hợp với nhiệm vụ mới. Điều này giúp chúng ta tiết kiệm đáng kể thời gian và công sức. Ngoài ra, transfer learning cũng cho phép chúng ta khai thác kiến thức và thông tin quý giá từ tác vụ trước đó, làm tăng khả năng tổng quát hóa và cải thiện hiệu suất của mô hình trong tác vụ mới.

Tóm lại, transfer learning đóng vai trò quan trọng trong lĩnh vực học máy, cho phép chúng ta tận dụng kiến thức đã học được từ các tác vụ trước đó để cải thiện hiệu suất và tiết kiệm thời gian huấn luyện của mô hình. Việc áp dụng transfer learning mang lại lợi ích lớn cho việc xây dựng các mô hình học máy trong các ngữ cảnh thực tế, đồng thời mở ra nhiều tiềm năng nghiên cứu và ứng dụng trong tương lai.



Hình 11:Kết quả của việc sử dụng Transfer Learning trước và sau khi áp dụng vào mô hình

### **2.4.5 Hạn chế của Transfer learning**

Transfer learning có những hạn chế cần được xem xét. Đầu tiên, khả năng áp dụng transfer learning phụ thuộc vào sự tương đồng giữa tác vụ nguồn và tác vụ đích. Khi hai tác vụ khác nhau quá nhiều, transfer learning có thể không mang lại kết quả tốt. Thứ hai, kích thước dữ liệu huấn luyện cũng quan trọng. Trong trường hợp dữ liệu huấn luyện đích hạn chế, transfer learning có thể gặp khó khăn trong việc học các đặc trưng của tác vụ đích. Cuối cùng, transfer learning cần xử lý vấn đề "catastrophic forgetting" khi nhiều tác vụ được thực hiện tuần tự.

## **2.5 Máy học véc-tơ hỗ trợ (SVM)**

### **2.5.1 Giới thiệu về Support Vector Machine - SVM**

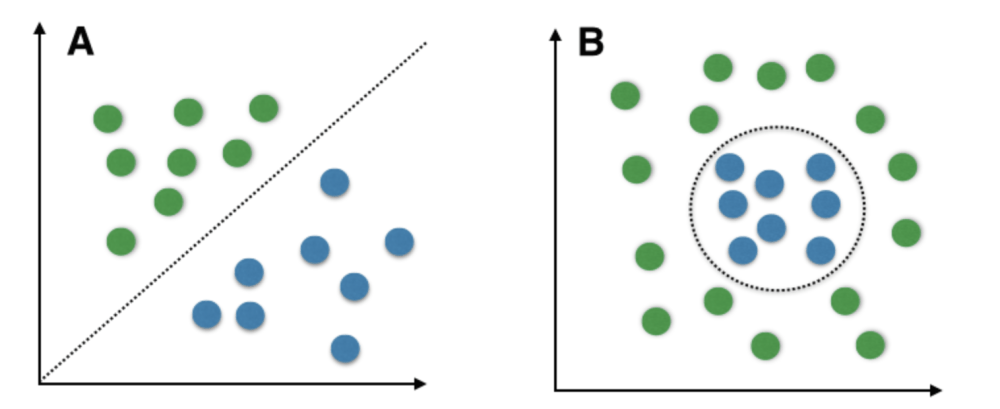
Máy vector hỗ trợ (Support Vector Machines - SVMs) là một loại thuật toán học có giám sát được sử dụng để phân loại hoặc dự đoán giá trị (regression). Ý tưởng chính đằng sau SVM là tìm một siêu phẳng (hyperplane) tách biệt tối đa các lớp khác nhau trong dữ liệu huấn luyện. Điều này được thực hiện bằng cách tìm siêu phẳng có ranh giới lớn nhất, được định nghĩa là khoảng cách giữa siêu phẳng và các điểm dữ liệu gần nhất từ mỗi lớp. Khi siêu phẳng được xác định, dữ liệu mới có thể được phân loại bằng cách xác định nằm ở phía nào của siêu phẳng. SVMs đặc biệt hữu ích khi dữ liệu có nhiều đặc trưng và/hoặc khi có một ranh giới rõ ràng trong dữ liệu.

Máy vector hỗ trợ (Support Vector Machine - SVM) là một thuật toán học máy giám sát đơn giản được sử dụng cho phân loại và/hoặc hồi quy. Nó thích hợp hơn cho việc phân loại, nhưng đôi khi cũng rất hữu ích cho việc hồi quy. Đơn giản, SVM tìm ra một siêu phẳng tạo ra ranh giới giữa các loại dữ liệu. Trong không gian hai chiều, siêu phẳng này chính là một đường thẳng.

Trong SVM, chúng ta đặt mỗi mục dữ liệu trong tập dữ liệu vào một không gian N-chiều, trong đó N là số lượng đặc trưng/đặc điểm trong dữ liệu. Tiếp theo, tìm ra siêu phẳng tối ưu để phân chia dữ liệu. Do đó, bạn đã hiểu rằng theo bản chất, SVM chỉ có thể thực hiện phân loại nhị phân (tức là lựa chọn giữa hai lớp). Tuy nhiên, có các kỹ thuật khác nhau để sử dụng cho các vấn đề đa lớp.

Để thực hiện SVM trên các vấn đề đa lớp, chúng ta có thể tạo một bộ phân loại nhị phân cho mỗi lớp dữ liệu. Hai kết quả của mỗi bộ phân loại sẽ là:

* Điểm dữ liệu thuộc lớp đó
* Điểm dữ liệu không thuộc lớp đó



Hình 12:Hình ảnh mô tả dữ liệu tuyến tính và dữ liệu phi tuyến tính

Chúng ta sử dụng SVM với Kernel cho dữ liệu không tách tuyến tính. Ví dụ, giả sử chúng ta có một tập dữ liệu không tách tuyến tính trong một chiều. Chúng ta có thể biến đổi tập dữ liệu này thành hai chiều và dữ liệu sẽ trở thành tách tuyến tính trong hai chiều. Việc này được thực hiện bằng cách ánh xạ mỗi điểm dữ liệu 1D thành một cặp 2D tương ứng. Vì vậy, với bất kỳ dữ liệu không tách tuyến tính nào trong bất kỳ chiều nào, chúng ta chỉ cần ánh xạ dữ liệu vào một số chiều cao hơn và sau đó biến nó thành tách tuyến tính. Đây là một phép biến đổi rất mạnh mẽ và tổng quát. Một kernel không gì khác là một đo lường sự tương đồng giữa các điểm dữ liệu. Hàm kernel trong SVM với Kernel cho biết, với hai điểm dữ liệu trong không gian đặc trưng ban đầu, độ tương đồng giữa các điểm trong không gian đặc trưng đã được biến đổi. Có nhiều hàm kernel khác nhau có sẵn, nhưng hai trong số đó rất phổ biến:

* **Hàm kernel hạt nhân Radial Basis Function (RBF):** Sự tương đồng giữa hai điểm trong không gian đặc trưng đã được chuyển đổi là một hàm giảm mạnh theo mức độ cách xa giữa các vector và không gian đầu vào ban đầu như được hiển thị dưới đây. RBF là kernel mặc định được sử dụng trong SVM.
* **Hạt nhân đa thức (Polynomial Kernel):** Hạt nhân đa thức sử dụng một tham số bổ sung là 'degree' để điều khiển mức độ phức tạp và chi phí tính toán của quá trình biến đổi của mô hình.

Một sự thật rất thú vị là SVM thực sự không cần thực hiện quá trình biến đổi thực tế này trên các điểm dữ liệu để chuyển sang không gian đặc trưng mới có số chiều cao hơn. Điều này được gọi là kernel trick. Kernel Trick: Bên trong, SVM với kernel có thể tính toán những biến đổi phức tạp này chỉ dựa trên tính toán độ tương đồng giữa các cặp điểm trong không gian đặc trưng có số chiều cao hơn, nơi biểu diễn đặc trưng đã được biến đổi là mặc định. Hàm tương đồng này, toán học là một loại nhân phức tạp, thực chất là kernel của SVM với kernel. Điều này làm cho việc áp dụng SVM trở nên khả thi khi không gian đặc trưng cơ bản là phức tạp hoặc thậm chí vô hạn chiều.

Các tham số quan trọng trong SVC với kernel (Support Vector Classifier):

* **Kernel:** Kernel, hay còn được gọi là hạt nhân, được lựa chọn dựa trên loại dữ liệu và cũng loại biến đổi. Mặc định, kernel được sử dụng là Radial Basis Function Kernel (RBF).
* **Gamma:** Tham số này quyết định mức độ tác động của một ví dụ huấn luyện duy nhất trong quá trình biến đổi, và ảnh hưởng đến mức độ chặt chẽ của ranh giới quyết định xung quanh các điểm trong không gian đầu vào. Nếu giá trị gamma nhỏ, các điểm cách xa nhau được coi là tương tự. Vì vậy, nhiều điểm được nhóm lại với nhau và có các ranh giới quyết định mượt mà hơn (có thể không chính xác hơn). Các giá trị gamma lớn khiến cho các điểm gần nhau hơn (có thể gây ra hiện tượng overfitting, khớp quá mức).
* **Tham số 'C':** Tham số này điều khiển mức độ điều chuẩn được áp dụng vào dữ liệu. Giá trị C lớn đồng nghĩa với việc áp dụng điều chuẩn thấp, dẫn đến việc dữ liệu huấn luyện khớp rất tốt (có thể gây ra hiện tượng quá khớp). Giá trị C nhỏ đồng nghĩa với việc áp dụng điều chuẩn cao hơn, làm cho mô hình chấp nhận sai lệch (có thể dẫn đến độ chính xác thấp hơn).

### **2.5.2 Ưu và nhược điểm của Kernel Support Vector Machine**

**Ưu điểm**  
SVM có hiệu suất rất tốt trên nhiều tập dữ liệu khác nhau. Chúng rất linh hoạt: có thể xác định các hàm kernel khác nhau hoặc định nghĩa các kernel tùy chỉnh cho các loại dữ liệu cụ thể. SVM hoạt động tốt trên cả dữ liệu chiều cao và chiều thấp.

**Nhược điểm**

Hiệu suất (thời gian chạy và sử dụng bộ nhớ) giảm đi khi kích thước tập huấn luyện tăng lên. Yêu cầu việc chuẩn hóa cẩn thận dữ liệu đầu vào và điều chỉnh tham số. Không cung cấp trực tiếp một bộ ước lượng xác suất. Khó hiểu lý do dự đoán được đưa ra.

### **2.5.3 Một số kỹ thuật trong SVM (Support Vector Machine)**

**Phân tích thành phần chính (Principal Component Analysis - PCA)** là một phương pháp quan trọng trong lĩnh vực xử lý dữ liệu và học máy. PCA tập trung vào việc giảm chiều dữ liệu, giúp chúng ta khám phá những đặc trưng quan trọng và giảm số chiều của dữ liệu mà không mất đi thông tin quan trọng. Ý tưởng cơ bản của PCA là tìm những thành phần chính (principal components) của dữ liệu, tức là các vector trọng số tuyến tính tương ứng với các chiều mới. Các thành phần chính này sắp xếp theo độ giảm dần của phương sai, nghĩa là chúng giữ lại phần lớn thông tin của dữ liệu.

PCA cung cấp nhiều lợi ích. Đầu tiên, nó giúp giảm số chiều của dữ liệu, giúp tăng tốc các phép tính và giảm yêu cầu về bộ nhớ. Thứ hai, PCA có thể giúp khám phá mối quan hệ và cấu trúc trong dữ liệu, từ đó giúp hiểu rõ hơn về mối liên hệ giữa các đặc trưng. Cuối cùng, PCA cũng có thể được sử dụng trong quá trình trực quan hóa dữ liệu, đặc biệt khi chúng ta muốn biểu diễn dữ liệu nhiều chiều trong không gian hai hoặc ba chiều để dễ dàng quan sát. Tuy nhiên, cần lưu ý rằng PCA có một số hạn chế. Một trong số đó là PCA không xem xét mối quan hệ giữa các lớp dữ liệu trong các bài toán phân loại. Ngoài ra, nếu dữ liệu không được chuẩn hóa hoặc các tham số không được điều chỉnh cẩn thận, PCA có thể không đạt được kết quả tốt.

Chart, scatter chart

Description automatically generated

Hình 13:Hình ảnh mô tả PCA giữ lại các thông tin quan trọng của ảnh

**Grid search** là một kỹ thuật quan trọng trong học máy để tìm kiếm và chọn lựa các tham số tốt nhất cho một mô hình. Nó đơn giản nhưng mạnh mẽ, cho phép chúng ta khám phá không gian tham số của mô hình và xác định các giá trị tối ưu. Grid search hoạt động bằng cách xác định một lưới các giá trị tham số tiềm năng và thử tất cả các tổ hợp có thể của chúng. Mỗi tổ hợp sẽ được đánh giá thông qua quá trình đánh giá chéo (cross-validation) hoặc tập kiểm tra riêng biệt để đo lường hiệu suất của mô hình. Kết quả là chúng ta có thể xác định bộ tham số tối ưu mang lại kết quả tốt nhất cho mô hình.

Grid search đòi hỏi một tập tham số và khoảng giá trị để tìm kiếm. Thông thường, chúng ta xác định một lưới đều với các bước nhất định. Tuy nhiên, điều này cũng có thể dẫn đến việc kiểm tra quá nhiều tổ hợp, ảnh hưởng đến hiệu suất và thời gian chạy. Do đó, việc sử dụng grid search cần cân nhắc để tránh tốn thời gian và tài nguyên không cần thiết.

Chart

Description automatically generated

Hình 14:Hình ảnh mô tả Gridsearch thực hiện việc tìm kiếm tham số tốt nhất cho dữ liệu

**MTCNN (Multi-task Cascaded Convolutional Networks)** là một thuật toán phổ biến trong lĩnh vực nhận diện khuôn mặt và phát hiện điểm mốc khuôn mặt. Được phát triển bởi công ty Trung Quốc là Megvii (trước đây là Face++) năm 2016, MTCNN đã đạt được sự công nhận rộng rãi nhờ khả năng chính xác và hiệu suất tốt trong các ứng dụng thực tế. MTCNN có khả năng phát hiện khuôn mặt trong các hình ảnh với độ chính xác cao. Nó sử dụng một mô hình mạng neural tích chập được thiết kế để thực hiện ba tác vụ chính: phát hiện khuôn mặt (face detection), xác định điểm mốc khuôn mặt (facial landmark localization) và phân loại khuôn mặt (face classification). Quá trình hoạt động của MTCNN được thực hiện theo ba giai đoạn liên tiếp. Giai đoạn đầu tiên là phát hiện khuôn mặt. Mô hình tạo ra một tập hợp các vùng đề xuất (proposals) có khả năng chứa khuôn mặt trong hình ảnh đầu vào. Giai đoạn thứ hai là xác định điểm mốc khuôn mặt. Nó sử dụng các mạng neural để định vị các điểm như mắt, mũi và miệng trong khuôn mặt. Cuối cùng, giai đoạn cuối cùng là phân loại khuôn mặt, tức là xác định liệu khuôn mặt đó có phải là khuôn mặt của một người hay không.

MTCNN đặc biệt hữu ích trong các ứng dụng nhận diện khuôn mặt, như nhận diện khuôn mặt trong ảnh, video, hệ thống giám sát an ninh và nhận dạng khuôn mặt trong các ứng dụng di động. Với khả năng chính xác và hiệu suất mạnh mẽ, MTCNN đã trở thành một công cụ quan trọng để nâng cao các ứng dụng nhận diện khuôn mặt và an ninh trong thế giới số hiện đại.

Diagram

Description automatically generated with medium confidence

Hình 15:Cấu trúc của MTCNN

**FaceNet** là một mô hình học sâu nổi tiếng trong lĩnh vực nhận dạng khuôn mặt. Được phát triển bởi nhóm nghiên cứu tại Google vào năm 2015, FaceNet đã mang lại những tiến bộ đáng kể trong việc giải quyết các thách thức trong nhận dạng khuôn mặt và tạo ra sự đột phá trong lĩnh vực này. FaceNet sử dụng mạng neural sâu để học và trích xuất các đặc trưng chính từ khuôn mặt. Điều đặc biệt của FaceNet là nó không chỉ tạo ra một biểu diễn số học cho mỗi khuôn mặt, mà còn xác định một không gian đặc trưng nơi các khuôn mặt cùng thuộc một người có khoảng cách gần nhau, trong khi các khuôn mặt thuộc các người khác nhau có khoảng cách xa nhau. Điều này giúp FaceNet có khả năng phân biệt và nhận dạng các khuôn mặt với độ chính xác cao. FaceNet có nhiều ứng dụng quan trọng, bao gồm nhận diện khuôn mặt, xác minh danh tính và phân loại giới tính. Nó đã được sử dụng rộng rãi trong các hệ thống nhận dạng khuôn mặt tự động, bảo mật, quản lý thư viện ảnh và cả trong lĩnh vực thị giác máy tính.

Với khả năng học và biểu diễn khuôn mặt vượt trội, FaceNet đã đạt được kết quả ấn tượng trong các thử thách nhận dạng khuôn mặt, vượt qua cả con người trong một số trường hợp. Điều này đặt nền tảng cho sự phát triển của công nghệ nhận dạng khuôn mặt và đóng góp tích cực vào các lĩnh vực liên quan đến an ninh và giải trí.

FaceNet có một số đặc điểm tối ưu đáng chú ý:

* **Biểu diễn không gian đặc trưng:** FaceNet tạo ra một không gian đặc trưng nơi mỗi khuôn mặt được biểu diễn dưới dạng một vector số học. Điều này giúp giảm chiều dữ liệu và giữ lại những đặc trưng quan trọng nhất của khuôn mặt, tạo ra một biểu diễn đồng nhất cho các khuôn mặt của cùng một người.
* **Học sâu và trích xuất đặc trưng:** Mô hình FaceNet sử dụng mạng neural sâu để học và trích xuất các đặc trưng chính từ khuôn mặt. Việc sử dụng mạng neural sâu cho phép nắm bắt được những đặc trưng phức tạp và tạo ra biểu diễn chất lượng cao cho mỗi khuôn mặt.
* **Tích hợp học đa nhiệm:** FaceNet kết hợp các mục tiêu học tương ứng với việc nhận dạng khuôn mặt, đặt điểm mốc và phân loại khuôn mặt trong một mô hình duy nhất. Điều này giúp mô hình học được các đặc trưng phổ quát và đa nhiệm, cải thiện độ chính xác và hiệu suất.
* **Độ chính xác cao:** FaceNet đạt được độ chính xác cao trong việc nhận dạng khuôn mặt và xác định đặc trưng. Với việc tạo ra một không gian đặc trưng độc đáo và khả năng phân biệt tốt giữa các khuôn mặt, mô hình đảm bảo độ chính xác trong các tác vụ nhận dạng và phân loại khuôn mặt.
* **Tính di động và hiệu quả:** Mô hình FaceNet được tối ưu để có kích thước nhỏ gọn và tốc độ xử lý nhanh. Điều này làm cho nó rất thích hợp cho các ứng dụng di động và nhúng, nơi tài nguyên hạn chế và yêu cầu tính di động cao.

A person in a suit

Description automatically generated with low confidence

Hình 16:Mô tả FaceNet xử lý dữ liệu

CHƯƠNG 3: NỘI DUNG VÀ KẾT QUẢ THỰC NGHIỆM CỦA VIỆC XÂY DỰNG MÔ HÌNH NHẬN DIỆN KHUÔN MẶT TRÊN RASPBERRY PI 4

## **3.1 Xây dựng mô hình với mạng CNN**

Các bước xây dựng mạng CNN:

* **Chuẩn bị dữ liệu:** Bắt đầu bằng việc thu thập dữ liệu bằng cách cắt các ảnh khuôn mặt từ video. Mỗi người là 200 ảnh để chuẩn bị dữ liệu cho mô hình. Đảm bảo dữ liệu đã được gán nhãn đúng và chia thành tập huấn luyện và tập kiểm tra với tỉ lệ là 80% – 20%.
* **Xây dựng kiến trúc mô hình:** Xác định kiến trúc mô hình CNN phù hợp cho bài toán của bạn. Kiến trúc bao gồm các lớp convolutional, pooling, và fully connected layers.
* **Định nghĩa lớp convolutional:** Thêm các lớp convolutional vào mô hình để học các đặc trưng từ dữ liệu đầu vào. Chúng áp dụng các bộ lọc (filters) để trích xuất thông tin quan trọng từ ảnh hoặc dữ liệu đầu vào.
* **Áp dụng lớp pooling:** Thêm các lớp pooling để giảm kích thước không gian của dữ liệu và giảm thiểu lượng tính toán. Các phép lọc pooling thông thường là MaxPooling hoặc AveragePooling.
* **Thêm lớp fully connected:** Sau khi dữ liệu đã được lấy mẫu và giảm kích thước, thêm các lớp fully connected để kết nối các đặc trưng đã học với các lớp đầu ra.
* **Định nghĩa lớp đầu ra:** Thêm lớp đầu ra phù hợp với bài toán của bạn. Đối với bài toán phân loại, sử dụng softmax layer để dự đoán xác suất thuộc về từng lớp.
* **Định nghĩa hàm mất mát (loss function):** Chọn hàm mất mát thích hợp cho bài toán của bạn, ví dụ như categorical cross-entropy cho bài toán phân loại.
* **Tối ưu hóa mô hình:** Chọn thuật toán tối ưu hóa như Stochastic Gradient Descent (SGD) hoặc Adam và điều chỉnh các tham số để tối thiểu hóa hàm mất mát.
* **Huấn luyện mô hình:** Đưa dữ liệu huấn luyện vào mô hình và thực hiện quá trình huấn luyện để tối ưu hóa các tham số của mô hình.
* **Đánh giá mô hình:** Sử dụng tập kiểm tra để đánh giá hiệu suất của mô hình.

**Graphical user interface, text, application

Description automatically generated**

Hình 17:Hình ảnh mô tả việc tạo cơ sở dữ liệu để lưu thông tin của người nhận diện

**Graphical user interface, text

Description automatically generated**

Hình 18:Hình ảnh mô tả việc thêm thông tin hoặc cập nhật thông tin trong cơ sở dữ liệu

**Text

Description automatically generated**

Hình 19: Hình ảnh mô tả việc hiển thị thông tin và cắt ảnh khuôn mặt

Text

Description automatically generated

Hình 20:Hình ảnh mô tả việc lấy dữ liệu và chuyển đổi ảnh xám

Text

Description automatically generated

Hình 21:Hình ảnh mô tả việc tạo dữ liệu cho tập huấn luyện

A picture containing text

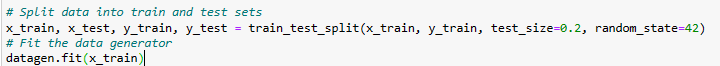
Description automatically generated

Hình 22:Hình ảnh mô tả việc gán nhãn cho dữ liệu

Text, letter

Description automatically generated

Hình 23:Sử dụng tăng cường dữ liệu



Hình 24:Chia dữ liệu thành train và test, sau đó gọi datagen cho tập huấn luyện

Text

Description automatically generated

Hình 25:Xây dựng mô hình và tiến hành huấn luyện

## **3.2 Xây dựng mô hình sử dụng Transfer Learning**

Các bước xử lý dữ liệu của mô hình này giống với các mô hình mạng CNN, chỉ khác việc sử dụng lại mô hình máy học đã được huấn luyện trên tập dữ liệu lớn để tinh chỉnh cho phù hợp với dữ liệu thực tế của đề tài.

Sử dụng bốn mô hình học sâu của mạng VGGNet (VGG16 – VGG19) và MobileNet – MobileNetV2

Text

Description automatically generated

Hình 26:Tạo ra 3 kênh màu với giá trị của ảnh xám để đưa vào mô hình để huấn luyện

Text

Description automatically generated

Hình 27:Tinh chỉnh mô hình VGG16

Text

Description automatically generated

Hình 28:Tinh chỉnh mô hình VGG19

Text

Description automatically generated

Hình 29:Mô hình MobileNet

Text

Description automatically generated

Hình 30:Mô hình MobileNetV2

A picture containing graphical user interface

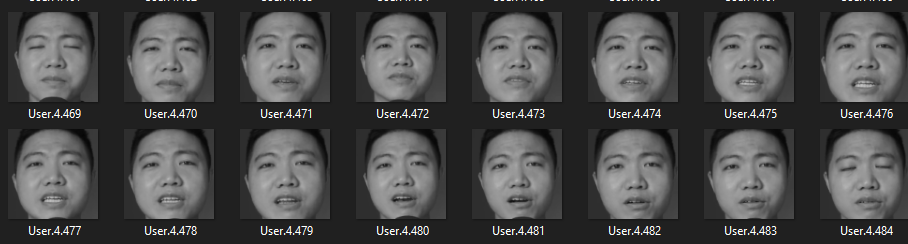
Description automatically generated

Hình 31:Hình ảnh dự đoán sau khi sử dụng các model trên

## **3.3 Đánh giá mô hình mạng CNN**

Do dữ liệu lấy từ các video thực tế, nhằm để có thể có nhiều các kiểu khuôn mặt khác nhau để mô hình huấn luyện tốt hơn nhưng vẫn còn hạn chế về mặt dữ liệu vì chưa thể tổng quát hết các trường hợp, do các vấn đề về môi trường, riêng tư nên việc lấy dữ liệu có giới hạn, vì lượng dữ liệu để thực hiện trên Raspberry Pi 4 là khá ít do phần cứng của thiết bị nhúng này có bị giới hạn và không đủ tài nguyên để xử lý thêm các tác vụ khác, nên em đã tiến hành sàng lọc và tinh chỉnh mô hình với tập dữ liệu vừa phải để Raspberry Pi 4 có thể chạy được với phần cứng của nó.

Lượng dữ liệu trên Raspberry đã cắt còn lại 5 người, mỗi người là 200 ảnh khuôn mặt.



Hình 32:Dữ liệu khuôn mặt đã qua xử lý và resize về cùng kích thước 256x256

Hình 33:Biểu đồ đánh giá độ chính xác và thời gian huấn luyện của các mô hình

VGG16: thời gian huấn luyện là 1h16’16s – độ chính xác là 100%

VGG19: thời gian huấn luyện là 1h23’41s – độ chính xác là 100%

MobileNet: thời gian huấn luyện là 1h12’22 – độ chính xác là 99%

MobileNetV2: thời gian huấn luyện là 24’02s – độ chính xác là 99%

CNN: thời gian huấn luyện là 2h55’49s – độ chính xác là 87.5%

**Nhận xét:** từ kết quả trên cho thấy các mô hình học sâu và một mô hình tự xây dựng có khả năng nhận diện được khuôn mặt với độ chính xác khá cao và thời gian huấn luyện tuỳ thuộc vào các kiến trúc của mô hình khác nhau với các độ sâu khác nhau nên sẽ nhanh hoặc chậm hơn.

## **3.4 Sử dụng SVM kết hợp PCA và Gridsearch cho nhận diện khuôn mặt**

Text

Description automatically generated

Hình 34:Thực hiện việc giảm chiều ảnh

Graphical user interface, text

Description automatically generated

Hình 35:Xử lý các nhãn và đặc trưng của hình ảnh

Text, letter

Description automatically generated

Hình 36:Chia dữ liệu thành 2 tập huấn luyện và kiểm tra đồng thời xử lý các thành phần của ảnh khi huấn luyện

Text

Description automatically generated with medium confidence

Hình 37:Các thành phần chính của ảnh sau khi dùng PCA để xử lý

Text

Description automatically generated

Hình 38:Tiến hành huấn luyện

A picture containing text, indoor, person

Description automatically generated

Hình 39:Tiến hành nhận diện bằng SVM

## **3.5 Sử dụng SVM kết hợp FaceNet và MTCNN cho nhận diện khuôn mặt**

Graphical user interface, text, application

Description automatically generated

Hình 40:Tiến hành load ảnh

Text

Description automatically generated

Text, letter

Description automatically generated

Hình 41:Nhận diện khuôn mặt và lưu trữ kết quả

A picture containing text, screenshot, monitor

Description automatically generated

Hình 42:Sử dụng kết quả nhận diện và vẽ khung hình

Graphical user interface

Description automatically generated

Hình 43:Tách và điều chỉnh kích thước khuôn mặt

Text

Description automatically generated

Hình 44:Thực hiện việc xử lý ảnh

Table

Description automatically generated

Hình 45:Gọi hàm và xử lý ảnh

Graphical user interface, text, application, chat or text message

Description automatically generated

Hình 46:Hiển thị ảnh để kiểm tra

Text

Description automatically generated

Hình 47:Sử FaceNet để Embedding các ảnh khuôn mặt thành vector



Hình 48:Lưu lại quá trình Embedding

Text

Description automatically generated

Hình 49:Tiền xử lý nhãn 'Y'

Diagram, schematic

Description automatically generated with medium confidence

Hình 50:Chia tập dữ liệu huấn luyện



Hình 51:Huấn luyện mô hình

Graphical user interface, text, application

Description automatically generated

Hình 52:Đánh giá mô hình

A picture containing text

Description automatically generated

A picture containing text

Description automatically generated

A picture containing text

Description automatically generated

A picture containing text

Description automatically generated

Hình 53:Thực hiện việc kiểm tra và lưu lại mô hình

Graphical user interface, application

Description automatically generated with medium confidence

Hình 54:Thực hiện việc nhận diện bằng camera

## **3.6 Đánh giá việc sử dụng nhận diện khuôn mặt bằng SVM và mạng CNN**

Dữ liệu dùng để thực hiện việc dùng CNN và SVM là khác nhau về mặt số lớp và kích thước cũng như hình ảnh, SVM sử dụng bốn lớp trong đó gồm có một lớp là dữ liệu của bản thân, các lớp còn lại sử dụng dữ liệu của LFW (Label Wild Face) để tiến hành nghiên cứu và thực hiện đề tài.

Từ các kết quả trên, đã cho thấy được việc nhận diện khuôn mặt kết với các kỹ thuật học máy hầu như đều cho ra kết quả khá tốt và có thể ứng dụng vào các hệ thống nhỏ nhận diện khuôn mặt trong thực tế. Các kết quả nghiên cứu đã chứng minh được việc xây dựng mô hình nhận diện khuôn mặt trên các thiết bị nhúng hoàn toàn được và ở đây là Raspberry Pi 4 mặc dù tài nguyên còn nhiều hạn chế nhưng vẫn có thể nhận diện được trong phạm vi nhỏ, và tối ưu về các chi tiết trong hình ảnh để giảm độ phức tạp.

## **4.1 Đánh giá**

Do việc lấy dữ liệu chưa đủ tính tổng quát, chỉ tập trung lấy những ảnh có tỉ lệ khác nhau nhiều hơn để phục vụ cho việc xây dựng mô hình chạy trên Raspberry Pi 4, lượng dữ liệu có hạn và chưa xử lý được nhiều hình ảnh chạy trên Raspberry Pi 4 nếu có số lượng ảnh nhiều hơn.

Qua việc lấy dữ liệu vừa đủ trên, em đã nghiên cứu và ứng dụng được vào bài toán nhận diện mặc dù vẫn còn nhiều hạn chế và khuyết điểm cần điều chỉnh, các mô hình sử dụng mạng CNN cho thấy được kết quả khá cao trong khi sử dụng SVM, việc sử dụng thêm SVM vào bài toán nhận diện với mục đích để so sánh thêm giữa CNN và SVM đồng thời học thêm được kiến thức mới.

Mô hình trên chỉ đủ để ứng dụng vào một thời điểm nhất định, có thể áp dụng vào các hộ gia đình nhỏ hoặc các nơi giám sát với quy mô nhỏ và mỗi ngày cần phải lấy mẫu và sử dụng nơi có điều kiện tương thích như trong phòng kín chỉ dành riêng cho việc nhận diện. Dù mô hình trên vẫn còn nhiều hạn chế nhưng vẫn có thể áp dụng vào các nơi vừa và nhỏ để nhận diện.

## **4.2 Hướng phát triển**

Mô hình trên cần được cải thiện về độ chính xác hơn và đặc biệt là phần xử lý các mẫu ảnh cũng như việc xử lý lấy mẫu cho tập dữ liệu, đồng thời tối ưu về mô hình để phù hợp hơn trên thiết bị nhúng. Đồng thời nâng cấp thêm phần cứng cho raspberry pi 4.

**TÀI LIỆU THAM KHẢO**

[1] J.-H. Kim, R.-G. Huang, S.-H. Jin, and K.-S. Hong, “Mobile-Based Flower Recognition System,” in *2009 Third International Symposium on Intelligent Information Technology Application*, Nov. 2009, pp. 580–583. doi: 10.1109/IITA.2009.407.

[2] M. Wieczorek, J. Siłka, M. Woźniak, S. Garg, and M. M. Hassan, “Lightweight Convolutional Neural Network Model for Human Face Detection in Risk Situations,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4820–4829, Jul. 2022, doi: 10.1109/TII.2021.3129629.

[3] “Geoffrey Hinton - IEEE Xplore Author Profile.” https://ieeexplore.ieee.org/author/37270925500 (accessed Apr. 18, 2023).

[4] Y. Zhou, Y. Liu, G. Han, and Y. Fu, “Face Recognition Based on The Improved MobileNet,” in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, Dec. 2019, pp. 2776–2781. doi: 10.1109/SSCI44817.2019.9003100.

[5] I. Gupta, V. Patil, C. Kadam, and S. Dumbre, “Face detection and recognition using Raspberry Pi,” in *2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, Dec. 2016, pp. 83–86. doi: 10.1109/WIECON-ECE.2016.8009092.

[6] S. A. Bleha and M. S. Obaidat, “Dimensionality reduction and feature extraction applications in identifying computer users,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 2, pp. 452–456, Mar. 1991, doi: 10.1109/21.87093.

[7] T. S. Gunawan, M. H. H. Gani, F. D. A. Rahman, and M. Kartiwi, “Development of Face Recognition on Raspberry Pi for Security Enhancement of Smart Home System,” *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, vol. 5, no. 4, Art. no. 4, Dec. 2017, doi: 10.52549/ijeei. v5i4.361.

[8] “A Flower Auto-Recognition System Based on Deep Learning - IOPscience.”https://iopscience.iop.org/article/10.1088/1755-1315/234/1/012088/meta (accessed Apr. 18, 2023).